Technical University of Košice



Faculty of Electrical Engineering and Informatics

Electrical Engineering and Informatics

15

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

ISBN 978-80-553-4682-3

2024

Electrical Engineering and Informatics 15
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:	July 2024
Edition:	1^{st}
Available on:	eei.fei.tuke.sk
Languague:	English, Slovak
Pages:	524

Editors:
Karol Kyslan
Emília Pietriková
Lukáš Pancurák

Proceedings reviewers:
Norbert Ádám
Frantisek Babic
Ján Bačík
Anton Baláž
Matej Bereš
Samuel Bucko
Gabriel Bugar
Peter Butka
Matúš Čavojský
Zsolt Čonka
Milos Drutarovsky
Juraj Ďurišin
František Ďurovský
Jozef Dziak
Daniel Dzivý
Peter Feciľak
Branislav Fecko
Pavol Fedor
Jana Fortes
Ján Genči
Peter Girovský
Renát Haluška
Máté Hireš
Eva Chovancová
Antónia Jusková
Ondrej Kainz
Slavomír Kardoš
Jozef Király
Štefan Korečko
Ondrej Kováč
Karol Kyslan
Milan Lacko
Ľubomír Livovský
Oliver Lohaj
Branislav Madoš
Marian Mach
Kristína Machová
Stanislav Ondas
Jakub Oravec
Ľuboš Ovseník
Lukáš Pancurák
Ján Paralič
Marek Pástor
Daniela Perduková
Ján Perháč
Alena Pietriková
Emília Pietriková
Marek Ružička
Branislav Sobota
Ján Staš
William Steingartner
Matúš Sulír
Csaba Szabó
Slavomír Šimoňák
Viktor Šlapak
Igor Vehec
Ján Zbojovský
Jaroslava Žilková

Copyright © 2024 FEI TUKE

ISBN 978-80-553-4682-3

Contents

Ján Bačík, Daniela Perduková, Pavol Fedor
Návrh pozemnej stanice pre podporu vývoja bezpilotných prostriedkov 7

Kristína Machová, Štefan Kožík
Analýza sentimentu s uvažovaním dôveryhodnosti autora 11

Kristína Machová, Ján Ofčák
Gradient posilňované rozhodovacie stromy verzus náhodné lesy 17

Kristína Machová, Laura Pončáková
Systém podpory dôveryhodnosti online obsahu 23

Kristína Machová, Jakub Šimoňák
Web aplikácia pre orientáciu v dezinformačnom obsahu 28

Kristína Machová, Lucia Grodecká
Slovníkový prístup k detekcii falošných správ o CoViD-19 32

Daniela Perduková, Marek Fedor
Laboratórny model malej vodnej elektrárne 36

Pavol Fedor, Peter Radvani
Fuzzy odhad kritických parametrov Klossovho vzťahu pre modelovanie asynchrónneho motora 42

Julián Gula, Branislav Madoš
Social network design with emphasis on security and privacy 48

Patrik Hurný, Branislav Madoš
Procedurálne generovanie obsahu počítačových hier 54

Peter Poprík, Branislav Madoš
Procedurálne generovanie obsahu v oblasti počítačovej grafiky 60

Martin Ondo-Eštok, Marián Mach
Numerická optimalizácia pomocou aritmetického optimalizačného algoritmu 67

Eva Križalkovičová, Ondrej Kainz, Miroslav Michalko, Miroslav Murin
Monitoring the occupancy of parking spaces via UAVs 72

Martin Novysedlák, Ondrej Kainz, Miroslav Michalko, Miroslav Murin
Data Recovery/Retrieval Using Forensics Analysis 78

Daniel Slanina, Michal Solanik
Leveraging Distributed Systems for Enhanced Parallel Calculations of Cosmic Ray Simulations in the Heliosphere 83

Dávid Lacko, Marián Mach
Digitálne dvojča hracej dosky v jednoduchých hrách 88

Martin Perignáth, Marián Mach
Štrukturálne regularizačné metódy pre hlboké siete 94

Pavol Dzurnak, Martin Hasin, Dominik Falis
Implementation of universal management for controlling smart devices 99

Michal Staš
Grafy na piatich vrcholoch 106

Vladimír Szomosi, Dušan Medveď
Návrh zariadenia pre výrobu vodíka 111

Daniel Čorňák, Dušan Medveď
Možnosti regulácie výkonu v ostrovnej prevádzke 116

Jozef Kompan, Anton Baláž
Digitalizácia analógových meračov 121

Tomáš Sivák, Anton Baláž
Systémy IDS cez proaktívne vyhľadávanie zraniteľnosti v sieťovom prostredí 127

Dominik Knieža, Branislav Madoš
Procedurálne generovanie modelu prostredia pre Unreal engine 135

Daniel Gordan, Marek Pastor, Tadeáš Kmecik, Tomáš Basarik
Výber polovodičov pre usmerňovač mostíkového meniča s fázovým posunom 142

Peter Girovský, Tadeáš Kmecik, Daniel Gordan, Tomáš Basarik
Porovnanie metód riadenia BLDC motora s využitím umelej inteligencie 147

Tomáš Basarik, Milan Lacko, Tadeáš Kmecik, Daniel Gordan
Určenie parametrov planárneho transformátora z nameraných priebehov 155

Barbora Viski, Peter Lukács
The Technological Aspects of Flexible Solar Cells 160

Yana Krasnova, Marián Mach
Image generation using diffusion deep models 166

Alžbeta Sklenáriková, Branislav Madoš
Procedurálny generátor grafickej reprezentácie rozsiahlych modelov riek 171

Tomáš Lenger
Analýza spoľahlivosti OhmegaPly rezistorov rôznych tvarov 178

Štefan Hričák, Ondrej Kováč
Úvod do kompresného snímania 184

Lukáš Ferenc, Ondrej Kováč
Prehľad metód kalibrácie vnútorných parametrov kamery 190

Miroslav Herbert, Ondrej Kováč
Štandardné videokodeky 196

Jaroslav Pajkoš, Miloš Drutarovský
Minimalistická implementácia TLS klienta pre vstavané aplikácie 202

Livovský Ľubomír
Možnosti použitia polymérnych kompozitných materiálov v konštrukcii FSR senzorov 209

Martin Mati, Peter Lukács
Aplikácia grafénu vo flexibilnej elektronike 215

Alena Pietrikova
Vplyv opakovaného reflow procesu na zmáčavosť povrchov pred osadením DPS 222

Miroslava Matejová, Ján Paralič, Adam Kováč
Nenávistné prejavy na sociálnych sieťach a ich detekcia pomocou vysvetliteľných modelov 227

Štefan Korečko, Michal Pustý, Branislav Sobota
The process of digitalization of the territory and DSML for the application domain of spatial planning and environment 234

Daniel Marcin, Milan Lacko
Aktívne balansovanie so spínanými kondenzátormi 241

Daniel Dzivý
Vplyv povrchových úprav dosiek plošných spojov na fraktografickú morfológiu a mechanické vlastnosti spájkovaných spojov 246

Kristián Eliáš, Ľubomír Beňa
Určenie spätných vplyvov obnoviteľných zdrojov energie na distribučnú sústavu 251

Anastasiia Tovstolis, Branislav Madoš, Heidar Khorshidiyeh
Vplyv modifikácie počtu a dĺžky ukazovateľov na dcérske uzly na veľkosť binárnej reprezentácie SV-DAG 257

Oliver Lohaj, Ján Paralič, Daniel Onofrej
Quantitative and qualitative properties of LLM-based data analysis models in terms of their usability 264

Igor Vehec, Lukáš Garnek
Možnosti vnorenia polymérnych hrubovrstvových rezistorov do 3D tlačených štruktúr 270

Ján Kaňuch, Peter Girovský, Filip Šulek
Inteligentné nabíjanie elektromobilov 275

Juraj Durisin
Test Stand for 3 Phase Asynchronous Electric Motor 285

Peter Girovský, Ján Kaňuch, Jaroslava Žilková
Neurónový regulátor rýchlosti pre jednosmerný cudzobudený motor 289

Stanislav Guskov, Miroslav Murin, Ondrej Kainz, Miroslav Michalko
Application for design and creation of graphics 296

Ján Kaňuch, Peter Girovský, Jaroslava Žilková
Súčasný stav možnosti nabíjania elektromobilov na Slovensku 304

Ján Papaj, Hai Tran Ngoc
Implementácia SD-MANET siete do simulátora Mininet-WiFi 311

Jozef Avuk, Matúš Sulír
Automated Blocking of Elements Overlapping Web Content 317

Kristián Stutiak, Branislav Madoš, Heidar Khorshidiyeh
Návrh hierarchickej dátovej štruktúry pre ukladanie geometrie časovo-variabilných 3D scén 324

Ivan Yatskiv, Miroslav Murin, Ondrej Kainz, Miroslav Michalko
Gathering Information about Events in Košice 329

Martin Janitor, Miloš Drutarovský
Kompaktný senzorový uzol s TCP/IP konektivitou a protokolom TLS 1.3 336

Branislav Sobota, Štefan Korečko, Nikoleta Dimunová
Virtual reality environment for support of art education 343

Daniel Novák, Marián Mach
Učenie neurónových sietí typu MLP pomocou algoritmu simulovaného žíhania 350

Ľuboš Ovseník, Jakub Oravec, Štefan Lazorčák
Návrh siete pre viac-pobočkovú firmu 355

Ľuboš Ovseník, Jakub Oravec, David Lechner
Monitorovanie poveternostných podmienok pre FSO systémy s využitím drona 362

Ľuboš Ovseník, Jakub Oravec, Slavomír Štosel
Skúmanie vlastností optických filtrov v DWDM prenosovom systéme 368

Zuzana Liščinská, Ľuboš Ovseník, Jakub Oravec
Využitie neurónových sietí na prepínanie liniek v hybridných FSO/RF systémoch 374

Patrik Gajdoš, Eva Kupcová, Matúš Pleva, Renát Haluška
Moderné metódy prevencie online obťažovania detí na Slovensku 381

Milan Lacko, Andrej Klein
HMI rozhranie a komunikácia s Master zariadením pre ovládanie žalúzií 386

Simona Kirešová, Milan Guzan
Vyhodnocovanie nameraných údajov a ich vizualizácia pomocou MATLAB aplikácie 393

Miroslava Matejová, Ján Paralič, Viktória Sanislová
Merateľné indikátory kvality vzdelávania získané z voľne dostupných dát 401

Ján Papaj, Tomáš Petrilák
Implementácia algoritmu PSO do MANET-IoT siete v simulátore NS-3 408

Šimon Gans, Ján Molnár
Analytic system for solving the magnetic field distribution in anisotropic magnetic materials ... 414

Milan Guzan, Simona Kirešová
Meranie koncentrácie CO2 a možný jeho vplyv v pedagogickom procese 420

Tomáš Maniak, Marek Pavlik
Vplyv bypass diód na zatienenie fotovoltických panelov 424

Peter Girovský, Jaroslava Žilková, Ján Kaňuch
Analýza vplyvu vlastností elektrického vozidla na dojazd 429

Diana Hozova, Jan Zbojovsky
Návrh elektroinštalácie pre medicínske centrum 437

Peter Bober, Želmíra Ferková
Meranie indukčnosti fáz spínaného reluktančného motora 444

Slavomír Kardoš
Akcelerometrická inerciálna jednotka v rôznych stupňoch voľnosti pre bezkontaktné spínacie pole 448

Lenka Kališková, Peter Butka
Využitie GAN sietí pre augmentáciu dát a zlepšenie klasifikácie dát 454

Lenka Kališková, Peter Butka
Využitie evolučných a heuristických princípov pri trénovaní konvolučných neurónových sietí ... 461

Lenka Kališková, Peter Butka, Adam Majirsky
Segmentácia koronálnych dier a aktívnych regiónov 468

Renát Haluška, Dalibor Šinaľ
Príklad využitia programovateľného drona pre monitorovanie nehodovej udalosti 474

Natalia Kurkina, Ján Papaj
Analýza smerovacích protokolov pre MANET v kontexte rôznych modelov mobility 478

Renát Haluška, Filip Pajkoš
Návrh letových vlastností drona pre monitorovanie miesta vzniknutého požiaru 483

Lenka Kališková, Peter Butka, Kamila Jenčiková
Detekcia meteorov na rádiových spektrogramoch 488

Renát Haluška, Adam Reňák, Matúš Pleva, Eva Kupcová
Príklad aplikácie pre podporu logopedických hier 495

Lenka Kališková, Peter Butka, Lívia Potočňáková
Segmentácia erupčných vlákien 502

Daniel Halas, Dušan Medveď
Analýza elektromagnetického poľa v agrokomplexe 509

Dominik Vranay, Laura Pituková, Peter Sinčák
Unveiling the Power of Capsule Networks: A Comparative Study with Convolutional Neural Networks in Feature Extraction and Transfer Learning 515

Author's Index 522

Návrh pozemnej stanice pre podporu vývoja bezpilotných prostriedkov

[1]*Ján BAČÍK,* [2]*Daniela PERDUKOVÁ,* [3]*Pavol FEDOR*

[1,2,3] Katedra Elektrotechniky a Mechatroniky, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenská republika

[1]jan.bacik.2@tuke.sk, [2]daniela.perdukova@tuke.sk, [3]pavol.fedor@tuke.sk

***Abstract* — V tomto príspevku je opísaný návrh a realizácia univerzálnej pozemnej stanice určenej pre podporu vývoja bezpilotných prostriedkov. Realizovaná pozemná stanica obsahuje výpočtové, komunikačné a navigačné moduly, ktorých úlohou je poskytnúť bezpilotnému prostriedku korekciu jeho aktuálnej polohy, s čo najvyššou presnosťou, a prenášať signálny od operátora do bezpilotného prostriedku a naopak. V príspevku je opísaný mechanický návrh pozemnej stanice, návrh časti elektronických modulov a ich vzájomné prepojenie.**

***Keywords* — mobilná robotika, senzorický systém, multikoptéra, RTK**

I. ÚVOD

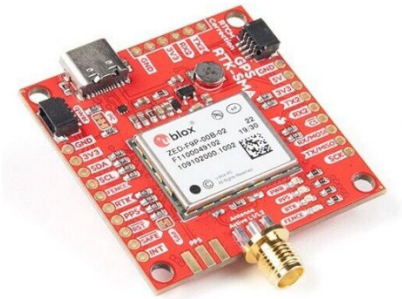
Dnešná doba prináša obrovský rozmach robotiky a robotických systémov ako takých. Tieto systémy dávno nenájdeme iba priemysle, ale aj v domácnostiach vo forme automatických vysávačov, kosačiek, či inteligentných domácich hračiek v podobe robotických zvieratiek a dronov [1]. Okrem toho sa mnohé „domáce“ robotické platformy stali obľúbenou platformou pre rôzne výskumné inštitúcie, ktoré sa venujú vývoju mobilných robotických systémov [2]. Ich výhodou je ich všestrannosť a možnosť rozšírenia o rôzne typy senzorov a výpočtové jednotky. V priemyselnom sektore sa mobilné robotické systémy využívajú ako logistické vozíky vo veľkých výrobných halách, kde zabezpečujú autonómny transport materiálu [3].

Okrem, už spomínaných, klasických mobilných robotov sa stále viac pozornosti venuje využitiu rôznych typov multikoptér alebo lietadiel s pevným krídlom a to nielen na komerčné účely, ako je napríklad autonómny prieskum alebo presun materiálu, ale aj ako tvorivý nástroj umelcov, ktorý nimi vytvárajú fascinujúce videá alebo dronové predstavenia. Pre tieto účely je nevyhnutné, aby tieto zariadenia disponovali presným navigačným systémom, ktorý umožňuje, v prípade potreby, plne autonómny let.

V tomto článku sme sa zameriavali na vývoj pozemnej stanice, ktorá je nevyhnutnou súčasťou komplexných autonómnych bezpilotných prostriedkov, nakoľko nezabezpečuje iba prenášanie cieľov misie do samotného prostriedku, ale často aj na základe dát z telemetrie a ďalších externých senzorov poskytuje bezpilotnému prostriedku korekciu polohy v priestore. Takéto presné korekcie sú napríklad nevyhnutné pri letoch viacerých bezpilotných prostriedkov v tesnej a synchronizovanej formácii.

II. ELEKTRONICKÉ MODULY POZEMNEJ STANICE

Základným senzorickým modulom nami navrhnutej pozemnej stanice je RTK (Real Time Kinematics) modul SparkFun Electronics GPS-1648 (Obr. 1) [4][5]. Vo všeobecnosti, RTK modul obsahuje GNSS snímač, ktorý prijíma signály z globálnych navigačných systémov a zároveň obsahuje určitý typ komunikačného kanálu pre príjem korekčných dát. Zdrojom korekčných dát môže byť RTCM korekčný tok vysielaný poskytovateľom tejto služby, napr. UNAVCO, alebo naopak, zdrojom korekčného signálu môže byť aj samotný sekundárny RTK modul, ktorý je presne umiestnený zemskom povrchu s presne známymi koordinátmi.



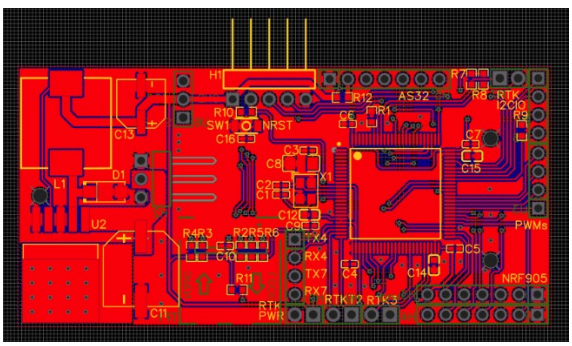
Obr. 1 SparkFun Electronics GPS-1648 GPS-RTK-SMA Breakout - ZED-F9P

Pre účely zabezpečenia obojsmernej komunikácie medzi pozemnou stanicou a bezpilotným prostriedkom je pozemná stanica vybavená komerčne dostupným modulom obsahujúci komunikačný čip NRF905 (Obr. 2). Tento komunikačný vysielač/prijímač je možné nastaviť tak, aby pracoval na frekvencií 433, 868 alebo 915Mhz. Používa moduláciu FSK/GMSK pri maximálnom výstupe +10dBm. Rýchlosť prenosu komunikačného modulu môže byť až do 76.8Kbit/s. Tento modul bol uprednostnený pred modulmi s pracovnou frekvenciou 2.4GHz, poprípade 5.8GHz a to z dôvodu lepšej penetrácie signálov z dlhšou vlnovou dĺžkou.

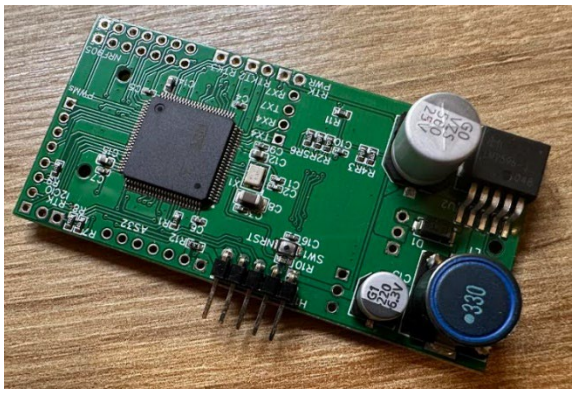


Obr. 2 NRF905 433 Mhz komunikačný modul

Jednotlivo spomenuté komerčne dostupné moduly sme potrebovali vzájomne komunikačne a energeticky prepojiť. Využitie rôznych komerčných napájacích, či procesorových modulov sme v tomto momente vývoja nepovažovali za správny nápad a to z dôvodu rôznych obmedzení a kompromisov, ku ktorým by sme museli pristúpiť, oproti riešeniu na mieru. Riešením teda bolo navrhnúť vlastnú PCB dosku obsahujúcu okrem napájaco distribučnej časti, aj časť riadiacu. Ako riadiaci procesor sme si vybrali 32 bitový procesor s označením STM32H743.



Obr. 3 Prepojovacia doska s procesom STM32H743 - návrh



Obr. 3 Prepojovacia doska s procesom STM32H743 - realizácia

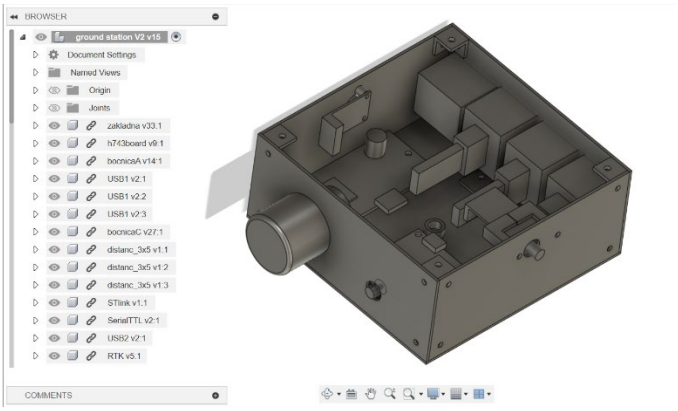
Tento procesor disponuje okrem dostatočného výpočtového výkonu aj dostatočným množstvom komunikačných periférii, najmä však vysokým počtom UART-ov (Universal asynchronous receiver-transmitter), ktoré sú potrebné pre nezávislú point to point komunikáciu medzi modulmi.

III. MECHANICKÝ NÁVRH POZEMNEJ STANICE

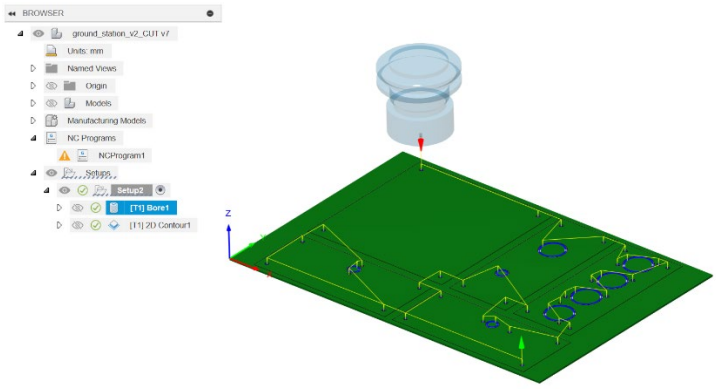
V prvok kroku, pri návrhu šasi pre pozemnú stanicu, bolo potrebné pre jednotlivé elektronické komponenty vytvoriť ich 3D reprezentáciu. Pri vytváraní jednotlivých 3D modelov nebol kladený dôraz na jednotlivé detaily, ale na základné rozmery a mechanické úchopy. Úlohou CAD modelu nebolo vytvoriť vernú kópiu reálneho sveta ale objektovú reprezentáciu, pomocou ktorej bolo možné navrhnúť jednotlivé spojovacie diely šasi.

Pre návrh celej stanice bol zvolený program Fusion 360. V ňom sme vytvorili jednotlivé komponenty a následne sme začali s návrhom jednotlivých spojovacích a krycích dielov. Pre krycie diely sme sa rozhodli použiť karbónové dosky hrúbky 2mm. Nevýhoda karbónu je nutnosť jeho rezania vo vodnom kúpeli, nakoľko karbónový prach je považovaný za karcinogénny materiál. Vodný kúpeľ má zabrániť uvoľňovaniu sa súčiastok prachu počas rezania do ovzdušia. Pre spojenie jednotlivých karbónových častí k sebe sme navrhli 3D tlačené spojovacie rohové uholníky. Tie umožňujú prichytiť jednotlivé karbónové diely k sebe pomocou skrutiek. Jednotlivé elektrické komponenty sú prichytené priamo do montážnych otvorov na jednotlivých karbónových dielcov alebo sú prichytené pomocou distančných stĺpov.

Výsledný 3D návrh s rozložením jednotlivých komponentov je na Obr. 4. Okrem v texte spomínaných modulov sa v návrhu nachádzajú aj pomocné komponenty, akými sú anténové konektory, či tlačidlo centrálneho vypnutia systému.

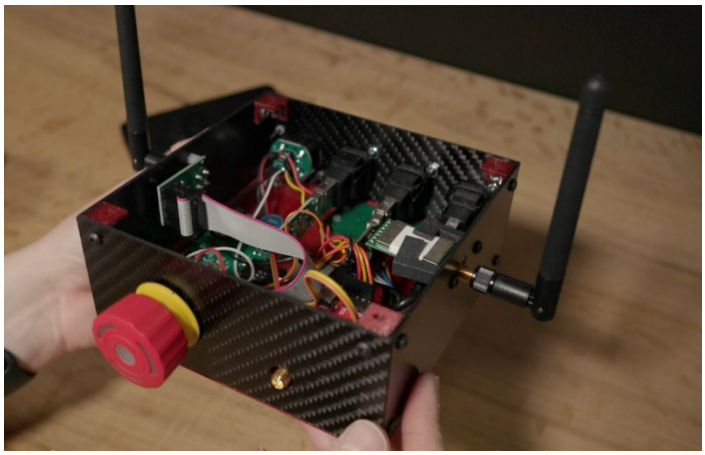


Obr. 4 CAD návrh mechanického usporiadania komponentov pozemnej stanice v šasi



Obr. 5 Rezný plán karbónovej dosky

Po ukončení samotného 3D modelu pozemnej stanice bolo potrebné pripraviť rezné plány (Obr. 5) pre CNC frézu. Ako bolo spomínané, samotné frézovanie jednotlivých dielov bolo realizované vo vodnom kúpeli, aby sa zabránilo šíreniu prachových častíc do ovzdušia. Výsledná pozemná stanica je na Obr. 6.



Obr. 5 Výsledná realizácia pozemnej stanice

IV. ZÁVER

Príspevok stručne popisuje návrh pozemnej stanice z mechanického hľadiska a jednotlivých použitých modulov. Táto platforma umožní ďalší vývoj v oblasti bezpilotných prostriedkov nakoľko predstavuje základný kameň, ktorý je potrebný pre vytvorenie komplexného systému schopného autonómnej prevádzky.

POĎAKOVANIE

Tento príspevok vznikol s podporou Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied v rámci projektu VEGA 1/0363/23.

REFERENCIE

[1] J. Forlizzi, C. DiSalvo, Service robots in the domestic environment: a study of the roomba vacuum in the home, proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, New Yourk, 2006.
[2] M. Dekan, L. Chovanec, A. Babinec, A. Vitko, New modules for the iRobot Create platform, Procedia Engieering, 2012.
[3] Ceit Logistic Trucks: availeble on web: http://www.ceittechinnovation.eu/index.php/sk/agv-system
[4] What is RTK, anyway?: available on web: https://www.sparkfun.com/rtk
[5] D.H. Titterton, J.L. Weston, Strapdown Inertial Navigation Technology, second edition, Institution of Engineering & Technology, Reston, 2005.

Analýza sentimentu s uvažovaním dôveryhodnosti autora

[1]*Kristína MACHOVÁ,* [2]*Štefan KOŽÍK*

[1] Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita Košice, Slovenská republika

[1]kristina.machova@tuke.sk, [1]stefan.kozik@student.tuke.sk

Abstrakt — **Článok sa zameriava na analýzu sentimentu v online príspevkoch na sociálnych sieťach. Cieľom práce je vyhodnotiť a klasifikovať dané názory diskurzu z hľadiska sentimentu. Pri analýze sentimentu sa kumuluje sentiment jednotlivých príspevkov do sumarizovaného názoru celej diskusie. Tento proces sumarizácie bol rozšírený o váhu prispievateľa, teda ide o váženú analýzu sentimentu. Váha sa odvádza od miery autoritatívnosti autorov daných príspevkov. Práca prezentuje aj prístup k predikcii autoritatívnosti prispievateľov a taktiež porovnáva výsledky váženej a neváženej klasifikácie názorov.**

***Kľúčové slová* — autoritatívnosť prispievateľa, konverzačný obsah, slovníkový prístup k analýze názorov, vážená analýza sentimentu**

I. Úvod

Vážená analýza názorov sa od klasickej analýzy názorov líši tým, že daný text hodnotíme nielen podľa sentimentu vyjadreného v texte, ale aj podľa váhy autoritatívnosti konkrétneho autora daného príspevku. Daná problematika sa s postupným rastúcim množstvom informácií a textu na sociálnych médiách stáva čoraz aktuálnejšou. Hlavným zámerom článku je použiť známe metódy na klasifikáciu názorov a skombinovať ich vo váženej analýze názorov a následne zhodnotiť a porovnať dosiahnuté výsledky medzi váženou a neváženou analýzou názorov. Príspevok taktiež popisuje návrh a implementáciu riešenia a na konkrétnom príklade ilustruje dosiahnuté výsledky.

S klasifikáciou názorov sa v bežnom svete stretávame pomerne často. Každý z nás dokáže pri akejkoľvek komunikácii, alebo diskusii podvedome určiť polaritu daného názoru, alebo textu. V praxi to znamená, že každý z nás dokáže rozpoznať, či sa z hľadiska oponenta jedná o príspevok kladný, záporný, alebo neutrálny. Výsledkom tejto klasifikácie je teda kategorizácia textu, alebo názoru, práve do jednej z troch spomínaných skupín. To, ako vnímame a hodnotíme názory a postoje iných ľudí priamo súvisia s našimi rozhodnutiami a budúcimi názormi k danej problematike. Klasifikácia nám umožňuje vyjadriť súhlas, alebo nesúhlas k danej problematike, utvoriť si vlastný názor k diskutujúcim a pod. V online podobe sa s ňou bežne stretávame a využívame ju pri rôznych recenziách, komentároch, online diskusiách a podobne. Aj v tomto prípade sa však jedná o klasifikáciu človekom. Vo svete IT technológií a online komunikácie sa nám preto naskytá nový problém a príležitosť klasifikovania príspevkov automaticky, čo môže byť veľkou pomocou pre moderátora diskusie.

II. Slovníkový prístup k analýze názorov

V princípe je možné použiť dva prístupy ku analýze názorov a to slovníkový prístup a prístup založený na strojovom učení. V tomto článku prezentujeme riešenie založené na slovníkovom prístupe.

Tak ako pri bežnom vnímaní textu a diskusie, tak aj pri automatickej klasifikácii sa výsledná polarita názoru odvádza od polarity slov. Pomocou známej polarity slov je možné po implementácii správnej metódy určiť polaritu a presnú hodnotu polarity celého textu. Základné úlohy klasifikácie názorov podľa [1] je možné zhrnúť do troch nasledujúcich bodov.

- Výber slov z analyzovaného textu vyjadrujúcich postoj, alebo kladný, či záporný význam.
- Triedenie slov do skupín podľa polarity, sily polarity, negácie, alebo intenzifikácie.
- Výsledný výpočet hodnoty polarity celého textu.

Vo všeobecnosti platí, že každý názor, alebo príspevok, musí patriť konkrétnemu autorovi, ktorý tento názor vyjadril, tzv. držiteľ názoru. Dôveryhodnosť tohto autora má takisto veľký

vplyv na výslednú hodnotu polarity príspevku, keďže náš prístup sa zameriava na váženú analýzu názorov.

Slová sme zaradili do dvoch skupín a ku každej sme vytvorili osobitný slovník. V prvej boli prídavné mená (skvelý, slabý, excelentný, nedostačujúci...), podstatné mená (hlúposť, krása, efektivita...) a slovesá (milovať, nenávidieť, klamať, pomáhať...) [2]. Slová boli zaradené do troch pozitívnych kategórií a troch negatívnych spolu s hodnotou polarity, čo ilustruje Obr. 1.

Kategória	Hodnota polarity	Slovo
silno pozitívne	**3**	excelentný, vynikajúci, bezchybný
pozitívne	**2**	super, krása, milovať
mierne pozitívne	**1**	dobrý, dostačujúci, pekný
mierne negatívne	**-1**	zlý, nedostačujúci, klamať
negatívne	**-2**	hlúposť, sklamaný, blbosť
silno negatívne	**-3**	príšerný, katastrofálny, nenávisť

Obr. 1 Ilustrácia zloženia prvého slovníka v rámci slovníkovej analýzy textov.

Druhá skupina obsahovala negácie (nie, nikdy, anti...), alebo intenzifikátory (veľmi, trochu, žiadny, maximálne...), teda slová, ktoré samy o sebe nevyjadrujú názor, ale v spojení so slovami zo skupiny číslo jeden nadobúda táto nová fráza inú hodnotu polarity ako bola polarita pôvodného slova. Príklad: *„Máš **pekné** tričko.“* sa môže zmeniť na *„Máš **veľmi pekné** tričko.“* alebo na *„Máš **nie veľmi pekné** tričko.“* Obr. 2 ilustruje stavbu druhého slovníka.

Kategória	Hodnota polarity	Slovo
silné zosilnenie	**1**	veľmi, úplne, maximálne, naj
mierne zosilnenie	**0,5**	mierne, naozaj, podstatne
zoslabenie	**-0,5**	trochu, zbytočne, málo
negácia	**-2**	nie, anti, nikdy, ne-

Obr. 2 Ilustrácia zloženia druhého slovníka v rámci slovníkovej analýzy textov.

Slová do slovníkov boli pridávané manuálne a anotované manuálne autormi. Slovníky obsahovali nielen spisovné slová ale aj nespisovné slová, slangy, hovorové slová, pejoratívne výrazy, zdrobneniny a pod. Každé slovo v slovníkoch spadalo do jednej zo spomínaných dvoch kategórií a malo kladný, alebo záporný význam. Výslednú klasifikáciu textu modeluje nasledovná rovnica (1).

$$P = \Sigma v(w_i^1)\cdot[1+\Sigma v(w_j^2)] \quad (1)$$

Kde:
P: ...je presná hodnota polarity pozorovaného textu
$v(w_i^1)$: ...je hodnota polarity slova w_i zo skupiny č. 1
$v(w_j^2)$: ...je hodnota polarity slova w_j zo skupiny č. 2
$\Sigma v(w_i^1)$: ...je súčet všetkých hodnôt slov v texte zo skupiny č. 1
$\Sigma v(w_j^2)$: ...je súčet všetkých hodnôt slov v texte zo skupiny č. 2.

Proces spracovania textu začína prehľadávaním každého slova v texte a porovnáva ho so slovami v slovníku. V prípade, že sa slovo zhoduje so slovom v slovníku, zaznamená sa skupina a hodnota polarity tohto slova. Táto hodnota sa súčasne cyklicky sčítava s predchádzajúcimi hodnotami polarity slov tých istých skupín. Ako výsledok po prehľadaní celého textu sú dve numerické hodnoty a to konkrétne súčet všetkých hodnôt polarít slov zo skupiny č. 1 $\Sigma v(w_i^1)$ a súčet všetkých hodnôt slov polarít zo skupiny č. 2 $\Sigma v(w_j^2)$. Tieto hodnoty sú následne využité vo výpočte polarity textu P pomocou vzorca (1). Za výslednú hodnotu polarity klasifikovaného textu považujeme premennú *P*, ktorú dostaneme vynásobením súčtov hodnôt z oboch skupín. Vo vzorci sa nachádza aj hodnota 1. Táto hodnota „1“ slúži ako neutrálna hodnota v prípade, že by súčet hodnôt skupiny č. 2 nadobudol hodnotu 0. V opačnom prípade by bola aj výsledná hodnota *P* rovná nule a výsledná klasifikácia by bola neúspešná.

III. PREDIKCIA MIERY AUTORITATÍVNOSTI

Tak, ako sa v diskusiách stretávame s nespočetným množstvom informácií, komentárov či názorov, takisto sa stretávame a reagujeme na rôzne typy autorov, ktorí do tejto diskusie prispievajú. Mieru autoritatívnosti týchto autorov sme v našom výskume hodnotiť podľa nasledovných vlastností – atribútov:

- Karma autora (*K*)
- Hodnotenie autora (*H*)
- Priemerný počet reakcií na jeden príspevok (*PR*)
- Priemerný počet znakov na príspevok (*PZ*)
- Priemerná poloha príspevku v strome (*PP*)
- Počet príspevkov (*PoP*)

Pre samotný výpočet a vyčíslenie autority existuje viacero postupov a metód. Na základe už vykonaných testov sme na výsledný výpočet autoritatívnosti zvolili najlepší model naučený pomocou lineárnej regresie z dát labelovaných takzvanou múdrosťou davu, keď prispievateľa hodnotia ostatní používatelia vo webovej diskusii [3]. Model je reprezentovaný nasledovným vzorcom (2) na výpočet autoritatívnosti *A*.

$$A = 0{,}43854*H + 0.325*K + (-0.0853)*PR + 0{,}002*PZ + (-0{,}2928)*PP + 1{,}0728*PoP \quad (2)$$

Takto vypočítaná autorita bude slúžiť ako váha na váženie hodnoty polarity názorov príspevkov daných autorov.

IV. METODOLÓGIA

Návrh prístupu ku váženej klasifikácii názoru sa dá zhrnúť do nasledovných krokov:

- zozbierať a vyextrahovať príspevky s menami autorov zo zvolenej diskusie
- príprava slovníkov
- zozbierať a vyextrahovať hodnoty atribútov autorov
- vypočítať autoritatívnosť jednotlivých prispievateľov
- analyzovať a klasifikovať text z hľadiska sentimentu,
- realizovať váženú analýzu názorov,
- porovnať a vyhodnotiť rozdiely medzi váženou a neváženou analýzou názorov.

Celý návrh a následne aj vývoj samotného programu bol realizovaný pomocou programovacieho jazyka JAVA v prostredí „NetBeansIDE“.

A. Extrakcia príspevkov a generovanie slovníkov

Po extrakcii príspevkov s menami autorov zo zdrojového kódu stránky diskusie bolo potrebné tieto príspevky ďalej očistiť a pripraviť pre analýzu sentimentu a odstrániť diakritiku z textu príspevkov. Niektoré príspevky boli v pôvodnom znení písané s diakritikou a iné bez. Príprava textu príspevkov zahŕňala takisto aj prekonvertovanie všetkých veľkých písmen na malé. Posledným krokom bolo manuálne odstránenie príspevkov s nulovou polaritou - príspevky, ktoré nevyjadrovali ani kladný, ani negatívny názor. Príspevky, ktoré boli napísané nezrozumiteľne boli takisto odstránené namuálne. Finálna verzia dátovej množiny aj s menami autorov bola uložená do textového súboru „final_nazory.txt“. Celý súbor, ako i následné operácie s textom v programe, sú vykonávané s ANSI kódovaním súboru

Slovník bol vytvorený manuálne zo slov, ktoré obsahovali jednotlivé príspevky v dátovej množine. Niektoré slová v slovníku boli upravené a napísané tak, aby presne vyhovovali konkrétnym slovám v texte príspevkov. Takýto spôsob by mal zaručiť, že výsledné odchýlky v presnosti analýzy budú dané analýzou samotnou v porovnaní s ľudskou klasifikáciou, nie chybne zaradenými slovami do slovníka. Vytvorený slovník v sebe obsahoval príslušné slovo, hodnotu polarity slova a skupinu slova. Podobne ako v predošlom prípade, slovník obsahoval slová bez diakritiky s malými písmenami pre čo najlepšiu zhodu so slovami v texte príspevkov. Aj v tomto prípade je slovník uložený v textovom súbore s kódovaním ANSI pod názvom „final_slovnik.txt“.

B. Extrakcia atribútov pre odhad autoritatívnosti

Prvým krokom k úspešnému výpočtu autoritatívnosti autorov zo zvolenej diskusie, bola správna extrakcia hodnôt parametrov autorov z webstránky diskusie. Aj v prípade tejto extrakcie sme postupovali podobným spôsobom, ako v prípade extrakcie príspevkov, a to transformáciou zdrojového kódu stránky do čitateľnej podoby numerických hodnôt. K odhadu autoritatívností

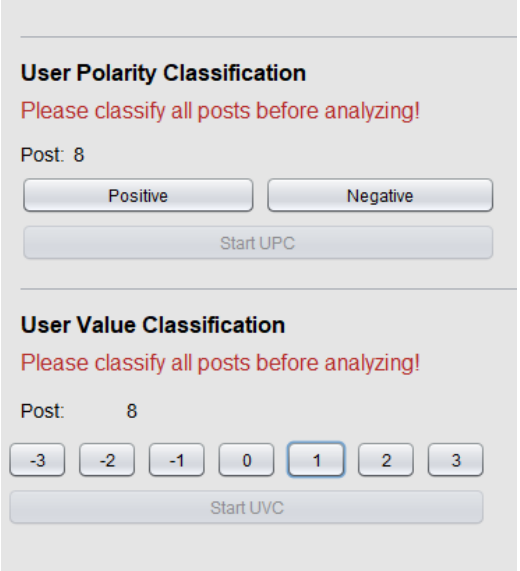
autorov boli potrebné hodnoty nasledovných atribútov – K, H, PR, PZ PP a PoP, pre každého prispievateľa.

Ďalším krokom po úspešnom vyextrahovaní dát bola úprava a normovanie získaných hodnôt pre finálny výpočet autoritatívnosti autorov. V tomto kroku bolo potrebné normovať všetky hodnoty atribútov príspevkov s rovnakým autorom okrem atribútu „karma“. Po úprave hodnôt atribútov bolo možné správne vypočítať miery autoritatívnosti autorov na základe predom zvolenej metódy výpočtu „lineárna regresia DAV“. Vypočítané autority sú uložené pod názvom „Autority.xlsx“.

V. VÁŽENÁ KLASIFIKÁCIA NÁZOROV

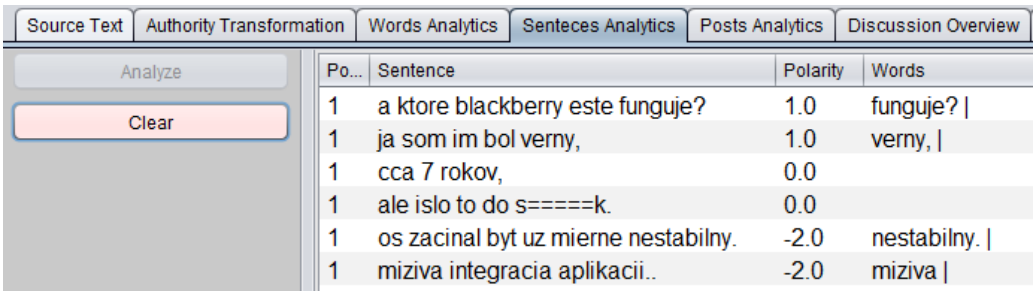
A. *Klasifikácia polarity príspevkov*

V našom výskume sme použili dva spôsoby klasifikácie polarity, jedna dvojhodnotová a druhá sedem hodnotová, čo ilustruje Obr. 3.



Obr. 3 Ilustrácia tried v dvoch spôsoboch klasifikácie polarity textov.

Ukážka samotnej klasifikácie použitím slovníkového prístupu je na Obr. 4.



Obr. 4 Ilustrácia samotnej klasifikácie polarity textov.

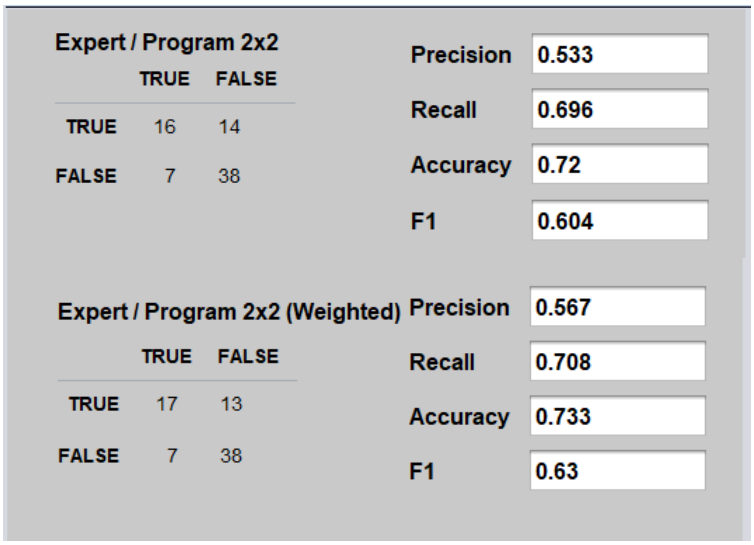
B. *Klasifikácia polarity príspevkov s uvažovaním váhy*

Po klasifikovaní príspevkov klasickou neváženou analýzou názorov sme pokračovali váženou analýzou názorov, ktorá do finálnej klasifikácie polarity príspevkov berie do úvahy aj vypočítané miery autoritatívnosti autora každého príspevku, teda WO (weight opinion) bol vypočítaný podľa vzorca (3).

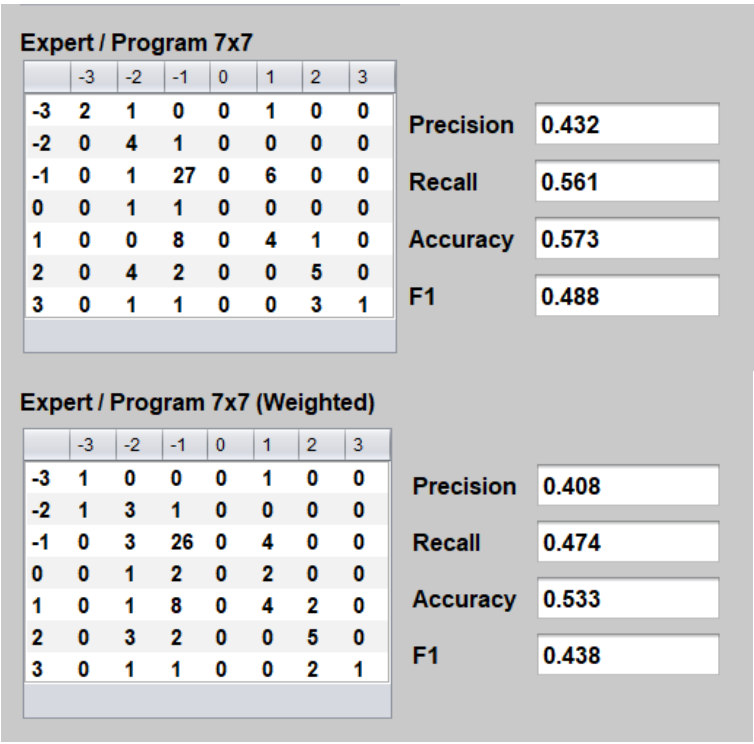
$$WO = \sum_k [\, w_k \sum_j o_{kj} \,] \tag{3}$$

Kde:
$w_k \in \langle 0,1 \rangle$; $o_{kl} \in \{-1,+1\}$ (resp. $\langle -3,+3 \rangle$).

Výsledky testov neváženej aj váženej analýzy názorov pre dve triedy true/false spolu s kontigenčnou maticou sú uvedené na Obr. 5 a pre sedem tried, respektíve sedem stupňov polarity (-3, -2, -1, 0, 1, 2, 3) na Obr. 6.



Obr. 5 Výsledky testov neváženej aj váženej analýzy názorov pre dve triedy true/false spolu s kontigenčnou maticou.



Obr. 6 Výsledky testov neváženej aj váženej analýzy názorov pre 7 tried polarity (-3, -2, -1, 0, 1, 2, 3) spolu s kontigenčnou maticou

VI. Záver

Hlavným cieľom práce prezentovanej v tomto článku bolo navrhnúť systém pre váženú klasifikáciu názorov a porovnať túto váženú analýzu názorov s neváženou klasifikáciou názorov. Výsledky ukázali, že vážená analýza v konečnom dôsledku má význam hlavne pri klasifikácií príspevkov do dvoch tried pozitívny/negatívny. Avšak pri klasifikácií príspevkov do siedmich kategórií sú výsledky horšie aj keď je to relatívne, keďže pri 7 triedach je náhodný výber 14,3 %, zatiaľ čo pri dvoch triedach je 50%.

V budúcnosti by bolo možné zdokonaliť proces váženej analýzy sentimentu tým, že automatizuje a implementuje do online prostredia. Celý proces slovníkovej metódy by sa v budúcnosti dal posunúť na úroveň, kedy by klasifikácia vedela správne vyhodnotiť aj slová s viacerými významami, vedela by zohľadňovať polaritu slovných hračiek. Takisto sa dá celá diskusia rozobrať viac do hĺbky a skúmať polaritu jednotlivých príspevkov vo vzťahu k diskusii, či dokonca k iným príspevkom na ktoré reagujú. Váženú analýzu názorov v recenziách produktov od bežných užívateľov, by bolo možné použiť ako nástroj na odhad najrelevantnejšej recenzie.

POĎAKOVANIE

Táto práca bola podporená Vedeckou grantovou agentúrou Ministerstva školstva Slovenskej republiky a SAV s číslom projektu. 1/0685/21 a Slovenskou agentúrou pre vedu a výskum s číslom projektu APVV-22-0414.

REFERENCES

[1] K. Machova, K., M. Mikula, X. Gao, X. and M. Mach "Lexicon-based Sentiment Analysis Using the Particle Swarm Optimization". Electronics, Vol. 9, no.1, 2020, ps. 1317.
[2] K. Machova, L. Marhefka „Opinion Classification in Conversational Content Using N-grams". In: Badica A., Trawinski B., Nguyen N. (eds) Recent Developments in Computational Collective Intelligence. Studies in Computational Intelligence, vol 513. Springer, Cham, 2014.
[3] J. Štefánik „Aproximácia závislosti autority od parametrov štruktúry webovej diskusie". Technická univerzita v Košiciach, 2017.
[4] L. Wu, F. Morstatter, H. Liu „SlangSD: Building and Using a Sentiment Dictionary of Slang Words for Short-Text Sentiment Classification."[10.5.2024]. Dostupné na internete: <https://arxiv.org/pdf/1608.05129.pdf>

Gradientovo posilňované rozhodovacie stromy verzus náhodné lesy

[1]*Kristína MACHOVÁ,* [2]*Ján OFČÁK*

[1] Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita Košice, Slovenská republika

[1]kristina.machova@tuke.sk, [1]jan.ofcak@student.tuke.sk

Abstrakt — **Článok sa zameriava na porovnanie výkonnosti dvoch metód strojového učenia a to gradientovo posilňovaných rozhodovacích stromov a náhodných lesov rozhodovacích stromov. Modely boli učené na dátach obsahujúcich toxické komentáre. Článok sa venuje aj popisu problému toxicity na sociálnych sieťach, ako vplýva na používateľov sociálnych sietí ako sa dá detegovať. Článok popisuje aj použité metódy, gradientovo posilňované rozhodovacie stromy a náhodné lesy rozhodovacích stromov. Tieto dva typy modelov boli trénované na dátach extrahovaných z portálu sme.sk, kde boli sumarizované údaje o vybratých používateľoch, ktorí sa zapojili do diskusií. Napokon boli diskutované výsledky experimentov a porovnaná efektívnosť a vhodnosť oboch prístupov na detekciu toxicity v online priestore.**

***Kľúčové slová* — toxicita v online priestore, rozhodovacie stromy, gradientové posilňovanie, náhodné lesy**

I. Úvod

Strojové učenie a umelá inteligencia sú v dnešnej dobe často používané v rôznych oblastiach, či už na riešenie klasifikačných alebo regresných úloh v rôznych oblastiach a taktiež sú vhodné na detekciu toxických komentárov alebo detekciu chatbotov na sociálnych médiách. Každým rokom sa metódy strojového učenia vylepšujú a málokedy nájdeme oblasť, kde by sa nevyužívali.

V tomto článku sa zameriavame na porovnanie dvoch metód strojového učenia, a to náhodné lesy a gradientovo posilňované rozhodovacie stromy. Náhodné lesy sa používajú dlhšiu dobu a sú často zvoleným prístupom pre svoju efektívnosť. Gradientové posilňovanie je novšia metóda a v súťažiach na kaggli, sa súťažiaci s touto metódou umiestňujú na vysokých priečkach. Obe metódy sú založené na princípe budovania rozhodovacieho stromu. Podobné metódy aj boosting a bagging.

Tieto metódy boli použité na učenie modelov na detekciu nedôveryhodných, resp. dôveryhodných používateľov a s tým súvisí problematika toxicity na internete, ktorá môže mať rôzne podoby ale v každom prípade vplýva na používateľov. Tieto detekčné modely je možné ponúknuť používateľom sociálny sietí na lepšiu orientáciu v konverzačnom obsahu a taktiež by ich bolo možné použiť na moderovanie online diskusie.

Obe spomínané metódy boli použité na detekciu dôveryhodných respektíve nedôveryhodných používateľov. Následne boli získané výsledky vizualizované a nakoniec boli porovnané výsledky oboch metód.

II. Online toxicita

A. Toxické komentáre

Sociálne média spôsobili zásadný prevrat v tom, ako používatelia medzi sebou komunikujú prostredníctvom internetu, pretože umožňujú jednoduchú a rýchlu výmenu informácií. Online platformy ako sú Youtube, Reddit, Facebook alebo Twitter, umožňujú používateľom diskutovať o rôznych témach budovaním online komunít na základe spoločných záujmov. S nárastom zapojenia používateľov v týchto online komunitách, moderácia online diskusií sa stáva čoraz náročnejšou. Táto problematika vyplýva z ťažkostí pri kontrole toxických komentárov, ktoré sa zameriavajú na rôznych používateľov s rôznymi typmi obťažovania a hrubosti. Toxické komentáre v online diskusiách odrádzajú používateľov od zdravých interakcií a od vyjadrenia svojho názoru.

Existuje viacero dôvodov, prečo dochádza k toxicite, v súvislosti s motiváciami spojenými s kyberšikanou, trollingom a socio-dynamickými efektmi, ako je napríklad skupinová polarizácia.

Jeden toxický komentár má potenciál prilákať viacero používateľov s cieľom zanechať podobný komentár v diskusii, čo vedie k znehodnoteniu online diskusie. Na vyjadrenie toxického komentára neexistuje presná definícia, no väčšinou sa jedná o komentár ktorého úmysel je zosmiešniť, uraziť, ponížiť alebo inak negatívne ovplyvniť určitého používateľa alebo skupinu ľudí. Obeťami toxicity môžu byť skupiny ľudí s určitým vierovyznaním, politickým názorom alebo aj náhodný používateľ [1].

Podľa prieskumu od spoločnosti McAfee Tweens Teens and Technology 2014 report skoro polovica mládeže v Indii používajúci sociálne média, zažili toxické komentovanie. Okrem toho prieskum naznačuje, že hlavným dôvodom vysokých čísel samovrážd tínedžerov sú negatívne komentáre adresované na ich osobu na sociálnych sieťach. Často to vedie k emocionálnej traume [2].

Ďalšou problematickou oblasťou sú videohry, v súčasnosti je video-herný trh považovaný za najrýchlejšie rastúci trh voľného čas a stal sa z jedným z najrozšírenejších symbolov dnešnej populárnej kultúry. Súčasné odhady naznačujú, že polovica populácie v západných krajinách zrejme hrá videohry. Tento rast popularity možno do značnej miery pripísať dizajnovému prvku súťaženia v hrách v reálnom čase (t.j. režimov hry pre viac hráčov). So zvýšeným využívaním hier sa však spájajú aj nové negatívne javy. Jedným z týchto javov je práve toxické správanie. Priemysel videohier (napr. Riot Games, Epic) sa pokúsili riešiť tento problém založením aliancie Fair Play. No bohužiaľ toxicita v hrách stále zostáva vážnym problémom v rôznych online videohrách pre viac hráčov, kde hráči majú možnosť medzi sebou komunikovať [3].

B. Detekcia toxicity

Sociálne média často zlyhávajú pri filtrovaní toxických príspevkov. Väčšinou existujú systémy nahlasovania, kde používateľ je schopný nahlásiť obsah ako zneužitie a takýto obsah potom môže byť odstránený zo sociálnej platformy. Napríklad Facebook má počet zamestnancov, ktorí pracujú s obsahom, ktorý je denne nahrávaný na profiloch, ako komentáre atď. Títo zamestnanci môžu manuálne overiť obsah a určiť, či je tento obsah škodlivý. Lenže tu vzniká problém, kde na platforme je množstvo dát a manuálne overovanie nedokáže stíhať overovať toto množstvo. Twitter tiež žiada svojich používateľov nesledovať ľudí, ak považujú ich obsah ako nevhodný. Ani jedna z vyššie uvedených stránok neposkytuje bezpečnostné mechanizmy na strane servera na obmedzenie negatívnych príspevkov. Mnohé platformy sa usilujú nájsť najlepšie riešenie ako obmedziť alebo zastaviť toxicitu na sociálnych sieťach a odborníci z oblasti informačných technológií hľadajú vhodnú metódu na detekciu toxických komentárov, napríklad vytvorenie vhodnej aplikácie ktorá umožní detegovať nevhodný obsah, označiť ho a upozorniť administrátora danej stránky aby ho v budúcnosti mohol moderátor stránky ľahko nájsť.

Autori Yadav a Manwatkar sa zamerali na automatickú detekciu toxicity a zisťovali urážlivé slová pomocou Aho-Corasick String Pattern algoritmu. Presnosť správnej detekcie ofenzívneho slova bola 90%. Táto metóda vedela automaticky detegovať urážlivé slová priamo po pridaní komentára, čo je aj nevýhodou tejto metódy, keďže nevie detegovať určité slovné spojenia [2].

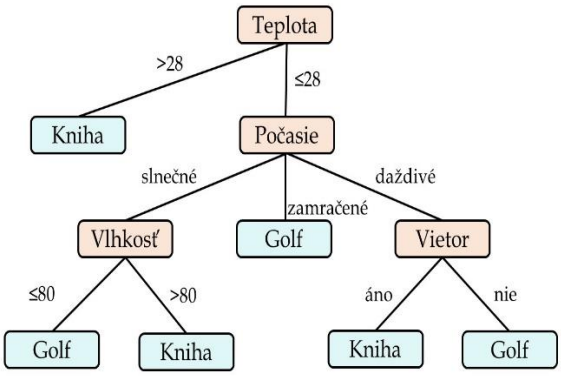
III. STROJOVÉ UČENIE

A. Rozhodovacie stromy

Rozhodovacie stromy sú klasickou metódou strojového učenia, ktorá sa dodnes často používa. Existuje celá rada stromových modelov, ako aj náhodné lesy rozhodovacích stromov a gradientovo posilňovaných rozhodovacích stromov, ktoré boli použité v našom výskume.

Algoritmus rozhodovacích stromov je založený na princípe delenia priestoru príkladov na podpriestory dovtedy, kým v každom podpriestore sú iba príklady jednej triedy. To je ukončovacia podmienka pre perfektnú klasifikáciu. Priestory príkladov sa delia podľa hodnôt testovacieho atribútu. Keby sa testovacie atribúty volili náhodne, strom by nebol optimálny. Optimálny strom znamená, že pri samotnej klasifikácii neznámeho príkladu sa bude overovať minimálny počet podmienok. Preto sa proces výberu testovacích atribútov riadi minimálnou entrópiou novovytváraných podpriestorov. Najčastejšie sa používa algoritmus C4.5 a jeho modifikácie pre klasifikačné problémy a CART pre regresné úlohy [4].

Príklad rozhodovacieho stromu učeného algoritmom C4.5 na známej golfovej množine ilustruje Obr. 1. Ide o riešenie klasifikačného problému, teda klasifikáciu do dvoch tried: ísť hrať golf/neísť hrať golf – lepšie čítať knihu kvôli počasiu.

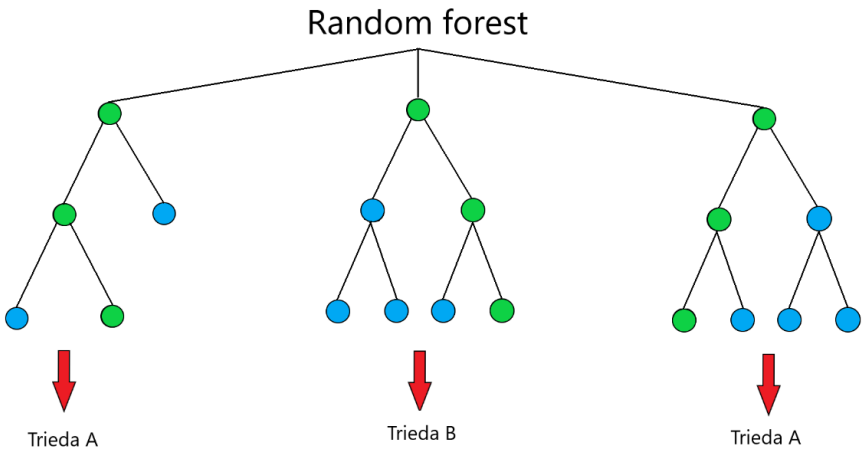


Obr. 1 Rozhodovací strom učený algoritmom C4.5 na známej golfovej množine.

B. Náhodné lesy rozhodovacích stromov

Náhodné lesy rozhodovacích stromv (Random Forests of Decision Trees) je jedna z metód učenia súborom metód [4]. Používa sa na zvýšenie efektívnosti slabých klasifikátorov. Učenie súborom metód znamená, že trénovacia množina sa rozdelí na viac častí a nad každou sa učí partikulárny klasifikátor. Pri finálnom rozhodnutí partikulárne klasifikátory hlasujú o výsledku čo modeluje vzorec (1). V prípade regresnej úlohy nejde o hlasovanie ale o spriemerňovanie výsledných hodnôt za každý partikulárny strom – viď vzorec (2).

Konkrétne náhodné lesy sú metódou, keď partikulárne klasifikátory môžu byť iba stromové modely. Používa sa tu náhodný výber podmnožiny atribútov pri hľadaní testovacích atribútov, čo zabezpečí, že jednotlivé stromy budú dekorelované, nepodobné a budú dávať lepšie výsledky na testovacej množine. To je ilustrované na Obr. 2.



Obr. 2 Ilustrácia hlasovania troch rozhodovacích stromov medzi triedami A a B v rámci náhodného lesa.

Tento princíp hlasovania je modelovaný nasledovným vzorcom (1) pre klasifikačný problém a vzorcom (2) pre regresný problém.

$$\hat{C}_{rf}^{B}(x) = majoritný\ hlas\ \{\hat{C}_b(x)\}_1^B \tag{1}$$

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{2}$$

C. Gradientovo posilňované rozhodovacie stromy

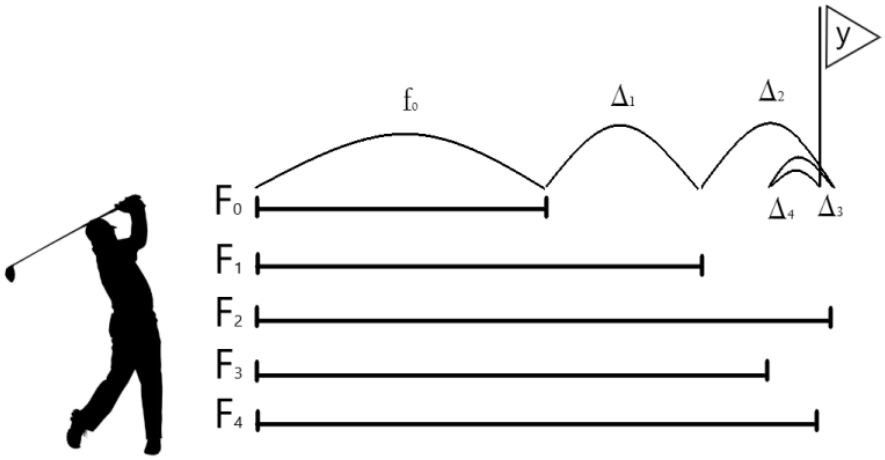
V poslednej dobe sa medzi stromovými modelmi začína presadzovať metóda gradientovo posilňované rozhodovacie stromy (gradient boosting decision tree - GBDT). Je to model rozhodovacích stromov, ktoré sú vytvárané a trénované sekvenčne, čiže postupne. V každej iterácii, sa gradient boosting decision tree učí rozhodovacie stromy, ktoré sú prispôsobené pomocou negatívnych gradientov, ktoré môžeme aj inak nazvať ako zvyškové chyby [5].

Formálne, vzhľadom na konvexnú stratovú funkciu **l** a súbor údajov s **n** inštanciami a **d**

funkciami $\{(x_i,y_i)\}$ pre $i = 1 \ldots n$ kde ($x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$), GBDT minimalizuje nasledujúci regulovaný cieľ (3):

$$\tilde{\mathcal{L}} = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{3}$$

Hlavný rozdiel oproti random forests (náhodným lesom), kde sa tiež vytvára viacero rozhodovacích stromov, je ten, že pri random forests sa vytvárajú stromy paralelne a pri gradient boostingu sa vytvárajú sekvenčne po každej predikcii, v reálnom svete tento proces môžeme prirovnať napríklad ku golfu, ako môžeme vidieť na Obr. 3, kde hráč sa vždy približuje k dierke, čiže v tomto prípade k odhadovanej hodnote y, hráč pri každom výstrele má určitú chybu, čiže každý hráčov výstrel na dierku je vlastne iterácia.



Obr. 3 Ilustrácia podobnosti gradient boosting rozhodovacieho stromu s hraním golfu.

Najväčší problém nastáva v učení sa rozhodovacích stromov, kde najviac časovo náročná časť je nájsť najlepšie deliace body vo vlastnostiach dát. Najpopulárnejším algoritmom je takzvaný „pre-sorted algorithm“, čo je algoritmus, ktorý vymenúva všetky deliace body na vopred zoradených (pre-sorted) hodnotách vlastností. Ďalším populárnym algoritmom je algoritmus, ktorý je založený na stĺpovom diagrame, takzvaný „histogram-based algorithm“, ten počas trénovania namiesto toho aby vyhľadával deliace body v zoradených hodnotách vytvára zásobníky atribútov na základe ktorých vytvára stĺpcové diagramy s atribútmi. Táto metóda má výhodu v tom, že je viac efektívna čo sa týka využitia pamäte zariadení a rýchlosti. Takéto algoritmy využíva napríklad metóda XGBoost alebo pGBRT [5].

IV. Implementácia

Na implementáciu dát sme sa rozhodli použiť už spracované verejne dostupné dáta, z diskusného portálu www.sme.sk, z ktorého boli extrahované dáta na odhad nedôveryhodnosti používateľa. Údaje boli vyberané najmä podľa obľúbenosti diskusií, čiže boli vyberané diskusie, ktoré boli dostatočne bohaté aby poskytli dostatok údajov. Dôležitým kritériom bola serióznosť portálu. Dataset obsahuje základných 6 atribútov, ktorých hodnoty boli extrahované pre 120 rôznych používateľov:

- Karma používateľa
- Priemerné hodnotenie používateľa -, toto hodnotenie udáva ako priemerne ostatní užívatelia portálu súhlasia alebo nesúhlasia s daným používateľom
- Priemerná poloha príspevku v strome komentárov, udáva ako hlboko sa nachádzal príspevok používateľa v diskusii
- Počet reakcií na príspevok daného používateľa
- Počet príspevkov daného používateľa v diskusii,
- Priemerný počet znakov, ktoré obsahujú príspevky používateľa

Dáta taktiež obsahujú aj nickname používateľov. Atribúty plnia úlohu nezávislé premenné Xi. Tieto dáta sú rozšírené o dve závislé premenné, ktoré môžeme. označovať ako Yj, a to finálne

hodnotenia autoritatívnosťou používateľa expertom a múdrosťou davu, ktoré bolo odvodené z hodnotení ostatných účastníkov diskusie.

Na hodnotenie efektívnosti implementovaných riešení bola použitá stredná kvadratická chyba (Mean Squared Error – MSE), stredná absolútna chyba (Mean Absolute Error – MAE) a R kvadrát, ktorý predstavuje podiel rozptylu pre závislú premennú y a vysvetľuje, do akej miery rozptyl jednej premennej vysvetľuje rozptyl druhej premennej. Ak je teda R^2 modelu 0.75, potom približne 75% pozorovanej variácie možno vysvetliť vlastnosťami modelu. Počíta sa podľa vzťahu (4):

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{4}$$

Na implementáciu GBDT sme použili GradientBoostingRegressor z knižnice sklearn, kde táto používa GBDT na regresiu. Pri GBDT v dátach nie všetky parametre výrazne ovplyvňovali výsledok. Medzi parametre GradientBoostingRegreesor, ktoré najviac ovplyvňovali výsledok patria patria nasledovné (v zátvorke uvedená optimálna hodnota získaná ladením):

- *Learning_rate (0.3)*, riadi, ako veľmi sa má zmeniť model v reakcii na odhadovanú chybu pri každej aktualizácii váh modelu, všeobecne sa odporúča používať hodnota medzi 0.1 a 0.3, predvolená hodnota v knižnici je 0.1
- *n_estimators (100),* tento parameter udáva, koľko krát sa má boosting vykonať, inými slovami môžeme povedať, že je to počet iterácií, predvolená hodnota 100
- *min_samples_split (3)*, minimálny počet vzoriek požadovaných na rozdelenie vnútorného uzla, predvolená hodnota je 2
- *max_depth (5),* je maximálna hĺbka rozhodovacieho stromu, predvolená hodnota je 3.

Podobne sme implementovali RF. Tu výsledok najviac ovplyvňovali dva parameter a to *n_estimators* (optimálna hodnota 80) a *min_samples_leaf* (6). Tabuľka 1 prezentuje najlepšie získané výsledky pre oba algoritmy a pre oba spôsoby labelovania – hodnotenie expertom a múdrosťou davu.

Tabuľka 1
Výsledky experimentov s učením pomocou GBDR a RF na dátach labelovaných expertom a múdrosťou davu.

	MAE	MSE	R2
GBDT, múdrosť davu	3.3036	28.46	92.97%
GBDT, expert hodnotenie	17.2560	457.74	21.08%
Random Forets,múdrosť davu	3.3016	23.69	91.85%
Random Forests, expert hodnotenie	15.10	359.69	27.86%

Pri náhodných lesoch boli už pri predvolených parametroch dosiahnuté lepšie výsledky ako pri GBDT ale nakoniec GBDT model dosiahol väčšiu úspešnosť. Problémom pri GBDT bolo nájsť správnu kombináciu parametrov na dosiahnutie hranice úspešnosti. Pri náhodných lesoch sme hľadali kombináciu parametrov kratšie, no nakoniec náhodné lesy mali menší R kvadrát aj keď výsledky boli dosť podobné a náhodné lesy mali aj menšiu priemernú absolútnu chybu a strednú štvorcovú chybu s výnimkou odhadovania hodnotenia vytvoreného expertom. Môžeme konštatovať, že obe metódy sú relatívne rovnako výkonné, aj keď náhodné lesy vedeli lepšie vybrať atribúty podľa ktorých sa stroj mal rozhodovať. GBDT vytvára rozhodovacie stromy sekvenčne a ich vytváranie trvá dlhšie ako to je pri náhodných lesoch, s čím sa ale vie vysporiadať modifikácia GBDT XGBoost, kde sa posilňovanie vykonáva paralelne.

V. Záver

Tento článok bol venovaný porovnaniu výkonnosti dvoch metód strojového učenia, gradientovo posilňovaných rozhodovacích stromov a náhodných lesov na na učenie modelov na detekciu nedôveryhodného, resp. dôveryhodného prispievateľa do online diskusie. Zistili sme, že obe metódy sú skoro rovnako výkonné, rozdiely vo výsledkoch sú minimálne, náhodné lesy získali najvyššie skóre 91,85% a gradientovo posilňované rozhodovacie stromy 92.97%, avšak ich priemerná absolútna chyba bola takmer rovnaká s rozdielom len 0.020, náhodné lesy však dokázali lepšie vybrať atribúty podľa ktorých sa rozhodovali. Ukázalo sa, že náhodné lesy sa dajú jednoduchšie implementovať ako GBDT ale pri dostatočnom ladení GBDT vieme dosiahnuť lepšie výsledky.

Poďakovanie

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0685/21 a Agentúry pre podporu vedy a výskumu v rámci projektu APVV-22-0414.

Referencie

[1] H. Almerekhi, et al. "Are These Comments Triggering? Predicting Triggers of Toxicity in Online Discussions". In *Proceedings of The Web Conference 2020* [online]. Taipei Taiwan: ACM, 2020. s. 3033–3040. .

[2] P. Parekh, H. Patel „Toxic Comment Tools: A Case Study". In *International Journal of Advanced Research in Computer Science* . 2017. Vol. 8, no. 5, s. 964–967.

[3] B. Kordyaka, et al. „Towards a unified theory of toxic behavior in video games". In *Internet Research* . 2020. Vol. 30, no. 4, s. 1081–1102.

[4] K. Machová „Nové trendy v strojovom učení. Štatistický prístup". FEI TU Košice, Košice, ps. 1-96, ISBN 978-80-553-2602-3.

[5] G. Ke, et al. "Lightgbm: A highly efficient gradient boosting decision tree". In *Advances in neural information processing systems,* 2017. Vol. 30, no.1., 2017, ps. 1-9.

[6] Q. Li, et al. "Privacy-preserving gradient boosting decision trees". In *Proceedings of the AAAI Conference on Artificial Intelligence,* 2020. ps. 784–791.

Systém podpory dôveryhodnosti online obsahu

[1]*Kristína MACHOVÁ,* [2]*Laura PONČÁKOVÁ*

[1] Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita Košice, Slovenská republika

[1]kristina.machova@tuke.sk, [1]laura.poncakova@student.tuke.sk

Abstrakt — **Článok sa zameriava na problematiku dôveryhodnosti sociálneho webu a na to, čomu je dôležité venovať pozornosť, ak čerpáme informácie z internetu. Popisuje ako vnímajú dôveryhodnosť webových informácií študenti, ale aj tí, ktorí už neštudujú. V príspevku je popísaná aj tvorba modelu založeného na slovníkovom prístupe, ktorého dôležitou časťou je predspracovanie textu, ale najmä vytvorenie slovníka. Súčasťou prezentovanej práce je aj vytvorenie interaktívnej hry ako aj aplikácie na webovom portáli. Sú vyhodnotené výsledky modelu založeného na slovníkovom prístupe a náučnej interaktívnej hry.**

***Kľúčové slová* — toxicita v online priestore, rozhodovacie stromy, gradient posilňovanie, náhodné lesy**

I. Úvod

Sociálne médiá založené na najnovších technológiách zaznamenali zvýšené používanie ako zdroj informácií. Používajú sa hlavne na vyhľadávanie informácií o závažných témach, ako je napríklad prepuknutie cholery na Haiti. Taktiež sa zaznamenalo veľké využitie u tých, ktorí vyhľadávajú informácie o zdraví. Ľudia teda využívajú sociálne „nástroje“ na zhromažďovanie informácií, zdieľanie príbehov, ale aj diskusiu o problémoch. Podobne aj zdravotnícke organizácie vidia výhody v sociálnych médiách, pretože im umožňujú prístup k informáciám o zdravotnej starostlivosti. Existuje veľa príkladov sociálnych médií napr. Digg, Facebook, Youtube, Flickr, ale veľký potenciál pre získavanie informácií predstavuje aj Twitter. Twitter je služba, ktorej podstatou je mikroblogerstvo, a ktorá bola spustená v marci 2006. V rámci využívania sociálnych médií predstavuje riziková situácia problém na ktorý sa zameriavajú. . Informácie z „prvej línie“ pochádzajú od očitých svedkov, ktorí informujú o nedávnych udalostiach. V mnohých prípadoch dokonca aj spravodajské agentúry, zhromažďujú informácie z týchto zdrojov skôr, ako dostanú najnovšie správy a poskytnú informácie. Technické problémy v miestach, ktoré sú zasiahnuté krízovou situáciou (nefunkčné satelitné spojenie a pod.) môžu spomaliť získavanie informácií od tradičných zdrojov, avšak informácie zo sociálnych médií môžu byť distribuované rýchlejšie. Napríklad počas haitského zemetrasenia v januári 2010 zohrávali sociálne médiá kľúčovú úlohu pri šírení informácií o tejto tragédii. Ako naznačujú Sutton, Palen a Shklovski (2008), sociálne média sa dostávajú do popredia ako zdroj informácií v čase katastrofy a rizikových situácií, aj keď presnosť informácií, ktoré sú zdieľané prostredníctvom kanálov je nejasná. Preto je nevyhnutné dozvedieť sa viac o tom, ako ľudia hodnotia informácie, ktoré získavajú na webových stránkach sociálnych médií, najmä z hľadiska ich dôveryhodnosti.

II. Dôveryhodnosť online obsahu

Jedným z kritérií, ktoré používame na filtrovanie informácií je ich dôveryhodnosť a uveriteľnosť. Je veľmi dôležité vziať do úvahy, že dôveryhodnosť je „vnem“, a teda nie je vlastnosťou, ktorá je podstatou samotného kanála alebo zdroja. Podľa Piercea je všeobecná definícia spoľahlivosti informácií ako „rozsah, v akom sa môžeme spoľahnúť na zdroj údajov, a teda aj na údaje samotné“. Pierce tiež uvádza, že spoľahlivé údaje sú dôveryhodné, autentické, originálne a renomované. Na dôveryhodnosť online informácií môže mať preto vplyv veľa faktorov. Jedným z modelov vytvorených na vyjadrenie a vysvetlenie procesu vytvárania úsudkov o dôveryhodnosti v online prostredí je MAIN model. Tento model poskytuje ľuďom heuristické spracovávanie podnetov pri rozhodovaní o dôveryhodnosti online zdroja. Podľa

Sundara (2008) je jednou z často používaných heuristík Strojová heuristika, podľa ktorej ľudia pripisujú väčšiu dôveryhodnosť informáciám, ktoré sú overené alebo vybrané strojom či počítačom ako osobou. Je to preto, lebo sa ľudia na stroj pozerajú ako na niečo, čo nemá žiadne myšlienky, emócie ani iné predsudky, a preto sa o ňom predpokladá, že nie je zaujatý. Keď sa táto Strojová heuristika spojí s aktuálnosťou správ, má to výrazný vplyv na posudzovanie dôveryhodnosti [1].

Štúdia [2] predpovedala, že frekvencia aktualizovania informácií bude mať pozitívny lineárny vplyv na vnímanú dôveryhodnosť zdrojov, takže rýchlejšie frekvencie aktualizácie povedú k zvýšeniu dôveryhodnosti zdrojov. Súčasný výskum taktiež predpovedal pozitívny vzťah medzi dôveryhodnosťou a kognitívnym spracovaním informácií. Zistenia však naznačujú, že frekvencia aktualizácií informácií nemusí mať priamy vplyv na dôveryhodnosť zdrojov, ale namiesto toho je to kognitívne spracovanie informácií tzv. prostredníkom vo vzťahu medzi frekvenciou aktualizovania a dôveryhodnosťou. To znamená, že kognitívne spracovanie informácií je dôležitou premennou, ktorú je potrebné zohľadniť v budúcich štúdiách dôveryhodnosti zdrojov.

A. Analýza dôveryhodnosti online informácií

Dôveryhodnosť informácií na internete je čoraz dôležitejšia, aj preto sa vyvinul webový analytický systém s názvom WISDOM (webové informácie rozumne a diskrétne usporiadané a zoradené), ktorý podporuje ľudský úsudok o dôveryhodnosti informácií analýzou webových stránok na základe rôznych kritérií, ako je odosielateľ, obsah a vzhľad. Informácie o vzhľade predstavujú štýl dokumentu a povrchovú charakteristiku, ako je rozloženie strany, počet inzerátov a či je alebo nie je uvedená kontaktná adresa. Ľudia môžu posúdiť užitočnosť a dôveryhodnosť webovej stránky pomocou informácií o vzhľade bez pochopenia obsahu. Napríklad, keď používateľ vyhľadáva na webových stránkach účinky zdravého jedla, tak preskočí stránku, ktorá obsahuje pár viet a veľa veľkých obrázkov. Ak nájde kontakt/adresu na webovej stránke, usúdi, že stránka je dôveryhodnejšia a naopak, ak nájde viacero reklám na webovej stránke tak usúdi, že stránka je menej dôveryhodná. Cieľom bolo vyvinúť modul pomocou informácií týkajúcich sa vzhľadu [3].

B. Analýza dôveryhodnosti online informácií

Práca [4] analyzuje informačnú dôveryhodnosť šírených správ prostredníctvom populárnej služby mikroblogovania. Na platforme mikroblogov používatelia uverejňujú krátke správy, ktoré sa zobrazujú ich sledovateľom (alebo používateľom, ktorí ich odoberajú) v reálnom čase. Tieto krátke správy sú známe ako príspevky mikroblogu, aktualizácie statusu alebo tweety – termín označujúci populárnu platformu mikroblogov, Twitter. Správy z mikroblogu zobrazujú širokú škálu obsahu, rovnako ako všetko ostatné na webe. Väčšina správ zodpovedá buď položkám konverzácie, ktoré sú cenné pre používateľa a jeho bezprostredný okruh priateľov (napr. aktualizácie o osobných aktivitách alebo mieste pobytu, klebety, chatovanie,...) alebo informáciám, novinkám, ktoré majú potenciál byť cenné pre širšiu komunitu (napr. oznámenia alebo komentáre k aktuálnych témam všeobecného záujmu). Okrem pomoci pri každodennej komunikácii o dôležitých udalostiach môže byť mikroblogovanie obzvlášť užitočné počas mimoriadnych udalostí a/alebo krízových situácií. Podrobná analýza tejto štúdie ukázala, že dôveryhodné tweety zvyčajne zahŕňajú odkazy na URL adresy, ktoré patria medzi 10000 najnavštevovanejších domén na webe. Všeobecne platí, že dôveryhodné tweety majú tendenciu obsahovať viac URL adries a sú dlhšie ako nedôveryhodné tweety. Nedôveryhodné tweety patria medzi tie pozitívnejšie, na rozdiel od dôveryhodných tweetov, ktoré majú tendenciu vyjadrovať negatívne pocity. Niektoré veľmi zaujímavé fakty možno pozorovať, keď vezmeme do úvahy otázniky a výkričníky, ktoré sa skôr objavujú pri nedôveryhodných tweetoch, ktoré sú písané v prvej a tretej osobe. V závere tejto štúdie je spomenuté, že do veľkej miery sa ignorovala minulá história používateľov, či už šlo o takých, ktorí si v minulosti získali dobrú povesť alebo tých, ktorí v minulosti šírili dezinformácie alebo spam, preto by bolo dobré tento faktor do budúcej štúdie taktiež zahrnúť.

III. ANALÝZA TEXTU

Analýza textu sa dá realizovať pomocou metód strojového učenia alebo slovníkovým prístupom. V tejto práci sme sa zamerali na slovníkový prístup.

A. Slovníkový prístup

Tak ako pri bežnom vnímaní textu a diskusie, tak aj pri automatickej klasifikácii sa výsledná polarita názoru odvádza od polarity slov. Pomocou známej polarity slov je možné po

implementácii správnej metódy určiť polaritu a presnú hodnotu polarity celého textu. Základné úlohy klasifikácie názorov podľa [5] je možné zhrnúť do troch nasledujúcich bodov.

- Výber slov z analyzovaného textu vyjadrujúcich postoj, alebo kladný, či záporný význam.
- Triedenie slov do skupín podľa polarity, sily polarity, negácie, alebo intenzifikácie.
- Výsledný výpočet hodnoty polarity celého textu.

Vo všeobecnosti platí, že každý názor, alebo príspevok, musí patriť konkrétnemu autorovi, ktorý tento názor vyjadril, tzv. držiteľ názoru. Dôveryhodnosť autora má takisto veľký vplyv na výslednú hodnotu polarity príspevku.

Kľúčovou vecou pre úspešnú analýzu textu slovníkovým prístupom je vytvorenie slovníka slov typických pre danú doménu a riešený problém, čo bola detekcia hoaxov, a potom ho presne anotovať, čo je podrobnejšie opísané v článku [5].

Tvorba slovníka bola časovo náročná, pretože slovník bol vytváraný manuálne v slovenskom jazyku, teda bolo potrebné prekladať anglické slová a určovať daným slovám ich váhu. Váha slov je v rozmedzí od 1 do 5, kde platí, čím je vyššie číslo, tým je slovo viac toxické. Slovník je zostavený z kombinácií rôznych slovníkov, ktoré obsahujú negatívne a vulgárne slová, ich rôzne tvary a taktiež aj ich synonymá. Okrem toho boli do slovníka pridané ešte bežne používané skratky, ktoré majú negatívny význam. Ďalej bol tento slovník obohatený o frázy ako: „je to iba môj názor“, „len som premýšľal“, „iba som sa zamyslel“ a iné, ktoré boli označené na facebookovej stránke Hoaxy a podvody – Polícia SR za používanú štandardnú komunikačnú techniku pre šírenie hoaxov. Taktiež odvolávanie sa na (vieru v) Boha je typické pre tvorcov dezinformácií. Aj prostredníctvom takýchto metód sú potom šírené prokremeľské naratívy, ktorých cieľom je zničenie solidarity celého sveta voči Ukrajine. Finálny slovník sa skladá z 1240 slov a fráz, a je vytvorený v textovom dokumente. Spôsob vytvárania slovníka, teda jeho tvar je zameraný na jeho využitie v programovacom jazyku Python. Ukážka vytvoreného slovníka je na Obr. 1. Na tvorbu slovníka bol použitý WADER lexikón obsahujúci viac ako 7000 slov. Synonymá a frázy boli brané zo zdroja „frazy.sk“.

```
otrasne: 4, obava: 3, obavy: 3, bojazlivo: 3, chapavy: 3, tajomny: 3, namahave: 3
namahavo: 3, arogancia: 3, arogantny: 3, arogantne: 3, sialene: 4, sialeny: 4
uskok: 3, nadavat: 4, vycitky: 3, napadnut: 4, vrah: 5, zavrazdit: 5, napadnutie: 4
napadat: 4, bludit: 1, rozdelovat: 1, ukrutny: 4, krutosti: 4, zverstvo: 4, utok: 4
utokov: 4, samovladca: 3, opito: 4, hrabivost: 3, lakomy: 3, lakomo: 3, lakomstvo: 3
hrozne: 4, strasne: 4, prisernost: 4, trapne: 2, nesikovnost: 1, ohovaranie: 2
zly: 3, zlo: 3, navnada: 3, banalne: 1, vyhnat: 2, vyhnanie: 2, bankrot: 2, barbar: 3
barbarsky: 3, barbarstvo: 3, prekliatie: 4, bity: 3, bitie: 3, bastard: 5, bastardi: 5
zobrak: 4, oklamat: 3, zradit: 3, zrada: 3, zrady: 3, zradca: 3, hasterenie: 2, mrcha: 5
bizarne: 3, vinny: 3, vydieranie: 4, blazon: 3, krvacat: 4, krv: 4, kravaca: 4
krvacajuca: 4, zablokovanie: 3, krviprelievanie: 4, krvilacny: 4, krvavy: 4, omyl: 3
omylom: 3, omyly: 3, tupy: 3, falosny: 3, bomba: 5, bombardovat: 5, bombardovanie: 5
otroctvo: 5, obtazovat: 3, obtazujuce: 3, drzy: 3, drzo: 3, drzost: 3, zlomit: 3
podplacanie: 4, zlomil: 3, brutalny: 5, brutalita: 5, brutality: 5, brutalne: 5
odpocuvanie: 3, tyranie: 5, tyran: 5, sikana: 5, sikanovat: 5, sikanovanie: 5, horiet: 3
spaleny: 3, palenie: 3, popaleniny: 3, kalamita: 4, kalamity: 4, katastrofa: 3
katastrofalne: 3, rakovina: 5, kanibal: 5, masaker: 4, odsudit: 4, odsudeny: 4, obet: 4
katastrofy: 3, chaoticky: 2, podvod: 3, podvadzat: 3, odvedeny: 3, podvodnik: 3
detinsky: 1, chladny: 2, dusit: 3, krik: 3, hlucny: 2, hruby: 3, namysleny: 3, donutit: 3
nutit: 3, natlak: 3, chladne: 2, kolaps: 3, odsudit: 3, odsudeny: 3, odsudzuje: 3
odsudenie: 3, priznanie: 3, priznania: 3, konflikt: 3, konflikty: 3, konfliktne: 3
konfrontacny: 3, konfrontacia: 3, konfrontuje: 3, sprisahanie: 4, sprisahania: 4
sprisahanecky: 4, konspirator: 4, konspiracie: 4, konspirovat: 4, kontaminovat: 3
```

Obr. 1 Ilustrácia vytvoreného slovníka pre účely odhaľovania hoaxov

B. Vyhodnotenie modelu založenom na slovníkovom prístupe

Vytvorený slovník bol použitý pri tvorbe modelu na rozpoznávanie hoaxov. Model bol testovaný na toxických a netoxických správach a výsledky testov sú uvedené v Tabuľke 1. Toxické správy, ktorých počet bol 35 boli získané z facebookovej stránky Hoaxy a podvody – Polícia SR. Netoxické správy – tiež v počte 35 boli získané z rôznych stránok a novín ako napríklad https://www.aktuality.sk/, https://www.sme.sk/a iné. Miera toxicity textu bola určovaná ako celé číslo od 1 do 5. Lepšie výsledky boli dosiahnuté v prípade keď stredná hodnota 3 bola považovaná tiež za toxickú. Tento model sa stal súčasťou webového portálu, čo ilustruje Obr. 2.

Tabuľka 1
Výsledky experimentov s modelom na detekciu nedôveryhodných postov založenom na slovníkovom prístupe.

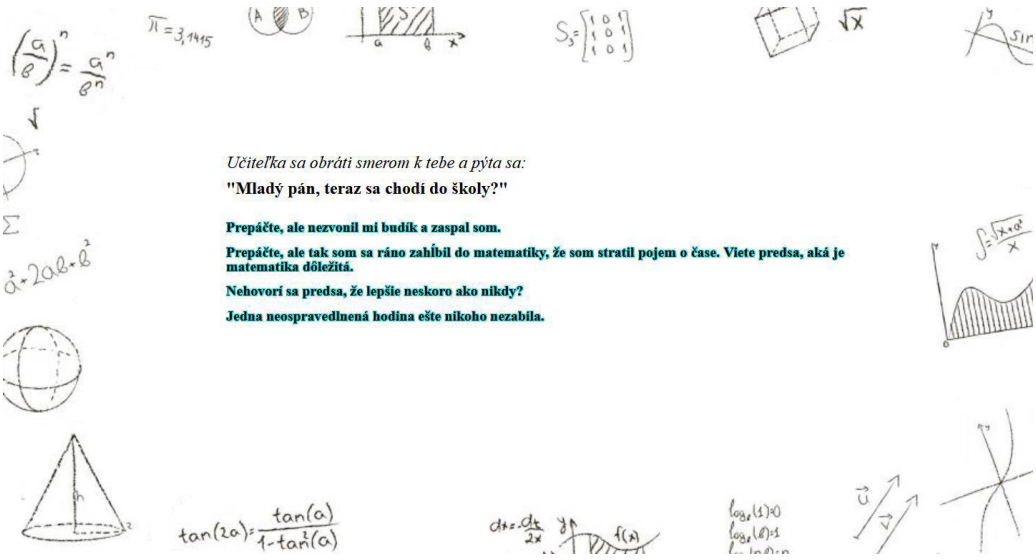
	Hodnota 3 netoxická	**Hodnota 3 ako toxická**
Presnosť	0.800	1.000
Návratnosť	0.933	0.814
Accuracy	0.871	0.886



Obr. 2 Ilustrácia modelu na odhaľovanie hoaxov, ktorý je súčasťou webového prtálu.

IV. WEBOVÝ PORTÁL

Model založený na slovníkovom prístupe bol aplikovaný na Google Cloud Platform, ako aj Model.exe, ktorý je dostupný na webovom portáli: https://oversi.onrender.com/index.html. Okrem tohto modelu obsahuje webový portál aj interaktívnu hru, ktorá bola vyvinutá v nástroji Twine. Po začatí hry sa hráč nachádza v izbe ráno po prebudení. Príbeh sa uberá takým smerom ako sa hráč rozhodne, teda podľa toho akú odpoveď z ponúknutých odpovedí si hráč vyberie. Hra je vytvorená tak aby bola zábavná pre tínedžerov. Začína oknom „izba po prebudení“ pokračuje oknom „škola s meškaním“ (môžeme to vidieť na Obr. 3.



Obr. 3 Ilustrácia interaktívnej hry.

Po hravom začiatku nasleduje test znalostí o antisociálnom chovaní. Každá odpoveď je ošetrená informáciou, či bola alebo nebola správna a prečo. Hra bola vyhodnotená pomocou dotazníka a to ukázalo, či v súvislosti s overovaním informácií, ktoré nájdu na internete študenti najčastejšie odpovedali „Skôr nie“, pri otázke, či nadobudli po odohraní hry nové poznatky bola typická odpoveď “Skôr áno”, až 69 % študentov (40% Skôr áno + 29 % Áno) si po odohraní hry bude viac všímať z akých zdrojov pochádzajú správy/príspevky, ktoré nájdu na internete, a čo sa týka celkového hodnotenia portálu, 71 % študentov (31% Skôr áno + 40 % Áno) odpovedalo, že im daný webový portál pomohol sa zorientovať v problematike dezinformácií.

V. ZÁVER

Tento článok bol zameraný na vytvorenie webového portálu, ktorého súčasťou je predovšetkým model slovníkového prístupu na analýzu dôveryhodnosti sociálneho webu, ale taktiež aj náučná interaktívna hra, ktorá je zameraná na dezinformácie. Hlavným cieľom pri tvorení modelu bolo generovanie slovníka, ktorý musel byť dostatočne reprezentatívny, aby daný model bol čo najpresnejší. Presnosť nášho modelu bola najvyššia, keď sme pri hodnote 3 považovali text za toxický. Avšak vyššiu návratnosť sme dosiahli, keď sme pri hodnote 3 považovali text za netoxický. Nevýhodou slovníkového prístupu je ale to, že nemusí obsiahnuť všetky potrebné slová, a teda, ak aj je text toxický, ale dané slovo/slová sa nenachádzajú v slovníku, nevie model tento text dobre vyhodnotiť. Taktiež môže text často obsahovať sarkazmus, ktorý je skutočne ťažké automaticky odhaliť a môže skresliť výsledky modelu.

Pokiaľ ide o interaktívnu hru, tak výsledky jej hodnotenia poukazujú, že bola úspešná, a teda dosiahla to, že 69 % študentov si po odohraní hry bude viac všímať z akých zdrojov pochádzajú správy/príspevky, ktoré nájdu na internete. Takisto aj väčšina študentov uviedla, že po odohraní hry nadobudli nové poznatky. Čo sa týka celkového vyhodnotenia webového portálu, tak ten mal takisto úspech, pretože až 71 % študentov uviedlo, že im daný webový portál pomohol zorientovať sa v problematike dezinformácií.

POĎAKOVANIE

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0685/21 a Agentúry pre podporu vedy a výskumu v rámci projektu APVV-22-0414.

REFERENCIE

[1] Pop, I.M., Pelau, C. Stanescu, M. „Reliability of social media platforms and online news as source of information for consumers. Conference“. In: BASIQ-New Trends in Sustainable Business and Consumption, Italy June 2019, 711-717 .

[2] Westerman, D., Spence, P.R., Heide, B. „Social Media as Information Source: Recency of Updates and Credibility of Information“. Journal of Computer-Mediated Communication 19, (2014) 171–183, DOI:10.1111/jcc4.12041 [cit 2021-05-05].

[3] Akamine, S., Kato, Y., Inui, K., Kurohashi, S. „Using Appearance Information for Web Information Credibility Analysis“. 2008 IEEE, DOI:10.1109/ISUC.2008.80 [cit 2024-05-12]

[4] Castillo, C., Mendoza, M., Poblete, B. „Predicting information credibility in time“. Sensitive socia media, Vol. 23, no. 5, 2012, pp. 560-588, DOI:10.1108/IntR-05-2012-0095 [cit 2021-06-29].

[5] K. Machova, K., M. Mikula, X. Gao, X. and M. Mach “Lexicon-based Sentiment Analysis Using the Particle Swarm Optimization”. Electronics, Vol. 9, no.1, 2020, ps. 1317.

Web aplikácia pre orientáciu v dezinformačnom obsahu

[1]*Kristína MACHOVÁ,* [2]*Jakub ŠIMOŇÁK*

[1] Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita Košice, Slovenská republika

[1]kristina.machova@tuke.sk, [1]jakub.simonak@student.tuke.sk

***Abstrakt* — Článok sa zameriava na pomoc používateľom internetu, aby sa dokázali lepšie orientovať vo veľkom množstve informácií, ktoré tam môžu nájsť. Cieľom je podporiť to, aby sa užívatelia naučili rozlišovať pravdivé informácie od tých, ktoré sú nejakým spôsobom klamlivé, podvodné či zavadzajúce. Článok popisuje základné pojmy tykajúce sa škodlivých informácií, ako aj ich delenie. Ďalej sú predstavené modely strojového učenia, ktoré sa často používajú na spracovanie textov. Článok taktiež prezentuje vytvorený webový portál, ktorého účelom je edukácia mladých ľudí v oblasti dezinformácií a ich šírenia. Modely strojového učenia boli testované na dvoch datasetoch, nevyváženom a vyváženom. Čo sa týka úspešnosti tohto testovania, vyššia bola dosiahnutá pri vyváženom datasete. Webový portál bol odprezentovaný študentom na zvolenom gymnáziu a jeho vyhodnotenie bolo uskutočnené prostredníctvom dotazníka.**

***Kľúčové slová* — dezinformácie v online priestore, strojové učenie, dôveryhodnosť webu, webový portál**

I. Úvod

Pravdivosť informácií je v dnešnej dobe často predmetom mnohých diskusií. Mnohé aktuálne témy často polarizujú spoločnosť. V mnohých prípadoch je veľmi zložité dopátrať sa k úplným, jednoznačným a pravdivým informáciám. Významný vplyv na tento stav má internet. Pomocou internetu je možne rozširovať akékoľvek informácie, ktoré sa vo veľmi krátkom čase môžu rozšíriť k veľkému množstvu ľudí. V dnešnej dobe človek počas dňa príde do kontaktu s množstvom informácii a preto nie je možne každú z nich overovať. Hoci slova ako hoax, trollovia či konšpiračne teórie patria k pomerne novým pojmom, so šírením dezinformácii sa ľudstvo stretáva už veľmi dlhu dobu. Článok sa zameriava na dve oblasti, webový portál a aplikáciu modelov strojového učenia na tvorbu detekčného modelu. .

Pre účely tejto prace bol vytvorený webový portál, ktorý poskytuje všetky podstatne informácie o problematike na jednom mieste a poskytuje návštevníkom užitočné rady ako odhaliť nepravdivý, podvodný či zavadzajúci text. Webový portál je obohatený o videonahrávky odborníkov z rôznych oblasti sociálneho života. Jednotliví experti v daných videách opisujú vplyv dezinformácii na ich odbornú oblasť. V snahe zvýšiť povedomie o tejto problematike sa uskutočnila aj edukácia na vybranom gymnáziu.

Modely strojového učenia boli aplikované na výroky politikov a političiek. Jednotlivé výroky boli zozbierané z webovej stránky Demagog.sk. Komunita, ktorá stojí za touto stránkou vyhodnocuje jednotlivé výroky politikov a političiek, ktoré zazneli v politických reláciách.

II. Online toxicita

A. *Základné typy*

Rozpoznávame tri základné typy škodlivých informácií:

- Dezinformácie sú falošné alebo nepresné informácie poskytované s cieľom zavádzať alebo klamať daného používateľa [1]. Patria tam:
    - Fake news – sú to spravodajské články, ktoré sú preukázateľne nepravdivé a majú za úlohu zavadzať používateľov.
    - Hoax – v origináli pochádza tento výraz zo slova „ho-kus“, ktoré môžeme poznať z kúzelníckych vystúpení, a znamená podviesť alebo klamať. Hoax je správa, ktorá je šírená medzi veľkým množstvom ľudí za účelom manipulácia a zmeny názoru na určité udalosti, skutočnosti okolo nás.
    - Konšpirácie – sú to teórie, ktoré popierajú určitú časť danej skutočnosti a nahrádzajú ju vlastným – možným vysvetlením. Vo všeobecnosti sú tieto správy vo veľkej miere spoločnosťou ignorovane.

- Misinformácie – sú nepravdivé informácie zdieľane osobou, ktorá dané informácie nepovažuje za klamlivé alebo zavadzajúce.
    - Klebety – sú príbehy, ktoré sú zvyčajne šírene medzi ľuďmi o iných ľuďoch, ktorí sa ocitli v nejednoznačných situáciách, kde pravdivosť daných príbehov je zvyčajne neoverená až pochybná.
- Malinformacie – sú pravdivé informácie, ktoré sú zdieľané za účelom uškodiť jednotlivcovi alebo skupine.
    - Spamy – sú nevyžiadané správy zasielané veľkému počtu príjemcov, ktoré často krát obsahujú irelevantne informácie.

B. Škodlivé účty na sociálnych sieťach

V dnešnej dobe má viacero účtov na sociálnych sieťach pochybnú podobu a za mnohými prípadmi nestoja skutoční ľudia. Jednoduchý prístup k vytváraniu účtov na sociálnych sieťach negatívne prispieva k tvorbe škodlivých účtov ako sú sociálni boti, kyborgovia alebo trollovia [2].

Sociálny bot označuje účet na sociálnej sieti, ktorý je riadený počítačovým algoritmom na automatickú tvorbu obsahu a interakciu s ľuďmi (alebo inými botmi). Slúži hlavne na manipuláciu verejnej mienky a šírenie falošných správ. Mnohé štúdie poukazujú na fakt, že v roku 2016 sociálni boti narušili online diskusie o prezidentských voľbách, kde ich v týždni pred samotnými voľbami bolo aktívnych takmer 19 miliónov, pričom tweetovali či už na podporu Trumpa alebo Clintonovej.

Trollovia na sociálnych sieťach sú často krát označovaní ako skupina účtov, za ktorými stoja skutoční používatelia, ktorých cieľom je narušiť online konverzácie za účelom vyprovokovať určitú stranu k emočnej reakcii na danú skutočnosť. Dôkazy naznačujú, že približne 1000 platených ruských trollov šírilo falošne správy o Hillary Clintonovej. Správanie trollov vie do veľkej miery ovplyvniť nálady ľudí a kontext online diskusii, čo umožňuje daným jedincom ľahké šírenie falošných správ medzi „normálnymi“ online komunitami. Hlavný efekt trollingu je vyvolanie negatívnych emócii ľudí, ako sú napríklad hnev, nenávisť a strach, ktoré vedú k pochybnostiam, nedôvere a v konečnom dôsledku k iracionálnemu správaniu.

Špeciálnym typom škodlivých účtov na sociálnych sieťach sú kyborgovia, ktorí predstavujú skupinu účtov fungujúcich na princípe počítačového algoritmu s priamym vstupom človeka. Takéto úcty sú zvyčajne registrované ako bežní legitímni používatelia, čo im slúži na maskovanie ich nekalých praktik a zároveň sú zahrnuté aj automatizovane programy na vykonávanie určitých akcii na sociálnych sieťach. Ľahké prepínanie medzi botom a ľudským prístupom dáva tomuto typu škodlivého účtu veľkú výhodu pri zdieľaní falošných správ [3].

III. MODELY NA ROZPOZNÁVANIE TOXICITY

A. Dáta

Dáta bolo potrebne manuálne extrahovať z webovej stránky Demagog.sk. Táto stránka vychádza z projektu Inštitútu pre dobre spravovanú spoločnosť. Komunita, ktorá stoji za týmto webovým portálom kontroluje jednotlivé výroky politikov, političiek a iných verejne aktívnych osôb. Princíp overovania vyjadrení prebieha na základe hodnotenia overiteľných tvrdení politikov, ktoré sú porovnávané s faktami v rôznych podobách:

- numerické informácie,
- minulé činy osôb (účasť pri príprave zákonov…),
- historické udalosti.

Tieto fakty je možné dohľadať vo verejne dostupných databázach či tlačových správach štátnych, mimovládnych ale aj súkromných inštitúcii. Tieto dáta pochádzajú z televíznych a rozhlasových diskusných relácii z obdobia rokov 2020 až po súčasnosť. Dataset pozostáva z 1177 záznamov a 7 atribútov, z toho 5 kategorických a 2 numerických.

V prvotnej fáze predspracovania dát sme nastavili text výrokov na malé písmena a odstránili sme všetky nadbytočne špeciálne znaky. Za tým nasledovala lematizacia slov. Nakoľko v súčasnej dobe neexistuje automatizovaný lematizer pre slovensky jazyk, bolo potrebne vytvoriť algoritmus, ktorý by prehľadával jednotlivé slova vo výrokoch politikov a menil ich na základný tvar. Algoritmus využíva voľne dostupnú morfologickú databázu slovenského národného korpusu, ktorá obsahuje slova s priradením základného tvaru slova (lemy) ku každému konkrétnemu tvaru slova. V rámci tejto databázy sa tu vyskytuje približne 99 tisíc slov (lem) vybraných z dostupných slovníkov a doplnených o najfrekventovanejšie slová z korpusu. Tato databáza pozostávala z 3,2 milióna záznamov nakoľko bola obohatená aj o homonyma. Pre účely našej prace bolo potrebne odstrániť duplikáty, aby po aplikovaní algoritmu sa pri prehľadávaní tohto morfologického slovníka vrátil nie viac ako jeden záznam. Po odstránení duplikátov sme dostali okolo 1,1 milióna unikátnych slovných tvarov. Predtým ako tento algoritmus pridá do

zoznamu slovo upravene do základného tvaru, prehľadáva ďalší súbor, a to súbor s neplnovýznamovými slovami ako sú napríklad spojky, predložky a pod. Súbor pozostáva zo 418 slov a ak sa slovo vyskytuje v tomto zozname, tak algoritmus nepridá dane slovo, teda ho vynechá a pokračuje s kontrolou ďalšieho slova z výroku v rámci datasetu. Po aplikovaní lematizácie a odstránenia neplnovýznamových slov bolo potrebne odstrániť "šum" v podobe diakritiky. V závere nášho predspracovania sme sa riadili ideou, ktorá považuje každý výrok, ktorý nie je ohodnotený ako pravdivý, že je prinajmenšom podozrivý a teda spojili sme nepravdivé, zavádzajúce a neoveriteľne výroky do kategórie "podozrivé". Pre potreby modelovania sme naše výroky vektorizovali a vytvorili polia výrokov, ktoré priamo vstupujú do jednotlivých modelov.

B. Modelovanie

V rámci nášho modelovania detekcie toxicity sme pracovali s 5 algoritmami strojového učenia na 2 datasetmi: BuzzFeed (7406 falošných a 7316 pravdivých správ) a PolitiFact (18899 falošných a 4437 pravdivých správ). Pri prvom sme zachovali pôvodný počet záznamov a v druhom sme zduplikovali niektoré výroky v snahe docieliť vyváženú množinu údajov. Na trénovanie sme použili nasledovné algoritmy, ktorých podrobnejší popis je možné nájsť v [4]:

- *MultinomialNB* - Multinomiálny Naivný Bayes klasifikátor
- *DecisionTreeClassifier* - Rozhodovacie stromy s parametrom criterion = 'entropy' a random_state = 0
- *LogisticRegression* - Logistická Regresia s parametrom random_state = 0
- *RandomForestClassifier* - Random Forest s parametrami n_estimators = 10, criterion ='entropy' a random_state = 0
- *SVC - Support Vector Machine* s parametrami C=1.0, kernel='linear', degree=3 a gamma='auto'

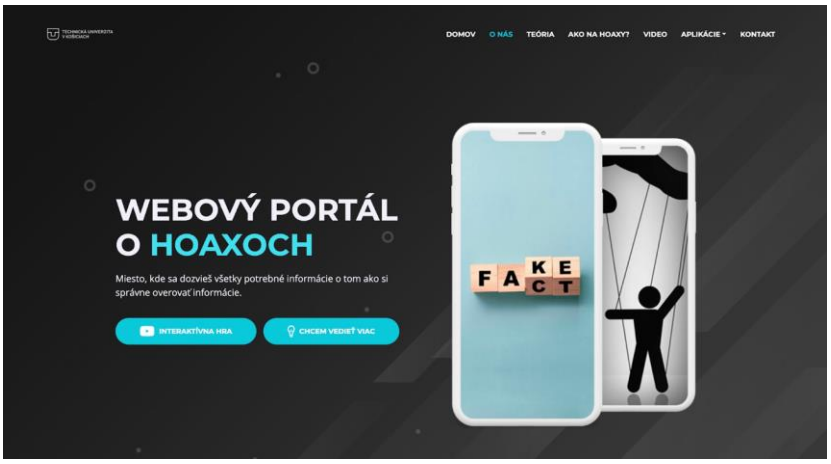
Pracovali sme s nevyváženým datasetom, kde bolo 69,07% pravdivých a 30,93% podozrivých výrokov a s vyváženým datasetom, kde bolo 50,03% pravdivých a 49,97% podozrivých výrokov. V tomto prípade sa nám v porovnaní predchádzajúcimi experimentmi presnosť prekvapivo zhoršila až na dva modely, kde sa naopak zlepšila. Na tomto datasete najlepšie obstal algoritmus Decision Tree s presnosťou 81,85%. Výsledky experimenrov sú uvedené v Tabuľke 1.

Tabuľka 1
Výsledky experimentov s učením pomocou 5 uvedených algoritmov strojového učenia na nevyváženej a vyváženej dátovej množine.

	PRESNOSŤ MODELU		
Modely strojového učenia	Nevyvážený dataset	Vyvážený dataset	Δ
Multinomiálny Naive Bayes	72,03%	53,85%	▼
Decision Tree	60,59%	81,85%	▲
Logistická regresia	72,03%	54,46%	▼
Random Forest	61,86%	75,38%	▲
Support Vector Machine	72,03%	51,38%	▼

IV. WEBOVÝ PORTÁL

Bola vytvorená webová stránka, ktorej súčasťou je aj najefektívnejší model strojového učenia. Bola vytvorená aj za účelom poskytnutia bližších informácii o pojmoch v oblasti antisociálneho chovania na sociálnych sieťach a o tom ako si správne overovať informácie v online priestore. Predtým ako sme začali vytvárať samotný webový portál bolo potrebne zozbierať tie najpodstatnejšie informácie z problematiky šírenia dezinformácii na internete. Pri budovaní webovej stránky bola použitá voľne dostupná šablóna z internetu, ktorá pozostávala z viacerých .css, .js a .html suborov. Tieto boli následne upravené na základe vlastného návrhu rozloženia webového portálu. V rámci .css suborov boli pozmenené farby a odsadenia jednotlivých častí. Boli vytvorené viaceré podstránky pre potreby lepšej orientácie, ktoré reprezentujú jednotlivé sekcie portálu. Nasadenie bolo realizované prostredníctvom platformy Render, ktorá okrem iného dokáže čítať súbory uložene na GitHube, kde boli uložene podklady a súbory k aplikácii. Úvodná stránka webového portálu (https://oversi.onrender.com/index.html) je uvedená na Obr. 1.



Obr. 1 Úvodná stránka webového portálu.

Webový portál pozostáva z viacerých sekcii. Sekcia „Teória“ obsahuje základné pojmy oblasti a ich vysvetlenie. Ide o dezinformácie (hoax, fake news, konšpiračná teória, deep fake), malinformácie (hate speech, harasment), misinformácie (omyl, klebety), škodlivé účty (troll, sociálny bot, kyborg) a iné pojmy ako spam, influencer, clickbait, manipulácia a propaganda. V sekcii „Ako na hoaxy“ sú uvedené kroky ako postupovať pri overovaní informácií a odkazy na web stránky nápomocné pri hľadaní pravdy v online priestore. V sekcii „Video“ sú audiovizuálne nahrávky odborníkov z oblasti médií, histórie, psychológie, zdravotníctva, výchovy mládeže, ktorí sa vyjadrovali k antisociálnemu chovaniu zo svojho odborného pohľadu. Sekcia „Aplikácie“ obsahovala konkrétne modely na overenie textu alebo autora príspevku.

Správne rozloženie a esteticky design webovej stránky ešte nemusí priniesť úspech v podobe zodpovednosti ľudí nešíriť klamlivé informácie naprieč internetom, preto bolo potrebne overiť úspešnosť. Na hodnotenie sme použili formu dotazníka o aktuálnych vedomostiach z oblasti, ktorý študenti vypĺňali tak pred výučbou pomocou web stránky ako aj po jej ukončení. Študenti dosahovali úspešnosť v rámci vedomostného dotazníka 60,6% na začiatku a 86,3% po ukončení práce s webovým portálom, z čoho vyplýva nárast zo 6 skoro na 9 správne zodpovedaných otázok z celkového počtu 10. Viac ako dve tretiny študentov uviedlo, že im daný webový portál/prezentácia pomohli zorientovať sa v problematike dezinformácii.

V. Záver

V článku bol prezentovaný navrhnutý prístup ku obmedzeniu šírenia dezinformácii v online priestore pomocou štruktúrovanej webovej stránky, kde sa prechádza jednotlivými krokmi od tých najvšeobecnejších až po tie najšpecifickejšie informácie o probléme. V úvode sú to základné pojmy a teória šírenia dezinformácii, ktorá nás vovádza do centra problematiky. Preskúmali sme existujúce riešenia čo nám pomohlo vytvoriť aplikáciu hodnotenie používateľských profilov na socialnych sieťach a taktiež model strojového učenia pre automatizovanú detekciu falošných sprav na internete. V budúcnosti by sme radi napĺňali víziu organizovania workshopov na základných a stredných školách, ktoré by mohli viesť mládež ku kritickému mysleniu a zodpovednému overovaniu informácii v online priestore.

Poďakovanie

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0685/21 a Agentúry pre podporu vedy a výskumu v rámci projektu APVV-22-0414.

Referencie

[1] W. Liang, F. Morstatter, M.K. Carley, H. Liu „Misinformation insocial media: Definition, manipulation, and detection,“ ACM SIGKDD Explorations Newsletter, 21 (2), 2019, pp. 80–90.
[2] E. Ferara, et al. „The rise of social bots,“ Communications of the ACM, 59(7), 2016, pp. 96–104.
[3] CHU, Z., et al. „Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?“, IEEE Transactions on Dependable and Secure Computing, 9(6, 2012, pp. 811–824.
[4] H. Belyadi, A. Haghighat „Machine Learning Guide for Oil and Gas Using Python, Chapter 5 - Supervised learning, 2021, pp. 186-211.

Slovníkový prístup k detekcii falošných správ o COVID-19

[1]*Kristína MACHOVÁ,* [2]*Lucia GRODECKÁ*

[1] Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita Košice, Slovenská republika

[1]kristina.machova@tuke.sk, [1]lucia.grodecka@student.tuke.sk

***Abstrakt* — Článok je zameraný na nájdenie vhodného prístupu k detekcii poplašných a falošných správ o ochorení COVID-19. Výskum sa zameriava na slovníkové prístupy. Popisuje päť existujúcich prístupov k analýze textu. V rámci vývoja detekčného systému bol vytvorený slovník pre danú doménu. Slovník bol anotovaný a použitý v analýze textu algoritmom hľadania zhody so slovníkom. Tento implementovaný prístup bol testovaný a článok uvádza analýzu výsledkov prostredníctvom rôznych metrík.**

***Kľúčové slová* — falošné správy, slovníkový prístup, analýza online obsahu, COVID-19**

I. Úvod

Dnešná doba je veľmi ovplyvnená internetom. Falošné a poplašné články majú rôzny charakter a týkajú sa skoro všetkého čo je súčasťou nášho života. S celosvetovými pandémiami sa stretli hlavne staršie generácie a v roku 2020 aj naša generácia. Ľudia začali vymýšľať a písať články o ochorení COVID-19. Od vypuknutia pandémie, sa začali internetom šíriť správy, ktoré chceli a vyvolávali paniku, zmätok a nedôveru či už voči vláde alebo lekárom. Najbežnejším typom falošných správ sú teórie o neexistencii vírusu. Inde zase píšu, že vírus vytvorila vláda, ktorá mala v úmysle ovládať svojich občanov. Druhým typom falošných správ sú informácie o prenose vírusu, jeho liečbe a symptómoch. V kontexte toho sa ukazuje ako mimoriadne dôležité aby si čitateľ akéhokoľvek článku informácie v ňom overil. Existujú stránky, ktoré obsahujú len falošné články, no mnohí ľudia im uveria a šíria tieto nepravdivé informácie ďalej. Svetové organizácie sa od začiatku pandémie snažia šíreniu falošných a nepravdivých článkov zamedziť a oboznámiť ľudí s postupom overovania si informácií.

II. Falošné a poplašné správy o COVID -19

COVID-19 je ochorenie spôsobené vírusom SARS-CoV-2, ktorý bol prvý krát objavený v decembri 2019 v meste Wuhan a odvtedy sa svet zmenil. Ľudstvo sa rozdelilo na dve časti. Prvá verí, že COVID-19 existuje a sú ochotní očkovať sa a dodržiavať opatrenia a nariadenia vlády a svetových zdravotných organizácií. Druhá časť je naopak proti všetkému, neveria, že ochorenie existuje alebo odmietajú akékoľvek opatrenie proti zamedzeniu jeho šírenia. Poslednou veľkou pandémia vo svete bola Španielska chrípka, ktorá svet zasiahla v dobe prvej svetovej vojny [1].

Od začiatku pandémie bolo vytvorených a napísaných množstvo falošných a poplašných správ o ochorení COVID-19. Dôvodom rozšírenia falošných a poplašných správ o ochorení COVID-19 je aj nedostatočná informovanosť ľudí. Výskum zdravotnej negramotnosti opísaný v článku [2] spočíval v odhaľovaní falošných správ na Twitteri. Bol vytvorený z datasetu, ktorý obsahoval informácie z mesiacov január až 10 máj v rokoch 2019 a 2020. Vzorkou pre výskum bola skupina ľudí, ktorá sa otrávila po prečítaní článkov na Twitteri o požití dezinfekčných prostriedkov, tie mali dopomôcť k ochráneniu ľudského organizmu od COVID-19, prípadne k uzdraveniu.

Dnes už existujú nástroje na odhaľovanie falošných správ a manipulovaných obrázkov (ktoré často sprevádzajú poplašné správy), ako napr. tieto:

- *Claimbuster* - webový nástroj pre kontrolu článkov. Využíva spracovanie prirodzeného jazyka. Patrí medzi modely generované kontrolovaným učením. Nástroj nájde podobné alebo rovnaké články a vráti odpoveď o pravdivosti článku. Ak nástroj nevie nájsť podobné alebo rovnaké články aké používateľ zadá, vypýta si ďalšie informácie o článku, na databázu a na umiestenie článku. Claimbuster monitoruje najnovšie poznatky. Podľa algoritmu udeľuje článkom skóre medzi 0,0 a 1. Články so skóre bližším k 1 sú najviac pravdivé. Algoritmus na výpočet skóre je založený na klasifikačnom skórovacom modeli (trénovaný ľuďmi). Na základe faktov prehľadá archívne články

podľa tokenov alebo podľa sémantickej podobnosti. Algoritmus vráti podobný článok a skóre dokumentu [3].

- *TinEye* - nástroj na spätné vyhľadávanie obrázkov. Využíva programovacie rozhranie API MulticolorEngine. Vyhľadávanie obrázkov funguje na princípe analyzovania pixelov. Extrahuje farbu v jednotlivom pixeli, podobné farby sa klasifikujú a vráti sa hodnota RGB, poradie podobnosti, trieda farieb a váhový vektor. Váha farby určuje dôležitosť farby vzhľadom na ostatné farby na obrázku. Používateľovi umožňuje nahrať obrázok a vyhľadať jeho zmeny. Pomocou rozhrania API objaví odkiaľ obrázok pochádza, či existujú upravené podoby. Vyhľadávanie je možné uskutočniť pomocou presunutia obrázka na stránku alebo skopírovaním URL adresy obrázka [4].

III. Analýza textu slovníkovým prístupom

Analýza textu sa dá realizovať pomocou metód strojového učenia alebo slovníkovým prístupom. V tejto práci sme sa zamerali na slovníkový prístup.

Pri slovníkovom prístupe sa tak ako pri bežnom vnímaní textu a diskusie, aj pri automatickej klasifikácii výsledná polarita názoru odvádza od polarity slov. Pomocou známej polarity slov je možné po implementácii správnej metódy určiť polaritu a presnú hodnotu polarity celého textu v zmysle toho, či reprezentuje alebo nie falošnú správu. Základné úlohy klasifikácie názorov podľa [5] je možné zhrnúť do troch nasledujúcich bodov.

- Výber slov z analyzovaného textu vyjadrujúcich postoj, alebo kladný, či záporný význam.
- Triedenie slov do skupín podľa polarity, sily polarity, negácie, alebo intenzifikácie.
- Výsledný výpočet hodnoty polarity celého textu.

Kľúčovou vecou pre úspešnú analýzu textu slovníkovým prístupom je vytvorenie reprezentatívneho slovníka slov typických pre danú doménu a riešený problém, čo bola detekcia falošných správ, a potom ho presne anotovať, čo je podrobnejšie opísané v článku [5]. Na výber reprezentatívnych slov sa používajú váhovacie techniky. Existujú tri typy rozdelenia algoritmov na extrakciu slov:

- Metóda založená na jednoduchých štatistikách. Existuje niekoľko typov extrakcie slov. V procese extrakcie majú stále niečo spoločné. Sem patrí napríklad N-gram alebo váhovacia technika TF-IDF.
- Metóda založená na grafe. Slová sa berú ako uzly v grafe. Uzly čiže slová sa spoja podľa určitých pravidiel a vytvoria tzv. sémantický graf. Slová sú získané určením takzvanej „hmotnosti“ každého jedného uzla spolu s jeho klasifikáciou. K metóde patria algoritmy SingleRank, TextTank, PositionRank a podobne.
- Metóda založená na reprezentatívnom slove. K metóde patria algoritmy ako single-TPR alebo TopicRank [6].

My sme v našom prístupe použili váhovaciu techniku TF-IDF [7]. Ukážka časti slovníka je v Tabuľke 1. Ide o prvých desať najvyššie skórovaných slov.

Tabuľka 1
TF-IDF váha pre prvých 10 najvyššie skórovaných slov vo vytvorenom slovníku.

slovo	váha	slovo	váha
irony	0.266135	steroids	0.112285
idylla	0.158747	skype	0.108692
discard	0.154677	scrub	0.099450
prank	0.140834	dung	0.097297
believeth	0.139209	beget	0.092806

Slovník sme uložili ako .csv a nazvali ho „old_dictionary“ na rozlíšenie slovníka pred anotáciou a po anotácii. Prvým krokom anotácie bolo načítanie csv, s názvom „old_dictionary“, aby sme vedeli rozlíšiť to, že slovník s týmto názvom obsahuje len slová, ktoré sme vybrali do slovníka. Anotácia spočívala v priradení hodnoty z rovnakého rozsahu pre všetky slová, ktoré sa nachádzajú v slovníku. Hodnoty sme priradili kvôli neskoršiemu použitiu, sčítaniu a kvôli zisteniu, ktoré slová majú väčšiu váhu ako ostatné. Rozsah hodnôt, ktoré sme priradili slovám v slovníku boli od 1 po 5. Číslo 1 sme priradili poslednému slovu „secret“, keďže malo najmenšiu váhu TF-IDF z celkového počtu 1038 slov, ktoré boli extrahované z textov dátovej množiny

„dataset-covid19-fake-news“. Číslo 5 sme priradili prvému slovu a to slovu „irony“, keďže malo najväčšiu váhu TF-IDF a do slovníka sme ho zaradili na prvé miesto. Ostatným slovám sme priradili hodnotu väčšiu ako 1 a menšiu ako 5. Hodnoty boli rozpočítané podľa lineárnej funkcie. Čím vyššie skóre TF–IDF slovo má, tým vyššia hodnota.

Po ukončení procesu anotovania sme slovník s hodnotami uložili pre ďalšie použitie ako csv s názvom „dictionary.csv“. V uloženom súbore sú tri stĺpce. V prvom stĺpci s názvom „word“ sme uložili slová, ktoré sa nachádzajú v slovníku. V druhom stĺpci s názvom „score“ je vypísané skóre z priemerných váh TF-IDF. Váha pre každé slovo je uložená pre porovnanie. Tretí nový stĺpec s názvom „label“ obsahuje hodnotu anotovania. Na Obr. 1 je ilustrovaných prvých 5 slov so skóre (hodnota TF-IDF váhy) a hodnotou ako výsledku samotného anotovania.

	word	score	label
0	irony	0.266135	5.000000
1	idylla	0.158747	4.918367
2	discard	0.154677	4.836735
3	prank	0.140834	4.755102
4	believeth	0.139209	4.673469
5	steroids	0.112285	4.591837

Obr. 1 Ilustrácia prvých 5 slov slovníka so skóre (TF-IDF) a anotačnou hodnotou.

IV. Implementovanie a testovanie slovníkového prístupu

Prvým krokom pre testovanie implementáciu vytvoreného slovníka je načítanie knižníc, datasetov a predspracovanie dát. V rámci predspracovania bola vykonaná:

- Zmena na čistý text
- Tokenizácia a segmentácia
- Lematizácia a morfológia
- Vyradenie neplnovýznamových slov.

Po načítaní a predspracovaní dát sme načítali anotovaný slovník s názvom „dictionary.csv“. Následne sme články rozdelili v každom riadku na slová keďže potrebujeme vyhľadávať konkrétne slová zo slovníka. Pre lepšiu orientáciu sme vytvorili stĺpec, ktorý obsahoval slová v analyzovanom texte, ktoré sú totožné so slovami zo slovníka. Definovali sme novú funkciu s názvom „matching_word“, ktorá má ako vstup zoznam „words“, z dát „combined_train“. Táto funkcia kontroluje každý riadok, či sa niekde nachádza slovo zo zoznamu slov slovníka. Pri nájdenom slove sa slovo uloží do nového stĺpca s názvom „matching_words“. Ak sa slovo vyskytuje v jednom riadku viac krát pre naše neskoršie použitie sa to zaznamená, keďže táto informácia je dôležitá pre presnejšie určenie miery „falošnosti“ analyzovaného textu.

Definovali sme funkciu s názvom „sum_values“, ktorá funguje na princípe prijímania zoznamu slov ako vstup a vracia súčet hodnôt slov. Na zrátanie výsledných hodnôt sme použili súčet a to funkciu „sum()“. Vytvorenú funkciu sme aplikovali na dáta a pre stĺpec s názvom „matching_words“. Výslednú sumu hodnôt sme zapísali do nového stĺpca s názvom „sum_of_values“. Následne sme dáta uložili do súboru „final_version.csv“. Pre lepšiu orientáciu sme definovali hodnotu podľa ktorej budeme zaraďovať články k falošným alebo skutočným správam. Definovali sme funkciu s názvom „calc_credibility“, ktorá počíta hodnotu na základe čísla v stĺpci „sum_of_values“. Môžu nastať dve možnosti. Ak je v stĺpci „sum_of_values“:

1. hodnota 0 – real (pravdivý text) – v tomto prípade priradíme novú hodnotu do stĺpca „credibility_value“ a to číslo 0
2. hodnota väčšia ako 1 – false (nepravdivý text) – v tomto prípade priradíme novú hodnotu do stĺpca „credibility_value“ na základe podielu. Číslo v stĺpci „sum_od_values“ vydelíme počtom slov v riadku v ktorom ho delíme.

Ďalej sme klasifikáciu falošných článkov rozdelili na tri triedy a to little fake news, medium fake news, very fake news nasledovne:

- ak je hodnota vyššia ako 0 ale zároveň nižšia alebo rovná ako 2.3 – nastaví sa hodnota stĺpca „inclusion“ na „little fake news“
- ak je hodnota vyššia ako 2.3 ale zároveň nižšia alebo rovná ako 3.6 – nastaví sa hodnota stĺpca „inclusion“ na „medium fake news“
- ak je hodnota vyššia ako 3.6 – nastaví sa hodnota stĺpca „inclusion“ na „very fake news“

Implementovaný prístup založený na slovníku sme testovali 30 percentách textov dátovej množiny „dataset-covid19-fake-news“, pričom 70 percent bolo použitých na generovanie slovníka. Táto dátová množina obsahuje 3123 príspevkov. Testovacia množina bola vyvážená s rovnakým počtom falošných a pravdivých správ. Výsledky testovania sú uvedené v Tabuľke 2.

Tabuľka 2
Výsledok testovania slovníkového prístupu k detekcii falošných správ o COVID-19.

metriky	presnosť	návratnosť	správnosť	F1
výsledky	61,58	94,59	67,79	74,61

Výsledok presnosti je 61,58 % má nadpriemernú hodnotu. V našom prípade je presnosť dobrá, nakoľko hodnotíme vyskytujúce sa slová zo slovníka v článkoch. Nevieme s istotou povedať, či autor článku využije slová, ktoré sme zaradili do slovníka. Výsledok návratnosti je 94,59% čo považujeme za veľmi dobré. V našom prípade 94% predpovedí bolo správnych. Výsledok správnosti je 67,79% je nadpriemerná hodnota. Model je dobrý v identifikácií falošných správ ako naozaj falošných správ. Výsledok F1 je 74,61% čo hodnotíme pozitívne. Výsledný model je dobrý v oblasti identifikovania falošných správ a vyhýba sa chybnej klasifikácii pri určovaní pravdivých článkoch ako falošných.

V. Záver

Tento článok bol zameraný na experimentovanie so slovníkovým prístupom k analýze textu a jeho použitím pri detekcii falošných správ o COVID-19. Dosiahnuté výsledky sú sľubné ak berieme do úvahy, že vytvoriť slovník slov typický pre falošné správy je oveľa ťažšie ako vytvoriť slovník pozitívnych a negatívnych slov v analýze sentimentu. Najlepší výsledok sme dosiahli pri mierke Návratnosť, čo znamená, že náš prístup je dobrý v znižovaní chyby FN (false negative), to znamená že je málo fake news, ktoré sme nenašli.

Poďakovanie

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0685/21 a Agentúry pre podporu vedy a výskumu v rámci projektu APVV-22-0414.

Referencie

[1] O.D. Apuke, O. Bahiyah „Fake News and COVID-19: Modelling the Predictors of Fake News Sharing among Social Media Users,“ In: Telematics and Informatics, vol. 56, no. 101475, 2021, pp. 16, ISSN: 0736-5853, Doi: https://doi.org/10.1016/j.tele.2020.101475.

[2] K. Gupta, K. Potika „Fake News Analysis and Graph Classification on a COVID-19 Twitter Dataset,” In: 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService). 23-26.8.2021, Oxford, United Kingdom. IEEE, 2021, pp.60–68 s. ISBN: 978-1-6654-3484-3. Doi: https://doi.org/10.1109/BigDataService52369.2021.00013].

[3] A .F. Adrakatti, R. S. Wodeya „Search by Image: A Novel Approach to Content Based ImageRetrieval System,“ In: International Journal of Library Science, vol. 14, no.3, 2016, pp. 41–47, ISSN: 0975-7546.

[4] H. Naeemul, et. Al. „ClaimBuster: the first-ever end-to-end fact-checking system,“ In: Proceedings of the VLDB Endowment. 2017, pp. 1945-1948, ISSN: 2150-8097, Doi: 10.14778/3137765.3137815.

[5] K. Machova, K., M. Mikula, X. Gao, X. and M. Mach “Lexicon-based Sentiment Analysis Using the Particle Swarm Optimization”. Electronics, vol. 9, no.1, 2020, ps. 1317.

[6] Y. Li, H. Ning „Keywords extraction method based on two-way feature fusion,“ In: International Conference on Machine Learning and Intelligent Systems Engineering(MLISE), 5-7.8. 2022, Guangzhou, China. IEEE, 2022, pp. 127-130, ISBN: 978-1-6654-9247-8. Doi: 10.1109/MLISE57402.2022.00033.

[7] K. Machová „Machine Learning in the Information Processing Systems,“ ELFA s.r.o., 2009, Kosice, 85s, ISBN 978-80-8086-130-8.

Laboratórny model malej vodnej elektrárne

[1]*Daniela PERDUKOVÁ,* [2]*Marek FEDOR*

[1] [2]Department of Electrical Engineering and Mechatronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]daniela.perdukova@tuke.sk, [2]marek.fedor@tuke.sk

***Abstract* — Článok popisuje realizáciu laboratórneho modelu malej vodnej elektrárne (MVE) a verifikáciu vlastností jeho hydromechanického podsystému. Porovnanie dosiahnutých výsledkov potvrdzuje jeho správnosť a tým možnosť jeho využitia pre spoľahlivé a systematické projektovanie takýchto systémov a výskum zameraný predovšetkým na ich riadenie.**

***Keywords* — hydromechanická časť MVE, laboratórny model MVE, malá vodná elektráreň, simulačný model MVE**

I. Úvod

Malé vodné elektrárne (MVE-výkon do 10 MW) sú relatívne jednoduché, technicky nenáročné energetické stavby, ktorých význam pri výrobe elektrickej energie v posledných desaťročiach celosvetovo neustále narastá. Pre MVE sú charakteristické vysoké investičné a nízke prevádzkové náklady [1]. Napriek tomu sa s nimi spájajú významné environmentálne a sociálno-ekonomické vplyvy. Tieto sú závislé hlavne od umiestnenia a typu danej vodnej elektrárne. Všeobecne je potrebné zdôrazniť, že popri negatívnych vplyvoch ide aj o vplyvy vysoko pozitívne. Využívanie hydroenergetického potenciálu vodných tokov na výrobu elektrickej energie je vysoko pozitívnym globálnym, dlhodobým opatrením na znižovanie významných negatívnych vplyvov na životné prostredie a zdravie, pretože predstavuje environmentálne najprijateľnejší spôsob získavania elektrickej energie. Výroba elektrickej energie vo vodných elektrárňach zaťažuje životné prostredie neporovnateľne menej, než výroba na báze tradičných fosílnych palív.

Pre systematické a racionálne projektovanie malých vodných elektrární a vyšetrovanie ich dynamických vlastností bol navrhnutý a zrealizovaný laboratórny model, ktorý umožňuje efektívne (t.j. v krátkom čase, s vysokým stupňom realizačnej spoľahlivosti a minimom nákladov) vytvárať kvalitné realizačné projekty takýchto MVE.

II. Simulačný model malej vodnej elektrárne

Model malej vodnej elektrárne (MVE) je pomerne zložitý nelineárny dynamický systém, ktorý sa skladá zo štyroch podsystémov [2] (Obr.1).

Základným podsystémom celej MVE je „*Hydraulická turbína*“, v ktorej sa premieňa energia vody na mechanickú energiu. Pre riadenie tejto premeny je ako akčný člen určený podsystém s názvom „*Servopohon*“, ktorým je možné ovládať prietok vody cez turbínu, prípadne tlak v turbíne. Riadenie turbíny má v podstate dve základné funkcie:

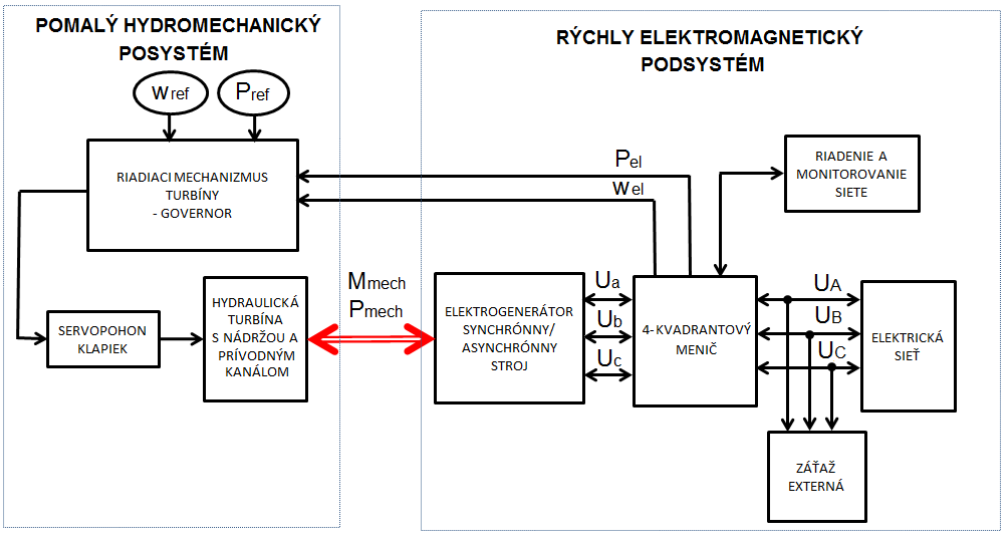
1. rozbeh turbíny na mechanické otáčky blízke elektrickej frekvencii siete,
2. po prifázovaní do siete riadenie dodávaného výkonu.

Toto zabezpečuje podsystém riadenia turbíny, označovaný v literatúre ako „*Governor*“ [3].

Premenu mechanickej energie turbíny na elektrickú zabezpečuje podsystém označený ako „*Elektrogenerátor*“, ktorým môže byť buď synchrónny alebo asynchrónny elektrický stroj. Pri mikroelektrárňach môže byť táto premena zabezpečovaná aj statickými meničmi s batériovým medziobvodom. Pri použití synchrónneho generátora je potrebné modelovať aj jeho budenie samostatným podsystémom.

Výstupom z elektrogenerátora je trojfázové elektrické napätie, ktoré je možné využívať priamo do vlastnej záťaže alebo cez blok „*Transformátora*“ je možné pripojiť MVE aj na externú elektrickú sieť s externou záťažou.

Úplný model MVE sa získa modelovaním jej jednotlivých podsystémov a ich vzájomným prepojením.



Obr. 1 Bloková schéma malej vodnej elektrárne

Každá malá vodná elektráreň je v princípe sústava dvoch základných podsysystémov [4]:

3. pomalého hydromechanického podsystému, ktorého hlavnou časťou je relatívne pomalá hydraulická turbína, ktorá slúži na premenu energie vody na mechanickú energiu;
4. a rýchleho elektromechanického podsystému s generátorom elektrickej energie, ktorý slúži na premenu mechanickej energie na elektrickú energiu vhodného typu, napr. trojfázovú striedavú elektrickú sieť.

Uvedené podsystémy sú mechanicky zviazané cez hriadeľ medzi turbínou a generátorom, ako je to ukázané na Obr.1 (červenou farbou).

Úplný simulačný model MVE bol navrhnutý a verifikovaný v programe MATLAB autormi článku v [4]. Získané výsledky budú použité pre verifikáciu laboratórneho modelu.

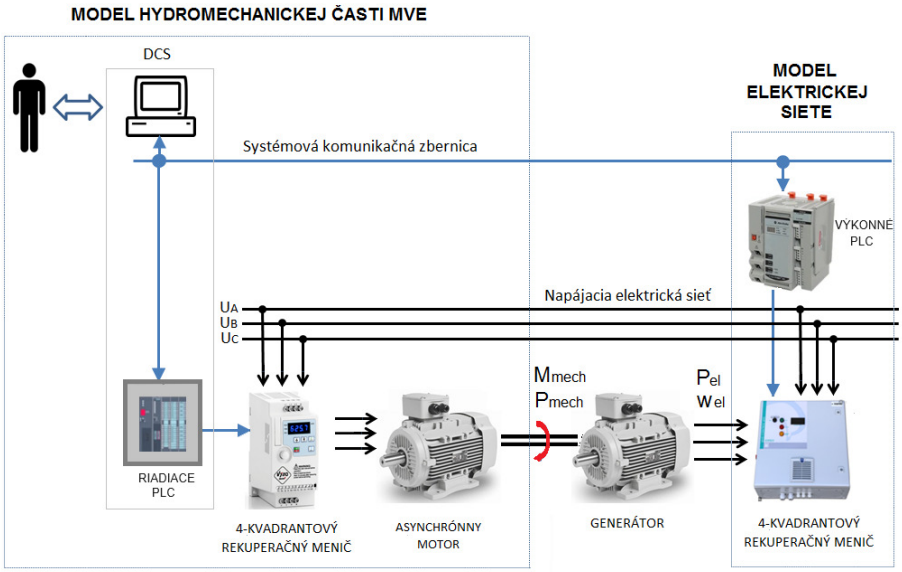
III. Laboratórny model hydromechanickej časti malej vodnej elektrárne

Jednotlivé hardvérové prvky laboratórneho modelu MVE boli definované na základe konzultácií s firmami zaoberajúcimi sa problematikou riadenia a prevádzkovania malých vodných elektrární.

Model pomalej hydromechanickej časti má výstup mechanický moment, resp, výkon na hriadeli turbíny. Všetky nelineárne zložité hydromechanické závislosti sú softvérovo modelované na štandardizovanom riadiacom systéme na báze PLC. Súčasne tento systém realizuje všetky štandardné funkcie pre riadenie technológií ako sú vizualizácia, styk s obsluhou, trendovanie, logovanie a spracovanie alarmov a pod. Pretože výstupom reálnej turbíny je mechanický moment na hriadeli, PLC generuje na základe svojho modelu žiadaný moment pre kvalitný dynamicky náročný pohon so štvorkvadrantovým meničom, ktorého motor emuluje moment hydraulickej turbíny (výkon).

Prechodové deje v elektrickej sieti sú podstatne rýchlejšie, preto sú emulované reálnym elektrogenerátorom (synchrónnym alebo asynchrónnym strojom), pripojeným cez štvorkvadrantový rekuperačný menič na reálnu 3-fázovú sieť. Riadenie a monitorovanie činnosti elektrogenerátora a výkonové toky medzi elektrickou sieťou a elektrogenerátorom realizuje vysoko výkonné štandardizované PLC s dobou vzorkovania 125 mikrosekúnd. Riadiace systémy na báze PLC vzájomne komunikujú prostredníctvom systémovej komunikačnej zbernice. Rekuperačné meniče oboch podsystémov umožňujú emulovať MVE s minimálnou spotrebou energie z externej siete.

Úplná štruktúra a koncepcia laboratórneho modelu pre modelovanie MVE je zobrazená na Obr.2.



Obr. 2 Koncepcia laboratórneho modelu MVE

Konkrétne hardwarové vybavenie laboratórneho modelu hydromechanickej časti MVE spolu s generátorom asynchrónneho typu je uvedené v Tab.1.

Table1
Špecifikácia zariadení pre laboratórny model

Pozícia	Zvolený typ
vizualizačný panel	*B&R PowerPanel PP500 series*
PLC	*B&R X20 series*
riadený menič napätia	*B&R ACOPOSinverter 8I66 series*
riadený asynchrónny motor	*ABB 3GAA101520-ADE662*
asynchrónny generátor	*ABB 3GA101520-ADE*
typ spojenia motor - generátor	*pevný spoj – spoločný hriadeľ*

Riadiaci systém na báze PLC zabezpečuje chod riadeného asynchrónneho motora tak, aby priebeh jeho otáčok, priamo zodpovedal správaniu sa simulovanej MVE.

K tvorbe riadiaceho programu bol použitý softwarový nástroj spoločnosti B&R – *Automation Studio V4.10.2.37* a simulačný model [4] s overenými parametrami získanými z meraní na reálnej MVE Dobšiná III. Riadenie asynchrónneho motora bolo realizované prostredníctvom inštalovaného riadeného meniča napätia *ACOPOSinverter*.

Riadiace PLC, riadený menič napätia, komunikačné karty a vizualizačný panel pre ovládanie laboratórneho modelu boli umiestnené do rozvádzača DT2.

Zrealizovaný laboratórny model hydromechanickej časti MVE je na Obr.3.



Obr. 3 Laboratórny model hydromechanickej časti malej vodnej elektrárne

IV. PREVÁDZKA LABORATÓRNEHO MODELU MVE

Základný princíp prevádzky navrhovaného laboratórneho modelu vychádzal zo simulácie jednotlivých prevádzkových stavov MVE s cieľom zabezpečiť odozvu, respektíve otáčky inštalovaného asynchrónneho motora nahradzujúceho správanie sa hydraulickej turbíny tak, aby zodpovedali simulovanému systému. V zmysle dostupných údajov poskytnutých prevádzkou MVE Dobšiná III. boli definované možné prevádzkové stavy malej vodnej elektrárne ako:

I. Stav, kedy MVE nie je v prevádzke a jej mechanické otáčky hriadeľa sú nulové
II. Stav, kedy MVE nie je v prevádzke, ale hriadeľ sa stále točí
III. Stav, kedy sa MVE rozbieha na želané otáčky a nie je pripojená k sieti
IV. Stav, kedy je MVE pripojená k sieti
V. Stav, kedy je MVE pripojená k sieti a dodáva výkon

Spôsob ovládania laboratórneho modelu vychádza z definície jeho prevádzkových stavov a bol realizovaný prostredníctvom HMI rozhrania a tlačidiel umiestnených na paneli rozvádzača DT2 (Obr.4). Pomocou tlačidiel 1, 2 a 3 je možné ovládať zapnutie, vypnutie , alebo úplné zastavenie laboratórneho modelu. Tlačidlo 1 je v laboratórnom modeli pripojené na digitálny vstup riadiaceho PLC a predstavuje vstupný signál od používateľa o požiadavke na štart/rozbeh MVE na sieťou definované otáčky. Stlačenie tohto tlačidla má za cieľ dostať laboratórny model zo stavu I., prostredníctvom stavu III., automaticky do stavu IV., na základe uvedeného popisu stavov. Ak užívateľ požaduje v čase prevádzky odstavenie malej vodnej elektrárne, model umožňuje dve alternatívy v závislosti od charakteru požiadavky používateľa:

1. Stlačenie tlačidla 2 - má za následok nastavenie referenčných hodnôt uhlovej rýchlosti a výkonu na nulové hodnoty, avšak s neustálym prepočtom aktuálnej uhlovej rýchlosti MVE tak, ako by to bolo v reálnej prevádzke za predpokladu nevyužitia externých bŕzd. Tento stav zodpovedá vyššie spomenutému prevádzkovému stavu II.
2. Stlačenie tlačidla 3 - umožňuje používateľovi absolútne zastavenie prevádzky modelu, respektíve simulácie s tým, že všetky premenné modelu sú nastavené na ich inicializačné hodnoty, čo uvádza model okamžite do stavu I.



Obr. 4 HMI rozhranie pre ovládanie laboratórneho modelu MVE

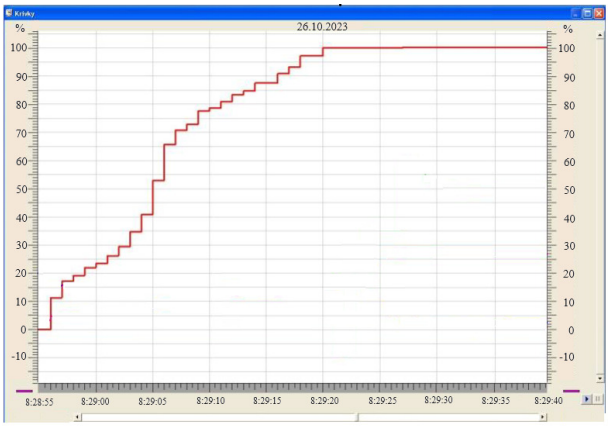
V. VERIFIKÁCIA HYDOMECHANICKEJ ČASTI LABORATÓRNEHO MODELU MVE

Správnosť laboratórneho modelu bola overená porovnaním priebehu mechanických uhlových otáčok ω_m turbíny počas jej rozbehu na otáčky siete získaného zo simulačného modelu [4] s parametrami z MVE Dobšiná III. s priebehom mechanických uhlových otáčok turbíny získaných priamo z laboratórneho modelu, kde boli zadané rovnaké parametre ako v simulačnom modeli.

V procese verifikácie bol laboratórny model uvedený do stavu I., t.j. MVE nie je v prevádzke a jej otáčky sú nulové. Zadaním hodnoty referenčnej uhlovej rýchlosti $\omega_{ref} = 0,6$; a stlačením Tlačidla 1 (ŠTART, Obr.4) sa turbína laboratórneho modelu rozbehla na definované otáčky.

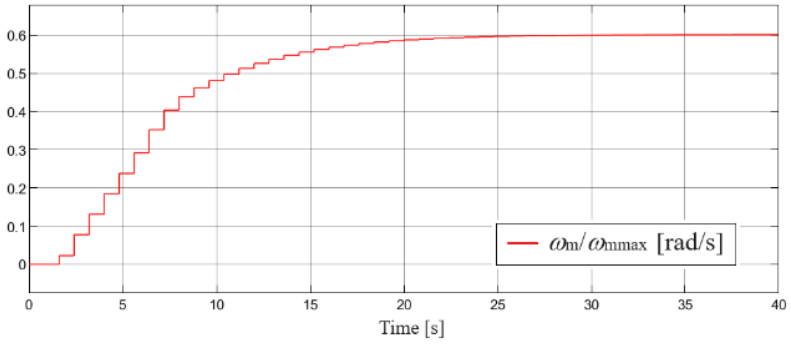
Na Obr. 5 je znázornený priebeh nábehu turbíny laboratórneho modelu na otáčky

odpovedajúce sieťovej frekvencii f_1=50 Hz (100%). Z grafu je vidieť, že turbína sa rozbehne na otáčky odpovedajúce sieťovej frekvencii za čas t=25 sekúnd, t. j. prechodový dej mechanickej uhlovej rýchlosti ω_m sa za daný čas ustáli.



Obr. 5 Priebeh mechanickej uhlovej rýchlosti turbíny pri nábehu laboratórneho modleu MVE.

Priebeh mechanických uhlových otáčok ω_m turbíny počas jej rozbehu k otáčkam siete získaným zo simulačného modelu v programe MATLAB/Simulink je znázornený na Obr. 6, kde vzhľadom na rozdielnu normalizáciu veličiny ω_m , jej hodnota 60 % zodpovedá hodnote 100 % na Obr. 5.



Obr. 6 Nábeh MVE na otáčky zodpovedajúce sieťovej frekvencii - 60 % (f_1 =50 Hz) získané z navrhnutého simulačného modelu MVE [4]

Z Obr. 5 a Obr. 6 vidíme, že dynamika prechodu mechanickej uhlovej rýchlosti ω_m turbíny počas jej rozbehu na sieťovú rýchlosť je rovnaká. To znamená, že turbína sa rozbehne na sieťové otáčky a je pripravená na pripojenie generátora k elektrickej sieti v čase t = 25 sekúnd. Týmto spôsobom sme overili presnosť hydromechanickej časti laboratórneho modelu, ktorá zahŕňa všetky vlastnosti jeho podsystémov, najmä samotnej turbíny, ako aj podsystému regulácie Governor (regulácia otáčok). Výsledné rozdiely pri prechode mechanických uhlových otáčok ω_m môžu byť spôsobené napríklad výstupom snímača otáčok laboratórneho modelu.

VI. ZÁVER

V článku je popísaný návrh a realizácia laboratórneho modelu MVE, ktorého hydromechanická časť bola verifikovaná na základe priebehu mechanických otáčok turbíny pri jej rozbehu na otáčky siete pre laboratórny a simulačný model.

Výhody takéhoto laboratórneho modelu pre konkrétneho používateľa definujú aplikačnú úroveň jeho nasadenia v praxi a sú nasledovné:

- výrazné skrátenie tvorby realizačných projektov,
- možnosti energetickej optimalizácie výkonových podsystémov a dynamickej optimalizácie riadiacich podsystémov energetických blokov,
- podstatné zvýšenie spoľahlivosti a kvality realizačných projektov,
- skrátenie doby realizačnej fázy energetických zariadení.
- podstatné zníženie nákladov na zmenové konania pri realizácii zariadenia,
- výrazné šetrenie ekonomických a časových nákladov pri celkovej realizácii MVE,

- efektívne analytické možnosti uchovávania získaného know-how a jeho využívania pri realizácii podobných zákaziek firmy v budúcnosti.

POĎAKOVANIE

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0210.

LITERATÚRA

[1] S. Overshoote, *Renewable energy: Power for sustainable future*, Oxford University Press, 2018 , p. 656. ISBN 9780198759751

[2] A. Acakpovi, E.B. Hagan, and F.X. Fifatin, "Review of Hydropower Plant Models," in: *International Journal of Computer Applications*, vol.108, no.18, 2014, pp. 33-38, doi.10.5120/19014-0541

[3] M. Sattouf, "Simulation Model of Hydro Power Plant Using Matlab/Simulink," in: *International Journal of Engineering Research and Applications*, vol. 4, no.1, 2014, pp 295-301. ISSN 2248-9622.

[4] D. Perduková,, M. Fedor, P. Fedor, „Various Approaches for a Design of the Hydraulic Turbine Model," in: *11th International Scientific Symposium on Electrical Power Engineering*, Elektroenergetika 2022, 12.-14.9.2022, Stara Lesna, Slovakia. ISBN 978-805534104-0.

[5] N. Kishor and J. Fraile-Ardanuy, *Modeling and Dynamic Behaviour of Hydropower Plants*, London: The Institution of Engineering and Technology, 2017. p. 279. ISBN: 978-1-78561-195-7.

Fuzzy odhad kritických parametrov Klossovho vzťahu pre modelovanie asynchrónneho motora

[1]*Pavol FEDOR,* [2]*Peter RADVÁNI*

[1] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika
[2] EnergoControl s.r.o., Pri plynárni 2, 040 01 Košice, Slovenská republik

[1]pavol.fedor@student.tuke.sk, [2]radvani@energocontrol.sk

***Abstract* — Predložený príspevok sa zaoberá návrhom zjednodušeného modelu asynchrónneho motora (AM) s využitím fuzzy odhadu kritických parametrov Klossovho vzťahu. K tvorbe takéhoto modelu je potrebné realizovať niekoľko meraní skokových odoziev na motore, pričom nie je potrebná znalosť jeho parametrov. Získaný jednoduchý model AM je možné využiť v nadradených regulačných obvodoch pre určenie hodnôt vnútorného momentu motora alebo pre identifikáciu jeho parametrov, napr. rotorového odporu z aktuálnych dynamických odoziev pohonu**

***Keywords* — asynchrony motor, fuzzy logika, Klossov vzťah, kritické parametre**

I. INTRODUCTION

Použitie fuzzy prístupu k popisu a riadeniu dynamických sústav zaznamenáva v poslednom období značný rozvoj. Filozofia popisu sústav na základe ich kvalitatívnych vlastností umožňuje nachádzať pomerne jednoduché a prakticky použiteľné modely aj zložitých nelineárnych sústav, pravda za cenu určitej nepresnosti v ich popise [1]-[3]. V oblasti elektrických pohonov je typickým predstaviteľom takejto silne nelineárnej sústavy trojfázový asynchrónny motor, ktorého analitický popis za určitých zjednodušujúcich predpokladov tvorí sústava piatich nelineárnych diferenciálnych rovníc s tromi vstupmi. Predložený príspevok sa zaoberá zostavením jeho zjednodušeného fuzzy modelu a verifikáciou jeho vlastností vo vybraných prevádzkových stavoch.

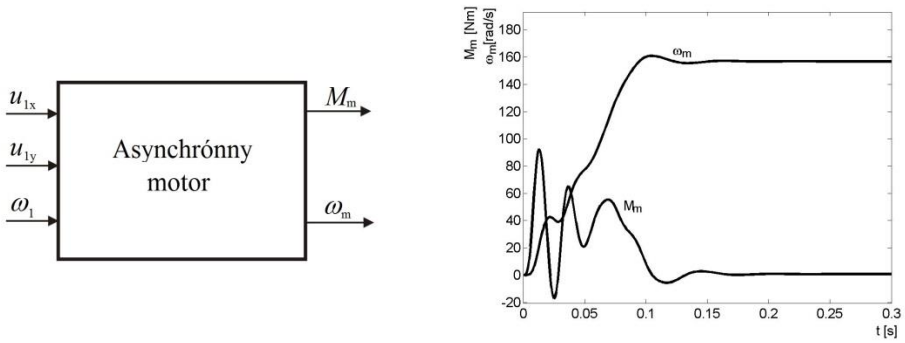
II. POPIS PROBLÉMU

Trojfázový asynchrónny motor nakrátko (ďalej len AM) môžeme popísať napr. sústavou piatich nelineárnych diferenciálnych rovníc v tvare [4]:

$$\frac{d}{dt}\begin{bmatrix} i_{1x} \\ i_{1y} \\ \Psi_{2x} \\ \Psi_{2y} \end{bmatrix} = \begin{bmatrix} -\omega_0 & \omega_1 & -K_{12}.\omega_g & -K_{12}.n_p.\omega_m \\ -\omega_1 & -\omega_0 & K_{12}.n_p.\omega_m & -K_{12}.\omega_g \\ M.\omega_g & 0 & -\omega_g & \omega_2 \\ 0 & M.\omega_g & -\omega_2 & -\omega_g \end{bmatrix} . \begin{bmatrix} i_{1x} \\ i_{1y} \\ \Psi_{2x} \\ \Psi_{2y} \end{bmatrix} + \begin{bmatrix} K_{11} & 0 \\ 0 & K_{11} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} . \begin{bmatrix} u_{1x} \\ u_{1y} \\ 0 \\ 0 \end{bmatrix}$$

$$n_p \frac{M_m}{L_2}\left(\psi_{2x}.i_{1y} - \psi_{2y}.i_{1x}\right) - m_z = J\frac{d\omega_m}{dt} \qquad (1)$$

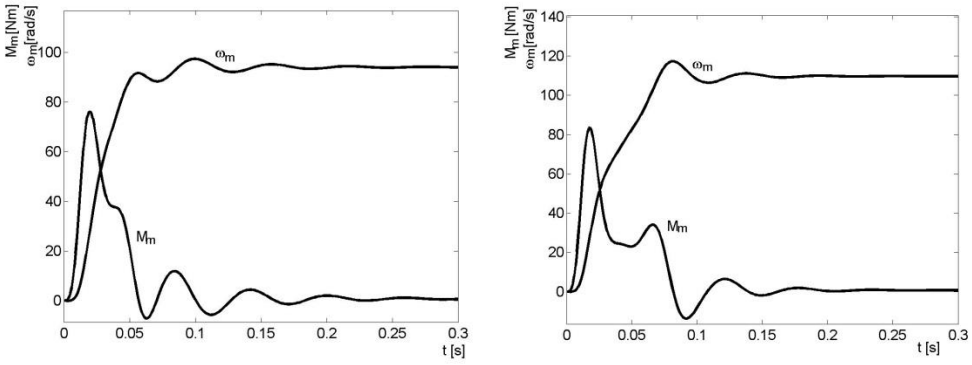
Jedná sa o prúdovo-tokový model, ktorý z hľadiska vstupov a výstupov predstavuje sústavu podľa Obr.1a , kde u_{1x} , u_{2x} sú zložky vektora statorového napätia U_1 v rotujúcej súradnicovej sústave *x, y* a $\omega_1=2\pi f_1$ je jeho kruhová frekvencia. Príklad dynamického priebehu výstupných veličín pohonu s AM pre jeho priame pripojenie na sieť s napätím $U_1=220V$ a frekvenciou $f_1=50$ *Hz* je na Obr.1b.



a) b)

Obr. 1 Asynchrónny motor a jeho priame pripojenie na sieť

Ľahko sa presvedčíme, že pre iné pracovné stavy motora je jeho odozva podstatne odlišná, ako je to ukázané na Obr.2.



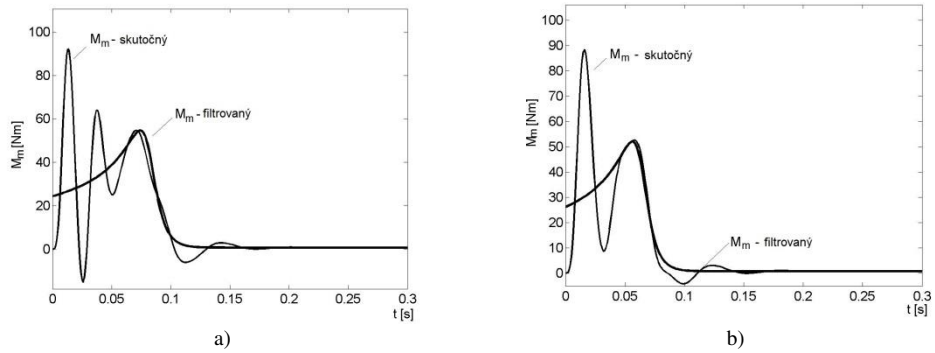
a) b)

Obr. 2 Odozvy AM pri pripojení na rôzne vstupné napätie a frekvenciu:
a) skok na 30% rýchlosti, b) skok na 35% rýchlosti

Z uvedených priebehov momentu a uhlovej rýchlosti asynchrónneho motora je zrejmé, že popísať kvalitatívne „správanie sa" AM, teda zostaviť jeho fuzzy model v tvare niekoľkých pravidiel typu

AK vstupy motora sú ... **POTOM** výstupy motora sú ...

je na základe iba nameraných vstupov a im zodpovedajúcich výstupov veľmi obtiažne. Z nameraných odoziev AM ako aj zo znalostí jeho fyzikálnej podstaty vieme, že AM obsahuje rýchly elektromagnetický podsystém (v odozvách reprezentovaný kmitaním momentu s vysokou frekvenciou, rádove desiatky Hz) a pomalší elektromechanický podsystém (reprezentovaný priemernou strednou hodnotou jeho momentu). Za predpokladu, že sa dopustíme určitej nepresnosti a zanedbáme pri vytváraní jeho popisu rýchle elektromagnetické deje, môžeme pre vytvorenie jeho fuzzy modelu použiť „vyfiltrované" náhradné odozvy podľa Obr.3.



a) b)

Obr. 3 Príklady zjednodušených odozviev AM: a) skok na 50% rýchlosti, b) skok na 40% rýchlosti

Statickú závislosť medzi vnútorným momentom asynchrónneho motora a sklzom popisuje tzv. Klossov vzťah v tvare:

$$M_m = \frac{2.M_k.s_k.s}{s_k^2 + s^2} \tag{2}$$

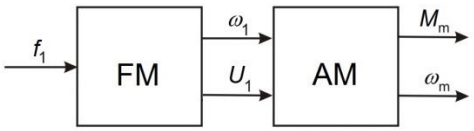
v ktorom M_k je maximálny (kritický) moment motora a s_k je maximálny (kritický) sklz. Sklz motora je definovaný ako pomerný rozdiel uhlovej rýchlostí statorového elektrického poľa ω_1 a mechanickej uhlovej rýchlosti ω_m podľa vzťahu:

$$s = \frac{\omega_1 - \omega_m}{\omega_1} \quad ; \qquad \omega_1 \neq 0 \tag{3}$$

Vzťah (2) môže teda najlepšie aproximovať zjednodušené momentové chovanie sa AM za predpokladu, že sú dobre určené jeho aktuálne parametrické body M_k a s_k . Tieto parametre sú silne nelineárnymi funkciami hlavne uhlovej rýchlostí statorového poľa ω_1 a veľkosti statorového napájacieho napätia U_1 (závislosť od mechanickej uhlovej rýchlosti ω_m je menej výrazná, preto ju kvôli zjednodušeniu modelu v ďalšom texte zanedbávame) .

$$M_k = f_M(\omega_1, U_1, \omega_m) \quad , \quad s_k = f_s(\omega_1, \omega_m) \tag{4}$$

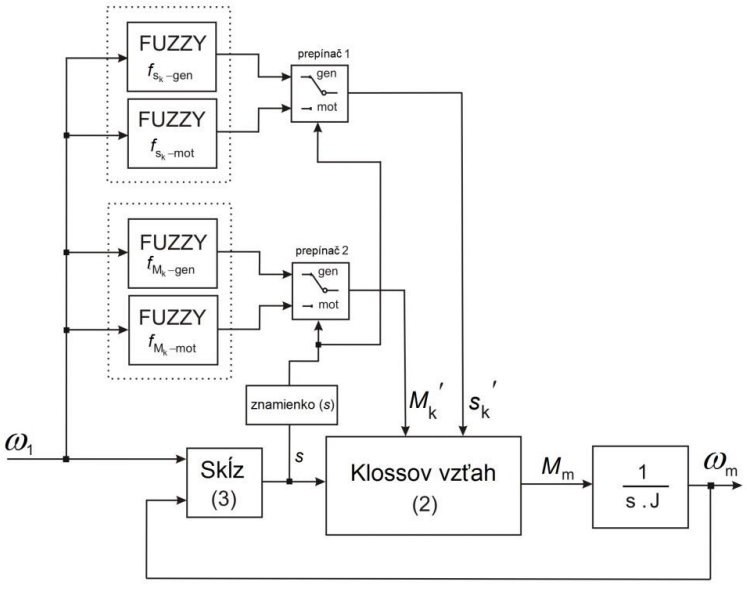
Predpokladajme ďalej, že AM bude napájaný zo štandardného skalárneho frekvenčného meniča (FM), ktorého jediným vstupom je žiadaná hodnota frekvencie statorového napätia $\omega_1=2\pi f_1$ a ktorý vnútorne zabezpečuje podľa predpísanej závislosti funkciu $U_1=f(\omega_1)$. Predmetom kvalitatívnej identifikácie pre zostavenie fuzzy popisu potom bude striedavý pohon s jedným vstupom ω_1 a výstupom M_m, resp. ω_m podľa Obr.4.



Obr. 4 Štruktúra sústavy FM-AM pre zostavenie fuzzy modelu

III. Postup identifikácie parametrov AM

Pre identifikáciu parametrov AM bola použitá fuzzy logika [5]-[7]. Základná štruktúra fuzzy modelu AM je ukázaná na Obr.5. V tomto modeli je výpočet sklzu a momentu AM realizovaný pomocou Klossovho vzťahu a odhad jeho aktuálnych parametrov M_k a s_k pomocou fuzzy systému, aproximujúceho konkrétne funkcie (4) pre daný pohon.



Obr. 5 Základná štruktúra modelu AM s fuzzy odhadom M_k a s_k

Fuzzy časť modelu sa skladá zo štyroch jednoduchých SISO podsystémov, ktorých vstupom je uhlová rýchlosť ω_1 a výstupom je aktuálna hodnota M_k, resp. s_k. Riadiacim signálom odvodeným od polarity sklzu sa prepínajú prepínače 1, 2 a tým sa rozlišuje motorická a generátorická prevádzka motora. Fuzzy pravidlá pre jednotlivé fuzzy podsystémy modelu budú v jednoduchom tvare:

AK ω_1 je ... **POTOM** M_k je ..., resp. s_k je (5)

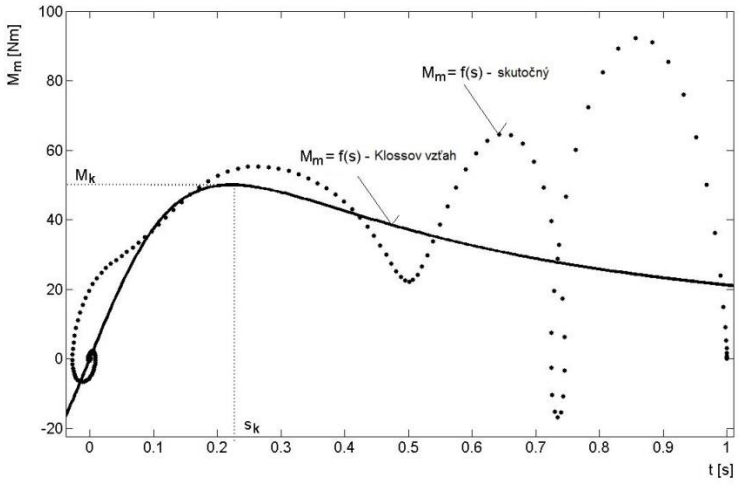
Pretože odozvy AM pre rôzne hodnoty statorovej frekvencie f_1 sú veľmi odlišné, je potrebné zmapovať celý jeho pracovný priestor. Priestor vstupných frekvencií rozdelíme na niekoľko pásiem. Merania robíme postupne tak, že skokom zmeníme vstupnú frekvenciu f_1 (resp. ω_1) z nuly na žiadanú hodnotu a vždy počkáme na ustálený stav ω_m pre nastavenú ω_1.

Odhad parametrov M_k a s_k pre danú odozvu je robený z príslušnej charakteristiky $M_m=f(s)$. Takto získame databázu parametrov M_k a s_k pre motorický režim AM. Podobne urobíme databázu aj pre generátorický režim, v ktorom samozrejme hodnoty parametrov M_k a s_k zmenia znamienko. Na rozlíšenie motorickej a generatorickej prevádzky motora slúžia v modeli prepínače 1, 2. Ak je sklz kladný, potom M_k a s_k sú kladné čísla, tzn. sme v motorickej prevádzke. Ak je sklz záporný, potom M_k a s_k sú záporné čísla, tzn. sme v generátorickej prevádzke.

Pretože pri každom skoku ω_1 je počas celej odozvy až do ustáleného stavu ω_1 konštantné, je potrebné nájsť pre konkrétnu odozvu tiež konštantné M_k, resp. s_k pre Klossov vzťah tak, aby odozva modelu AM z Obr.5 čo najlepšie aproximovala reálnu odozvu pohonu pre daný skok ω_1. K tomuto účelu je možné použiť rôzne všeobecne známe metódy regresie z matematiky a analýzy údajov, ale často stačí jednoduchý odhad M_k a s_k a jeho následné overenie simuláciou konrétneho prevádzkového stavu pohonu.

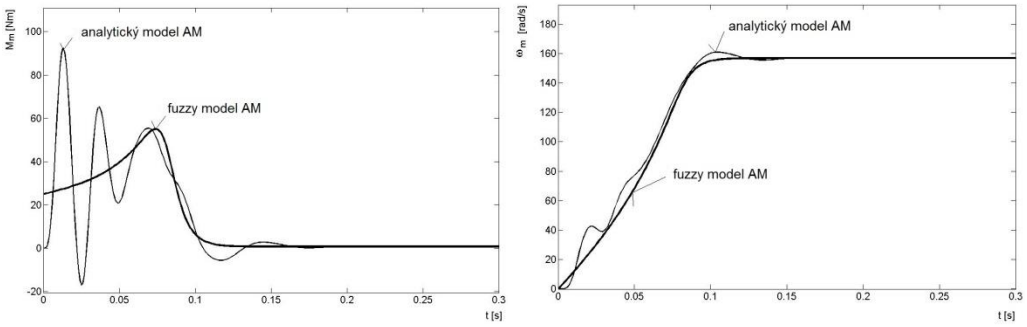
IV. VERIFIKÁCIA FUZZY MODELU AM

Pre verifikáciu uvedenej metodiky bol zvolený trojfázový asynchrónny motor s parametrami podľa Dodatku. Na tomto motore boli simulované merania jeho odoziev s dobou vzorkovania T=1 ms pre vstupné frekvencie f_1=10, 20, 30, 40, 50 [Hz]. Závislosť momentu motora od sklzu pri skoku na nominálne otáčky, čomu odpovedá vstupná frekvencia 50Hz je na Obr.6. Na tomto obrázku je zobrazený aj priebeh, ktorý je daný Klossovým vzťahom a zároveň najlepšie aproximuje reálnu odozvu pohonu. Z tohto priebehu sa dajú vyčítať hľadané parametre M_k a s_k. Takýmto spôsobom je potrebné zmapovať celý pracovný priestor f_1 pre motorický režim aj pre generátorický režim motora.

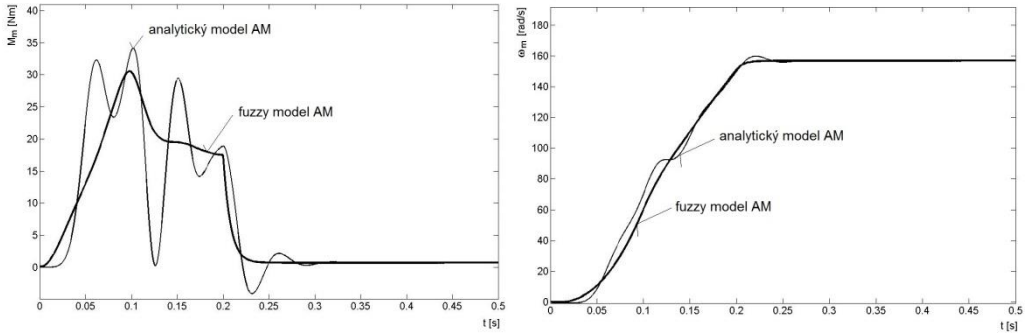


Obr. 6 Určenie parametrov Mk a sk pre skok z f1 = 0Hz na f1 = 50Hz

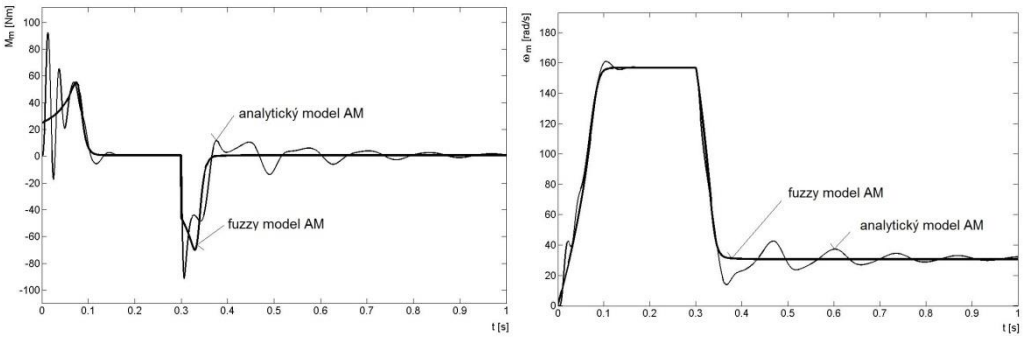
Na základe takto získanej databázy bol zostavený model AM podľa Obr.5 s fuzzy identifikáciou parametrov M_k a s_k, Fuzzy odhad týchto parametrov bol realizovaný jednoduchým fuzzy systémom typu Sugeno s piatimi pravidlami. Porovnanie odoziev identifikovaného motora a jeho fuzzy modelu pre vstupné frekvencie podľa identifikačného merania je ukázané na Obr.7. Pre lineárne rastúci vstupný signál f_1 sú uvedené výsledky fuzzy modelovania AM na Obr.8. Na Obr.9 je porovnaný s analytickým modelom skok vstupnej frekvencie z f_1=50Hz na f_1=10Hz.



Obr. 7 Porovnanie AM a jeho fuzzy modelu pre priame pripojenie na sieť:
a) priebeh momentov, b) priebeh uhlových rýchlostí



Obr.8 Porovnanie AM a jeho fuzzy modelu pre lineárne rastúci vstupný signal:
a) priebeh momentov; b) priebeh uhlových rýchlostí



Obr.10 Porovnanie AM a jeho fuzzy modelu pre skok otáčok odpovedajúcej 50Hz a následne na 10Hz:
a) priebeh momentov; b) priebeh uhlových rýchlostí

V. ZÁVER

V predloženom článku bol ukázaný postup na báze fuzzy logiky, ktorý umožňuje výrazne zjednodušiť popis AM, pričom je potrebné nameriať na tomto motore iba niekoľko jednoduchých skokových prechodových charakteristík. Tento postup nahradil sústavu štyroch nelineárnych diferenciálnych rovníc jednoduchou statickou fuzzy nelinearitou a analyticky popísaným Klossovým vzťahom bez toho, aby bolo potrebné poznať parametre motora. Ako potvrdzujú výsledky simulácií, odhad vnútorného momentu motora je aj v dynamických stavoch pohonu veľmi dobrý a preto v princípe použiteľný v ľubovoľných nadradených regulačných štruktúrach, uvedných napr. v [8], [9]. Odhadnuté hodnoty parametrov M_k a s_k naviac umožňujú určiť aktuálne reálne parametre pohonu v dobe merania bez potreby špeciálnych meraní (naprázdno, nakrátko a pod.).

DODATOK

Parametre AM použité pre simuláciu:

$P_N = 3$ kW	$M_N = 20$ Nm	$U_{1N} = 220$ V
$J = 0{,}022$ kgm^2	$I_{1N} = 6{,}4$ A	$r_1 = 1{,}83\ \Omega$
$n_N = 1420$ ot/min	r_2 ‘$= 2{,}26\ \Omega$	$n_p = 2$
$L_h = 216.44$ mH	$M = 0{,}144$ H	$R_1 = 1.22\ \Omega$
$L_{S1} = L_{S2} = 12{,}036$ mH	$R_2 = 1.508\ \Omega$	$K_{11} = 63{,}999$ H^{-1}

$\omega_2 = \omega_1 - \omega_m$ $K_{12} = -60{,}627\ H^{-1}$ $\omega_g = 9{,}902\ s^{-1}$
$\omega_0 = K_{11}\ (R_1 + (\ M^2/\ L_2\ ).\omega_g) = 164.723\ s^{-1}$

POĎAKOVANIE

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0210.

LITERATÚRA

[1] T. Takagi, M. Sugeno, "Fuzzy identificationof systems and its application to modeling and control," *IEEE Trans. Systems*, *Man and Cybernetics*, 15, 1985, pp. 116-132.
[2] T. A. Bennassar, A. Abbou, M. Akherraz, and M. Barara, "Sensorless sliding mode control of induction motor using fuzzy logic Luenberger observer," Int. J. of Fuzzy Systems and Advanced Applications, vol. 9, 2022, pp. 20-26. DOI: 10.46300/91017.2022.9.4.
[3] Ch. Ben Regaya, A. Zaafouri, and A. Chaari, "A new sliding mode speed observer of electric motor drive based on fuzzy-logic," Acta Polytechnica Hungarica, vol. 11, no. 3, 2014, pp. 219-232.
[4] C. D. Tran, P. Palacky, M. Kuchar, P. Brandstetter, and B. N. Dinh, "Current and speed sensor fault diagnosis method applied to induction motor drive," IEEE Access, vol. 9, 2021, pp. 38660-38672. DOI: 10.1109/ACCESS.2021.3064016.
[5] J. Abynyi, R. Babugkao, B. Feil, „Structure selection for nonlinear input-output models based on fuzzy cluster analysis," IEEE Int. Conf. on Fuzzy Systems, 2003, pp.464-469. DOI: 10.1109/FUZZ.2003.1209408.
[6] T. A. Bennassar, A. Abbou, M. Akherraz, and M. Barara, "Sensorless sliding mode control of induction motor using fuzzy logic Luenberger observer," Int. J. of Fuzzy Systems and Advanced Applications, vol. 9, 2022, pp. 20-26. DOI: 10.46300/91017.2022.9.4.
[7] J. Shaofei, D. Yubo, H. Shun, T. Bowen, P. Xiang, and J. Liting, "Fuzzy concept evaluation based on prospect theory and heterogeneous evaluation information," J. of computing and information science in engineering, vol. 22, no. 4, 2022. DOI: 10.1115/1.4053673.
[8] Y. Ren, R. Wang, S. J. Rind, P. Zeng, and L. Jiang, "Speed sensorless nonlinear adaptive control of induction motor using combined speed and perturbation observer," Control Engineering Practice, vol.123, 2022. DOI: 10.1016/j.conengprac.2022.105166.
[9] D. Perduková, P. Palacký, P. Fedor, P. Bober, V. Fedák, "Dynamic Identification of Rotor Magnetic Flux, Torque and Rotor Resistance of Induction Motor," IEEE Access, Vol.8, 2020, pp. 142003-142015. ISSN 21693536. DOI: 10.1109/ACCESS.2020.3013944

Social network design with emphasis on security and privacy

[1]*Julián GULA,* [2]*Branislav MADOŠ*

[1, 2]Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

[1]julian.gula@tuke.sk, [2]branislav.mados@tuke.sk

Abstract **– This paper presents a custom social network solution designed with a focus on enhancing user security and privacy. The work identifies key challenges in current social networks and proposes measures such as end-to-end encryption using the RSA algorithm, private key encryption for database storage, and secure hashed password storage. The solution also incorporates a custom captcha system and a limit on login attempts to protect against automated attacks. Additional features include a user-friendly interface, server hijacking protection, regular background tasks, and user notifications. This work contributes significantly to the field of secure social networking and lays a foundation for future research and development.**

Keywords **– design, privacy, security, social network**

I. Introduction

This work proposes a secure social network design, addressing the prevalent security and privacy issues in existing platforms. It includes end-to-end encryption, secure hashed password storage, and protection against potential attacks. The result is a prototype that prioritizes user security, serving as a foundation for future research in secure social networking. Ethical considerations and user privacy protection are integral to this design.

II. Related work

As part of our analysis, we reviewed several existing applications and works that address social networking issues, with an emphasis on security and privacy.

Signal, Wickr, Threema, and Dust prioritize security with end-to-end encryption and self-destructing messages. Ricochet uses peer-to-peer and Tor for anonymous messaging. Briar operates decentralized and offline. Telegram offers encryption and self-destructing messages, although centralized. Snapchat specializes in temporary visual content without broad text or group features.

A. Advantages of existing solutions

Security and privacy: these existing approaches place a strong emphasis on security, encryption and privacy of communications.

Peer-to-peer communication: The use of peer-to-peer communication allows direct message exchange between users, eliminating the need for a central server and increasing resilience to network outages.

B. Disadvantages of existing solutions

Complexity: Implementing a peer-to-peer network for communication and security can be technically challenging and require specific expertise.

Speed and reliability limitations: In some cases, the use of a peer-to-peer network can be limiting in speed and reliability compared to centrally managed systems.

User willingness: Using these applications requires users to become familiar with the new environment and possibly install additional software.

In summary, these applications have their limitations and drawbacks, and we decided to design and implement our social network that would address the identified problems and shortcomings of existing social networks. It is important to weigh all of these advantages and disadvantages and find a solution that places a strong emphasis on security while still being mindful of privacy, user-friendliness, and flexibility of use.

III. Analysis of the security features

A. Security attacks and threats

SQL Injection: A SQL injection attack is executed when an attacker injects malicious code into an application's database through user input fields [1].

Cross-Site Scripting (XSS): Cross-site scripting (XSS) attacks involve injecting malicious code or a malicious script into a website. The website then executes the script, allowing the attacker to steal sensitive user data, like session tokens and cookies, or perform other actions [1].

Cross-Site Request Forgery (CSRF): CSRF attacks allow a malicious user to execute actions using the credentials of another user without that user's knowledge or consent [2].

Unauthorized Access and Data Theft: They typically cause problems like denial of access to computers and networks, unauthorized access to and usage of corporate networks, theft and exposure of private data, and unauthorized changes to computers and networks [3].

B. Protection and implementation of security features

Password hashing: Instead of storing passwords in plaintext, it's a good idea to hash them. We'll use cryptographic hashing to safeguard sensitive user information [4].

Authentication and authorization: Django provides robust user and permission management tools that can help protect your application [2].

Protection against CSRF attacks: Django has built-in protection against most types of CSRF attacks, providing you have enabled and used it where appropriate [2].

Protection against XSS attacks: When using Django templates, you are protected against most XSS attacks. However, it is important to understand what protection it provides and what its limitations are [5].

Encryption: Encrypt and decrypt important information and user data to increase the security of sensitive data stored in the database.

It is important to understand that security should be a priority when developing web applications. As a developer, ours primary goal should be to protect your user's data and your system's integrity from a variety of potential attacks [6].

IV. Technology analysis for social network development

In this section, we focus on the technologies we used in the development of our social network. Together, these technologies form the core architecture of our social network and allow us to build secure, scalable, and powerful web applications.

A. Django

Our social network is built on Django, a web framework based on the MVT model, which is a variation of the MVC model. Django provides a declarative URL routing system that allows developers to define URL patterns in a clean and expressive way. Django also includes a powerful ORM system that allows developers to interact with the database using Python code instead of SQL queries.

B. Peer-to-peer networking

Communication in our social network is done through a P2P network. Peer-to-peer (P2P) is a decentralized communications model in which each party has the same capabilities and either party can initiate a communication session [7]. This model allows direct communication between two devices (users), thus avoiding connection and communication with the server.

C. End-to-end encryption

End-to-end encryption (E2EE) is a method that ensures that messages are encrypted at the sender's device and decrypted at the receiver's device using an appropriate encryption algorithm. This means that no one else, including the service provider, can read these messages [8]. To secure the communication and message exchange between users of our social network, we implemented E2EE using RSA algorithm1. This method ensures that messages are encrypted on the sender's device and decrypted only on the receiver's device. It is important to note that the key must be securely shared between the sender and the recipient. If the key is compromised, anyone with that key can decrypt the messages [9].

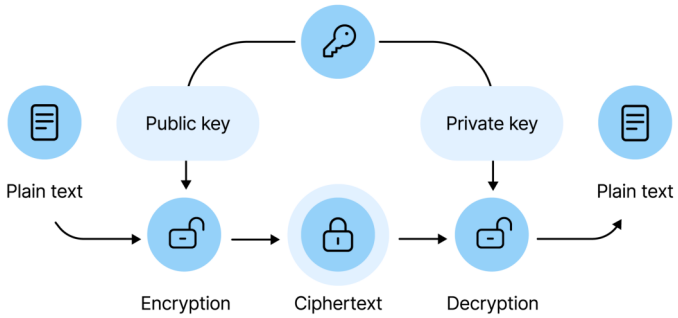


Fig. 1 Illustration of how RSA works [10]

D. Redis

Remote Dictionary Server (Redis) is a key component in creating a peer-to-peer chat. Redis is an in-memory database that supports a variety of data structures, such as lists, sets, sorted sets, hashes, etc. Redis has a featurePub/Sub messaging, which allows developers to scale the backend by creating multiple server instances [11].

E. Celery

To handle the huge amount of reports, we used Celery2, a distributed system that provides the tools needed to maintain such a system. Celery is an excellent tool for processing background tasks in Django applications, allowing the application to continue the main process without waiting for background tasks to complete.

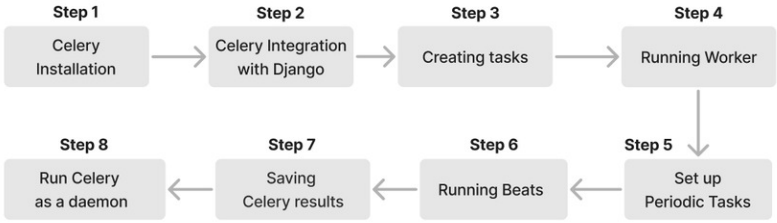


Fig. 2 Flow Diagram Celery [12]

V. Design of the Social Network Structure

The application's design, as depicted in the Django web application component diagram3, delivers a secure and user-friendly platform with a focus on privacy and real-time communication, while also illustrating the interconnected structure built on Django, Redis, and Celery, and the clear relationships between these technologies.

A. Application structure

The application's architecture utilizes the Django framework alongside a PostgreSQL database, incorporating Django Channels and Redis for real-time peer-to-peer communication, and leveraging Celery with Redis for background task processing. Encrypted with end-to-end RSA, the system's structure spans multiple Django modules, with PostgreSQL handling data storage and Redis facilitating short-term data and caching.

B. Design of visual elements

The custom logo was designed to embody simplicity and symbolism, with careful selection of colors and fonts to enhance its visual appeal and brand identity. The user interface (UI) and user experience (UX) were thoughtfully designed to ensure a consistent and intuitive interaction, with an emphasis on consistency of colour palette, typography and responsive design for different screen sizes, including mobile devices.

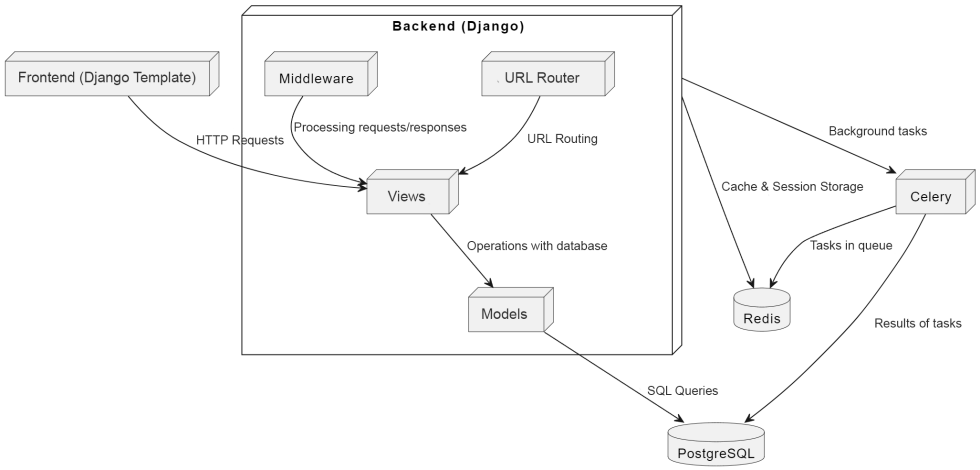


Fig. 3 Django web application component diagram

VI. Security features implementation

A. Custom CAPTCHA Mechanism

Our custom CAPTCHA mechanism operates by presenting users with a rotating key image. The task is to align the key's edge precisely with a red dot, serving as a human verification process. Successful alignment within a tolerance of ±10 degrees confirms human interaction.

- Key Image Generation: We generated a rotatable key image for CAPTCHA challenges.
- Red Dot Creation: A reference red dot was added to aid key orientation.
- User Interaction: Users rotate the key image to align with the red dot.
- Validation: The system verifies the alignment within the tolerance range.
- Handling Incorrect Attempts: Users receive guidance upon incorrect attempts.

B. Account Lockout Functionality after Three Unsuccessful Login Attempts

This security measure mitigates brute-force attacks by locking accounts for 5 minutes after three unsuccessful login attempts. Potential Improvements would be dynamic lockout time adjustments and logging login attempts for security analysis are potential enhancements.

- Lockout Time Control: Accounts remain locked until the specified time elapses.
- Password Verification: Failed attempts trigger account lockout.
- Account Lockout: Exceeding three unsuccessful attempts results in a 5-minute account lock.

C. Application Protection against Server Takeover through Extensive Attacks

We employed Django Ratelimit to safeguard against extensive attacks, limiting request rates per user or IP address within set time intervals.

- Code Segment: Python code snippets demonstrate rate limiting for GET and POST requests.
- Protection: Rate limiting prevents excessive load or DoS attacks, ensuring platform stability.

D. Registration without Email Address Usage

This registration method, which does not require an email address, reduces personal data exposure and spam risks. It offers security benefits such as risk reduction and spam protection, prioritizing user privacy and safety.

VII. Chat room implementation

Our platform offers secure peer-to-peer message exchange without intermediaries, with options for direct or database-stored communication, message retention settings, and notifications for message expiration. Users can also select self-destructing messages with automated deletion for enhanced privacy. RSA encryption provides robust security for all message exchanges.

Figures4 and 5 depict the chatroom's active user list and the primary chat interface, respectively, highlighting the platform's interactive features.

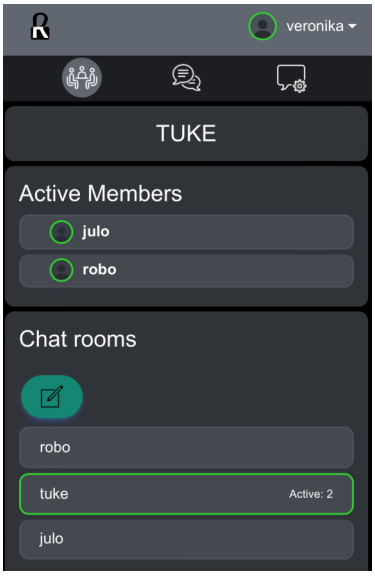


Fig. 4 Chatroom - active users and chat rooms

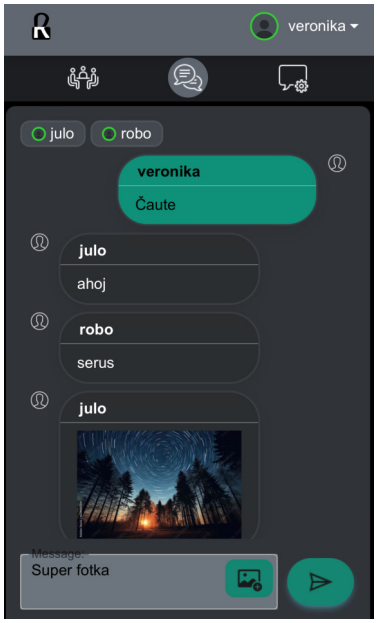


Fig. 5 Chat room - main chat room screen

- **Chat architecture:** Our chat operates on a peer-to-peer basis with end-to-end encryption, leveraging the RSA algorithm. Channels library facilitates WebSocket-based peer-to-peer communication, running on a Redis server. Celery manages background tasks for message control, article checks, online status, and session expiry cleanup.
- **Chat room interface:** The chat interface features a central window for displaying conversations, complete with support for text entry and image uploads. A dynamic sidebar shows active members and chat rooms, while group chats also display a list of participants. Additionally, a settings menu allows users to customize options such as message expiration, notifications, and room names, and includes an exit button for leaving group chats.
- **ChatConsumer implementation:** ChatConsumer class, inheriting from WebsocketConsumer, manages connection, disconnection, message receipt, user count, and message storage.
- **ASGI application:** ASGI facilitates communication between the Django application and asynchronous web servers, supporting WebSocket connections for real-time interaction.
- **Message sending:** Messages are broadcasted to all active members, including inactive users if specified, using group_send function. Messages contain encoded content, image, recipient, sender, and code, ensuring comprehensive message delivery.
- **Image optimization and sending:** Images are sent within the peer-to-peer network, optimized for efficient transmission using resize_image function, preserving quality while minimizing network resources.
- **Message and image storage:** Messages, including encrypted content, are stored in the database if specified, facilitating message retention and retrieval. However, image storage is not implemented in the current version.

VIII. Social Network Feature Implementations

- **Transformation of social media links:** Social media links are transformed into embedded elements, enabling seamless content browsing within our platform. Regular expressions identify social media links, replaced with HTML elements for direct content embedding. Security measures, such as HTML escaping, prevent XSS vulnerabilities.
- **Session management for security:** Efficient session management is crucial for user security. Our application implements session extension based on user interactions, ensuring session continuity and mitigating session hijacking risks.
- **User notification implementation:** Notifications play a vital role in user security and engagement. Our platform sends notifications for friend requests, chat invitations, room exits, and inactive message alerts, enhancing user interaction and platform security.

IX. Testing

The solution was thoroughly tested for functionality, reliability, and security. Unit and integration tests were performed on various features of the social network to ensure correct and

consistent operation. Security checks were conducted to ensure all security measures, such as end-to-end encryption, password hashing, and protection against SQL Injection, XSS, CSRF, and unauthorized access, were functional and secure.

X. Solution summary and evaluation

The design and implementation of the social network with an emphasis on security and privacy were successful. The user interface of the application is intuitive, easy to use, and ensures a smooth and hassle-free user experience. The peer-to-peer chat rooms function correctly, and all security measures have been implemented effectively. The use of Python, Django, PostgreSQL, Redis server, and Celery has resulted in a robust and flexible application.

A. Future works

Future enhancements could include the ability to add attachments to messages, increased support for sending notifications, and account recovery options. Implementing encrypted transmission of parts for enhanced communication security could also be considered. For deployment into production, technologies like NGINX server and Docker could be utilized.

B. The benefits of work

The solution offers several advantages and innovations compared to other existing social networks focusing on security. It provides a wide range of features and security measures that go beyond what Signal and Telegram offer, making it attractive for users seeking a robust social network with a high level of security and privacy. The solution is a contribution to the field of social networks and pushes the boundaries in the area of security and privacy on the internet.

XI. Conclusion

This paper dealt with the analysis of social networking issues, with an emphasis on security and privacy aspects. We have designed and implemented a custom social network that addresses the identified problems and shortcomings of existing social networks. Our solution includes a wide range of features and security measures, including peer-to-peer chat with end-to-end encryption using RSA and Django Channels, and WebsocketConsumer for peer-to-peer communication. We've also implemented various security measures, including custom recaptcha, a limit to entering only 3 incorrect login attempts, ratelimits on API requests for both IP and username, and more. The results of our testing have shown that our solution is functional and well secured. In the future, we could work on improving and extending this social network, adding new features and improving existing features. In conclusion, we have successfully met all the set guidelines for the paper and we believe that our solution has contributed to the development of the field of social network security.

Acknowledgment

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

References

[1] StackHawk. (2023) 10 web application security threats and how to mitigate them. [Online]. Available: https://www.stackhawk.com/blog/10-web-application-security-threats-and-how-to-mitigate-them/
[2] Django. (2023) Security in django. [Online]. Available: https://docs.djangoproject.com/en/5.0/topics/security/
[3] Fortinet. (2023) 7 common web security threats for an enterprise. [Online]. Available: https://www.fortinet.com/resources/cyberglossary/web-security-threats
[4] J. Isah. (2023) Web security in django – how to build a secure web application. [Online]. Available: https://www.freecodecamp.org/news/how-to-build-a-secure-django-web-app/
[5] Sebastian. (2023) Xss protection in django: Essential techniques for secure web applications. [Online]. Available: https://clouddevs.com/django/cross-site-scripting-xss-protection/
[6] ——. (2023) Web defense 101: Enhancing security in django web applications. [Online]. Available: https://clouddevs.com/django/enhancing-security/
[7] L. Rosencrance. (2022) What is peer-to-peer (p2p)? [Online]. Available: https://www.techtarget.com/searchnetworking/definition/peer-to-peer
[8] Cloudflare. (2023) What is end-to-end encryption (e2ee)? [Online]. Available: https://www.cloudflare.com/learning/privacy/what-is-end-to-end-encryption/
[9] A. Poppe. (2023) End-to-end encryption: A modern implementation approach using shared keys. [Online]. Available: https://quickbirdstudios.com/blog/end-to-end-encryption-implementation-approach/
[10] L. Klusaitė. (2022) What is rsa encryption, and how does it work? [Online]. Available: https://nordvpn.com/blog/rsa-encryption/
[11] A. Raina. (2022) How to build a chat application using redis. [Online]. Available: https://developer.redis.com/howtos/chatapp/
[12] Episyche. (2022) How to run periodic tasks in django using celery ? [Online]. Available: https://episyche.com/blog/how-to-run-periodic-tasks-in-django-using-celery

Procedurálne generovanie obsahu počítačových hier

[1]Patrik HURNÝ, [2]Branislav MADOŠ

[1,2]Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]patrik.hurny@student.tuke.sk, [2]branislav.mados@tuke.sk

Abstract **– Tento článok sa venuje procedurálnemu generovaniu v počítačovej grafike a analyzuje súčasné metódy a literatúru. Cieľom výskumu je navrhnúť a implementovať algoritmy na generovanie detailných otextúrovaných 3D objektov. Článok predkladá prehľad existujúcich prístupov k procedurálnemu generovaniu, ich výhody, výzvy a podrobne opisuje navrhnuté funkcionality. Hlavným cieľom je implementácia Lindenmayerového systému pre generovanie stromov a rastlín v hernom engine Unity. Na zlepšenie realizmu a optimalizáciu boli generované modely vytvorené s použitím reálnych obrázkov ako referencií a s aplikovaním realistických textúr. Efektívnosť generovania modelov bola analyzovaná nastavovaním parametrov a meraním času potrebného na ich vytvorenie.**

Keywords **– procedurálne generovanie, počítačová grafika, textúrovanie, UV mapovanie, generovanie objektov**

I. Úvod

V oblasti počítačovej grafiky sa procedurálne generovanie obsahu (PCG) stalo kľúčovou technikou pre automatizované vytváranie digitálnych prostredí, objektov a textúr. Tento článok sa zameriava na procedurálne generovanie v kontexte tvorby realistických 3D modelov a má za cieľ prispieť k rozvoju techník a metód v tejto dynamicky sa rozvíjajúcej disciplíne.

Súčasný pokrok v počítačovej grafike a vizualizácii vytvára stále väčší dopyt po vysoko kvalitných, detailných a realistických 3D objektoch. Tradičné metódy tvorby týchto objektov, ktoré zahŕňajú ručnú prácu grafikov a dizajnérov, môžu byť časovo náročné a finančne nákladné. Procedurálne generovanie ponúka efektívnu alternatívu, ktorá môže tento proces zjednodušiť, avšak vyžaduje hlboké pochopenie algoritmických a dizajnových princípov.

V tejto štúdii sa analyzujú existujúce metódy procedurálneho generovania, ich aplikácie a výzvy v oblasti 3D počítačovej grafiky. Primárnym cieľom je vyvinúť a implementovať funkcie pre procedurálne generovanie rastlín a stromov, pričom sa kládol dôraz na detailné a realistické reprezentácie. Článok poskytuje prehľad súčasných techník a prístupov k procedurálnemu generovaniu rastlinstva, vrátane ich výhod a obmedzení.

Po poskytnutí prehľadu existujúcich techník a prístupov sa článok zameriava na vývoj nových algoritmov a metód pre procedurálne generovanie rastlín a stromov. Tento proces zahŕňa tvorbu realistických 3D modelov a detailnú simuláciu rôznych aspektov rastlinstva, ako sú textúra kôry, rozloženie listov a štrukturálne charakteristiky, špecifické pre rôzne druhy rastlín a stromov.

Ďalej sa článok venuje výzvam, ktoré procedurálne generovanie rastlín a stromov prináša, vrátane potreby optimalizácie pre efektívne využitie výpočtových zdrojov a zabezpečenia, aby generované modely boli vizuálne presvedčivé a realistické. Výsledkom bude súbor nástrojov a techník, ktoré môžu byť aplikované v praxi a poslúžia ako základ pre ďalší výskum a vývoj v tejto rýchlo sa rozvíjajúcej oblasti.

II. Procedurálne generovanie objektov

Procedurálne generovanie predstavuje techniku v oblasti počítačovej grafiky a herného dizajnu, ktorá umožňuje automatické tvorenie objektov využitím algoritmov a stanovených pravidiel, eliminujúc tak potrebu manuálneho modelovania každého jednotlivého detailu. Táto metóda je efektívnym nástrojom pre tvorbu komplexných objektov, ktoré by v prípade manuálneho modelovania vyžadovali značné úsilie a veľa času [1]. V dnešnej dobe je procedurálne generovanie kľúčovou súčasťou v oblastiach ako počítačové hry, vizuálne umenia a dokonca aj vedecký výskum. Táto technika dokáže generovať rôznorodé objekty, vrátane terénov, architektonických štruktúr, postáv a textúr, pričom využíva matematické vzorce, štatistické metódy a generatívne algoritmy.

A. L-Systémy

L-systémy, tiež známe ako Lindenmayerove systémy, predstavujú základný koncept v teoretickej informatike a matematike, ktorý sa objavil v 60. rokoch 20. storočia [2]. Tento inovatívny model bol navrhnutý s úmyslom skúmať a porozumieť rastu a vývoju rastlín a ďalších živých organizmov. Jeho rekurzívne orientovaný prístup priniesol nové poznatky v oblasti biologických procesov a rozšíril možnosti v počítačovej grafike a simulačných technikách.

L-systémy sa kategorizujú do dvoch hlavných typov: deterministických a nedeterministických. Deterministické L-systémy sa vyznačujú pevnými pravidlami, ktoré pre určitý vstup vždy produkujú rovnaký výstup. Na druhej strane, nedeterministické L-systémy, vrátane ich stochastických variánt ovplyvnených pravdepodobnostnými faktormi, ponúkajú väčšiu flexibilitu a rôznorodosť možných výstupov, čo je obzvlášť užitočné v simuláciách [3].

Jednou z najvýznamnejších aplikácií L-systémov je ich využitie v procedurálnom generovaní objektov v počítačovej grafike a hernom dizajne. Tieto systémy umožňujú vytvárať komplexné a detailné virtuálne prostredia, ako sú krajiny alebo vegetácia, bez nutnosti manuálneho modelovania každého objektu. Vďaka efektívnosti pri generovaní opakujúcich sa alebo fraktálne podobných štruktúr sú L-systémy ideálne na simulovanie prírodných vzorov a štruktúr. Kľúčovým aspektom L-systémov je ich pravidlová sada, definujúca transformáciu symbolov v reťazci v každej iterácii, zapisované vo forme predchodca $\rightarrow$ následník.

V praxi sa L-systémy používajú tak, že sa najprv definuje počiatočný reťazec, axióm, a potom súbor pravidiel pre jeho transformáciu. Počas každej iterácie sa symboly v axióme transformujú podľa týchto pravidiel, čím dochádza k postupnej evolúcii reťazca. Tento proces umožňuje L-systémom generovať zložité štruktúry z relatívne jednoduchých pravidiel, čo je ideálne pre modelovanie prírodných a umelých systémov. Dnes sú L-systémy neoceniteľným nástrojom v počítačovej grafike, kde umožňujú efektívne a realistické generovanie komplexných virtuálnych prostredí.

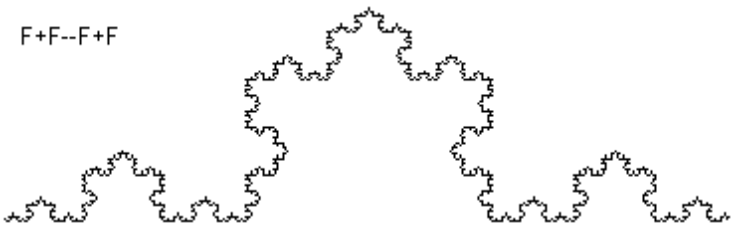


Fig. 1: Kochova vločka vytvorená pomocou L-systému

B. Kolonizácia priestoru

Ďalšou zaujímavou technikou procedurálneho generovania objektov je kolonizácia priestoru, tiež známa ako space colonization. Ide o techniku, ktorá sa najčastejšie používa na procedurálne generovanie stromov a rastlín, s dôrazom na realistické rozloženie vetiev a listov podľa environmentálnych podmienok. Tento prístup je inšpirovaný spôsobom, akým stromy v prírode rastú smerom k svetlu a živinám, čím simuluje prirodzený rast stromov.

Kľúčovým aspektom metódy space colonization je definícia priestoru, ktorý obsahuje „atraktory“, bodové zdroje, reprezentujúce ciele rastu stromu. Atraktory môžu byť distribuované náhodne alebo podľa určitého vzoru, čím odrážajú rozloženie svetla alebo živín potrebných pre strom. Rast stromu začína z definovaného koreňového bodu, pričom sa postupne rozvetvuje s cieľom dosiahnuť tieto atraktory. Vetvy bližšie k atraktorom majú pri raste prioritu, čo vedie k realistickému a efektívnemu rozloženiu vetiev [4].

Jedinečnosť tejto metódy spočíva v schopnosti vytvárať štruktúry, ktoré sú nielen vizuálne presvedčivé, ale aj biologicky pravdepodobné. Každý strom vytvorený touto technikou je jedinečný, čo odráža rozmanitosť nájdenú v prírode. Táto flexibilita umožňuje procedurálne generovaným stromom integrovať sa do rôznych scenárov, od virtuálnej reality až po simulácie ekosystémov.

Výzvou metódy je efektívne využitie výpočtovej sily a optimalizácia algoritmov, najmä pri generovaní veľkého množstva stromov alebo v komplexných ekosystémoch. Ďalej, hoci metóda space colonization poskytuje dobré výsledky pre určité typy stromov, môže byť menej vhodná pre stromy s osobitými rastovými vzorcami, ako sú tie s veľmi hustými alebo nepravidelnými korunami.

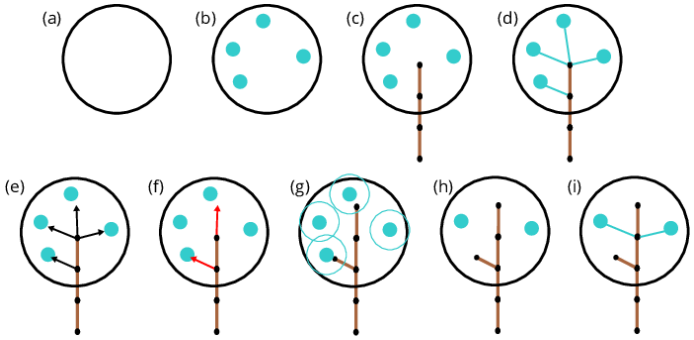


Fig. 2: Ilustrácia algoritmu kolonizácie priestoru

C. Fraktály

Fraktálne stromy sú grafické reprezentácie, ktoré napodobňujú vzhľad prírodných stromov a ich vetvenie. Princíp, na ktorom sú založené, využíva jednoduché algoritmy na vytvorenie vzorov, ktoré sa opakujú na rôznych úrovniach detailov. Každá vetva stromu sa delí na menšie vetvy, ktoré sa následne môžu ďalej deliť, čím vzniká sebapodobná štruktúra. Tento proces sa opakuje viackrát, až kým nedosiahne požadovanú úroveň detailu [5]. Táto metóda ilustruje, ako môžu byť zložité a detailné štruktúry vytvorené z relatívne jednoduchých pravidiel a opakovania, čím sa otvára priestor pre zaujimavé možnosti v oblastiach ako digitálne umenie, vedecká vizualizácia a architektonický dizajn.

Charakteristickou vlastnosťou fraktálov je ich schopnosť demonštrovať, ako môžu zložité a na prvý pohľad náhodné štruktúry v prírode vzniknúť z opakovanej aplikácie pomerne jednoduchých matematických pravidiel. Tento koncept nachádza uplatnenie v širokej škále oblastí, od počítačovej grafiky až po ekológiu, kde je kľúčové modelovať komplexné prírodné javy s vysokou úrovňou nepravidelnosti a detailu. V týchto oblastiach sa často vyžaduje modelovanie komplexných prírodných javov, charakterizovaných nepravidelnosťou a zložitosťou, čo znamená opakovanie s podobným zameraním na detail a nepravidelnosť.

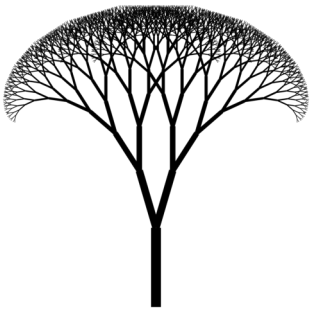


Fig. 3: Fraktálny strom

III. Návrh a implementácia riešenia

Navrh riešenia proceduralného generovania vegetácie pozostáva z dvoch základných komponentov. Prvým základným komponentom je samotný komponent na generovanie základných uzlov pomocou Lindenmayerových pravidiel. Tieto pravidlá budú uložené vo tzv. skriptovateľnom objekte v Unity, ktorý je možné ukladať ako súbor. Tento skriptovateľný objekt budeme volať Lindenmayer Preset. Každé pravidlo bude obsahovať začiatočný axiom, ktorý funguje ako identifikátor určujúci, aké pravidlá daný axiom obsahuje, a popis, ktorý bude slúžiť na jednoduché rozpoznanie účelu daných pravidiel. Pravidlá budú vo formáte reťazca, pomocou ktorého bude možné definovať akcie v jednom riadku, napríklad (F+F-F). Parameter pre iterácie bude určovať, koľkokrát sa má pravidlo zopakovať, čo je užitočné pri tvorbe vzoru z opakujúcich sa pravidiel. Nakoniec, pravidlo zahrnie rôzne nastavenia, ako je typ pravidiel určujúci, či sa má generovať mriežka konára alebo listu, a nastaviteľné parametre ako dĺžka, odklon, sklon a náklon. Každý z týchto parametrov je možné nastaviť na pevnú hodnotu alebo na náhodné číslo v špecifikovanom rozsahu, čo umožní systému generovať vždy iné modely závislé od seedu. Pri pravidlách typu list je možné nastaviť aj ID materiálu, ktorý sa má na mriežku aplikovať, alebo šírku listu. Jednotlivé pravidlá pozostávajú z nasledujúcich jednoduchých znakov, pričom každý znak reprezentuje jednu akciu, ktorá sa má vykonať.

Znak	Akcia
F	Krok dopredu
+	Odklon hore (yaw up)
-	Odklon dole (yaw down)
\	Sklon hore (pitch up)
/	Sklon dole (pitch down)
<	Naklon vľavo (roll left)
>	Naklon vpravo (roll right)
[	Uloženie aktuálneho stavu
]	Načítanie posledného stavu

Table 1: Tabuľka akcií Lindenmayerovho systému

Pomocou týchto akcií je možné vymodelovať kompletný systém pravidiel, ktoré neskôr použijeme pri generovaní samotného modelu. Na modelovanie týchto pravidiel bol implementovaný komponent, ktorý tieto pravidlá spracováva. V rámci tohto komponentu bola implementovaná funkcia *Generate*, ktorej cieľom je generovať zoznam segmentov. Každý segment predstavuje inštanciu triedy *SegmentType*, čo umožňuje priradenie špecifických funkcií jednotlivým segmentom. Funkcia *Generate* začína vytvorením slovníka, kde axiómy tvoria kľúče a príslušné pravidlá hodnoty. Tento mechanizmus zabezpečuje efektívne vyhľadávanie pravidiel asociovaných s konkrétnymi axiómami. Po inicializácii pravidiel začína generovanie L-systému, charakterizované postupným spracovaním znakov abecedy. Implementácia zahŕňa vytvorenie štruktúry *AxiomNode*, ktorá uchováva informácie o axióme, nastavenia pravidiel a zoznam potomkov axiómu. Tento modulárny prístup umožňuje presné modelovanie štruktúr podľa pravidiel L-systému.

Generovanie L-systémov prebieha prostredníctvom rekurzívneho procesu. Vstupnými argumentmi tejto rekurzívnej funkcie sú pravidlá priradené k základnému axiómu, ktorý je reprezentovaný štruktúrou *AxiomNode*. Funkcia systematicky prechádza každým znakom, ktorý môže reprezentovať ďalšie pravidlá alebo vykonávať určité akcie. Pre každý taký znak sa vytvára nová inštancia štruktúry *AxiomNode*, ktorá sa potom pridáva do zoznamu potomkov pôvodného axiómu. Tento prístup zabezpečuje efektívne modelovanie rôznych komplexných štruktúr v rámci L-systému.

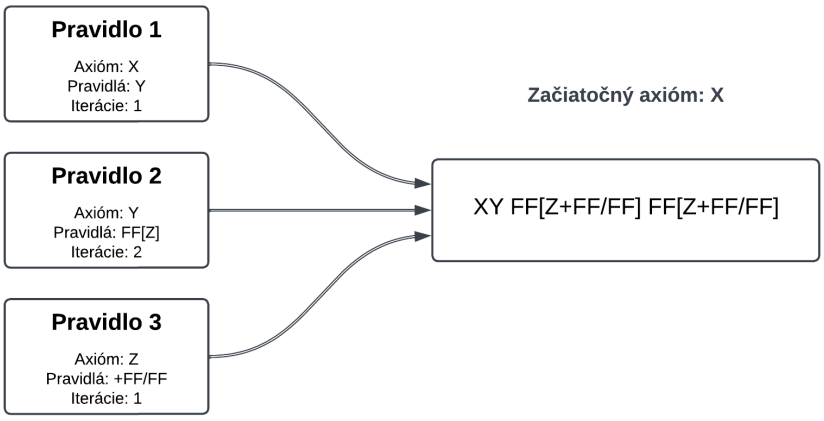


Fig. 4: Názorná ukážka generovania L-Systému

Druhým základným komponentom v procese generovania modelu je skript s názvom *Nature Generator*. Tento skript obsahuje odkaz na skriptovateľný objekt, kde sú definované všetky riadiace pravidlá. Ďalej obsahuje inštanciu komponentu pre generovanie uzlov podľa pravidiel Lindenmayerovho systému. Hlavné funkcie tohto generátora sú *Prepare* a *Generate*. Funkcia *Prepare* zahŕňa prípravu všetkého potrebného na vytvorenie modelu vrátane jeho vizualizácie pred fyzickým vytvorením mriežky, čo znižuje spotrebu pamäťových zdrojov a umožňuje úpravy. Na začiatku vykonania funkcie *Prepare* sa vytvorí nová inštancia komponentu *Lindenmayer* s odkazom na skriptovateľný objekt a seed. Po zavolaní funkcie *Generate* na komponente Lindenmayer sa vygenerujú uzly, s ktorými bude následne pracovať generátor. Funkcia vráti zoznam uzlov, ktoré sa ukladajú do komponentu generátora pre ľahší prístup. Keď sa následne zavolá funkcia *Generate* na samotnom generátore, ten vytvorí mriežku podľa zadaných pravidiel a priradí

materiály príslušným častiam modelu. Kvalita textúr sa zachová rozdelením vetiev a listov na dva oddelené objekty s rozdielnymi UV mapami a materiálmi.

Generátor na zlepšenie kontroly nad vzhľadom finálneho modelu využíva modifikátory uložené v skriptovateľnom objekte. Program obsahuje dva typy modifikátorov. Prvým z nich je *Segment Modifier*, ktorý umožňuje modifikáciu jednotlivých segmentov, ako je napríklad zabezpečenie nerovného rastu stromu. Modifikátor segmentu obsahuje funkcie, ktoré sa volajú v rôznych fázach procesu generovania, ako sú pred vetvou, na segmente, po segmente a po vetve. Druhý typ modifikátora, *Mesh Modifier*, pridáva ďalšie detaily modelu, ako je generovanie kmeňa stromu. Modifikátor sa spustí buď pred alebo po vytvorení mriežky modelu. Nasledujúci diagram znázorňuje, ako údaje prúdia v hlavnom komponente generátora.

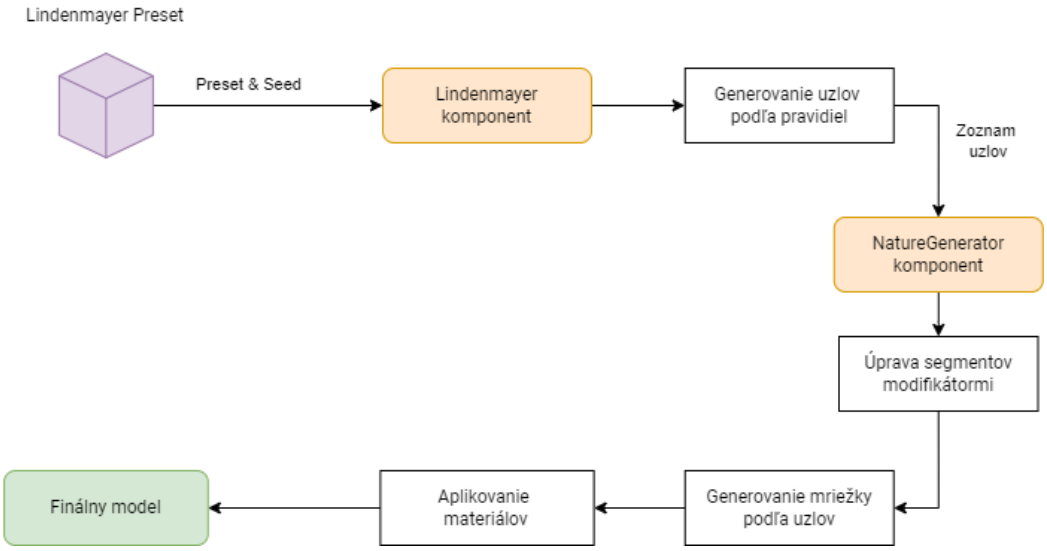


Fig. 5: Architektúra programu generovania obsahu

IV. Výsledky riešenia

Procedurálne generovanie modelov pomocou Lindenmayerovho systému poskytlo zaujímavé výsledky. Využitie Lindenmayerových pravidiel umožňuje vytvárať širokú škálu rôznych modelov, čo prináša nové možnosti v oblastiach modelovania a vizualizácie. Tento prístup umožňuje komplexné a detailné modelovanie objektov, ktoré majú potenciál nájsť uplatnenie v mnohých oblastiach, vrátane herného dizajnu a rôznych vedeckých aj umeleckých disciplín.

Na overenie efektívnosti generátora pri tvorbe modelov sme aplikovali predvoľbu s pravidlami pre generovanie modelu stromu. Počas testovania sme postupne menili parametre ktoré určujú kvalitu modelu. Tieto parametre zahŕňali počet radiálnych segmentov a vyhladzovanie modelu, ktoré ovplyvňujú celkový čas potrebný na generovanie modelu. Časové merania boli vykonávané v milisekundách. Testovanie zahŕňalo procedurálne generovanie so základnými nastaveniami. Uskutočnili sme dve série testov, pričom každá séria sa opakovala päťkrát. V prvej sérii bolo zapnuté vyhlazovanie mriežky, zatiaľ čo v druhej sérii bolo vyhlazovanie vypnuté. V každej sérii testov sme päťkrát zmenili nastavenie parametra *Branch Radial Segments* (BRS), ktorý menil počet radiálnych segmentov jednotlivých článkov modelu. Použité hodnoty tohto parametra boli 10, 20, 30, 40 a 100. Výsledky z testovania je možné vidieť v priloženej tabuľke. Z výsledkov je zrejmé, že vyhladzovanie mriežky modelu výrazne ovplyvňuje rýchlosť generovania modelu. Na druhej strane, vypnutie tejto funkcie vedie k viditeľným prechodom medzi segmentami.

	Vyhladzovanie zapnuté	**Vyhladzovanie vypnuté**	**Počet vrcholov**
BRS 10	14-15ms	1-2ms	6096
BRS 20	29-30ms	2ms	10416
BRS 30	47-48ms	3ms	14500
BRS 40	67-70ms	4ms	18612
BRS 100	258-270ms	7ms	44648

Table 2: Porovnanie časov základného testovania

A. *Náhľad modelov*

Procedurálne generovanie modelov pomocou Lindenmayerovho prístupu umožňuje generovať širokú škálu modelov. Pre vytvorenie modelu je potrebné vytvoriť Lindenmayer Preset skriptovateľný objekt v ktorom je potrebné špecifikovať pravidlá podľa ktorých sa systém bude riadiť počas generovacieho procesu. Tieto pravidlá môžu určovať smer, rotáciu alebo typ vetvy, čo systému umožňuje rozhodnúť, či má generovať konáre alebo listy. Na nasledujúcich dvoch obrázkoch je možné vidieť náhľady vygenerovaných modelov. Systém okrem stromov dokáže generovať aj modely s opakujúcimi sa vzormi, ako je napríklad papradie.



(a) Borovica (b) Papradie

Fig. 6: Náhľad vygenerovaných modelov

V. Záver

Hlavným prínosom tejto práce je implementácia metódy procedurálneho generovania, ktorá využíva Lindenmayerove systémy na tvorbu realistických stromov a rastlín. Táto metóda umožňuje rýchle a efektívne generovanie rôznych prírodných scén bez potreby ručného modelovania každého objektu. Výsledkom sú modely s veľkou mierou detailu a vizuálnej príťažlivosti, čo predstavuje významný krok v oblasti procedurálneho generovania. Systém bol otestovaný a optimalizovaný, takže sa dá používať ako plnohodnotný nástroj, ktorý uľahčí prácu pri modelovaní rôzneho rastlínstva.

Poďakovanie

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

Referencie

[1] Alba Amato. *Procedural content generation in the game industry*. Jan. 1, 2017, pp. 15–25. DOI: 10.1007/978-3-319-53088-8_2. URL: https://doi.org/10.1007/978-3-319-53088-8_2.

[2] Lila Kari, Grzegorz Rozenberg, and Arto Salomaa. *L Systems*. Jan. 1997, pp. 253–328. DOI: 10.1007/978-3-642-59136-5\{_}5. URL: https://doi.org/10.1007/978-3-642-59136-5_5.

[3] P. Prusinkiewicz and Aristid Lindenmayer. *The algorithmic beauty of plants*. Berlin, Heidelberg: Springer-Verlag, 1990. ISBN: 0387972978.

[4] Adam Runions, Brendan Lane, and Przemyslaw Prusinkiewicz. "Modeling Trees with a Space Colonization Algorithm". In: Jan. 2007, pp. 63–70. DOI: 10.2312/NPH/NPH07/063-070.

[5] David Mavrodiev. *Fractals: A Journey Through Mathematical Artistry*. 2024. URL: https://medium.com/@d.s.m/fractals-a-journey-through-mathematical-artistry-72302b0cded0 (visited on 02/05/2024).

Procedurálne generovanie obsahu v oblasti počítačovej grafiky

[1]Peter POPRÍK, [2]Branislav MADOŠ

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

[1]peter.poprik@student.tuke.sk, [2]branislav.mados@tuke.sk

Abstrakt **– Práca sa zaoberá technikami procedurálneho generovania obsahu v oblasti počítačovej grafiky, so zameraním na tvorbu trojrozmerných polygonálnych modelov stromov, pričom sú skúmané rôzne metódy, vrátane algoritmov kolonizácie priestoru a Lindenmayerových systémov, z hľadiska ich vhodnosti pre generovanie stromov. Boli vyvinuté dve metódy generovania: jedna pre rovnomerné vetvenie a druhá pre generovanie s primárnou vetvou. Oba moduly poskytujú možnosť nastavenia parametrov, ktoré ovplyvňujú tvar a štruktúru stromov. Výsledky ukazujú, že vyvinutý generátor stromov dokáže v reálnom čase vytvárať rôznorodé a realistické modely stromov, vhodné pre použitie v iných 3D aplikáciách. Zistenia naznačujú možnost budúcich vylepšení, ako napríklad zavedenie textúrovania a rozšírenie podporovaných druhov stromov.**

Kľúčové slová **– počítačová grafika, procedurálne generovanie, vykresľovanie scén, WaveFront OBJ formát,**

I. Úvod

Procedurálne generovanie obsahu je technika, ktorá umožňuje vytvárať obsah pomocou algoritmov namiesto tradičných metód, ktoré vyžadujú ručné vytváranie obsahu. Presnejšie môže byť procedurálne generovanie obsahu definované ako: *Algoritmické vytváranie obsahu, založené na náhodných procesoch, ktoré umožňuje generovanie obsahu len s obmedzeným alebo nepriamym vstupom používateľa.* Táto definícia bola vytvorená na základe definícií poskytnutých v [1], [2] a [3].

Procedurálne generovanie obsahu nachádza najväčšie uplatnenie v priemysle počítačových hier, kde v rámci herného dizajnu umožňuje vytvárať herné prvky. Medzi takto generované prvky patria nie len grafické objekty ale aj mechaniky hry, mapy, správanie a podobne. Podrobnú taxonómiu prvkov, ktoré je možné procedurálne generovať, uvádza [4].

V rámci tejto práce je skúmané procedurálne generovanie troj-dimenzionálnych, polygonálnych modelov stromov. Stromy v herných tituloch často slúžia na budovanie prostredia a vytváranie atmosféry. Z hľadiska herného dizajnu je dôležité, aby stromy zahrnuté v hernom prostredí čo najbližšie emulovali skutočné stromy a zároveň poskytovali dostatočnú variabilitu, keďže v prírode nie sú žiadne dva stromy rovnaké. Pre realistické herné prostredie je teda potrebné vytvoriť veľké množstvo, rôznych modelov stromov, čo je časovo náročné a nákladné. Z tohto dôvodu je procedurálne generovanie stromov vhodným riešením, ktoré umožňuje vytvárať stromy rôznych tvarov.

II. Procedurálne generovanie stromov

Strom z botanického hľadiska nie je jednoznačne definovaný, pretože pod tento pojem patrí veľká skupina rôznych druhov rastlín, ktoré sa od seba môžu výrazne líšiť. V [5] je strom definovaný ako: *Drevitá trvalka druhu, ktorý zvyčajne tvorí jednu samonosnú hlavnú stonku a má jednoznačnú korunu*. Na základe tohto širokého pojmu je však ťažké vytvoriť procedurálny generátor, ktorý by bol univerzálny pre všetky druhy stromov, a teda lepší prístup je bližšie špecifikovať, aký druh alebo skupinu druhov bude generátor podporovať. Súčasťou špecifikácie je aj určenie štruktúry a jednotlivých častí stromu, ktoré budú procedurálne generované.

Ďalším krokom je určenie pravidiel pre vetvenie stromu, ktoré budú tvoriť základ generovania stromu. Pod vetvením stromu sa rozumie proces, kedy sa z kmeňa oddeľujú vetvy, ktoré sa ďalej delia na ďalšie vetvy až kým sa nedosiahne požadovaná úroveň vetvenia. Typické spôsoby vetvenia sú bližšie popísané v [6], kde sú uvedené monopodiálne a sympodiálne typy vetvenia

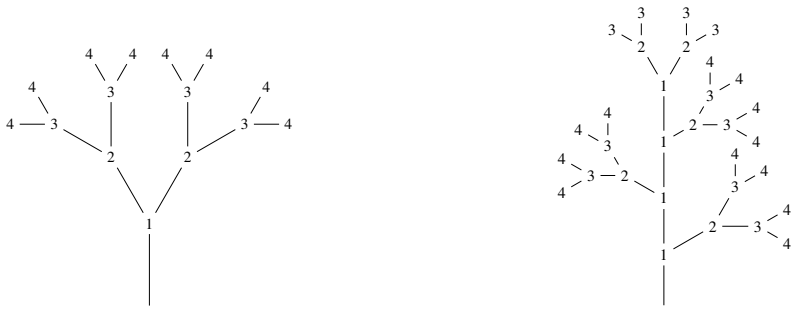
vyskytujúce sa v stromoch a iných rastlinách. Celkovú štruktúru stromu potom možno rozdeliť na jednotlivé časti, ktoré budú generované podľa zvolenej metódy a pravidiel.

Jeden z prístupov k procedurálnemu generovaniu stromov je využitie kolonizácie priestoru. Táto metóda využíva zoznam bodov v priestore, ktoré sú postupne kolonizované a na základe toho sú vytvárané vetvy stromu. Predstavuje to rast stromu, ktorý je založený na biologických princípoch rastu rastlín na základe súťaže o priestor a zdroje. Prístupy k procedurálnemu generovaniu s využitím kolonizácie priestoru sú bližšie popísané v [7] a [8].

Ďalšie prístupy k procedurálnemu generovaniu stromov sú založené na Lindenmayerových systémoch (L-systémy) [9]. L-systém je formálny model, ktorý je založený na pravidlách alebo gramatikách, ktoré sú aplikované na reťazec symbolov. Takýmto spôsobom sú reťazce prepisované, čím sú vytvárané nové reťazce s pravidelnými vzormi, ktoré sú podobné organickým štruktúram. Pri L-systémoch je pre realistické výsledky, blízke skutočným stromom, potrebné zvoliť vhodné pravidlá a parametre, ktoré budú určovať tvar a štruktúru stromu. Jednu z metód na procedurálne generovanie stromov s využitím L-systémov popisuje [10].

III. Návrh riešenia

V tejto práci bol zvolený algoritmický prístup k procedurálnemu generovaniu stromov, ktorý je založený na rekurzii, pričom boli definované pravidlá pre vetvenie stromu. Výsledný generátor stromov parametrizáciou umožňuje vytvárať stromy rôznych druhov, tvarov a veľkostí vhodným nastavením parametrov. V rámci riešenia boli vyvinuté dva moduly stromov s rôznymi pravidlami pre vetvenie. Spôsoby vetvenia pre jednotlivé moduly sú priblížené na obrázku 1.



(a) **(b)**

Obr. 1: Spôsoby vetvenia stromov: **(a)** rovnomerné vetvenie, **(b)** vetvenie s primárnou vetvou

Obrázok 1(a) ukazuje symetrický model vetvenia, kde každá vetva sa rozdeľuje na dve menšie vetvy. Na druhej strane, obrázok 1(b) aplikuje vetvenie použité v obrázku 1(a) na svoje bočné vetvy, zatiaľ čo hlavná vetva pokračuje priamo nahor. V obidvoch prípadoch ide o rozdelenie na dve vetvy, teda ide o rozdvojenie, ktoré je možné aplikovať rekurzívne na každú vetvu stromu.

Návrh riešenia pre troj-dimenzionálny priestor začína definovaním segmentu kmeňa stromu, ktorý je v tomto prípade reprezentovaný valcom. Vrcholy valca sú uložené na dvoch kruhoch, ktoré sú od seba vzdialené o určitú vzdialenosť v smere rastu. Následne sú medzi vrcholmi kruhov vytvorené steny pomocou polygónov, ktoré tvoria povrch valca.

A. Vytváranie segmentov

Vytváranie segmentov zahŕňa vytvorenie vrcholov a stien, ktoré tvoria povrch valca. Pozície vrcholov sú vypočítané:

$$x = r \cdot \cos(\theta) \qquad y = 0 \qquad z = r \cdot \sin(\theta)$$

kde r je polomer kruhu, θ je uhol určujúci pozíciu vrcholu na kružnici. Výpočet θ je závislý od počtu vrcholov kruhu a indexu vrcholu:

$$\theta_i = \frac{2\pi \cdot i}{n}$$

kde i je index vrcholu, ktorý sa pohybuje od 0 po $n-1$ a n je počet vrcholov kruhu. Kruh musí byť následne orientovaný na základe orientácie predchádzajúceho kruhu a taktiež naklonený kolmo od smeru rozdelenia predchádzajúceho kruhu. Tieto operácie sú vykonané pomocou kvaternionov.

Posunutie kruhu je potrebné vykonať v smere rastu stromu a v smere kolmom od smeru rozdelenia predchádzajúceho kruhu. Najprv je aplikované posunutie v smere rastu stromu tak,

že na každý vrchol kruhu je aplikovaný vektor posunutia. Následne je aplikované posunutie kolmo od smeru rozdelenia predchádzajúceho kruhu, kde výsledný vektor posunutia je vypočítaný vektorovým súčinom medzi smerom rozdelenia a smerom rastu predchádzajúceho kruhu.

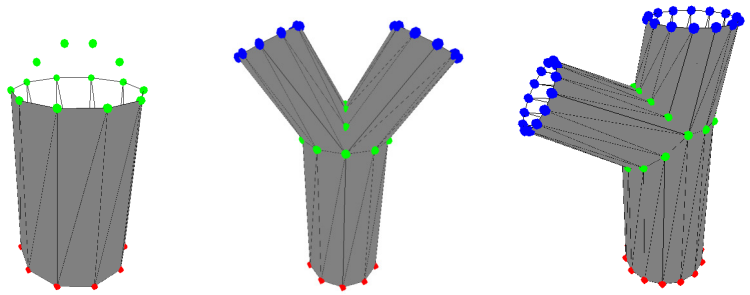
V prípade generovania vrcholov pre rozdeľujúci sa segment je potrebné vytvoriť dodatočné vrcholy, ktoré umožnia plynulý povrch medzi segmentmi. Pozície týchto vrcholov sú interpolované medzi začiatočným a koncovým bodom rozdelenia. Pri daných dvoch bodoch A a B v priestore, pozícia interpolovaného bodu P je daná vzorcom:

$$P = (1 - t) \cdot A + t \cdot B$$

kde t sa mení od 0 do 1. potom je aplikované vertikálne posunutie y:

$$y = \sin(\pi \cdot t) \cdot h$$

kde h konštanta určujúca vertikálne posunutie. Zmenou vertikálnej pozície každého interpolovaného vrcholu sú ich pozície upravené tak, že vytvárajú oblúk. Výber bodov A a B je uskutočňovaný náhodne, pričom sú vždy umiestnené oproti sebe. Tento výber je realizovaný pomocou deterministického pseudonáhodného generátora čísiel, ktorý využíva inicializačné číslo (seed).



(a) **(b)** **(c)**

Obr. 2: Vytváranie segmentov: **(a)** segment pred rozdelením, **(b)** segment po rovnomernom vetvení, **(c)** segment po vetvení s primárnou vetvou

Po vytvorení kruhov sú ich vrcholy indexované a sú vytvorené steny, ktoré tvoria povrch valca. Indexy sú uložené v zozname, ktorý je následne použitý na vytvorenie polygónov. Proces pri 6 vrcholových kruhoch, kde spodný kruh obsahuje vrcholy s indexami 0 až 5 a vrchný kruh obsahuje vrcholy s indexami 6 až 11 by prebiehal nasledovne:

$$\begin{array}{cccccc} (0,6,1), & (1,6,7), & (1,7,2), & (2,7,8), & (2,8,3), & (3,8,9), \\ (3,9,4), & (4,9,10), & (4,10,5), & (5,10,11), & (5,11,0), & (0,11,6) \end{array}$$

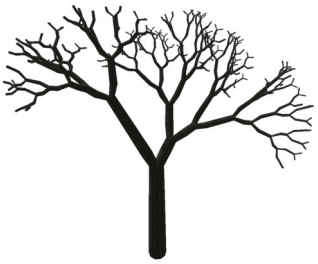
Najprv je spojený prvý spodný vrchol s prvým vrchným vrcholom a druhým spodným vrcholom. Potom je spojený druhý spodný vrchol s druhým vrchným vrcholom a tretím spodným vrcholom a takto pokračuje až kým nie sú vytvorené všetky steny.

B. Rovnomerné vetvenie

Parameter	Opis
numFaces	Počet stien kmeňa.
splitChance	Pravdepodobnosť rozvetvenia.
splitDepth	Hĺbka rekurzie kde začína rozvetvenie.
maxDepth	Maximálna hĺbka rekurzie.
radius	Polomer kmeňa.
radiusDecay	Zníženie polomeru podľa hĺbky rekurzie.
height	Výška segmentov.
heightDecay	Zníženie výšky podľa hĺbky rekurzie.
angle	Uhol rozvetvenia.
angleIncrement	Inkrementácia uhla rozvetvenia podľa hĺbky rekurzie.
leavesScale	Veľkosť listov.
seed	Základ pre generátor náhodných čísel.
color	Farba kmeňa.
leavesColor	Farba listov.
leavesTexture	Textúra listov.

Tabuľka 1: Parametre pre generovanie rovnomerne rozvetveného stromu.

Modul pre rovnomerné vetvenie stromu umožňuje nastavenie parametrov z tabuľky 1. V tomto prípade sú pri rozvetvení segmenty tvorené rovnomerne, tak ako je zobrazené na obrázku 2(b). Výsledky takto generovaných stromov sú zobrazené na obrázkoch 3 a 4.



Obr. 3: Rovnomerne rozvetvený strom bez listov.



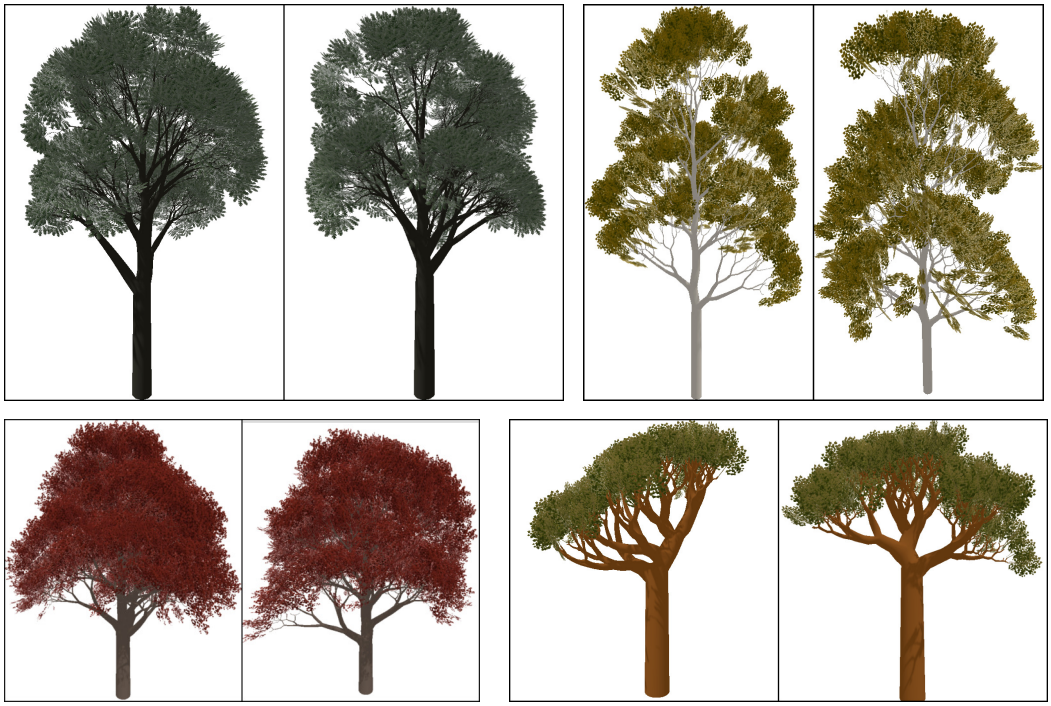
Obr. 4: Rovnomerne rozvetvené stromy.

C. Vetvenie s primárnou vetvou

Parameter	**Opis**
numFaces	Počet stien kmeňa.
splitChance	Pravdepodobnosť rozvetvenia.
radius	Polomer kmeňa.
radiusDecay	Zníženie polomeru kmeňa podľa hĺbky rekurzie.
branchRadiusDecay	Zníženie polomeru vetvy podľa hĺbky rekurzie.
height	Výška kmeňových segmentov.
heightDecay	Zníženie výšky kmeňových segmentov podľa hĺbky rekurzie.
branchHeight	Výška segmentov vetiev.
branchHeightDecay	Zníženie výšky segmentov vetiev podľa hĺbky rekurzie.
firstSegmentHeight	Výška prvého kmeňového segmentu.
angle	Uhol rozvetvenia pri prvom rozvetvení od kmeňa.
branchAngle	Uhol rozvetvenia vetiev.
branchAngleIncrement	Inkrementácia uhla rozvetvenia vetiev podľa hĺbky rekurzie.
branchDroopAngle	Uhol sklonu vetiev.
offsetDistance	Posun kmeňového segmentu od predchádzajúceho segmentu.
maxDepth	Maximálna hĺbka rekurzie pre kmeň.
branchMaxDepth	Maximálna hĺbka rekurzie pre vetvy.
leavesScale	Veľkosť listov.
seed	Základ pre generátor náhodných čísel.
color	Farba kmeňa.
leavesColor	Farba listov.
leavesTexture	Textúra listov.

Tabuľka 2: Parametre pre generovanie stromu s primárnou vetvou.

Modul pre vetvenie s primárnou vetvou umožňuje nastavenie parametrov z tabuľky 2. V tomto prípade je pri rozvetvení segmenty tvorené tak, že hlavná vetva pokračuje priamo nahor a bočné vetvy sú rozvetvené podľa zvoleného uhla, tak ako je zobrazené na obrázku 2(c). Po rozvetvení z hlavnej vetvy sú vytvorené nové vetvy, ktoré sú následne rozvetvené podobne ako pri rovnomernom vetvení. Výsledky takto generovaných stromov sú zobrazené na obrázku 5.



Obr. 5: Stromy s primárnou vetvou.

IV. Vyhodnotenie

Vyvinuté moduly umožňujú generovanie rôznorodých stromov. Na základe predchádzajúcich príkladov je možné vidieť, že vhodným nastavením parametrov je možné vytvoriť 3D modely stromov s veľkou variabilitou a náhodnosť určená generátorom náhodných čísel umožňuje vytvárať rôzne modely aj s rovnakými parametrami. Výkon aplikácie bol testovaný s rôznymi hodnotami parametrov, pričom bolo dokopy vykonaných 750 testov. Testovanie prebiehalo na počítači s nasledujúcimi systémovými parametrami:

- **Operačný systém:** Windows 11 Home 64-bit (10.0, Build 22631)
- **Výrobca:** ASUSTeK COMPUTER INC.
- **Model:** Strix GL704GW_GL704GW
- **BIOS:** GL704GW.312
- **Procesor:** Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz (12 CPUs), 2.2GHz
- **Pamäť:** 32768MB RAM

S Grafickou kartou NVIDIA GeForce RTX 2070, ktorej parametre sú:

- **Celková pamäť:** 24314 MB
- **VRAM:** 8006 MB
- **Zdieľaná pamäť:** 16308 MB

Čas potrebný na generovanie stromov bol meraný pomocou presných časových značiek. Meranie začalo pred generovaním stromu a skončilo hneď po generovaní, pričom rozdiel medzi týmito dvoma časovými značkami poskytuje čas generovania stromu. FPS bolo merané počas vykresľovania scény tak, že sa počítal počet vykreslených snímok počas stanoveného časového intervalu a následne sa vypočítal priemer z týchto hodnôt. Na meranie a zber údajov bol použitý vlastný skript v JavaScripte, ktorý využíval nástroje prehliadača a WebGL API na časovanie generovania stromov a meranie FPS počas vykresľovania.

A. Štatistická analýza

Štatistická analýza výsledkov ukázala, že zmena parametrov má vplyv na výkon aplikácie. V tabuľke 3 sú uvedené štatistiky pre čas generovania stromov a ďalej v tabuľke 4 sú uvedené štatistiky pre priemerné FPS (frames per second) alebo snímky za sekundu.

Modul	Priemerný čas	Min. čas	Max. čas	Št. odch.	25. percentil	75. percentil
evenSplitTree	129.5	10.6	5044.3	460.0	12.3	39.2
PrimaryBranchTree	239.2	14.5	5304.0	545.9	26.6	174.3

Tabuľka 3: Prehľad štatistík pre čas generovania. Uvedený čas je v milisekundách.

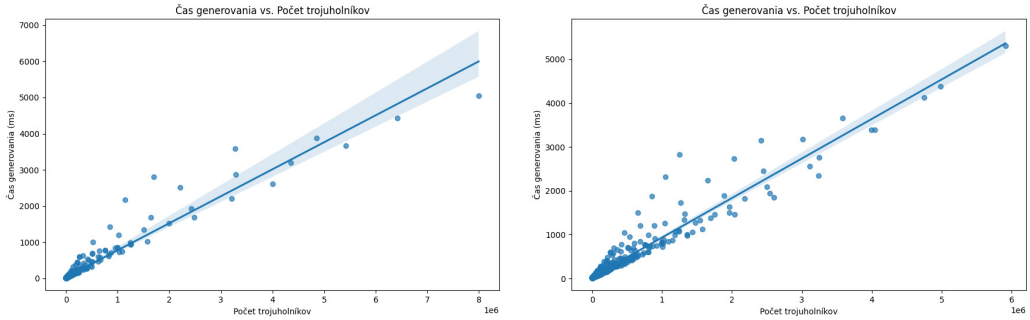
Modul	Priemer	Min	Max	Štandardná odchýlka	25. percentil	75. percentil
evenSplitTree	139.0	13.6	144.6	20.1	143.5	143.9
PrimaryBranchTree	133.8	15.6	144.4	27.2	142.3	143.8

Tabuľka 4: Prehľad štatistík pre priemerný počet snímok za sekundu.

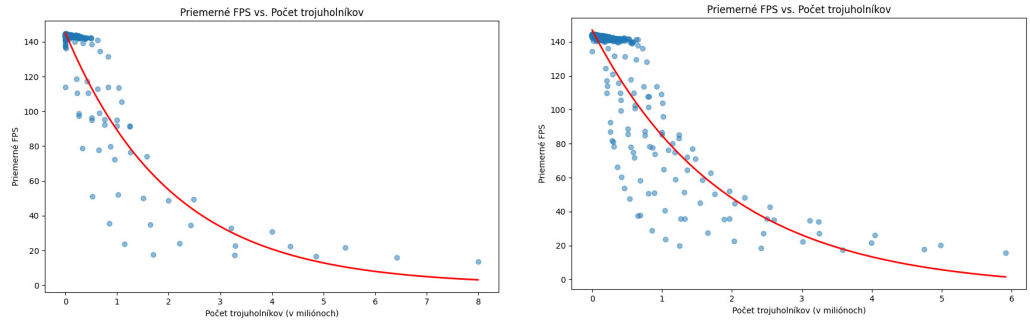
Z uvedených štatistík je zrejmé, že moduly majú podobný výkon s malými rozdielmi a väčšina stromov sa vykresľuje rýchlo, pričom v odľahlých prípadoch, môže byť vykresľovanie pomalšie.

B. Vizualizácia výsledkov

Grafy na obrázku 6 zobrazujú vzťah medzi časom generovania stromov a počtom trojuholníkov. Vizualizácia potvrdzuje údaje zo štatistík, že väčšina stromov sa generuje rýchlo, pričom len v odľahlých prípadoch je generovanie pomalšie. Extrémne prípady s dlhým časom generovania sú spôsobené takou kombináciou hodnôt parametrov, ktorá spôsobuje veľkú zložitosť stromu. Podobne grafy na obrázku 7 ukazujú, že výkon vykresľovania v počte snímok za sekundu sa u väčšiny stromov pohybuje v rozmedzí nad 120 FPS.



Obr. 6: Graf vzťahu času generovania stromov a počtu trojuholníkov pre oba typy stromov.



Obr. 7: Graf vzťahu priemerného FPS a počtu trojuholníkov pre oba typy stromov.

V. Záver

Riešenie v podobe generátora stromov umožňuje vytvárať rôznorodé 3D modely stromov s možnosťou konfigurácie parametrov prostredníctvom grafického rozhrania. Výsledná webová aplikácia je schopná generovať stromy v reálnom čase a zobrazovať ich v prehliadači, pričom je možné exportovať modely vo formáte OBJ pre ďalšie použitie v 3D grafických aplikáciách.

Vzhľadom na budúci vývoj existuje možnosť rozšírenia o ďalšie druhy stromov a pridanie pokročilejších techník pre vizualizáciu, ako je UV mapovanie pre textúrovanie stromov s bump mapami, čo by zvýšilo vizuálnu kvalitu výsledkov. Ďalším smerom vývoja by mohla byť implementácia pravidiel vetvenia pre ihličnaté stromy, ktorá by umožnila generovanie štruktúr typických pre túto skupinu druhov.

Acknowledgement

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

LITERATÚRA

[1] G. N. Yannakakis and J. Togelius, *Artificial Intelligence and Games*, 1st ed. Springer Cham, 2018.

[2] J. Togelius, E. Kastbjerg, D. Schedl, and G. N. Yannakakis, "What is procedural content generation? mario on the borderline," in *Proceedings of the 2nd International Workshop on Procedural Content Generation in Games*, ser. PCGames '11. New York, NY, USA: Association for Computing Machinery, 2011. [Online]. Available: https://doi.org/10.1145/2000919.2000922

[3] A. Amato, *Procedural Content Generation in the Game Industry*. Cham: Springer International Publishing, 2017, pp. 15–25. [Online]. Available: https://doi.org/10.1007/978-3-319-53088-8_2

[4] M. Hendrikx, S. Meijer, J. Van Der Velden, and A. Iosup, "Procedural content generation for games: A survey," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 9, no. 1, feb 2013. [Online]. Available: https://doi.org/10.1145/2422956.2422957

[5] T. Gschwantner, K. Schadauer, C. Vidal, A.
Lanz, E. Tomppo, L. di Cosmo, N. Robert, D. Englert Duursma, and M. Lawrence, "Common tree definitions for national forest inventories in europe," *Silva Fennica*, vol. 43, no. 2, pp. 303–321, 2009.

[6] O. Deussen and B. Lintermann, *Plants*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 9–26. [Online]. Available: https://doi.org/10.1007/3-540-27104-X_2

[7] A. Runions, B. Lane, and P. Prusinkiewicz, "Modeling trees with a space colonization algorithm," *Natural Phenomena*, vol. 7, no. 63-70, p. 6, 01 2007.

[8] R. Ratul, S. Sultana, J. Tasnim, and A. Rahman, "Applicability of space colonization algorithm for real time tree generation," 12 2019, pp. 1–6.

[9] A. Lindenmayer, "Mathematical models for cellular interactions in development i. filaments with one-sided inputs," *Journal of Theoretical Biology*, vol. 18, no. 3, pp. 280–299, 1968. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022519368900799

[10] R. Sun, J. Jia, and M. Jaeger, "Intelligent tree modeling based on l-system," in *2009 IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design*. IEEE, 2009, pp. 1096–1100.

Numerická optimalizácia pomocou aritmetického optimalizačného algoritmu

[1]*Martin ONDO-EŠTOK,* [2]*Marián MACH*

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

[1]martin.ondo-estok@student.tuke.sk, [2]marian.mach@tuke.sk

***Abstrakt* — Článok sa zameriava na problematiku použitia metaheuristických algoritmov pre riešenie úloh numerickej optimalizácie. Pozornosť je venovaná algoritmu AOA, ktorý je založený na matematickej inšpirácii, v rámci ktorej algoritmus využíva vlastnosti základných aritmetických operácií z hľadiska explorácie a exploatácie v priestore možných riešení. Článok sa zaoberá základnou podobou tohto algoritmu. V rámci experimentovania bolo realizovaných niekoľko experimentov s hľadaním vhodných hodnôt niektorých parametrov tohto algoritmu ako aj jeho porovnaním s vybranými metaheuristickými algoritmami v rámci optimalizácie množiny zvolených testovacích funkcií.**

***Kľúčové slová* — prírodne inšpirovaný algoritmus, AOA, metaheuristika, numerická optimalizácia, priestor prehľadávania**

I. Úvod

Metaheuristické algoritmy sa používajú pre účely hľadania riešení optimalizačných úloh pomocou prehľadávania veľkého množstva kandidátov na riešenie optimalizačného problému, ktorí spolu vytvárajú priestor prehľadávania. Tieto algoritmy sa inšpirujú vynálezmi prírody – mnohé biologické (napr. prirodzený výber a prežitie, rôzne vzory chovania, spôsoby komunikácie organizmov) alebo fyzikálne (napr. gravitácia, teplotné procesy) fenomény sa stali hlavnou ideou v pozadí viacerých metaheuristických algoritmov [1]. Jednou z možných inšpiračných oblastí je aj matematika. Do skupiny matematicky inšpirovaných prehľadávacích algoritmov patrí aj AOA (Aritmetický Optimalizačný Algoritmus) [2] a jeho modifikované varianty [3].

Tento článok sa zameriava na základnú podobu algoritmu AOA. Jeho štruktúra je nasledovná – v druhej sekcii sa venuje matematickej inšpirácii a popisu metódy, pričom pozornosť je venovaná voľbe a popisu exploračnej aj exploatačnej časti algoritmu. V tretej sekcii je uvedená séria experimentov s daným algoritmom na úlohe hľadania riešenia globálneho minima skupiny vybraných optimalizačných funkcií.

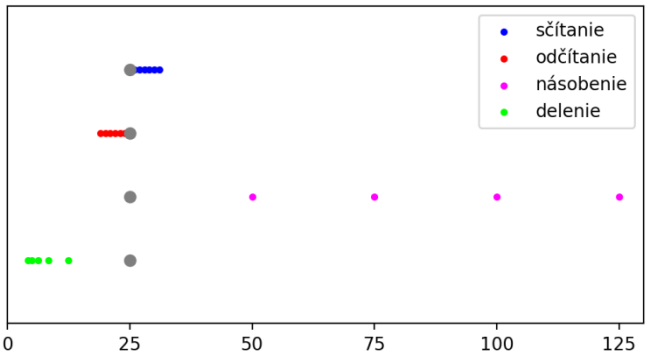
II. Aritmetický optimalizačný algoritmus

A. Matematická inšpirácia

Hlavná inšpirácia algoritmu vychádza z používania základných aritmetických operácií, ktoré sú tradičným začiatkom každej výučby počítania s číslami.

Ak na nejakú hodnotu aplikujeme aritmetickú operáciu (spolu s nejakým druhým operandom), tak výsledkom je nová hodnota. Pôvodná aj nová hodnota môžu reprezentovať pozíciu v priestore prehľadávania – aplikácia aritmetickej operácie bude mať za následok zmenu pozície v tomto priestore. A veľkosť tejto zmeny, pri fixovanej hodnote druhého operandu, závisí od samotnej operácie.

Uvažované vlastnosti týchto operácií sú ilustrované na obr. 1, ktorý zobrazuje posun pozície, reprezentovanej šedým bodom, pri použití niekoľkých druhých operandov (tých istých pri všetkých operáciách). Z obrázku je zrejmé, že operácia sčítania/odčítania spôsobuje menší posun ako operácie násobenia alebo delenia.



Obr. 1 Prehľadávacie vlastnosti aritmetických operácií.

Preto, ak sa algoritmus snaží o exploráciu, keď chce realizovať väčšie zmeny polohy v priestore prehľadávania kvôli preskúmavaniu aj vzdialenejších oblastí, tak použije operáciu násobenia alebo delenia. Ak naopak je jeho cieľom exploatácia nadobudnutých znalostí a teda chce preskúmať lokálne okolie nejakej polohy, tak naopak použije operáciu sčítania alebo odčítania.

B. Základná štruktúra algoritmu

Vo svojej podstate sa jedná o jednoduchý populačný algoritmus, ktorý pracuje s populáciou riešení. Táto populácia je na začiatku (v čase $t=0$) generovaná náhodným spôsobom a počas nasledujúcich iterácií sa vyvíja – v každej iterácii je vygenerovaná nová populácia N riešení, pričom nové riešenia sú pomocou aritmetických operácií odvádzané z predchádzajúcich riešení. Po vykonaní zadaného počtu iterácií T_{max} algoritmus končí s tým, že navonok poskytne najlepšie nájdené riešenie. Pseudokód algoritmu má tvar:

```
t = 0
inicializuj populáciu N riešení náhodne
while t <= Tmax
    vyhodnoť všetky riešenia
    urči doteraz najlepšie riešenie
    for i=1 až N
        for každú pozíciu riešenia
            generuj náhodnú hodnotu r1
            if r1 > Limit
                generuj riešenie exploračným spôsobom
            else
                generuj riešenie exploatačným spôsobom
            end if
            uprav polohu vo vnútri hraníc
        end for
    end for
    t = t + 1
end while
vráť najlepšie riešenie
```

Keďže každá pozícia generovaného riešenia sa generuje spôsobom nezávislým od spôsobu generovania iných pozícií a samotné generovanie môže byť realizované jednou zo štyroch základných aritmetických operácií, tak jedno riešenie môže byť zmenené na jedno zo 4^n riešení, odvoditeľných z daného riešenia.

O tom, či nové riešenie sa bude generovať v exploračnej alebo exploatačnej fáze algoritmu, sa rozhoduje náhodným spôsobom – náhodné číslo r1 je porovnávané voči limitu označovanému ako MOA (Math Optimizer Accelerated). Jeho hodnota závisí od čísla iterácie podľa vzťahu

$$MOA(t) = Min + t\frac{Max-Min}{T_{max}} \quad (1)$$

kde sa lineárne zvyšuje s narastajúcim číslom iterácie. To znamená, že s narastajúcim časom sa zvyšuje pravdepodobnosť prechodu do exploatačnej fázy a znižuje pravdepodobnosť prechodu do exploračnej fázy. Toto reflektuje často využívaný princíp na začiatku hľadania preferovať

exploráciu, zatiaľ čo ku koncu sa hľadanie viac sústreďuje na prehľadávanie identifikovaných sľubných lokálnych oblastí.

C. Exploračná fáza algoritmu

Táto fáza pre generovanie nových riešení používa dve stratégie – aritmetické operácie násobenia a delenia. Výber medzi nimi je realizovaný na pravdepodobnostnej báze s rovnakými pravdepodobnosťami. V prípade násobenia je hodnota generovaného riešenia *x* na pozícii *j* určovaná podľa vzťahu

$$x_j(t+1) = best_j \left[MOP\left(\left(ub_j - lb_j \right) b + lb_j \right) \right] \tag{2}$$

a v prípade delenia podľa

$$x_j(t+1) = \frac{best_j}{\left[MOP\left(\left(ub_j - lb_j \right) b + lb_j \right) \right]} \tag{3}$$

kde $best_j$ je hodnota najlepšieho riešenia na *j*-tej pozícii a lb_j a ub_j je dolná a horná hranica definičného oboru *j*-tej dimenzie. Parameter *b* je riadiacim parametrom, pre ktorý autori algoritmu doporučujú hodnotu 0.5 a MOP (Math Optimizer Probability) definovaný podľa

$$MOP(t) = 1 - \left(\frac{t}{T_{max}} \right)^{\frac{1}{\alpha}} \tag{4}$$

je parameter, ktorý s narastajúcim časom zmenšuje svoju hodnotu z počiatočnej hodnoty 1.0 ku konečnej 0.0.

D. Exploatačná fáza algoritmu

Táto fáza pre generovanie nových riešení používa dve stratégie – aritmetické operácie sčítania a odčítania. V prípade sčítania je hodnota generovaného riešenia *x* na pozícii *j* určovaná podľa vzťahu

$$x_j(t+1) = best_j + MOP\left(\left(ub_j - lb_j \right) b + lb_j \right) \tag{5}$$

a v prípade odčítania podľa vzťahu

$$x_j(t+1) = best_j - MOP\left(\left(ub_j - lb_j \right) b + lb_j \right) \tag{6}$$

Aj v tejto fáze je výber medzi stratégiami realizovaný na pravdepodobnostnej báze s rovnakými pravdepodobnosťami.

III. VYBRANÉ EXPERIMENTY

A. Testovacie problémy

Pre experimentovanie s algoritmom [4] bolo použitých niekoľko pomerne známych funkcií, ktoré sú často využívané pre testovanie a porovnávanie optimalizačných algoritmov [5, 6].

1. Ackleyho funkcia je definovaná vzťahom

$$f(x) = -a \exp\left(-b \sqrt{\frac{1}{d} \sum_{i=1}^{d} {x_i}^2} \right) - \exp\left(\frac{1}{d} \sum_{i=1}^{d} cos(c x_i) \right) + a + exp(1) \tag{7}$$

kde pre jednotlivé konštanty boli použité typicky používané hodnoty $a = 20$, $b = 0.2$ a $c = 2\pi$. Globálny extrém s hodnotou $f_{min} = 0$ sa nachádza v počiatku súradnicového systému. Pre testovanie bolo použité $d = 2$, definičný obor každej dimenzie bol [-32,32].

2. Shubertova dvojdimenzionálna funkcia je definovaná vzťahom

$$f(x) = \left(\sum_{i=1}^{5} i \ \cos\left((i+1)x_1 + i \right) \right) \left(\sum_{i=1}^{5} i \ \cos\left((i+1)x_2 + i \right) \right) \tag{8}$$

kde globálny extrém má hodnotu $f_{min} = -186.7309$, pričom funkcia má mnoho lokálnych aj globálnych miním. Pre testovanie bol pre každú dimenziu použitý definičný obor [-10,10].

3. McCormicova dvojdimenzionálna funkcia je definovaná vzťahom

$$f(x) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1 \tag{9}$$

kde globálny extrém tejto funkcie má hodnotu $f_{min} = -1.9133$ v bode [-0.54719, -1.54719. Pre testovanie bol pre každú dimenziu použitý definičný obor [-5,5]. Funkcia má síce iba jedno minimum, ale s veľmi pomalým klesaním k nemu.

B. Nastavenie parametrov algoritmu

Algoritmus má niekoľko parametrov, ktoré ovplyvňujú jeho chovanie: T_{max} (dobu prehľadávania priestoru riešení), N (veľkosť populácie), *Min* a *Max* (pre výpočet MOA), a (pre zmeny MOP) a b (riadiaci maximálnu hodnotu zmeny). Zamerali sme sa na parametre a a b, ostatné parametre boli nastavené na hodnoty doporučené autormi algoritmu. Dosiahnuté výsledky pri použití Shubertovej funkcie sú uvedené v tab. 1.

Tabuľka 1
Porovnanie výkonu AOA na Shubertovej funkcii pre rôzne hodnoty parametrov a a b

Parametre		Kvalita riešení		
B	A	Najlepšie	Najhoršie	Priemer
0.1	0.5	-30.95254194	-9.026329059	-19.861139306
0.1	5	-177.7875504	-3.59550927	-73.03583163
0.1	10	-179.4448735	-31.82743809	-93.29454365
0.5	0.5	-92.22663678	-11.28027307	-26.77594141
0.5	5	-111.2751592	-12.35601548	-38.81052131
0.5	10	-93.12857028	-3.681894319	-37.72030396
0.9	0.5	-64.45845307	-14.36868542	-33.5285611
0.9	5	-163.4331923	-5.2569613	-54.41393868
0.9	10	-185.2010166	-11.9022138	-103.3028232

Na tejto funkcii sa ako najlepšia kombinácia hodnôt parametrov ukázala posledná kombinácia. Na Ackleyho funkcii to bola kombinácia 4 a pre McCormicovu funkcia bola zase najlepšou kombinácia 3. Celkové výsledky sú v tab. 2.

Tabuľka 2
Porovnanie kombinácií parametrov na všetkých testovacích funkciách.

	Kombinácia parametrov								
Funkcia	1	2	3	4	5	6	7	8	9
Ackley	9	7	4	1	2	3	8	5	6
Shubert	9	3	2	8	5	6	7	4	1
McCormic	9	3	1	8	5	2	6	7	4
Celkové poradie	9	5	1	7	4	2-3	8	6	2-3

Tabuľka ukazuje poradia jednotlivých kombinácií parametrov (hodnotu 1 má najlepšia kombinácia a hodnotu 9 zase najhoršia kombinácia) pri použití testovacích funkcií. Výsledné poradie je určované na základe súčtu parciálnych poradí. Je zrejmé, že celkovo najlepšia kombinácia nie je najlepšou voľbou pre každú funkciu, avšak pri každej z funkcií patrí medzi tie lepšie voľby.

Povedané inak, nie je možné nájsť univerzálne najlepšie nastavenie parametrov algoritmu, ktoré by bolo optimálnym nastavením platným pre ľubovoľnú optimalizovanú funkciu. Ale na druhej strane je možné nájsť pomerne dobré kompromisné nastavenie, použiteľné pre viaceré funkcie.

C. Porovnanie s inými metaheuristickými algoritmami

Pre vzájomné porovnanie algoritmu AOA s inými algoritmami sme vybrali nasledujúce algoritmy: SCA (sínus-cosínus algoritmus), PSO (optimalizácia rojom častíc), BAT (netopierí algoritmus) a GA (genetický algoritmus). Výber bol realizovaný na základe podobnosti týchto algoritmov s prezentovaným algoritmom. Pre realizáciu týchto algoritmov sme použili voľne dostupný toolbox [7].

Pri optimalizácii daných funkcií sme menili veľkosť populácie a zodpovedajúcim spôsobom aj počet iterácií (pri väčšej populácii bol použitý vyšší počet iterácií). V prípade malej populácie obsahovala populácia iba 10 riešení a algoritmy realizovali iba 10 iterácií. Dosiahnuté výsledky sú uvedené v tab. 3.

Tabuľka 3
Porovnanie algoritmov na základe priemerných riešení – populácia 10.

	Algoritmy				
Funkcia	AOA	SCA	PSO	BAT	GA
Ackley	1	3	2	5	4
Shubert	5	3	4	1	2
McCormic	4	2	1	5	3
Celkové poradie	4	2	1	5	3

V tabuľke možno nájsť poradie jednotlivých algoritmov, dosiahnuté pre jednotlivé funkcie. Prezentovaný algoritmus AOA patril skôr medzi tie algoritmy, ktoré podávali slabší výkon. Je evidentné, že malá populácia tomuto algoritmu nesvedčí.

Postupne sme experimentovali s väčšími populáciami až do hodnoty 100. Výsledky pre takto veľkú populáciu v spojitosti so 100 iteráciami sú v tab. 4.

Tabuľka 4
Porovnanie algoritmov na základe priemerných riešení = populácia 100.

	Algoritmy				
Funkcia	AOA	SCA	PSO	BAT	GA
Ackley	1	2	3	5	4
Shubert	5	4	2	1	3
McCormic	2	3	1	5	4
Celkové poradie	2	3	1	4-5	4-5

Z tabuľky vidno, že výkonnosť algoritmu AOA vzrástla, väčšia populácia algoritmu prospieva. Aj keď neporazil všetky algoritmy, s ktorými sa porovnával, s väčšou populáciou sa radí k tým úspešnejším.

IV. ZÁVER

Vykonané experimenty potvrdili, že algoritmus založený na základných štyroch aritmetických operáciách je použiteľný v úlohe optimalizačného nástroja pre účely riešenia úloh numerickej optimalizácie. Aj keď v prípade malých populácií jeho výkon neoslnil, pri väčších populáciách sa prejavil ako užitočný nástroj. Jeho výhodou je zrozumiteľnosť vnútorných pochodov daná jednoduchosťou použitých matematických operácií, ktorá vedie na jednoduchú implementáciu algoritmu. Bonusom je explicitný výber medzi exploračnou a exploatačnou fázou, ktorý môže byť jednoduchým spôsobom modifikovaný pre účely prispôsobenia algoritmu tvaru optimalizovanej funkcie.

POĎAKOVANIE

Táto práca vznikla s podporou Vedeckej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0394/22.

LITERATÚRA

[1] M. Abdel-Basset, L. Abdel-Fatah, A.K. Sangaiah, Metaheuristic Algorithms: A Comprehensible Review, In: *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*, 2018, Academic Press, 185-231.
[2] L. Abualigah et al., The Arithmetic Optimization Algorithm, *Computer methods in applied mechanics and engineering*, 2021, vol. 376, 38 pages.
[3] A. Kaveh, K.B. Hamedani, Improved arithmetic optimization algorithm and its application to discrete structural optimization, *Structures*, 2022, vol 35, 748-764.
[4] M. Ondo-Eštok, *Aritmetický optimalizačný algoritmus pre numerickú optimalizáciu*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2022.
[5] Virtual Library of Simulation Experiments: Test Functions and Datasets. https://www.sfu.ca/~ssurjano/optimization.html [prístupné 17.5.2024]
[6] Test Functions Index, http://infinity77.net/global_optimization/test_functions.html [prístupné 17.5.2024]
[7] EvoloPy: An open source nature-inspired optimization toolbox for global optimization in Python, http://github.com/7ossam81/EvoloPy [prístupné 17.5.2024]

Monitoring the occupancy of parking spaces via UAVs

[1]Eva KRIŽALKOVIČOVÁ, [2]Ondrej Kainz, [3]Miroslav Michalko, [4]Miroslav Murin

[1,2,3,4]Department of Informatics, Faculty of Electrical Engineering and Informatics
Technical University of Košice, Slovak Republic

[1]eva.krizalkovicova@student.tuke.sk, [2]ondrej.kainz@tuke.sk,
[3]miroslav.michalko@tuke.sk, [4]miroslav.murin@tuke.sk

Abstract **– The main goal of this study was the creation of a software solution for monitoring free parking spaces using a UAV. In the analytical part of the work, the analysis of the available algorithms for the detection and recognition of objects in the image was performed. Subsequently, a solution proposal for the software part was developed. To detect free parking spaces, we used the YOLO algorithm. We have successfully trained the algorithm to detect free rectangular parking spaces, since testing the functionality of the pre-trained YOLO model for car detection was unsuccessful, as the model was not reliable for car detection. The user interface consists of two parts, the user interface for the parking lot owner and the user interface for the user.**

Keywords **– detection, UAV, user interface, YOLO**

I. Introduction

Unmanned aircraft, otherwise known as drones, have many uses. Most people may associate the term with bird's eye view videos, but there is much more to it than that. Unmanned aerial vehicles are being used in more and more areas. The use of unmanned aerial vehicles for rescue, agriculture, construction industry and traffic monitoring are less well known. In order to be able to use the data that the unmanned aircraft is able to provide us, it is necessary to use certain algorithms. An unmanned aircraft will provide us with an image, either photos or video, we can also improve such an aircraft with some sensors. In this work, however, we will focus on algorithms that are able to detect movement in real time, just from the already mentioned video. With these algorithms, we distinguish two types of methods, one-stage and two-stage, methods from both areas are mentioned in the work. At the end, we will evaluate which algorithm we chose to create the solution and why we chose that particular one.

We will deal with the specific use of unmanned aerial vehicles to monitor the occupancy of parking spaces. An unmanned aircraft scans the parking lot, then we train the algorithm model so that it is able to detect an empty space. After successfully mastering the training of the algorithm, we will focus on creating a suitable and intuitive user interface that will use the results from the already mentioned trained algorithm and display them to the user. The result of this work will have a positive impact on regulating problems with finding a parking space, thus saving time and problems for drivers in this area. The driver will be informed in advance of the number of free parking spaces. It will have available information about several parking lots and their occupancy at any time and anywhere. Thanks to the software solution, the driver will know where to park and save considerable time. Also, this solution will have a positive impact for parking lot owners, as thanks to information about availability, drivers will know whether to use their parking lot.

II. Analysis and comparison of algorithms for detection of objects in the image

A. Two-stage methods

1) R-CNN: As the paper [5] discusses the R-CNN, it is a method where we use a selective search to extract only 2000 regions from the image - region suggestions. Therefore, now instead of trying to classify a large number of regions, you can only work with 2000 regions. These 2000 region suggestions are generated using the selective search algorithm. These 2000 candidate region

suggestions are then fed to a convolutional neural network, which produces a 4096-dimensional feature vector as output. The CNN works as a feature extractor and the output dense layer consists of features extracted from the image, and the extracted features are fed to the SVM to classify the presence of the object within the candidate region proposal. In addition to predicting the presence of an object within the region designs, the algorithm also predicts four values that are offset values to increase the accuracy of the bounding box. For example, given a region suggestion, the algorithm would predict the presence of a person, but that person's face in that region suggestion might be halved. The offset values therefore help in adjusting the bounding box of the region design. There are several problems with this method. Training such a network requires a lot of time, since it involves the classification of 2000 suggested regions for one image. This algorithm cannot be implemented in real time, as the time for the test image is estimated to be 47 seconds. Last but not least, it is a selective search algorithm that is fixed, and thus no learning takes place at this stage. This could lead to the creation of poor candidate region designs.

2) Fast R-CNN: Gad in [4] states that this method was developed by Ross Girshick, a UI researcher at Facebook and a former researcher at Microsoft. Fast R-CNN solves several problems in R-CNN. It is clear from the name that it will be an increase in speed compared to the original solution. Ross Girshick, proposed a new layer called *ROI Pooling* that extracts feature vectors of equal length from all designs (i.e. ROIs) in the same image. Compared to R-CNN, which has multiple phases (region proposal generation, feature extraction, and SVM classification), fast R-CNN builds a network that has only one phase. Fast R-CNN shares computations (i.e. convolutional layer computations) across all designs (i.e. ROIs) instead of doing computations for each design independently. This is done using a new *ROI Pooling* layer that makes fast R-CNN faster than R-CNN. Fast R-CNN does not cache the extracted features and therefore does not need as much disk space compared to R-CNN which needs hundreds of gigabytes. Fast R-CNN is more accurate than R-CNN. Despite the advantages of fast R-CNN, there is still its critical drawback. Producing regions still relies on a time-consuming selective search algorithm. This method cannot be tailored to a specific object search task and therefore may not be accurate enough to detect all target objects in the dataset.

3) Faster R-CNN: We learned from [9]'s post that it is an extension of fast R-CNN. Again, this is an R-CNN speedup. In this case, it is thanks to the regional design network. The Regional Design Mesh is a fully convolutional mesh that generates designs with different scales and aspect ratios. This network implements neural network terminology with an emphasis on object detection (fast R-CNN) telling where to look. Instead of using image pyramids (i.e. multiple instances of an image but at different scales) or filter pyramids (i.e. multiple filters with different sizes), the concept of anchor boxes is introduced. An anchor box is a reference box with a specific scale and aspect ratio. With multiple reference anchor frames, there are then multiple scales and aspect ratios for a single area. It can be thought of as a pyramid of reference anchor boxes. Each area is then mapped to each reference anchor box, thus detecting objects at different scales and aspect ratios. Convolutional computations are shared across this network, and via fast R-CNN, reducing computation time.

4) Mask R-CNN: In the papers [12] and [6] was described how Mask R-CNN works. Mask R-CNN was created using faster R-CNN. Faster R-CNN has two outputs for each candidate object, namely the class label and the *offset* of the bounding box. Mask R-CNN adds a third branch that gives the mask of the object. The additional mask output is different from the class and bound outputs, requiring the extraction of a much finer spatial distribution of the object. Mask R-CNN is easy to train, it is a very efficient method. It adds only a small overhead to the faster R-CNN. Another advantage is that it is easy to generalize to other tasks. The key element of R-CNN masks is pixel-to-pixel alignment, which is the main missing element in the previously mentioned R-CNNs.

B. One-stage methods

1) SPP: The contribution of [13] and [14] describe the functioning of the SPP (Spatial Pyramid Pooling) algorithm. SPP was created to improve existing deep convolutional neural networks (CNN). These networks required a fixed image size to function properly, which can reduce accuracy and create detection problems. The SPP architecture should eliminate this requirement. SPP is also good due to its ability to reduce the deformation of objects. This method provides more than a hundred times speedup over SPP. In SPP it is about adding a SPP layer. This layer does not need a specific image size as mentioned earlier. It is proven that even training data on different image sizes has a positive effect on object detection. These spatial stacks are sized proportionally to the size of the image, so the number of stacks is fixed regardless of the size of the image. Therefore, if we want to use object detection from images of any size, it is necessary

to replace the last pooling layer, after the last convolutional layer, with SPP. The output of SPP is kM-dimensional vectors with the indicated number of stacks as M (k is the number of filters in the last convolutional layer). Vectors with fixed dimensions are entering the last layer.

2) SSD: The sources [2] and [7] describe functioning of the SSD algorithm. SSD (Single shot detector) is designed to detect objects in real time. SSD speeds up the process by eliminating the need for a network for region design. To recover from the drop in accuracy, SSD uses several enhancements, including multi-scale and default box functions. These improvements allow SSDs to match the accuracy of faster SSDs using lower resolution images, further increasing speed. The very name of the model tells us about its functionality. The SSD model detects an object in a single pass through the input image, unlike other models that pass the image more than once to obtain the output detection.

3) YOLO: From the articles [3] and [1] we learned about the functioning of the YOLO algorithm. YOLO (You Only Look Once) is a popular real-time object detection algorithm. YOLO combines what was once a multi-step process with a single neural network to perform classification and bounding box prediction for detected objects. It is heavily optimized for detection performance, and as a result can run much faster than running two separate neural networks that are separated for object detection and classification. It does this by redesigning traditional image classifiers for use in the regression task of identifying bounding boxes for objects. YOLO can work faster than real time because it can work at a speed of 45 frames per second. This speed makes it a very suitable algorithm for applications that require real-time detection. As the name suggests, it looks at the entire image at once and only once, allowing it to capture the context of detected objects. This feature reduces the number of false positive detections that happen with R-CNNs, since R-CNNs look at different parts of the image separately. In addition, YOLO can generalize representations of different objects, making it more usable in a variety of new environments. The YOLO algorithm works by dividing the image into N grids, each of which has the same SxS dimensional area. Each of these N grids is responsible for detecting and locating the object it contains. Correspondingly, these grids predict the B coordinates of the bounding box relative to their cell coordinates, along with the label of the object and the probability that the object is in the cell. This process greatly reduces the computation because both detection and recognition are provided by the cells from the image, but it introduces many duplicate predictions due to multiple cells predicting the same object with different bounding box predictions. YOLO uses non-maximum suppression to solve this problem. In Non Maximal Suppression, YOLO suppresses all bounding boxes that have a lower probability score. YOLO achieves this by first looking at the probability scores associated with each decision and selecting the largest one. It then suppresses bounding boxes that have the largest intersection with the current high-probability bounding box. This step is repeated until the final bounding boxes are obtained.

4) HOG: The contributions [11], [15], [8] and [10] describe detailed functionality of HOG (Histogram of Oriented Gradients). The HOG detection method, despite its age, is still a used method. It is mainly used for face and image detection. The HOG descriptor focuses on the structure or shape of the object. It is better than any edge descriptor because it uses both magnitude and gradient angle to calculate features. Generates histograms for image regions using gradient sizes and orientations. A feature descriptor is an image representation that simplifies the image by extracting useful information and removing redundant information. A flag descriptor typically converts an image of size (width times height times channels) into a vector of length n. For example, with the HOG method, the feature descriptor can have an output image of size 64 x 128 x 3, and the output feature vector has a length of 3780. Of course, the HOG descriptor can also be calculated for other sizes. The feature vector has important uses in tasks such as image recognition and object detection. One of the disadvantages of HOG is that its computation speed is delayed in object detection for large images because it uses the *sliding window* technique to extract features from each pixel of the image. Thus, the accuracy is not highly reliable compared to current convolutional neural networks. However, HOG is still preferred for object detection in the computer vision industry due to its speed and calculation accuracy. HOG is widely used for pedestrian detection and medical image analysis.

C. Used algorithm

Real-time recognition is required for the proposed solution, which will deal with the recognition of free parking spaces from drone video. For this reason, when deciding, we focused on algorithms that had this requirement. We have a choice of several one-step methods that work in real time. In the end, we decided on the YOLO algorithm, the reason for this choice was, as already mentioned, real-time detection, and we considered the disadvantages such as poor detection results for small objects. For our particular program, where it will be necessary to recognize primary car and an

empty parking space, these disadvantages should not be so limiting. The choice of this algorithm was also influenced by the knowledge that the YOLOv5 model already detects cars, and we will be able to use this in our work. Last but not least, there is very good documentation for this algorithm, which will help us with training it.

III. Design and implementation of the solution

We outline the comprehensive approach taken in the design and implementation of a software solution aimed at detecting free parking spaces using a trained algorithm. The primary objective is to facilitate the digitization of parking spaces and provide real-time occupancy information through an accessible web application. The solution integrates various components, including data collection, algorithm training, database creation, and user interface development, to ensure an effective and user-friendly system. The research focuses on the data output from unmanned aerial vehicles (UAVs) rather than any specific type of UAV. This allows for flexibility in data collection and ensures that the solution can be adapted to various UAV platforms, provided they can capture the necessary footage of parking spaces.

A. Designing solution and training algorithm

After choosing a suitable algorithm, we will focus on understanding its functioning in practice. It is necessary to train this algorithm for the most accurate detection of the object we have chosen, which in our case is a free parking space. First we need to get the data. We can get the data needed for detection, either from freely available sources on the Internet or by recording them ourselves. Since in our work we are considering the use of an unmanned aircraft to detect these places, we will try to obtain data recorded by an unmanned aircraft. Subsequently, we will use this data in the trained algorithm, part of the output of which will be the storage of this data in a database, with which we will be able to work further in the user interface. The user interface will have two purposes, one will be to make the digitization of parking lots available to the owners of such objects, and the second purpose will be a clear representation of the acquired data to users.

The difference in approach in user interfaces is in the use of a trained algorithm model. The user will use the data found in the database, while the client will use the trained algorithm model for their creation and subsequent storage in the database. In our work, the user represents a person who will use the software solution to obtain information about parking spaces, the prerequisite is a driver or a person entrusted with the organization of order in places where parking spaces are located. The client represents a person who is interested in digitizing the parking lot and thus providing users with information about the occupancy of his parking lot. The result will be a benefit for both the client and the user, as the client will have more customers thanks to users being informed about parking options. In this way, the user will be deprived of the stress and frustration arising from problems with parking in frequented places.

We chose a suitable procedure for creating a functional software solution. We collected data, which we further used for data annotation. The data consists of photos of parking lots with different parking options, but also of a video taken by an unmanned aircraft. We then divided this video into frames using a freely available program (*frames*). So this video was already suitable for annotation. Next, it is necessary to train the algorithm for detecting free parking spaces. Training will be performed until we can successfully detect free parking spaces from the available photos. After successfully training the model, the part of data storage follows. To store data, it is necessary to create a database in which we will store results based on the name of the object - parking lot. After the successful creation of the database, we will create a functional user solution in the form of a web application. The web application will have two functions namely providing a user interface for the client and also a user interface for the user. Solution procedure in steps:

1) Data collection.
2) Data annotation.
3) Algorithm training.
4) Creating a database.
5) Creating a user solution for the client.
6) Creating a user solution for the user.

We used the freely available Roboflow tool for data annotation, from which we then exported the data in yolov5 format.

For the database, we used firebase. We chose this solution because this database works in real time, which is a key attribute for our software solution.

Training the YOLO algorithm has one big advantage, and as already mentioned, very good documentation, which also weighed in the selection of available algorithms. A manual created in

Google Colab software is available for training the algorithm. The training of the algorithm will take place in several steps:

1) Installing requirements.
2) Building the data set.
3) Training our own YOLOv5 model.
4) Performance evaluation of our model.
5) We will run our trained model on specific data to test its accuracy.
6) Download the trained model.

B. Testing algorithm

We tested a model of the algorithm that was already trained from YOLO and already included car detection.

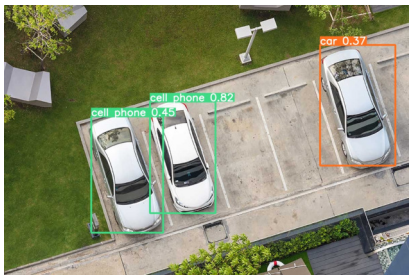


Fig. 1 Result of detection with YOLO model



Fig. 2 Result of detection with YOLO model

From figures provided 1 and 2, we can see, that for some cars the results were correct, but for other even when the bounding box was correct the label of the object was not. Due to these inconsistencies we concluded that this trained model is not suitable for our software solution.



Fig. 3 The result of our trained model image



Fig. 4 The result of our trained model video

On figure 3 and 4, we can see that the model we trained was successful. Figure 3 is the result of an image testing, and figure 4 is the result of testing on live camera. Since these are the results we decided it is ready to be used for the web application.

C. Creating web application

To make our trained model useful, we created a user interface for it. The current interface is designed for two users. One of the users is a client who has a parking lot and is interested in digitizing it in this form. The second user is the user of the application himself, who uses it to secure information about free parking spaces. The web application is created in Python using the Streamlit library.

The user interface for the client consists of a login page, after logging in the client will be taken to the main page, which consists of a form where information such as the name, location and path to the source of the image or video must be entered. There is also a preview of the card, which is identical to what the user will see, and the TEST, START and STOP buttons.

The user interface for the user does not include a login page, as the data is accessible to everyone. On the main page, the user sees cards that contain data about each parking lot, the most important data being the number of free spaces.

IV. Conclusion

The aim of the work was to analyze the available algorithms for object detection in real-time and the subsequent creation of a software solution for tracking free parking spaces. The research did not focus on any specific UAV but rather output from the UAV. After analyzing the available algorithms for object detection in real-time, we decided to use the YOLO algorithm in our work. This algorithm met our requirements, was able to detect objects in real-time, and had good documentation available to create our own trained model, which we further used in the detection of free parking spaces. First, it was necessary to collect data, which we subsequently annotated. After annotating the data, we used this annotated data to train the algorithm. We decided to focus on detecting standard rectangular parking spaces because training an algorithm to recognize all types of parking spaces would require too much data. After successfully training the algorithm, we continued with the creation of the user interface. We have divided the user interface into two user interfaces that differ in their functionality. One solution serves to obtain information such as the image source or photo, which was used to detect places and store them in the database. The second solution has the task of clearly visualizing this data to the user.

We consider the addition of the possibility of digitizing parking lots to be a major benefit of our work. Currently, we face a lot of traffic jams in busy places due to the lack of parking spaces and the subsequent search for free parking spaces. Our solution could at least partially solve this problem. The user could thus avoid these traffic jams and head straight to the parking lot that has available free spaces. The solution might also have a positive impact on parking lot owners, as thanks to the digitalization of their parking lots, the user could choose their parking lot. The organizers of various events would also have a significant benefit, as they could digitize their parking lot only in the event of an ongoing event thanks to the UAV. In this way, they would not have to invest in a static camera system and could make the parking experience more pleasant for visitors.

The resulting solution also has the great advantage of being flexible. It can be adapted to the client's needs. In this case, we want to outline a possibility where the digitalization of parking lots would not necessarily be the case. The algorithm can be trained on different objects, and the user interface can be used as a template.

V. Acknowledgment

This publication was realized with support of the Operational Programme Integrated Infrastructure in frame of the project: Intelligent systems for UAV real-time operation and data processing, code ITMS2014+: 313011V422 and cofinanced by the Europezn Regional Development Fund.

References

[1] Ani Aggarwal. "YOLO Explained". In: (2020).
[2] Wei Liu Dragomir Anguelov Dumitru Erhan Christian Szegedy Scott Reed Cheng-Yang Fu Alexander C. Berg. "SSD: Single Shot MultiBox Detector". In: (2016).
[3] Manish Chablani. "YOLO — You only look once, real time object detection explained". In: (2017).
[4] Ahmed Fawzy Gad. "Faster R-CNN Explained for Object Detection Tasks". In: (2021).
[5] Rohith Gandhi. "R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms". In: (2018).
[6] Kaiming He Georgia Gkioxari Piotr Dollar Ross Girshick. "Mask R-CNN". In: ().
[7] Jonathan Hui. "SSD object detection: Single Shot MultiBox Detector for real-time processing". In: (2018).
[8] Wai Yan Kyaw. "Histogram of Oriented Gradients". In: (2021).
[9] Huaizu Jiang Erik Learned-Miller. "Face Detection with the Faster R-CNN". In: (2017). DOI: 10.3390/rs12030389.
[10] Satya Mallick. "Histogram of Oriented Gradients explained using OpenCV". In: (2020).
[11] Karthik Mittal. "A Gentle Introduction Into The Histogram Of Oriented Gradients". In: (2020).
[12] Elisha Odemakinde. "Everything about Mask R-CNN: A Beginner's Guide". In: (2017).
[13] Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition". In: (2015).

Data Recovery/Retrieval Using Forensics Analysis

[1]Martin NOVYSEDLÁK, [2]Ondrej Kainz, [3]Miroslav Michalko, [4]Miroslav Murin

[1,2,3,4]Department of Informatics, Faculty of Electrical Engineering and Informatics
Technical University of Košice, Slovak Republic

[1]martin.novysedlak.2@student.tuke.sk, [2]ondrej.kainz@tuke.sk,
[3]miroslav.michalko@tuke.sk, [4]miroslav.murin@tuke.sk

Abstract **– The primary objective of this study is to explore the field of forensic analysis. The research includes an examination of various data storage methods, along with discussions on data protection and secure deletion techniques. Key findings from the synthetic components of the study are presented as follows. Firstly, a simulation of data loss will be conducted, followed by demonstrations of data recovery using open-source software. Secondly, the study investigates the process of dumping RAM from personal computers and smartphones. Additionally, methods for securely deleting data are analyzed, and a summary of the advantages and disadvantages of different approaches is provided. Finally, the study concludes with the partial implementation of a software tool for data recovery, illustrating the fundamental principles underlying similar recovery programs.**

Keywords **– data recovery, forensics analyis, RAM analysis, safe delete**

I. Introduction

This paper aims to contribute to the field of forensic analysis by conducting a comprehensive exploration of data storage systems and techniques for safeguarding data integrity. Through a investigation, it aims to explain the functionality of diverse data storage devices and evaluate strategies for data protection and secure deletion.

The study unfolds with analysis of potential scenarios leading to data loss, incorporating both theoretical and practical experiments utilizing open-source recovery tools. Furthermore, it delves into the intricate process of extracting RAM data from personal computers and smartphones, shedding light on the wealth of information these devices retain.

In addition to exploring methodologies for data recovery, the study devotes attention to the critical aspect of secure data deletion. It provides a thorough examination of various approaches, highlighting their respective advantages and disadvantages to inform best practices in data management.

As a culmination of its findings, the paper presents a partially implemented program for data recovery, designed to showcase the foundational principles underpinning such software solutions. Through this exploration, the paper aims to contribute valuable insights to the field of forensic analysis and data security.

II. Analysis

A. Introduction to forensic analysis

Forensic analysis is one of the crucial method for experts in so called blue team. These experts protect organizations against malicious activities from outside. Such attacks are analyzed using various tools to name a few there are decompilers, packet analyzer and last but not least data recovery tools.

Tools for data retrieval are main component in a uncovering various malicious activities that occurred in a system or to recover deleted data by standard user.

B. Data storage devices

In modern world there are existing couple of types of data storage devices. Mostly used in systems are hard disk and SSD. Their functionality is different; hard drive uses spinning disks whereas SSD are without moving parts. They also differ in a way that they store data based on physical principles. This could mean that data recovery may present different approach based on type of data storage. There are volatile and non-volatile types of memory. The volatile ones requires constant eclectic current to be present in order to work, for example RAM.

Hard disk drive

The main components is one or more rigid disks, whose spinning rate is measured in thousands of rounds per minute. The disks are coated in magnetic layer. Another main component is moving actuator arm that is, responsible for reading and writing data. The magnetic layer is divided in small cells which can be oriented such that north pole of magnet is pointing upwards or downwards. The change in polarity on border of two cells presents one and no change in polarity represents zero. The actuator is sensible to the difference of polarity and also it is capable of polarizing cells [1].

The necessity to actuator have to reach specific location with the randomization of blocks where data are located makes hard disk drive slower but offers more storage space than its counterpart SSD.

Flash memory

Te flash memory principle relies on electricity and capacitors. The block of memory is made out of couple of capacitors. When read or write operation occurs, the drive parts lets electricity into capacitor and store them based on quantum tunneling. This charged cells presents one. However, capacitors can get overused after time and thus reducing the storing capability of flash memory cells. On the other hand flash memory presents reliable data storage with multiple cells occupy relatively small space. This has advantage of saving space but at higher manufacture cost [2].

The most encountered devices with flash memory are SSD (or flash drive, NVMe) and RAM is among volatile type.

Other types

Other types of data storage are for example magnetic tapes, bar codes or even DNA [3]. However, in this work did not aim research for data recovery from other storage that flash (SSD and RAM).

C. Data recovery from non-volatile storage

When user deletes file by dragging it to bin and emptying it, the data that file was made of still occupy the space in storage; only the reference to file is deleted. That specific space in storage is then marked as writable device can rewrite it. Every file alongside metadata has headers that are based by definition of file's extension standard and they contain specific bit stream that describes start of file and its end. With this information we can find start and end of file in stream of bites to read content between them and recover file.

D. Data retrieval from volatile storage

Such storage that is, RAM need constant flow of electric current to operate, but it is made out of type of flash cells [4]. With this knowledge we must have the target device turned on and just then create memory dump. Although authors in [5] showed, that it is not explicitly needed when we would have special tools for cooling down RAM to temperatures way below freezing point but this method was not in scope of the work.

To create RAM dump we use the *LiME* [6] tool that builds kernel object for specific version of kernel and then insert it into running kernel and extend the functionality. This approach allows to create memory dump and using tool *volatility3* we will perform actual analysis to extract various data. This tool is modular and using plugins we can acquire various information about device and running processes at the time of memory dump creation.

E. Safe delete and safe storage

Although data can be recovered it is strongly recommended to use methods that can save user's time and effort. Such methods can be setting up RAID scheme [7] across multiple disks to eliminate risk of data loss. Another reliable method is to create backup of data on another device, for example USB drive, cloud, etc.

As we already stated the conventional deletion by dragging file to bin and emptying it is not sufficient to safe delete data. Some experts stand with authors of work [8] that for safe delete we need special algorithm safe delete data with 35 overwrites. The data shall not be recovered even with using method with magnetic force microscopes. On the other hand, experts in works [9] and [10] states that *Gutmann method* is exaggerated at this time and just three overwrites of data are sufficient. In this work we will also look at procedures to safe delete or safe move file.

F. Implementation of program for data recovery

Part of the work is to partially implement a software tool that will allow to restore deleted data. he primary goal is to show fundamental principles of programs that recover data.

III. Synthetic and implementation part

In this section we look how the process of data recovery take place. Every data recovery for consistency purposes must undergo environment simulation methods.

A. Data recovery from flash drive

As stated above we must prepare right conditions to show consistency and proper functionality of program usage. Firstly we insert USB drive to computer and zero the whole storage. Secondly, we must format the drive so that operating system would know how to read and write data to it. These steps secures the correct simulation of environment and now we can move on to actual data recovery

The process of data recovery must begin with choosing the right tool for it. In this work we have decided to use open-source program *foremost* [11] that is easily installed in Linux environment. With that done we see in documentation that foremost requires storage image as a storage input.

Then we must insert files that we want to be recovered. In our case we will move into USB drive files with suffix *.pdf, .png, .cpp*. We took into account that files with these suffixes are very common amongst daily users and programmers. Right after moving files into external storage we deleted it using conventional *Shift + Delete* shortcut, the equivalent to moving files to bin and emptying it.

To create storage image we will use *dd* native in Linux that provides low level of converting and copying files. As input we will use the path to external drive located in */dev/sdb*, since everything in Unix-like systems are files. The *sdb* in path can differ in devices in depending on how many storage devices are connected to computer.

After successful copy creation we can now proceed to usage of *foremost*. The usage requires some options to define to successfully recover files. After usage and some time the data recovery worked properly and data were recovered.

For purposes we have done data recovery in whole computer running Linux that is, in the root folder */*. What we found was manufacturer's document files about warranty and more interesting pictures from social media from long time ago. These picture formats were done by browser when it was saving it into its cache.

B. RAM data extraction and analysis

At first we must download and install open source programs that are *LiME* [6],*dwarf2json* [12] and *volatility3* with their dependencies. Firstly, we create *.ko* file which is kernel object that extends functionality of running kernel. This kernel module is depended of kernel version we want to make memory dump. Using *insmod* we insert kernel module into running kernel.

Depending on version of kernel of the device that is undergoing RAM dumping we must also create symbols, so that we can analyze it. Symbols for program *volatility3* [13] defined offsets where actual data is located within memory dump, such as NTLM hashes, running processes, etc. Every version sets these offsets differently. To build symbols for specific version of Linux kernel we use program *dwarf2json* which builds symbols based on debug file format *.dwarf*. When we accomplish this we can analyze the memory dump.

Using installed program *volatility3* and correctly imported symbols we can analyze the memory dump. The program offers lots of preinstalled plugins to get various information from memory dump. It is in the experts scope what date he wants to retrieve.

In out work we did analysis of existing memory dump that contains for the educational purposes called CTF (Capture The Flag) [14]. We also demonstrated how to build kernel module, symbols and how to use program *volatility3* to analyze memory dump.

In dump we can find various sensitive data such as running processes, NTLM hashes and cookies from browser which could be potentially exploited in attack known as *Session Hijacking*.

C. Safe deletion

In the section of safe deletion we used programs that deletes data not only by removing reference to them, but also overwrite the content in the place where data were located. After every safe deletion we did image of the drive and we tried to recover them again. Also for comparison the *Shift + Delete* method does not safely deletes the data.

ERASER

Open-source program *ERASER* [15] supported on Windows did offer lots of algorithms that safe delete data, Gutmann method included. *ERASER* provides more actions than a safe delete, fore example: safe defragmentation of the disk or safe moving of file. Program also offers scheduling these tasks and creates so called *cron job*. We decided for classical approach with three overwriting passes with random bytes.

The result was as expected; we did not manage to recover the data after safe deletion using program *ERASER*.

Command shred

Command *shred* comes native on Linux systems. By default it just overwrites the file but it does not delete it. By using options that are defined in documentation we can also safe delete the file. We can also define the number of overwriting passes and what bytes should program overwrite with, for example zeroes or random bytes.

However, the program does not offer the large amount of algorithms that overwrites the data. Its advantage is nativity of Linux system and easy usage.

Also the result with three overwriting passes was enough to safe delete data so that program *foremost* shall not recover.

IV. Partial implementation of the program

In out work we also did partial implementation of software tool that recover data. The tool was implemented to show fundamental principles of similar programs for data recovery. As we stated in the analysis, the program finds byte stream that is specific for each file type i. e. extension defined by the standard. In our case we decided to recover just *.jpg* files.

The program is written in *python3* programming language. Firstly we defined constants f. e. SOI, EOI and couple more for program flow. Secondly program open the image in binary read mode in which is loop and it is finding anything between SOI and EOI. When EOI is reached, the program saves the byte stream found. At the end it creates new file and write the found byte streams in them. With these steps we successfully recovered *.jpg* that we conventionally deleted.

The tool works only for *.jpg* and can be extended to find other file extensions. We can safely say that this program demonstrated functionality of similar tools for data retrieval.

V. Conclusion

In this paper, we have focused on the issue of digital forensics and proposed a solution to identify, retrieve and analyze digital evidence from various devices. The aim of the study was to comprehensively analyze the tools and techniques in this field, compare modern data storage devices and their data recovery capability, perform a practical analysis of real digital devices, investigate different methods for securely erasing data, and design a program for scanning a disk and locating erased files.

The contribution of the paper is not only the theoretical contribution to the understanding of digital forensics, but also practical guidelines and tools that can be used by practitioners.

The first major output was a data recovery process using open-source software. We successfully created a simulated example, and the recovered data of different formats confirmed to us that the classical deletion through the graphical interface is not sufficient. Next, we successfully performed a RAM write using open source programs. We have also demonstrated the procedure to perform this recording not only on the original device but also on external devices such as a smart phone. We analyzed the previously created record using open source programs and successfully extracted various data. We then dealt with the other side of the problem, where we tried to securely delete

the data. We used several approaches and programs and then attempted to recover the data.The result was successful; we successfully deleted the data and the recovery failed.

In addition, the implemented software tool has fulfilled its purpose. As a result of the implementation, the software successfully recovered seemingly lost data by searching for specific byte sequences in storage. Its main contribution was to demonstrate the basic logic of similar programs designed for data recovery.

For further pursuit, a tool can be implemented to scan a disk and find deleted files for various file types, continue research into methods for recovering data from corrupted or substandard data storage, compare commercial and open-source forensic tools, and create educational materials for forensic analysts.

VI. Acknowledgement

This work was supported by Cultural and Educational Grant Agency (KEGA) of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the projects No. 060TUKE-4/2022 and No. 004ŽU-4/2024.

References

[1] N. Zlatanov, *Hard Disk Drive and Disk Encryption*, 11 2015.
[2] R. Micheloni, *Solid-State-Drives (SSDs) Modeling*, 1st ed. Springer Cham, 3 2017.
[3] B. Hubert, 2 2021, citované dňa 24.3.2023. [Online]. Available: https://berthub.eu/articles/posts/amazing-dna/
[4] P. Haugen, I. Myers, B. Sadler, and J. Whidden, "A basic overview of commonly encountered types of random access memory (ram)," 5 2001. [Online]. Available: https://www.rose-hulman.edu/Class/ee/yoder/ece332/Papers/RAM%20Technologies.pdf
[5] J. A. Halderman, S. D. Schoen, N. Heninger, W. Clarkson, W. Paul, J. A. Calandrino, A. J. Feldman, J. Appelbaum, and E. W. Felten, "Lest we remember: Cold boot attacks on encryption keys," in *17th USENIX Security Symposium*. San Jose, CA: USENIX Association, 7 2008. [Online]. Available: https://www.usenix.org/legacy/event/sec08/tech/full_papers/halderman/halderman.pdf
[6] 504ensics, "Lime," 2024. [Online]. Available: https://github.com/504ensicsLabs/LiME
[7] D. A. Patterson, G. Gibso, and R. H. Katz, "A case for redundant array of inexpensive disks (raid)," Department of Electrical Engineering and Computer Sciences, University of California Berkeley, CA 94720, Tech. Rep., 6 1988. [Online]. Available: https://www2.eecs.berkeley.edu/Pubs/TechRpts/1987/CSD-87-391.pdf
[8] P. Gutmann, "Secure deletion of data from magnetic and Solid-State memory," in *6th USENIX Security Symposium (USENIX Security 96)*. San Jose, CA: USENIX Association, 7 1996. [Online]. Available: https://www.usenix.org/legacy/publications/library/proceedings/sec96/full_papers/gutmann/index.html
[9] V. Kanekal, "Data reconstruction from a hard disk drive using magnetic force microscopy," 2013. [Online]. Available: https://escholarship.org/uc/item/26g4p84b
[10] W. Craig, K. Dave, and S. Shyaam, "Overwriting hard drive data: The great wiping controversy," 12 2008, pp. 243–257.
[11] "Foremost." [Online]. Available: https://github.com/gerryamurphy/Foremost
[12] "Dwarf2json," 2024. [Online]. Available: https://github.com/volatilityfoundation/dwarf2json
[13] "Volatility3," 2024. [Online]. Available: https://github.com/volatilityfoundation/volatility3
[14] A. Kumar, "Memlabs," 2024. [Online]. Available: https://github.com/stuxnet999/MemLabs
[15] G. Trant, "Eraser," 2024. [Online]. Available: https://sourceforge.net/projects/eraser/

Leveraging Distributed Systems for Enhanced Parallel Calculations of Cosmic Ray Simulations in the Heliosphere

[1]Daniel Slanina, [2]Michal Solanik

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

[1]daniel.slanina@student.tuke.sk, [2]michal.solanik@tuke.sk

Abstract **– This paper describes current state and implemented concepts of Distributed System for Cosmic Ray Simulations in the Heliosphere, which has been developed in multiple iterations. In the latest version of the system, a support for custom simulation models, and redesign of a distributed task scheduler were introduced, resulting in much better flexibility and usability of the system. During testing of the latest version of this distributed system it has been proved, that this version is indeed the fastest, most flexible and most secure version among all iterations of this system. When compared to the first prototype of this distributed system, we've achieved an 25.26% speedup in calculation of testing simulation.**

Keywords **– demand-driven scheduling, distributed system, distributed task scheduling, heliosphere**

I. Introduction

Cosmic Ray Simulations in the Heliosphere are based on complex physical models. A large amount of computational resources are needed to calculate them in a reasonable time-frame. However, due to hardware limitations, it is impossible to vertically scale resources infinitely. For any further scaling and speedup we need to scale horizontally - by adding more independent computing machines, that work together. By doing so, the system becomes *distributed*, as described in multiple sources [1], [2]. This paper describes current state of Distributed System for Cosmic Ray Simulations in the Heliosphere and summarizes all contributions to this system, made since its inception.

II. Geliosphere

The first crucial component of this distributed system is the Geliosphere application. It acts as a parallel implementation of the calculation engine for the simulations. It was first designed and implemented in [3] and later gradually improved and extended in [4]. Currently, Geliosphere has support for calculations via both CPU and NVIDIA-based GPU. During compilation process, it selects the most optimal code-paths and implementations, based on the target hardware, resulting in highly optimized compiled executable, that is ready to run.

Geliosphere contains built-in implementations of following models:

- 1D Forward-in-time
- 1D Backward-in-time
- 2D Backward-in-time (Geliosphere)
- 2D Backward-in-time (SOLARPROPLike)

Each built-in model has its own set of parameters, which must be set, as defined in the Geliosphere documentation [5]. Apart of these implementations, Geliosphere also supports running simulations for custom models. However, in this case a custom parallel implementation of such model, which is compatible with Geliosphere API is required. Dynamic model implementation must also be as much optimized for specific hardware as possible. Therefore, it should be compiled the same way as Geliosphere itself, with the same rules applied.

III. Autosphere

To simplify the process of creating custom model implementations for Geliosphere, a new component, *Autosphere*, was created. Detailed proposal and implementation of this component is described in [4]. Autosphere includes a simple custom Domain-specific language, that enables users to easily define new simulation models via mathematical equations and relations. Based on the user input, which is written in this DSL, it generates an optimal implementation of the simulation model, which can be executed via Geliosphere. Autosphere itself consists of two core parts:

1) Autosphere Validator - responsible for ensuring correctness of the user input code. Validator is only performing static analysis and syntactic validation, therefore is not bound to target platform or hardware and can be run anywhere. Moreover, this step is fast and does not require any complex computation. It provides instant feedback to the user about input code.
2) Autosphere Generator - responsible for generating parallel implementation of simulation model, based on valid user input code. Generation and compilation is highly dependent on target hardware, so it is necessary to run this step on the same hardware and platform as compilation of Geliosphere. Due to this constraint, Autosphere is often run on computing machine as a sibling process to Geliosphere.

The main advantage of Autosphere is that user does not need to manually write a an implementation in any programming language, which must be compatible with Geliosphere API. It also takes care of selecting and generating the most optimal implementation, so the process is fully automated and transparent from the user perspective. It enables our main target group - physicians and other scientists to easily describe and create their custom simulation models.

IV. Distributed System

Both previously described components were only CLI tools, meant to run on a single computing machine. To enable horizontal scaling capabilities for those independent machines, they need to be orchestrated. In [6], an initial architecture of distributed system was proposed and implemented. Afterwards, in [7], this architecture was slightly adjusted and extended with some new components, for example to enable support for custom simulation models.

System architecture is divided into two core layers. *Centralized layer* contains all system services that manages system state. On this layer, there is also a component with implementation of distributed scheduler, which orchestrates and coordinates all distributed tasks within system. On the other hand, *distributed layer* represents all distributed computing machines that are responsible for executing different set of tasks, that are scheduled and assigned to them. Current top-level architecture of the distributed system is shown on figure 1. Each service within the system manages its own part of the domain.

A. Distributed task scheduler

Very important part of the distributed system is the *distributed task scheduler*. Its implementation resides within *Jobs service* and takes care of assigning and managing all distributed tasks for connected workers. Generic implementation of the scheduler allows it to manage different kinds of distributed tasks, without the need of changes in implementation. Future extensibility of the system by adding more kinds of distributed tasks is incredibly easy.

The scheduler uses *demand-driven scheduling* technique, so workers actively ask for a task. Only when a demand is risen, a job is assigned and scheduled. Distribution decision tree for scheduler is described in figure 2. Scheduler is highly inspired by *Gitlab CI* pipelines system [8].

Distributed task for simulation calculations uses *fine-grained parallelism*, so resources of connected distributed workers are utilized to the maximum. Many smaller calculations are preferred, so slower workers are not slowing down faster ones, and also whole system in general. However, an appropriate size of a single calculation chunk is selected, so it does not create additional communication overhead, resulting in losing performance. In case of very slow worker, that creates a risk of slowing the whole simulation calculation, scheduler has implemented rescheduling features, that automatically detects this case. After defined grace period expires, a task is scheduled to a second worker (and possibly even more, if this case repeats), with a mechanism of *redundant distribution*. Then, both assigned workers compete, who finishes the execution first. When a faster worker finishes task execution, a cancellation signal is sent to a slower worker, so it can free its resources for another tasks.

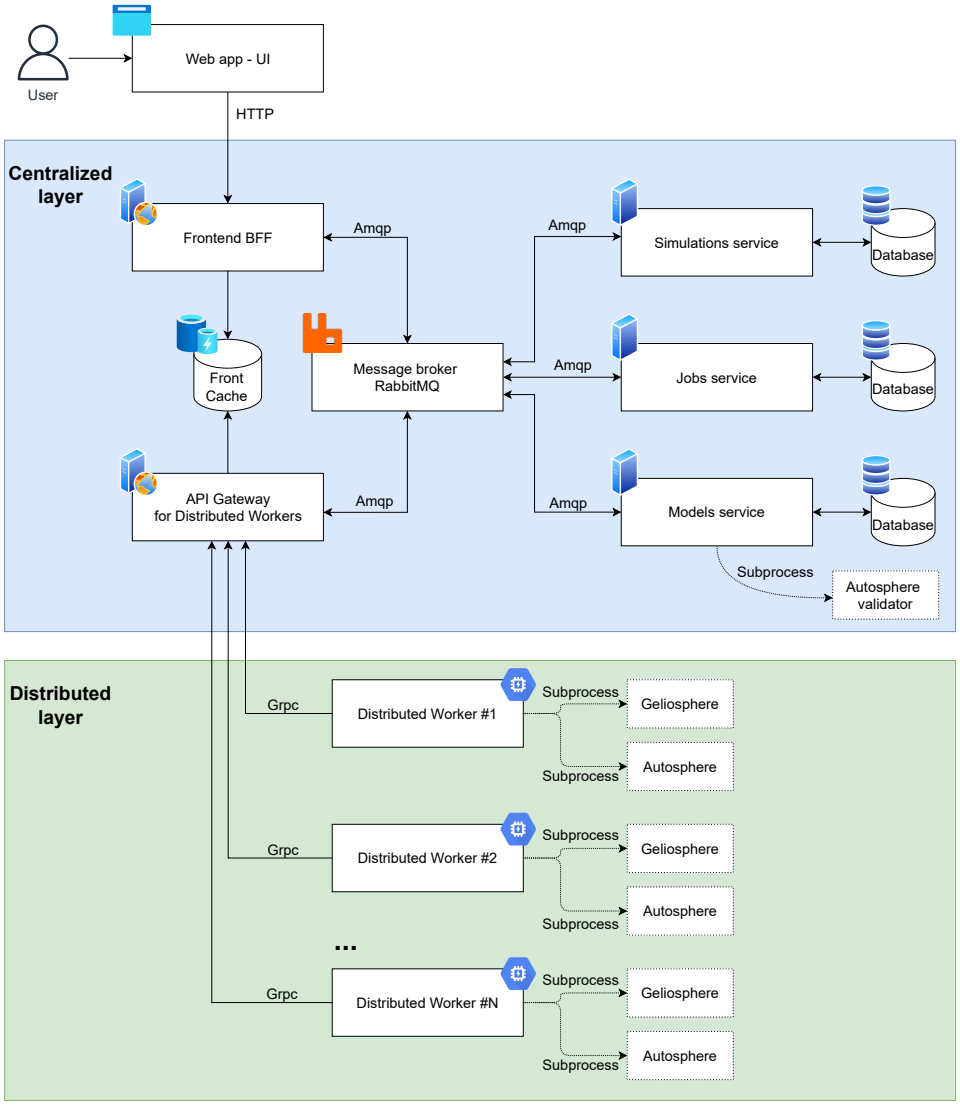


Fig. 1 Current overall architecture of the distributed system

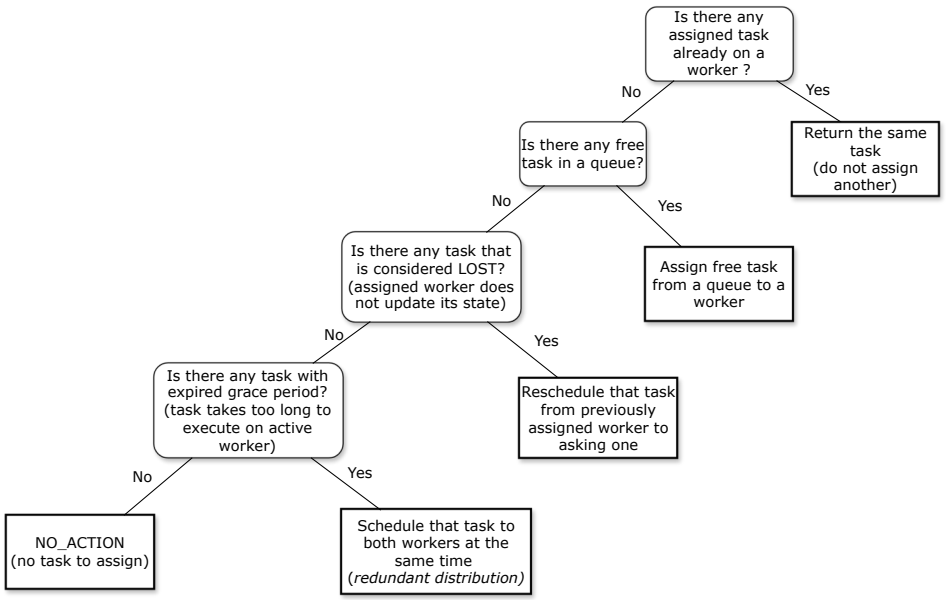


Fig. 2 Decision tree of distributed scheduler when scheduling a task

B. Security

Due to distributed nature of the system, workers cannot be fully trusted. They can run in any environment, any machine. In [7], multiple security mechanisms were designed and implemented to mitigate security threats, especially at communication layer between centralized and distributed layers. These security mechanisms span across multiple system components, depending on use-case.

Component *API Gateway for Distributed Workers* acts as a security firewall between centralized layer and distributed workers. It implements token-based authentication mechanism, so only authorized and allowed workers can communicate with the system backend services. Each worker is provided with a single, unique authentication token, which gives them identity. If incoming request is not authorized, communication is rejected automatically on API Gateway level.

On the level of *Jobs service* component, following authorization mechanisms are implemented, that improves system's security and integrity:

- Single worker can have *only one* task assigned at the time. With this mechanism, the system mitigates the risk of a single, malicious worker locking out all distributed tasks.
- Worker can only finish tasks that are assigned to them. This prevents a malicious worker from tampering with task result data.
- Worker cannot have locked (leased) task for infinite time. By introducing grace period time-frame, which is finite (and by default set to 180 seconds), the possible deadlock of whole simulation calculation is mitigated. Tasks are always finished by workers, even if any of them is malicious.

Internal communication layer, backed by RabbitMQ message broker, is fully hidden and protected from public access. Only trusted components within centralized layer have access to the underlying message broker. It is not needed to implement security mechanisms on this level, as it is considered trusted environment.

C. Deployment

To simplify deployment process for this distributed system, an automated tooling, .NET Aspire [9], for generating Kubernetes deployment manifests is used. Components from centralized layer are covered by those deployment manifests and therefore easily deployable to any Kubernetes cluster. Components are also pre-configured with default settings; additional configuration is not necessary (although recommended for custom usage).

Distributed workers have pre-prepared Docker images for CPU type, and also most commonly used GPU types (such as NVIDIA RTX 3060, NVIDIA GTX 1080Ti, etc.). Project also contains Dockerfile definition, so Docker image can be built on your own, optimized for target platform and hardware. However, worker needs some configuration to get started, such as valid authentication token, or URL of the *API Gateway* component.

V. Verification of distributed system

During evaluation phase of master's thesis [7], the system was verified and tested in an environment very similar to production. Whole centralized layer was deployed to *Azure Kubernetes Service*, within Microsoft Azure Cloud environment. Due to prepared generated deployment manifests it took only a single command to deploy all system components to Kubernetes cluster.

For distributed workers, with collaboration of Institute of Experimental Physics, Slovak Academy of Sciences, we received 2 computing machines for testing purposes. These two machines served as distributed workers for the system Hardware configurations of used computing machines are described in table 1.

Table 1
Hardware configurations of distributed workers used in system testing

	Worker #242	Worker #244
CPU	Intel Core i5-8400 @ 2.80GHz	Intel Core i5-7500 @ 3.40GHz
RAM	16GB	16GB
GPU	NVIDIA RTX 2060	NVIDIA GTX 1080 TI
VRAM	6GB	11GB

A. *Results of system testing*

Testing have brought following results:

- Latest iteration of this distributed system is the fastest one, when compared with previous prototypes. It is faster by 7.2%, when compared with previous iteration, or by 25.26%, when compared with the first prototype version. Full speed comparison is provided in table 2.
- Implemented security measurements and mechanisms are valid, correct and cannot be bypassed. Authorization mechanisms also protect system from different malicious actions.
- System is resilient to failures and can easily recover from them. This applies to either system components' failures on centralized layer, or failures of each distributed workers. After worker failure, deadlock has not occurred on the system, instead task was rescheduled to another worker.

Table 2
Total time needed to calculate testing simulation by different versions of distributed system

Distributed system version	Total calculation time [hours]	Speedup [%]
First prototype of the system from [3]	1.17	0
Second iteration prototype from [6]	0.942	21.36
Latest version of distributed system from [7]	0.874	25.26

VI. Conclusion

This paper summarizes designed and implemented concepts in Distributed System for Cosmic Ray Simulations in the Heliosphere. Leveraging distributed systems in this area is very beneficial, as they provide a massive speedup in execution times. Designed and implemented concepts can also be used in other similar scientific domain, that tackles with complex simulations, or any other form of complex computation.

Testing in production-like environment has proven, that the latest iteration of this distributed system is *the fastest* yet, in terms of single simulation calculation. In percentage, it is over 25% faster than the first iteration of the system and more than 7% faster than his predecessor.

Apart of the speed results, testing have also shown that communication between system layers is secured and cannot be bypassed with malicious thoughts. *API Gateway for Distributed Workers* plays a key role in maintaining security when connecting trusted backend services with untrusted workers.

Designed distributed task scheduler has a lot of potential in the future of this system. Its generic implementation enables extensibility without the need of either changing or adding additional implementation to support new types of distributed tasks. Scheduler logic is also completely stateless, due to demand-driven scheduling technique used. This makes it perfect candidate for a cloud service, that is optimized for cost and performance.

In the future, there is a plan to fully integrate Autosphere into system, after Autosphere will be released as stable version, since it is only a prototype, for now. In this version of the system, the user interface - Web application, is still only in a prototype phase. It is fully usable and working, but it could benefit from some UI improvements.

References

[1] A. S. Tanenbaum and M. Van Steen, *Distributed systems: principles and paradigms*. Prentice-hall, 2007.
[2] G. F. Coulouris, J. Dollimore, and T. Kindberg, *Distributed systems: concepts and design*. pearson education, 2005.
[3] M. Solanik, "Paralelizácia dvojrozmerných modelov modulácie kozmického žiarenia v heliosfére," Master's thesis, Technická univerzita v Košiciach, 2020.
[4] ——, "Methods and algorithms for acceleration / parallelization of physical model calculations," Ph.D. dissertation, Technical university of Košice, 2023.
[5] The Geliosphere Authors. Geliosphere documentation. [Online]. Available: https://github.com/msolanik/Geliosphere/blob/main/README.md
[6] D. Slanina, "Jemnozrnný paralelizmus v distribuovanom systéme," Bachelor's thesis, Technická univerzita v Košiciach, 2022.
[7] ——, "Optimalizácia architektúry distribuovaného systému pre výpočty simulácii v heliosfére," Master's thesis, Technická univerzita v Košiciach, 2024.
[8] GitLab B.V. Gitlab documentation. [Online]. Available: https://docs.gitlab.com/ee/
[9] Microsoft. .net aspire documentation. [Online]. Available: https://learn.microsoft.com/dotnet/aspire/

Digitálne dvojča hracej dosky v jednoduchých hrách

[1]*Dávid LACKO,* [2]*Marián MACH*

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

[1]david.lacko.4@student.tuke.sk, [2]marian.mach@tuke.sk

***Abstrakt* — Článok sa zameriava na problematiku vytvárania systému pre hranie jednoduchých doskových hier. Pozornosť je venovaná hre dáma hranej na šachovnicovej hracej doske, založenej na pomerne jednoduchých pravidlách. Článok sa zaoberá vytváraním digitálneho dvojčaťa fyzickej hracej dosky, poskytujúceho digitálnemu hernému agentovi prehľad o aktuálnej hernej situácii na reálnej hracej doske. V rámci experimentovania s vytvoreným riešením boli získané skúsenosti o použiteľnosti navrhnutého riešenia pri praktickom hraní.**

***Kľúčové slová* — hra dáma, digitálne dvojča, hrací agent, počítačové videnie, Harrisov detektor**

I. Úvod

Používanie počítačov v oblasti hier má veľmi dlhú tradíciu. Už v sedemdesiatych rokoch minulého storočia vytvoril Arthur L. Samuel program hry dáma (checkers) pre prvý komerčný počítač firmy IBM [1]. Odvtedy bolo vytvorených mnoho aplikácií zameraných na rôzne hry, ktoré je možné v zásade rozdeliť do dvoch hlavných kategórií: 1) snaha poskytnúť používateľovi príťažlivé hracie prostredie s prípadným hracím agentom nižšej až strednej výkonnosti, 2) snaha o vytvorenie hracieho agenta s výkonnosťou prevyšujúcou výkonnosť najlepších ľudských hráčov. Príkladom prvej kategórie je [2], príkladmi druhej kategórie sú systémy Deep Blue a jeho nasledovníci [3], AlphaGo [4] and AlphaStar [5].

S rozvojom v oblastiach počítačového videnia a robotiky začali snahy o plné nahradenie ľudského herného protivníka, sediaceho za hracou doskou – hrací program získava „oči“, ktorými hru sleduje, a „ruky“, pomocou ktorých hru ovplyvňuje. Samotný herný agent pracuje s digitálnym dvojčaťom [6] fyzickej hracej dosky, na ktorej daná hra reálne prebieha.

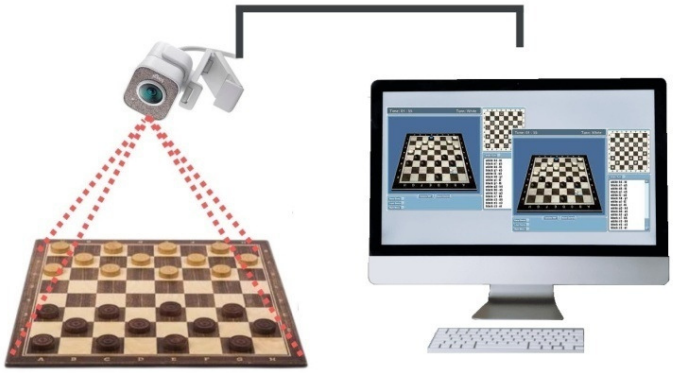
Tento článok sa zameriava na vytvorenie digitálneho dvojčaťa šachovnicovej hracej dosky. Jeho štruktúra je nasledovná – v druhej sekcii sa venuje celkovej architektúre vytvoreného hracieho systému. V tretej sekcii je uvedená séria krokov, potrebných pre vytvorenie digitálneho dvojčaťa. Štvrtá sekcia sa venuje niektorým skúsenostiam s robustnosťou vytvoreného riešenia pri praktickom využívaní.

II. Štruktúra hracieho agenta

V rámci predchádzajúcich aktivít sme vytvorili viacerých hracích agentov, medzi nimi agenta založeného na MCTS (Monte Carlo Tree Search) [7] ako aj agenta založeného na učení posilňovaním [8]. Ako ďalší krok v transformácii spôsobu komunikácie agenta s ľudským protihráčom do formy, umožňujúcej použitie fyzickej hracej dosky, sme sa pokúsili rozšíriť systém o možnosť snímania aktuálnej hernej situácie na doske. Toto je schematickým spôsobom ilustrované na Obr. 1.

Predpokladá sa hra s použitím reálnej hracej dosky, pričom ľudský hráč na doske vykonáva nielen svoje ťahy, ale aj ťahy agenta, ktorý svoje ťahy zobrazuje na monitore. Takéto usporiadanie principiálne umožňuje dva spôsoby vykonávania ťahov ľudským hráčom:

- Súťažná hra – ľudský hráč presne prenáša deklarované ťahy svojho protivníka do fyzického priestoru (s agentovou kontrolou, či v jeho mene boli skutočne vykonané tie ťahy, ktoré boli deklarované),
- Tréningová hra – ľudský hráč nerešpektuje agentom deklarované ťahy ale vykonáva iné ťahy za účelom získania špecifických postavení figúr na hracej doske (agent sa vždy prispôsobí aktuálnej situácii bez ohľadu na to, či situácia vznikla ako dôsledok jeho deklarovaných ťahov alebo nie).



Obr. 1 Ilustrácia štruktúry hracieho agenta.

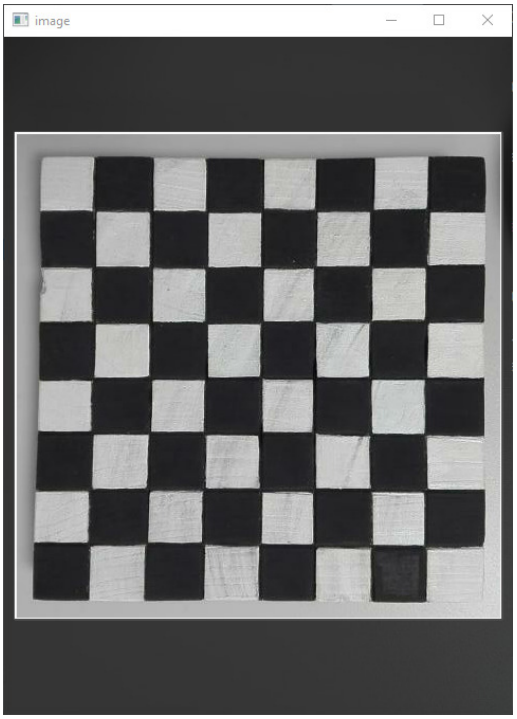
V oboch prípadoch je nutné systém doplniť o digitálne dvojča hracej dosky, ktoré je počas hry neustále aktualizované takým spôsobom, aby reflektovalo aktuálnu situáciu na doske. Samotní hrací agenti potom vychádzajú zo stavu tohto digitálneho dvojčaťa, na základe ktorého analyzujú stav hry a navrhujú jej pokračovanie deklarovaním navrhovaného ťahu.

III. Vytvorenie digitálneho dvojčaťa

A. Kalibrácia kamery

Nastavuje sa veľkosť snímanej plochy, rotácia kamery, nastavenia kamery a upravuje sa obraz pomocou rôznych manipulácií. Cieľom je umožniť presné umiestnenie hracej dosky do obrazu a tým napomôcť zabezpečeniu presného a spoľahlivého získania obrazových dát dosky pre ďalšie spracovanie (Obr. 2).

Príslušný kód zisťuje, ktoré kamery sú pripojené k počítaču, ich vlastnosti, nastavuje rôzne parametre vybranej kamery (vrátane veľkosti a hrúbky obrysových čiar na obrázku) a manipuluje s obrázkom tak, aby boli viditeľné rôzne časti scény.



Obr. 2 Snímaný obraz po kalibrácii kamery.

B. Nájdenie šachovnice

Nájdenie šachovnice sa vykoná nájdením pozícií vnútorných rohov šachovnice. Algoritmus funguje tak, že hľadá preddefinovaný vzor šachovnice v obrázku a snaží sa nájsť všetky rohy

šachovnice. Začne tým, že aplikuje škálovanie obrazu podľa nastavených parametrov. Potom sa použije detektor rohov – Harrisov detektor [9]. Tento detektor je založený na výpočte štrukturálneho gradientu v každom bode obrázka.

Je potrebné mať na pamäti, že algoritmus predpokladá správne zarovnanie šachovnice. To znamená, že rohy šachovnice sa v obraze budú nachádzať v zvislých a vodorovných riadkoch. Taktiež je dôležité, aby šachovnica bola dostatočne viditeľná a jasná, pretože v opačnom prípade by mohlo dôjsť k chybám detekcie rohov.

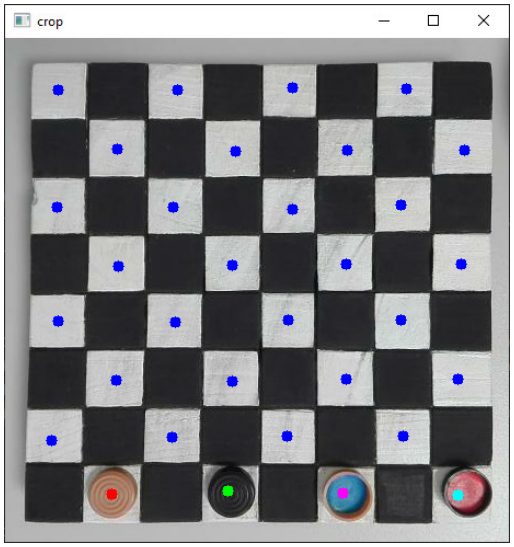
Po nájdení rohov šachovnice sa vypočíta stred každého políčka. Vzhľadom na to, že každý roh šachovnice generuje stredy políčok v každom smere, vzniká potreba odstrániť niektoré zo stredov a ponechať len jeden pre každé políčko.

C. Kalibrácia farieb

Po nájdení šachovnice je potrebné kalibrovať figúrky, aby systém mohol správne rozpoznať ich polohu na šachovnici. Kalibrácia sa vykonáva tak, že figúrky sa rozmiestnia na prednastavené pozície na šachovnici (v užívateľskom rozhraní sa zobrazí šachovnica so špecifikáciou políčok, na ktoré systém žiada umiestniť figúrky – Obr. 3). Konkrétne je potrebné umiestniť štyri figúrky – pešiaka protivníka, pešiaka agenta, dámu protivníka a dámu agenta. Umiestňujú sa na špecificky určené pozície, aby bola umožnená správna identifikácia figúrok (plus jedna špecifická pozícia musí ostať voľná, pretože z tejto pozície systém určuje farbu prázdneho políčka). Na základe zistenej informácie potom systém dokáže rozlišovať medzi voľným a obsadeným políčkom a v prípade obsadeného dokáže identifikovať hraciu figúrku na tomto poli a stranu, ktorej figúrka patrí.

Aby bola detekcia figúrok presnejšia, nesníma sa iba farba na pozíciách stredov určených políčok ale bolo aplikované spriemernenie v oblasti troch pixelov každým smerom. Táto technika pomáha odstrániť šum a zlepšiť detekciu pohybu figúrok. Vďaka tomuto spriemerneniu je algoritmus robustnejší a menej citlivý na šum.

Celkový proces kalibrácie a detekcie figúrok je kritický pre presné fungovanie šachovnicového systému. Ak je kalibrácia správne vykonaná a detekcia figúrok je spoľahlivá, systém bude schopný správne zaznamenať každý ťah a vytvoriť presnú kópiu skutočnej šachovnice.



Obr. 3 Kalibrácia farieb figúrok.

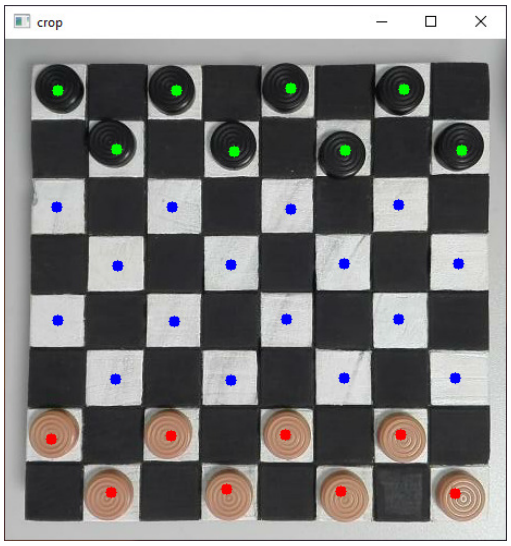
D. Vytvorenie digitálneho dvojčaťa

Po nakalibrovaní šachovnice a získaní referenčných farieb jednotlivých políčok je možné aplikovať algoritmus na digitálny obraz šachovnice a vytvoriť tak digitálnu kópiu reálnej šachovnice. Algoritmus prehľadáva jednotlivé body (identifikované stredy políčok) digitálneho obrazu šachovnice a porovnáva farby týchto bodov s referenčnými farbami kalibrovaných políčok. Ak sa farba bodu podobá farbe niektorej z kalibrovaných figúrok, tak algoritmus priradí danú figúrku na skúmanú pozíciu. Ak sa farba bodu podobá na farbu voľného políčka, tak algoritmus skúmanú pozíciu neobsadí žiadnou figúrkou (Obr. 4).

Pre odstránenie vplyvu šumu, prítomného v obraze, sa používa spriemernenie farby v oblasti troch pixelov v každom smere. Pre zisťovanie podobnosti farieb sa používa Euklidova

vzdialenosť. Algoritmus určí vzdialenosť aktuálne detegovanej farby na políčku voči farbe každej kalibrovanej figúrky a kalibrovaného voľného políčka a vyberie alternatívu s minimálnou vzdialenosťou.

Po aplikovaní tohto algoritmu sa celý obraz šachovnice prevedie do digitálnej podoby. Je to možné použiť na začiatku pre vytvorenie digitálneho dvojčaťa hracej dosky ale aj počas hry – periodicky sa situácia vyhodnocuje a pri detekcii zmeny je príslušne upravené aj digitálne dvojča.

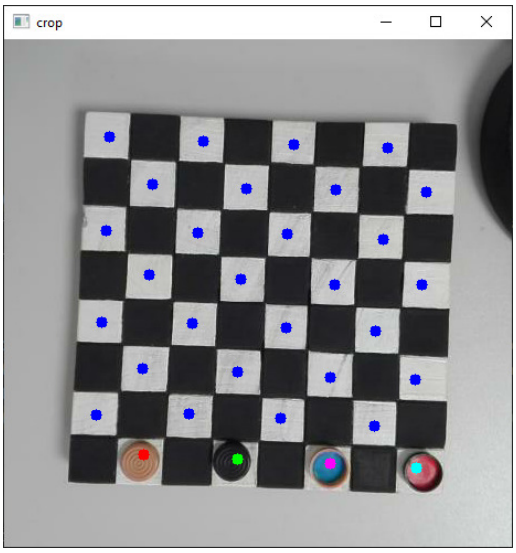


Obr. 4 Snímanie aktuálneho stavu rozloženia figúrok.

IV. Skúsenosti s robustnosťou riešenia

A. Umiestnenie šachovnice

Keďže rozmer hracej dosky nie je presne daný a samotná kalibrácia umiestnenia šachovnice slúži iba na identifikáciu stredov jednotlivých políčok, je možné použiť šachovnice rôznych veľkostí (resp. šachovnica môže v obraze zaberať rôzne veľkú plochu – Obr. 5). Oveľa dôležitejšie je zarovnanie hrán šachovnice v obraze. Empiricky sme zistili, že pre správne fungovanie by šachovnica nemala byť odklonená o viac ako 6° od ideálneho zarovnania.

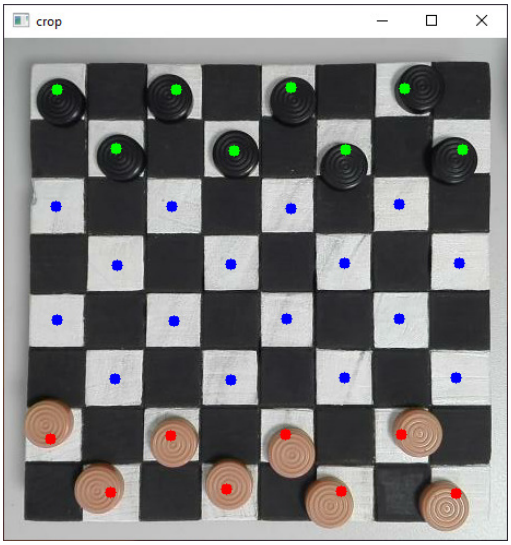


Obr. 5 Oddialená a odklonená šachovnica.

B. Umiestnenie figúrok

Keďže algoritmus zisťuje stav v strede každého políčka a jeho bezprostrednom okolí, figúrka nemusí byť umiestnená presne na stred políčka, avšak musí tento stred prekrývať – a to jedno

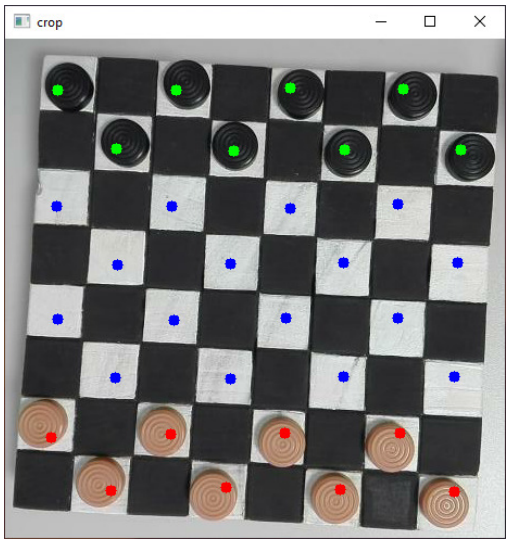
ktorou svojou časťou, či svojím stredom alebo svojím okrajom (Obr. 6). Vďaka tomu hráč nemusí dbať na to, aby jeho figúrka bola presne zarovnaná v príslušnom políčku šachovnice.



Obr. 6 Posunutie figúrok.

C. Posunutie a rotácia šachovnice

Ak sa šachovnica po kalibrovaní posunie alebo otočí (Obr. 7), detekcia polohy figúrok môže aj naďalej fungovať správnym spôsobom. Závisí to na tom, o akú veľkú zmenu polohy hracej dosky sa jedná. Podstatné pre správne ďalšie fungovanie algoritmu je, aby identifikované body, ktoré predtým boli identifikované ako stredy políčok, po posunutí alebo pootočení aj naďalej patrili k daným políčkam a nezasahovali do susedných políčok. A aby mierili do tých častí políčok, ktoré sú prekrývané figúrkami.



Obr. 7 Posunutá šachovnica.

V. Záver

Vykonané experimenty potvrdili praktickú použiteľnosť navrhnutého a implementovaného riešenia. Riešenie sa ukázalo dostatočne robustné pre bežne používané umiestňovanie figúrok na políčka bez dôrazu na presné umiestňovanie na stred. Rovnako je možné absorbovať aj posunutia resp. rotáciu hracej dosky v malom rozsahu. Nevýhodou je, že pri väčších posunutiach či rotácii je nutné zopakovať kalibračný proces, čo však nemôže byť realizované pri zachovaní aktuálnej pozície figúrok, pretože vyžaduje premiestnenie figúr do vopred očakávaných pozícií.

POĎAKOVANIE

Táto práca vznikla s podporou Vedeckej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0394/22.

LITERATÚRA

[1] Arthur Samuel: Pioneer in Machine Learning, *IBM Journal of Research and development*, 1992.
[2] Lichess.org – free, open-source chess platform, http://lichess.org/ [prístupné 17.5.2024]
[3] L. Greenemeier, 20 years after Deep Blue: How AI has Advanced Since Conquering Chess, *Scientific American*, 2017.
[4] D. Silver et al., Mastering the game of Go with deep neural networks and tree search, *Nature*, 2016, vol. 529, 484-489.
[5] O. Vinyals et al., Grandmaster level in StarCraft II using multi-agent reinforcement learning, *Nature*, 2019, vol. 575, 350-354.
[6] A. Sharma et al., Digital Twins: State of the art theory and practice, challenges, and open research questions, Journal of Industrial Information integration, 2022, vol. 30. 15 pages.
[7] D. Lacko, *Hrací agent založený na MCTS (Monte Carlo Tree Search)*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2021.
[8] D. Lacko, *Učenie posilňovaním v jednoduchých hrách*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2023.
[9] K.G. Derpanis, Harris corner detector, York university, 2004, http://www.cse.yorku.ca/~kosta/CompVis_Notes/harris_detector.pdf [prístupné 17.5.2024]

Štrukturálne regularizačné metódy pre hlboké siete

[1]*Martin PERIGNÁTH,* [2]*Marián MACH*

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

[1]martin.perignath@student.tuke.sk, [2]marian.mach@tuke.sk

***Abstrakt* — Článok sa zameriava na problematiku použitia regularizačných metód pri učení hlbokých sietí, snažiacich sa zvýšiť generalizačnú schopnosť týchto sietí. Pozornosť je venovaná štrukturálnej regularizácii, v rámci ktorej sú do štruktúry siete pridávané alternatívne prepojenia medzi vrstvami, čím vznikajú paralelné cesty, ktorých výsledky sú agregované. Článok sa zaoberá tromi štrukturálnymi metódami – Random Drop, Single-branch Shake a Shake Drop. V rámci experimentovania bolo realizovaných niekoľko experimentov s týmito metódami v kontexte klasifikácie obrazovej informácie s použitím známej dátovej množiny.**

***Kľúčové slová* — hlboké siete, štrukturálna regularizácia, Random Drop, Single-branch Shake, Shake Drop**

I. Úvod

Regularizačné metódy sa snažia o zvýšenie schopnosti modelu generalizovať, čím vlastne pôsobia proti preučeniu modelu. Zatiaľ čo učiaci algoritmus (napr. spätné šírenie chyby) sa snaží minimalizovať chybu na trénovacích dátach, regularizácia cieli na zníženie chyby na neznámych dátach, aj za cenu nárastu trénovacej chyby [1].

K najznámejším metódam patria dropout, L1 (Lasso regresia) a L2 (Ridge regresia) regularizácia. Ale možných metód je viac, pričom [2] ich delí do nasledujúcich piatich kategórií:

- tlmenie váh – rôzne spôsoby penalizácie váh s cieľom efektívne znížiť ich počet,
- augmentácia dát – pridávanie rôznych spôsobov skresľovania trénovacích dát s cieľom dosiahnuť, aby výsledky modelu boli invariantné voči týmto skresleniam,
- súborové metódy – kombinovanie predikcie viacerých modelov do jedného učiaceho procesu,
- štrukturálne metódy – úpravy štruktúr hlbokých sietí pridávaním alternatívnych ciest medzi vrstvami,
- implicitné metódy – metódy s regularizačným efektom.

V predkladanom článku sa sústreďujeme na skupinu štrukturálnych metód. Známymi reprezentantmi takýchto úprav sú siete ResNet (pridáva priame prepojenia paralelne k ceste cez konvolučné bloky) a Inception (vytvára paralelné cesty cez konvolučné filtre rôznych veľkostí). Na rozdiel od týchto notoricky známych prípadov sa venujeme niekoľkým menej známym metódam, patriacim do tejto kategórie. Všetky vychádzajú zo vzťahu zo siete ResNet

$$G(x) = x + F(x) \tag{1}$$

reprezentujúceho agregáciu výsledku viacerých (dvoch) ciest.

Tento článok sa zameriava na vybrané metódy štrukturálnej regularizácie. Jeho štruktúra je nasledovná – v druhej sekcii je uvedený opis sledovaných metód. V tretej sekcii je uvedená séria experimentov s vybranými metódami na úlohe klasifikácie obrazov.

II. Vybrané metódy štrukturálnej regularizácie

A. Random Drop

Túto metódu je možné považovať za určitú verziu metódy dropout. Hlavným rozdielom je to, že zatiaľ čo dropout vypúšťa signál jednotlivých neurónov, Random Drop vypúšťa signál vo vrstvách [3]. Vzťah (1) je transformovaný do podoby

$$G(x) = \begin{cases} x + b_l F(x) \\ x + b_l F(x) \\ x + E[b_l]F(x) \end{cases} \quad (2)$$

kde prvý výraz je využívaný pri doprednom šírení počas trénovania, druhý je použitý pre spätné šírenie počas trénovania a tretí je určený pre testovanie (mimo trénovania). Hodnota b_l predstavuje Bernouliho náhodnú premennú s pravdepodobnosťou

$$P(b_l = 1) = E[b_l] = p_l \quad (3)$$

Pravdepodobnosť p_l ponechania signálu je určená lineárnym spôsobom podľa počtu použitých blokov s touto metódou. Autori metódy odporúčajú nastaviť

$$p_l = 1 - \frac{l}{L}0.5 \quad (4)$$

kde L predstavuje celkový počet takýchto blokov a l zase poradie bloku.

B. Single-branch Shake

Táto metóda [4] transformuje vzťah (1) do podoby

$$G(x) = \begin{cases} x + a F(x) \\ x + b F(x) \\ x + E[a]F(x) \end{cases} \quad (5)$$

s rovnakým členením pre dopredné šírenie počas trénovania, spätné šírenie počas trénovania a pre testovanie. Hodnoty a a b sú vyberané nezávisle – sú to náhodne generované náhodné premenné podľa rovnomerného rozdelenia. To znamená, že $E(a) = 0.5$.

C. Shake Drop

Tento spôsob agregácie vetiev upravuje agregáciu podľa (1) do tvaru [4]

$$G(x) = \begin{cases} x + (b_l + \alpha - b_l\alpha) F(x) \\ x + (b_l + \beta - b_l\beta) F(x) \\ x + E[b_l + \alpha - b_l\alpha]F(x) \end{cases} \quad (6)$$

opäť s rovnakým členením pre dopredné šírenie počas trénovania, spätné šírenie počas trénovania a pre testovanie. Hodnota b_l opäť predstavuje Bernouliho náhodnú premennú s pravdepodobnosťou podľa vzťahov (3) a (4). Hodnoty α a β sú náhodné premenné generované podľa rovnomerného rozdelenia.

Hodnota b_l vlastne kontroluje správanie metódy počas fázy trénovania. Ak $b_l=1$, tak chovanie v doprednej a spätnej fáze učenia je podobné originálnej sieti ResNet

$$G(x) = \begin{cases} x + F(x) \\ x + F(x) \end{cases} \quad (7)$$

ale ak $b_l=0$, tak potom nastáva zmena

$$G(x) = \begin{cases} x + \alpha F(x) \\ x + \beta F(x) \end{cases} \quad (8)$$

a výstup konvolučnej vetvy je perturbovaný hodnotami koeficientov α a β.

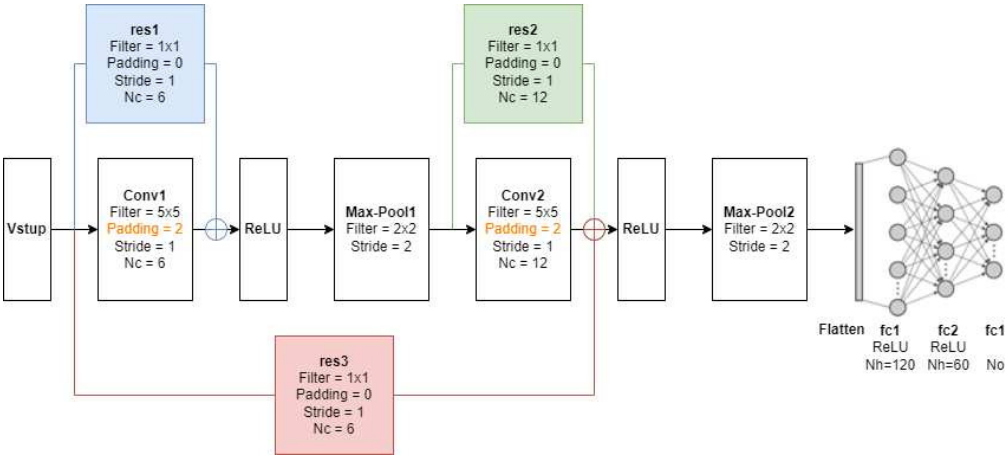
III. Vybrané experimenty s regularizačnými metódami

A. Testovací problém

Pre experimentovanie s metódami štrukturálnej regularizácie bola zvolená dátová množina obrazov FashionMNIST. Jedná sa o množinu 70000 šedotónových obrazov rozmeru 28x28 pixelov. Obrazy reprezentujú rôzne typy oblečenia klasifikované do desiatich tried (napr. tričko s krátkym rukávom, nohavice, pulóver, šaty, kabát, sandále, košeľa, ap.), pričom každá trieda obsahuje 7000 obrázkov. V tomto prípade sa jednalo o klasifikáciu obrazov do jednej z možných klasifikačných tried, použitých v danej dátovej množine.

B. Model pre klasifikáciu obrazov

Pre experimenty so štrukturálnymi regularizačnými metódami bola použitá jednoduchá konvolučná sieť pre klasifikáciu obrazov. Konvolučná časť (tvorená dvomi konvolučnými vrstvami a jednou poolingovou vrstvou) bola doplnená MLP časťou (pozostávajúcou z troch lineárnych vrstiev). Vytvorená štruktúra je znázornená na obr. 1, kde vytvára chrbtovú časť siete [5].



Obr. 1 Štruktúra použitej siete.

Uvedená základná štruktúra môže byť doplnená prepojovacími vetvami, ktoré sú na obrázku znázornené ako *res1*, *res2* a *res3*. Aby nejaká konvolučná vrstva mohla byť premostená, musí platiť:

- Tenzory na vstupe aj výstupe danej vrstvy musia mať rovnakú veľkosť, na výstupe nemôže byť tenzor menších rozmerov ako na vstupe. To bolo zabezpečené nastavením vhodného paddingu.
- Počet kanálov na vstupe konvolučnej musí byť zosúladený s počtom kanálov na výstupe danej konvolučnej vrstvy. Keďže sa tieto počty líšia, v prepojovacej vetve musí byť transformácia počtu kanálov, ktorá bola realizovaná pomocou 1x1 konvolúcie.

Uvedená štruktúra umožnila experimentovať so štyrmi alternatívami prepojovacích spojení: *res1*, *res2*, *res1+res2*, *res3*.

C. Základné nastavenie

Pred samotným experimentovaním s jednotlivými metódami bolo nutné nastaviť parametre procesu učenia siete. Pre toto bola použitá štruktúra bez prepojovacích vetiev. Každý experiment bol vyhodnocovaný pomocou štyroch tradičných metrík: správnosti, presnosti, návratnosti a F1. Jednotlivé parametre boli zoskupené do niekoľkých skupín. Pre všetky parametre nejakej skupiny boli definované možné hodnoty a boli vyskúšané všetky kombinácie hodnôt parametrov danej skupiny. Takýmto spôsobom sa brali do úvahy aj možné interakcie medzi parametrami [5]. Výsledné nastavenie je v tab. 1.

Tabuľka 1
Nastavenie hyperparametrov procesu učenia

Hyperparameter	Hodnota
Normalizácia dát	Áno
Dávka	32
Optimalizátor	Adam
Rýchlosť učenia	0.001
Ukončenie učenia	Early stopping
Chybová funkcia	Krížová entropia

Použitím tohto základného nastavenia bol zistený výkon siete bez prepojovacích vetiev (bez štrukturálnych regularizačných metód), ktorý je v ďalšom využitý ako základňa, voči ktorej je možné porovnať vplyv jednotlivých metód. Dosiahnutý výsledok je v tab. 2.

Tabuľka 2
Model bez prepojovacích vetiev

správnosť	presnosť	Návratnosť	F1
0.8958	0.8959	0.8958	0.8956

D. Výsledky experimentov

Výsledky pre metódu Random Drop sú v tab. 3. Všetky štyri alternatívy uvažovaných prepojení dosiahli výsledky lepšie než boli výsledky bez prepojovacích vrstiev, aj keď zlepšenie alternatív *res1* a *res3* nemožno považovať za signifikantné. Najlepšou alternatívou prepojenia sa ukázala *res2*.

Tabuľka 3
Výsledky dosiahnuté metódou Random Drop

Prepojenie	správnosť	presnosť	návratnosť	F1
Res1	0.8979	0.8984	0.8979	0.8972
Res2	0.9034	0.9041	0.9034	0.9032
Res1 + Res2	0.9005	0.9001	0.9005	0.9001
Res3	0.8978	0.8969	0.8978	0.8970

Výsledky pre metódu Single-branch Shake sú uvedené v tab. 4. Najlepšou alternatívou prepojenia bola *res1*. Avšak ani jedna z alternatív neprevýšila model bez prepojovacích vetiev, aj keď výkon víťaznej alternatívy *res1* veľmi nezaostal.

Tabuľka 4
Výsledky dosiahnuté metódou Single-branch Shake

Prepojenie	správnosť	presnosť	návratnosť	F1
Res1	0.8949	0.8949	0.8949	0.8947
Res2	0.8921	0.8921	0.8921	0.8919
Res1 + Res2	0.8808	0.8803	0.8808	0.8799
Res3	0.8788	0.8792	0.8788	0.8774

Výsledky metódy Shake Drop sú v tab. 5. Najlepšou alternatívou bola alternatíva *res1*, ktorá nebola horšia ako model bez prepojení – avšak ani lepšia. Ostatné tri alternatívy zaostali za očakávaním, pričom dosiahnutý výkon bol v prípadoch, ktoré obsahovali *res2*, signifikantne horší.

Tabuľka 5
Výsledky dosiahnuté metódou Shake Drop

Prepojenie	správnosť	presnosť	návratnosť	F1
Res1	0.8958	0.8969	0.8958	0.8955
Res2	0.8283	0.8316	0.8283	0.8248
Res1 + Res2	0.8239	0.8235	0.8239	0.8224
Res3	0.8677	0.8713	0.8677	0.8682

Pre úplnosť sme realizovali aj experiment s prepojeniami podľa vzťahu (1). Dosiahnuté výsledky sú uvedené v tab. 6. Aj v tomto prípade zvíťazila alternatíva *res1*, pričom však všetky alternatívy prepojenia poskytli vyššie hodnoty meraných metrík (aj keď v jednom prípade iba nepatrne) než model bez prepojenia.

Tabuľka 6
Výsledky dosiahnuté metódou ResNet

Prepojenie	správnosť	presnosť	návratnosť	F1
Res1	0.9071	0.9085	0.9071	0.9074
Res2	0.9028	0.9034	0.9028	0.9028
Res1 + Res2	0.8982	0.8989	0.8982	0.8982
Res3	0.9055	0.9070	0.9055	0.9059

IV. ZÁVER

Vykonané experimenty zaostali za očakávaniami. Aj keď v niektorých prípadoch bol dosiahnutý vyšší výkon než dosiahol model bez prepojení, rozdiely boli malé. Nedá sa povedať,

že by experimenty jednoznačne potvrdili prínos skúmaných metód štrukturálnej regularizácie. Otáznym zostáva, či to je všeobecnou vlastnosťou skúmaných metód alebo v prípade hlbšej siete o mnohých konvolučných vrstvách by sa tieto metódy prejavili viac. Zodpovedať túto otázku ostáva ďalším experimentom. Na druhej strane na základe získaných výsledkov je možné stanoviť hypotézu, že vplyv štrukturálnej regularizácie je tým vyšší, čím bližšie sú doplnené prepojenia ku vstupu siete. Dôkaz alebo vyvrátenie tejto hypotézy bude tiež predmetom ďalších experimentov.

POĎAKOVANIE

Táto práca vznikla s podporou Vedeckej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0394/22.

LITERATÚRA

[1] X. Shen, J. Lin, Consistency of neural networks with Regularization, ArXiv, 2022, https://arxiv.org/pdf/2207.01538 [prístupné 23.5.2024]
[2] A. Kadra et al., Well-tuned Simple Nets Excel on Tabular Datasets, in *Proc. of 35th Conference on Neural Information Processing Systems (NIPS)*, 2021.
[3] G. Huang et al., Deep networks with Stochastic Depth, in *Proc. of European Conference on Computer Vision (ECCV)*, Springer, 2016, 646-661.
[4] Y. Yamada et al., ShakeDrop Regularization for Deep Residual Learning, *IEEE Access*, 2019, vol 7, 186126-186136.
[5] M. Perignáth, *Využitie regularizačných metód v učení neurónových sietí*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2023.

Implementation of universal management for controlling smart devices

[1]Pavol Dzurnak [2]Martin Hasin [3]Dominik Falis

[1]Department of Engineering for Design of Machines and Transport Equipment, Faculty of Mechanical Engineering Technical University of Košice, Slovak Republic
[2]Department of computers and informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic
[3]Department of computers and informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

[1]pavol.dzurnak, [2]martin.hasin@tuke.sk [3]dominik.falis@student.tuke.sk

Abstract **– This article initially deals with analyzing API systems for controlling smart devices, from which important information is obtained. The goal is to test the implementation of network communication for different interfaces, which can be used later in the processes of overhead crane controlling, which can improve some crane's smart features like sway control, target positioning, etc. Testing is carried out for the sake of the simplest possible implementation of office smart devices The following is a proposal of a database that supports the mapping of values and attributes for the management and control of devices that are essential for such system. The article describes the design and implementation of the system, including the creation of elements such as an API provider, endpoints and a link to define an external interface for communication. The design of the device with connection to attributes and value mapping is designed for the control itself. The server interface is created in the Rust language and the user interface in the Angular framework. Finally, the article deals with testing the resulting application on assigned devices.**

Keywords **– smart devices, overhead crane, Rust, Angular, Docker**

I. Introduction

Nowadays, many homes, offices or certain buildings have smart devices that are required to be controlled from one place. Some smart device systems also have their own apps, but they may not support third-party devices, so there are apps like Google Home, Samsung SmartThings, and the like that support a wide range of popular devices. The problem arises when the device is old, or has specific access to the system, or support for control via these applications has not yet been added to it, there can be many reasons. In that case, adding an unsupported device is difficult, theoretically using some software development kits like Google Home has when you buy a smart speaker, but that's an additional cost in money and time. Another problem is the non-adaptive user interface and closed source code that prevent the implementation of new features [1].

As an alternative, Home Assistant offers a more flexible solution with tools like Node-RED that allow advanced configuration, but still require significant implementation effort, such as implementing logic using Node-RED and user interface using Home Assistant components. To mention examples, these solutions are not very suitable and therefore this article will deal with the design of a database for data storage and the implementation of such a system.

A universal device control system should allow easy addition of address, endpoints and attributes for different types of device data. An intuitive interface is also important, allowing quick access to the most important device functions, such as switching on or main attributes, as well as more detailed control.

II. Analysis of existing systems

Smart homes are becoming a common part of many homes, integrating various smart systems and devices into a unified interface. Such systems enable control from lighting to security elements

through a central system, increasing comfort and safety. The selected systems include both closed and open source systems [2].

Google Home acts as a central hub for controlling smart home devices, enabling the connection and management of various compatible devices. Users can add devices, create automation scenarios and control devices remotely [3] through the app. However, adding unsupported devices requires studying the documentation and logs, which can be difficult, or require the purchase of hub [4].

Samsung SmartThings provides an extensive ecosystem for the integration of smart devices, enabling management and automation through a single application. The platform supports development using APIs, SDKs and allows adding custom functions [5]. The main disadvantages are the dependence on cloud processing According to the SmartThings SDK documentation, users can control devices via API, but the devices must be supported by the application [6]. The documentation also mentions the possibility of creating an application with a driver in the system, but does not include a procedure for adding unsupported devices [7].

Home Assistant, an open source software, enables the integration and control of smart devices from a single platform [8]. It supports a wide range of devices and allows the creation of automations with Node-RED, which simplifies the process without the need for deep programming knowledge [9]. Nevertheless, the integration and configuration of devices can be time-consuming, especially for devices with many controllable attributes.

III. Analysis of API interfaces for control

In an effort to develop a universal application for controlling smart devices, an important initial step is to analyze the various API systems used by these devices. Ecosystems like Philips Hue are becoming increasingly complex and diverse, integrating different devices that communicate and work differently. By analyzing these APIs, key insights can be gained about how different devices work, what features they offer, and how they can be integrated into a universal control system. This analysis is important not only to ensure compatibility across a wide range of devices, but also to ensure that the driver application can exploit the full potential of each device [10].

A. Philips Hue

Philips Hue is an entire ecosystem of smart devices that is well-known for innovation in the smart home industry. It includes various devices that can be controlled digitally via a smartphone. This system supports a variety of devices such as smart bulbs, lamps, cameras, switches and many other types of accessories that are designed to work together seamlessly. Devices from Philips Hue primarily work through a wireless connection such as Wi-Fi, Bluetooth or by connecting to the central office using ZigBee. Communication is analyzed through the Philips Hue central hub, which supports communication through the REST API interface.

Each device contains its own attributes for control, which are specific to the given type of device, and also each attribute has its own structure and its own data type. These values can all be found in the official documentation from Philips Hue, which has several versions. Version 2 is newer and gradually replaces the older version. In the Philips Hue CLIP API version 2 documentation, it is possible to view which attributes, which values and which data types this API can work with [11]. The various attribute and value options are as follows:

- Brightness - The attribute is controlled in this system using the `dimming.brightness` attribute, which accepts a numeric value from 1 to 100.
- Color temperature - This attribute is specified in the documentation API using the JSON parameter `color_temperature.mirek`, which accepts an integer value from 153 to 500. Mostly, for example, light bulbs have color temperature values given in kelvin, and the value in the mirek unit can be calculated using the formula `mirek = 1000000 / (value in kelvin)`.
- Color - The color attribute is a bit more complicated, as this attribute consists of two values X and Y, and thus `color.xy.x` and `color.xy.y`. The value is in the interval between 0 and 1 as a decimal number.
- On state - This attribute is defined by the `on` attribute in the Philips Hue system. This attribute can take on the value `true` or `false`, so it is defined as a `boolean` data type.

The support of various attributes is very extensive in the Philips Hue system, but for the initial analysis, these most extensive attributes are sufficient to discover different configurations and options for setting values, their shape, structure and type of values.

Lighting control URL structure

Once verified, the API allows control of individual lights via specific URLs. GET and PUT requests can be made to this address. In the case of a GET request, it is possible to obtain data about the given lamp, and with a PUT request, it is possible to control the given lamp.

```
https://<ip_address>/clip/v2/resource/light/{id}
```

An example for controlling lighting can be turning on a lamp, with half brightness and cold white color:

```
PUT https://<ip_address>/clip/v2/resource/light/{id}
{
    "on": {"on": true},
    "dimming": {"brightness": 50}
    "color_temperature": {"mirek":200}
}
```

B. Air conditioner interface

As another system for analysis and testing, an air conditioner system was available, which has its own HTTP interface, which is a web application through which the air condition can be controlled. It is necessary to analyze this interface, what are its requirements, components, data types of device attributes and so on.

The interface is very simple and contains a few parameters such as on/off, temperature and tilt. It is accessed using an IP address, where the air conditioner has its own interface, so it is not part of any central system or interface where it would be possible to control several devices, and thus the device is also an end point directly.

Each attribute to control in this air condition has its own structure and specific data type. It was not possible to find the documentation for this interface, so it was necessary to analyze the HTTP requests sent by the given web application:

- On state - In this system, surprisingly, the on/off state is controlled by providing an “On” or “Off” string for the “wON” attribute.
- Temperature - This attribute is classically stored as an integer value named “wTEMP”. The interface also has a minimum value of 9 degrees and a maximum value of 33 degrees Celsius.
- Action attribute - An unknown attribute was also found, the meaning of which seems to be the type of action for the given request. This attribute was called “wSETd1” and it was necessary to specify “Set” as the value.

Another important factor is that when setting any attribute, the system expects the action attribute, “wSETd1”, to be sent as well. With this finding, it is necessary to implement a part in the system where it will be possible to set attributes whose current or static value will always be sent with each request.

Air conditioner Control URL Structure

This interface does not use classic communication using JSON, but uses the sending of attributes and values using the query parameter in the URL address. This request is executed as a POST request, with the request body being empty.

An example of such a request, where the power-on state is set to on and the temperature is set to 25 degrees, is for example:

```
POST http://<ip_address>/wSETd1=Set&wON=On&wTEMP=25
```

C. Evaluation of the analyzed systems

After analyzing the Philips Hue systems and the air condition system, important information was obtained. Philips Hue, with its extensive capabilities, provides useful insights into the control of various parameters and devices that are important when designing systems. The air condition system, although simpler, brings new perspectives and requirements for testing the universality of the system. For effective integration and creation of a flexible, universal system, it is necessary to include the following aspects:

- Defining the base URL of the API point,
- REST API endpoints, for getting and setting values,
- Authentication against an external REST API,
- Data structures that need to be sent to endpoints,

- Values of various data types, including minimum and maximum values for numeric types,
- The device can also be an endpoint directly,
- Ability to send specific data with each request,
- Data can also be sent as a query parameter for more flexible communication.

IV. Database design for the control system

Based on the data analysis that was performed on example systems, the following database structure was created in an iterative way, which is shown in the figure 1. The goal is to create a database solution that will be able to effectively manage various types of data coming from these systems.

When designing the database, the need for a flexible scheme was taken into account, which allows the addition of new types of devices and functions without the need for fundamental changes to the structure. This is achieved by using tables with generalized columns for devices and their attributes, which allows us to store different types of data in a uniform form.

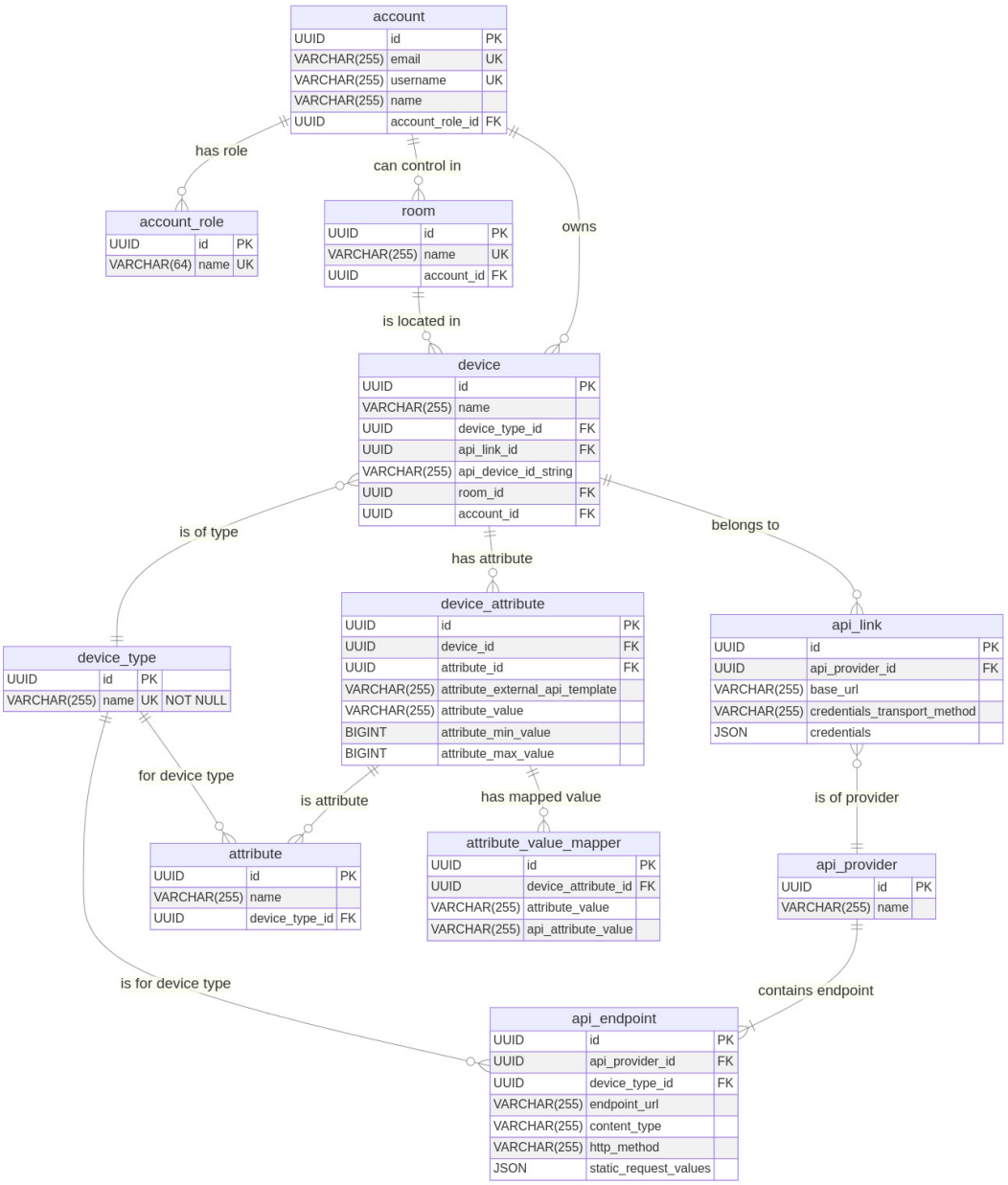


Fig. 1 Database model

The structure consists of tables:

- account - This table will contain basic information about system users. It includes data such as name, email address or even the role assigned to the user. It will make it possible to identify and manage individual users and their access to the system.

- account_role - Table intended for storing user roles that define access levels and authorizations in the system. Examples of roles are ”admin”, ”user”, each with different permissions. This table helps to implement security rules and manage access.
- room - Represents physical rooms or areas where smart devices are placed. Each record contains, for example, a room identifier and a name. It allows you to organize devices according to their physical location, so a device can contain this room identifier.
- device - This table stores information about specific smart devices managed by the system. It includes data such as device identifier, name, device type, API link to which it belongs, device identifier on the external system, or room identifier.
- device_type - The table defines the types of smart devices, such as ”light”, ”air conditioner”, etc. It stores information that characterizes each type, such as possible operations or specific attributes. It facilitates system expansion with new types of devices and simplifies their configuration.
- device_attribute - The table represents a link between specific devices and their attributes. It is used to assign values to these attributes for each device, such as the current room temperature or the set light intensity. Thus, the specific values of the attributes for each device are recorded here, as well as the min-max range if possible, enabling dynamic monitoring and control of the status of devices in the system. It also contains a JSON template for sending data to the external API, the method of which will be explained later.
- attribute - This table is used to store definitions of attributes that different devices in the system may have. Attributes can include properties such as temperature, humidity, on/off state, light intensity, and so on. Each attribute has its own unique name and information to which type of device it belongs.
- attribute_value_mapper - This table stores identifiers for device attributes that are defined in the system and associates them with corresponding attribute values from external systems. It is this table that allows the system to understand and translate which attribute value from the internal database corresponds to a specific attribute value from an external source. For example, an attribute value defined in this system as “true” may be represented as “On” in the external air condition API.
- api_provider - This table is a list of all API providers with which this system communicates. Each provider has a unique name, for example ”Philips Hue”. This name serves as a reference point to identify and group different instances or systems under a single provider. This approach enables centralized management and easier expansion of the system with new APIs without the need for duplicating configurations.
- api_link - Provides specific details for each instance of the system or service, such as basic URLs, methods of transferring login data (e.g. via HTTP header or cookies), and the login data itself in JSON format. Although storing credentials in JSON format may not be the most secure solution, and ideally sensitive data should be stored in a more secure repository such as Vault, this approach simplifies configuration for prototyping and development purposes. Each api_link is associated with a corresponding api_provider, allowing the system to dynamically access different instances of API providers under a single name.
- api_endpoint - This table is intended to hold specific endpoints for each API provider defined in the api_provider table. Each entry in api_endpoint contains a URL address that is specific to the given endpoint and defines the access path to the functionalities offered by the given API provider. This design allows the system to address various API functions such as data retrieval, status updates, or other actions that are specific to a particular device or service. It also contains which type of device the endpoint is intended for, the type of content in which to send data, i.e. JSON or query parameters, HTTP methods such as GET, PUT, POST and also static values for the given endpoint, if the device requires them to be sent in every request.

V. Implementation and testing of the system

As part of the implementation of the system for controlling intelligent devices, the main development steps were first the creation of database migrations, which ensured the basic structure of the database. Subsequently, a server interface was developed that automatically applies these migrations at startup and provides REST API endpoints for system management. These endpoints enable communication with the user interface, which contains subpages for creating and managing records in individual tables. These functionalities include adding new devices with appropriate attributes, defining endpoints, as well as provider APIs and lines that control these devices. This approach allows maintaining a high degree of configurability and universality of the system, which is essential for effective adaptation to different types of smart devices. The main part of the server

interface is also the processing of values from the database and the parsing of values into relevant data types, for example in the case of device attribute values, since for universality the values were stored as a string of characters. In the case of numerical values, there are minimum and maximum values, which again serves the server interface and thus ensures communication between the user interface and correct data storage in the database. The architecture of the application is shown in the Figure 2.

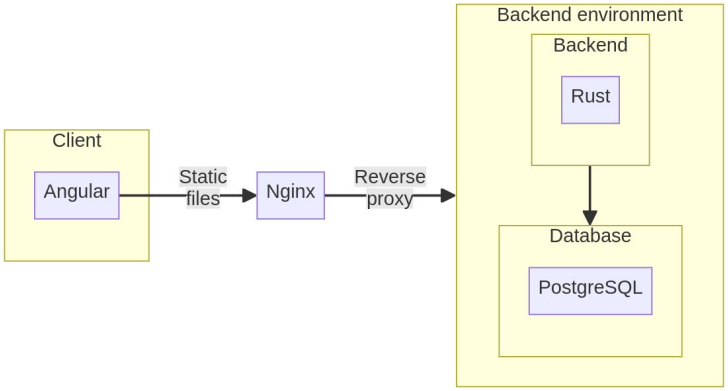


Fig. 2 System architecture

Tests were carried out on assigned devices, while we focused on controlling and managing various device attributes, the values of which were stored in the database. During the testing, we verified that the data is stored in the database correctly and that the links between the tables are correctly set, which ensures the correct functionality and integration of the various components of the system.

VI. Conclusion

At the end of this article, the key aspects and results of the designed system for controlling smart devices are evaluated. The main goal was to create a database model that information about devices is stored in the database with variable attributes and communication requirements via external APIs, which was achieved by using Rust and Angular technologies for reliable and correct system operation. Based on these experiences can we effectively develop smart features for overhead cranes, so the operation is saved, potential safety hazards are reduced, working conditions are intelligently monitored, and the working efficiency is improved. System functionality, including authentication, was successfully verified during testing. Future improvements could include adding functionality to both the system and the database to retrieve and store values of the current status of devices from external APIs, or implementing a Vault application into the system for better security.

ACKNOWLEDGMENT

This article was elaborated in the framework of the Grant Projects KEGA 044TUKE-4/2024

References

[1] J. Mocnej, A. Pekar, W. K. Seah, P. Papcun, E. Kajati, D. Cupkova, J. Koziorek, and I. Zolotova, “Quality-enabled decentralized iot architecture with efficient resources utilization,” *Robotics and Computer-Integrated Manufacturing*, vol. 67, p. 102001, 2021.
[2] J. Mocnej, A. Pekar, W. K. Seah, and I. Zolotova, *Network traffic characteristics of the IoT application use cases*. School of Engineering and Computer Science, Victoria University of . . . , 2018.
[3] Google Ireland Limited. A home that knows how to help. [Online]. Available: https://home.google.com/welcome/
[4] M. SCIACCO, “Integrating iot devices in 3rd-party smart-home ecosystems: local vs remote middleware solutions,” 2022.
[5] Samsung Electronics Co., LTD. Smartthings. [Online]. Available: https://www.samsung.com/sk/smartthings/app/
[6] L.-J. Hwang, T.-H. Tu, T.-H. Lin, I. Kuo *et al.*, “A learning environment for embedded systems.”
[7] Samsung Electronics Co., LTD. Create an app. [Online]. Available: https://developer.smartthings.com/docs/api/public#tag/Apps/operation/createApp
[8] Home Assistant. Awaken your home. [Online]. Available: https://www.home-assistant.io/
[9] A. Soler Mascarell, “Home automation platform based on node red,” B.S. thesis, Universitat Politècnica de Catalunya, 2021.

[10] J. Mocnej, W. K. Seah, A. Pekar, and I. Zolotova, "Decentralised iot architecture for efficient resources utilisation," *IFAC-PapersOnLine*, vol. 51, no. 6, pp. 168–173, 2018, 15th IFAC Conference on Programmable Devices and Embedded Systems PDeS 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896318308942

[11] Philips Hue. Hue clip api documentation version v2 get light. [Online]. Available: https://developers.meethue.com/develop/hue-api-v2/api-reference/#resource_light_get

Grafy na piatich vrcholoch

Michal STAŠ

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

michal.stas@tuke.sk

Abstrakt **– Tento príspevok sa zaoberá sumarizáciou výsledkov jednoduchých grafov na piatich vrcholoch v súčinovom spojení s diskrétnymi grafmi, cestami a kružnicami. Využitím grafov vieme reprezentovať rozličné štruktúry a problémy z najrôznejších odborov. Optimalizácia počtu krížení na hranách jednotlivých nakreslení grafov má pri takejto reprezentácii významný vplyv.**

Kľúčové slová **– sumarizácia, graf, priesečníkové číslo, súčinové spojenie dvoch grafov, diskrétny graf, cesta, kružnica**

I. Úvod

Priesečníkové číslo je dôležitým parametrom grafu, pretože poskytuje informácie o zložitosti grafu a taktiež o obtiažnosti jeho vizualizácie. Okrem toho počet krížení súvisí s mnohými ďalšími parametrami a algoritmami grafu, ako je testovanie rovinnosti a farbenie grafu. Grafy sa častokrát používajú na znázornenie zložitých sietí, akými sú napr. sociálne, komunikačné či dopravné siete (Obr. 1). Zníženie počtu krížení na hranách vo vizualizáciách siete napomáha lepšie pochopiť základnú štruktúru siete a identifikovať dôležité uzly a spojenia [1].



Obr. 1 Kríženia ciest

V pomerne rozsiahlej teórii grafov je znižovanie počtu pretínania hrán základným problémom aj pre teóriu planárnych grafov. Mnoho dôležitých grafových algoritmov a optimalizačných problémov je definovaných práve pre rovinné grafy, t.j. nakreslenia grafov bez prekríženia hrán [2]. V kartografii sa grafy používajú na znázornenie geografických prvkov ako sú cesty, rieky a politické hranice. Zníženie počtu preťatí hrán v takýchto grafoch napomáha k prehľadnejším a čitateľnejším mapám [3]. Azda najdôležitejšou oblasťou uplatnenia mimimalizácie kríženia hrán v nakresleniach grafov je VLSI (Very Large Scale Integration) technológia používaná pri navrhovaní a výrobe integrovaných obvodov (Obr. 2). Elektronické obvody sa skladajú z veľkého počtu tranzistorov a



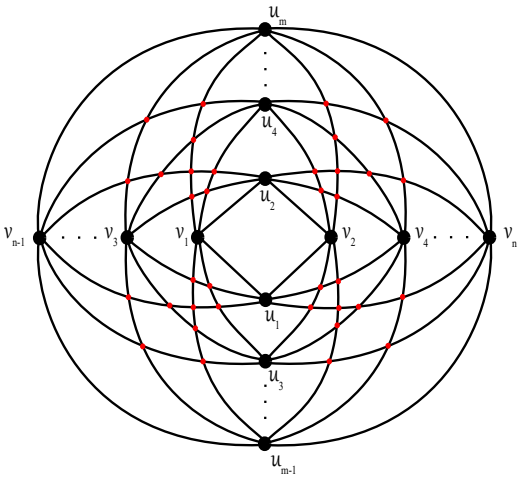
Obr. 2 VLSI technológia

iných komponentov, ktoré sú následne integrované do jedného čipu. Táto technológia umožňuje vytváranie vysoko komplexných a kompaktných integrovaných obvodov s možnosťou použitia v širokej škále aplikácií, od spotrebnej elektroniky a počítačov až po komunikačné systémy a zdravotnícke zariadenia. Pri návrhu elektronických obvodov je minimalizácia počtu krížení hrán dôležitá pre zníženie rušenia signálu a zlepšenie výkonu obvodu. Nakreslenia grafov s menším počtom krížení môžu viesť k efektívnejším a spoľahlivejším návrhom obvodov.

II. SÚČINOVÉ SPOJENIE DVOCH GRAFOV

Priesečníkové číslo $\mathrm{cr}(G)$ grafu G je minimálny počet hranových preťatí spomedzi všetkých možných nakreslení G v rovine. Súčinové spojenie dvoch grafov G_i a G_j, označované $G_i + G_j$, získame z vrcholovo disjunktných kópií G_i a G_j pridaním všetkých hrán medzi $V(G_i)$ a $V(G_j)$. Pre $|V(G_i)| = m$ a $|V(G_j)| = n$ je hranová množina $G_i + G_j$ zjednotením hranovo disjunktných množín grafov G_i, G_j a kompletného bipartitného grafu $K_{m,n}$. Ďalej nech D_n označuje diskrétny graf (niekedy nazývaný aj prázdny graf) na n vrcholoch, a P_n a C_n sú cesta a kružnica na n vrcholoch.

Určenie priesečníkového čísla kompletného bipartitného grafu $K_{m,n}$ je jedným z otvorených problémov v teórii grafov. Horným obmedzením $\mathrm{cr}(K_{m,n})$ je $Z(m,n) = \lfloor \frac{m}{2} \rfloor \lfloor \frac{m-1}{2} \rfloor \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor$, ktoré sa zvykne nazývať aj Zarankiewiczovo číslo. Zarankiewiczova hypotéza [4] o nadobudnutí rovnosti horného odhadu nebola dodnes potvrdená, no vďaka Kleitmanovi [5] bol tento dohad zatiaľ dokázaný pre všetky celé čísla $m, n \geq 1$ za predpokladu $\min\{m,n\} \leq 6$. Veľká časť už známych piesečníkových čísiel grafov je popísaná s využitím Zarankiewiczovho čísla $Z(m,n)$ (Obr. 3).



Obr. 3 Nakreslenie $K_{m,n}$ s $Z(m,n)$ priesečníkmi

Sumarizácia priesečníkových čísiel súčinových spojení grafov stupňa najviac štyri s diskrétnymi grafmi, cestami a kružnicami je známa vďaka Kleščovi a Schrötterovi [6], [7], [8]. Na základe viacerých publikovaných výsledkov z posledného obdobia môžeme predošlú sumarizáciu výsledkov zovšeobecniť aj pre jednoduché grafy na piatich vrcholoch. V tomto príspevku použijeme definície a zápis prisečníkových čísiel súvisých grafov, ktoré uvádza Klešč [6]. Kvôli jednoznačnosti zápisu použijeme špeciálne označenie rádu päť prostredníctvom horných indexov pre 13 nesúvislých grafov H_i^5 v Tabuľke 1 a 21 súvislých grafov G_i^5 v Tabuľke 2.

III. JEDNODUCHÉ GRAFY NA PIATICH VRCHOLOCH

Existuje trinásť nesúvislých grafov $H_1^5 = 5K_1$, $H_2^5 = 3K_1 \cup P_2$, $H_3^5 = 2K_1 \cup P_3$, $H_4^5 = K_1 \cup 2P_2$, $H_5^5 = K_1 \cup P_4$, $H_6^5 = P_2 \cup P_3$, $H_7^5 = 2K_1 \cup K_3$, $H_8^5 = K_1 \cup K_{1,3}$, $H_9^5 = K_1 \cup (K_{1,3} \cup \{e\})$, $H_{10}^5 = K_3 \cup P_2$, $H_{11}^5 = K_1 \cup C_4$, $H_{12}^5 = K_1 \cup K_{1,1,2}$ a $H_{13}^5 = K_1 \cup K_4$ rádu päť. Podľa Klešča [6] sú priesečníkové čísla súčinového spojenia $P_m + C_n$ jednoznačne určené ak $m \geq 2$, $n \geq 3$ a $\min\{m,n\} \leq 6$. Doplnením výsledkov z [9] pre všetky nesúvislé podgrafy cesty P_5 v súčinovom spojení s cestami a kružnicami získame $\mathrm{cr}(H_i^5 + D_n)$, $\mathrm{cr}(H_i^5 + P_n)$ a $\mathrm{cr}(H_i^5 + C_n)$ pre $i = 1, \ldots, 6$. Pri stanovení priesečníkových čísiel grafov $H_{10}^5 + D_n$, $H_{10}^5 + P_n$ a $H_{10}^5 + C_n$ v [10] boli silne využité viaceré paritné symetrie. Graf H_7^5 získame odobraním jednej hrany z grafu H_{10}^5 a tak sa predpokladá, že ich priesečníkové čísla v príslušných súčinových spojeniach budú rovnaké. Spoločný článok Klešča a Staša [11] je jednou z prvých publikácií, v ktorej boli využité vlastnosti cyklických permutácií. Jej výstupom sú $\mathrm{cr}(H_i^5 + D_n)$, $\mathrm{cr}(H_i^5 + P_n)$ a $\mathrm{cr}(H_i^5 + C_n)$ pre $i = 8, 9$. Grafu H_{11}^5 patrí prvenstvo nakoľko bol v [12] preskúmaný ako prvý nesúvislý graf obsahujúci kružnicu ako podgraf. Na základe myšlienky cyklických permutácií je publikovaný aj výsledok $\mathrm{cr}(H_{12}^5 + D_n)$ Klešča, Staša a Petrillovej [13], a neskôr doplnený o výstupy s P_n a C_n v [14]. Podobne to bolo aj v prípade grafu H_{13}^5, priesečníkové čísla súčinového spojenia s diskrétnymi grafmi D_n v [15] boli postupne doplnené o výstupy s cestami P_n [16] a kružnicami C_n [17].

Tabuľka 1
Sumarizácia priesečníkových čísiel $H_i^5 + D_n$, $H_i^5 + P_n$ a $H_i^5 + C_n$

i	H_i^5	$\mathrm{cr}(H_i^5 + D_n)$, $n \geq 1$	$\mathrm{cr}(H_i^5 + P_n)$, $n \geq 2$	$\mathrm{cr}(H_i^5 + C_n)$, $n \geq 3$
1		$Z(5,n)$	$Z(5,n)$	$Z(5,n)$
2		$Z(5,n)$	$Z(5,n)$	$Z(5,n)$
3		$Z(5,n)$	$Z(5,n)$	$Z(5,n)$
4		$Z(5,n)$	$Z(5,n)$	$Z(5,n)$
5		$Z(5,n)$	$Z(5,n)$	$Z(5,n)+1$
6		$Z(5,n)$	$Z(5,n)$	$Z(5,n)$
7		$?n^2 - 2n + \lfloor \frac{n}{2} \rfloor$ $(n \geq 2)?$	$?n^2 - 2n + \lfloor \frac{n}{2} \rfloor ?$	$?n^2 - 2n + \lfloor \frac{n}{2} \rfloor ?$
8		$Z(5,n) + \lfloor \frac{n}{2} \rfloor$	$Z(5,n) + \lfloor \frac{n}{2} \rfloor$	$Z(5,n) + \lfloor \frac{n}{2} \rfloor + 1$
9		$Z(5,n) + \lfloor \frac{n}{2} \rfloor$	$Z(5,n) + \lfloor \frac{n}{2} \rfloor$	$Z(5,n) + \lfloor \frac{n}{2} \rfloor + 1$
10		$n^2 - 2n + \lfloor \frac{n}{2} \rfloor$ $(n \geq 2)$	$n^2 - 2n + \lfloor \frac{n}{2} \rfloor$	$n^2 - 2n + \lfloor \frac{n}{2} \rfloor$
11		$Z(5,n) + \lfloor \frac{n}{2} \rfloor$	$Z(5,n) + \lfloor \frac{n}{2} \rfloor + 1$	$Z(5,n) + \lfloor \frac{n}{2} \rfloor + 2$
12		$Z(5,n) + 2\lfloor \frac{n}{2} \rfloor$	$Z(5,n) + 2\lfloor \frac{n}{2} \rfloor$	$Z(5,n) + 2\lfloor \frac{n}{2} \rfloor + 2$ $(n \geq 4)$
13		$Z(5,n) + n + \lfloor \frac{n}{2} \rfloor$	$Z(5,n) + n + \lfloor \frac{n}{2} \rfloor + 1$ $(n \geq 3)$	$Z(5,n) + n + \lfloor \frac{n}{2} \rfloor + 3$

Poznáme dvadsaťjeden súvislých grafov $G_1^5 = P_5$, $G_2^5 = K_{1,4}$, $G_3^5 = H_8^5 \cup \{e\}$ s tromi listami, $G_4^5 = H_{10}^5 \cup \{e\}$, $G_5^5 = G_1^5 \cup \{e\}$ s dvomi listami, $G_6^5 = K_{1,4} \cup \{e\}$, $G_7^5 = H_{11}^5 \cup \{e\}$, $G_8^5 = C_5$, $G_9^5 = K_1 + 2P_2$, $G_{10}^5 = K_{2,3}$, $G_{11}^5 = K_1 + (P_3 \cup K_1)$, $G_{12}^5 = G_4^5 \cup \{e\}$ s jedným listom, $G_{13}^5 = G_8^5 \cup \{e\}$, $G_{14}^5 = K_1 + P_4$, $G_{15}^5 = K_{1,1,3}$, $G_{16}^5 = H_{13}^5 \cup \{e\}$, $G_{17}^5 = 2K_1 + (P_2 \cup K_1)$, $G_{18}^5 = P_2 + (P_2 \cup K_1)$, $G_{19}^5 = K_1 + C_4$, $G_{20}^5 = K_5 \setminus e$, $G_{21} = K_5$ na piatich vrcholoch. Opäť podľa Klešča [6] sú priesečníkové čísla súčinového spojenia $C_m + C_n$ jednoznačne určené ak $m, n \geq 3$ a $\min\{m, n\} \leq 6$. Na základe oboch tvrdení získavame výstupy pre dva grafy G_1^5 a G_8^5. Medzi ďalšie významné výstupy skôr zmienenej publikácie [11] patrí určenie $\mathrm{cr}(G_i^5 + D_n)$, $\mathrm{cr}(G_i^5 + P_n)$ a $\mathrm{cr}(G_i^5 + C_n)$ pre $i = 3, 4, 5, 7, 13$. Je však potrebné poznamenať, že prvenstvo publikovania vlastností grafov G_7^5 a G_{13}^5 prináleží Kleščovi and Schrötterovi [18]. Huang a Zhao [19] sa s využitím myšlienky špeciálnej dekompozičnej techniky zaslúžili o stanovenie $\mathrm{cr}(G_i^5 + D_n)$ pre $i = 2, 6, 9, 11, 14$. Tieto výstupy boli s pomerne veľkým časovým odstupom doplnené aj o priesečníkové čísla súčinových spojení s P_n a C_n [20]. V prípade grafov G_{10}^5 a G_{17}^5 sa priesečníkové čísla súčinového spojenia s diskrétnymi grafmi D_n v [21] postupne doplnili o výstupy s cestami P_n [22] a kružnicami C_n [23]. Dôležitým míľnikom v teórii grafov bolo stanovenie priesečníkového čísla kompletného tripartitného grafu G_{15}^5 v súčinovom spojení s D_n [24]. O doplnenie výstupov pre G_{12}^5, G_{15}^5 a G_{18}^5 s cestami a kružnicami sa zaslúžili Staš a Timková [14] s odvolaním sa na už dokázanú hodnotu $\mathrm{cr}(G_{12}^5 + D_n)$ [13]. Podobne to bolo aj v prípade grafu G_{16}^5, kedy sa priesečníkové čísla súčinového spojenia s diskrétnymi grafmi D_n [15] postupne skompletizovali aj o výstupy s cestami P_n [16] a kružnicami C_n [17]. Taktiež významný výsledok Dinga a Huanga [25] o $\mathrm{cr}(G_{19}^5 + D_n)$ bol neskôr doplnený o výstupy s P_n a C_n [26]. Vďaka jednej z najvýznamnejších publikácií o kompletných viacpartitných grafov [27] poznáme presnú hodnotu priesečníkového čísla grafu G_{20}^5 v súčinovom spojenní s D_n a aj na tento výsledok boli po dlhšom období úspešne naviazané výstupy s P_n a C_n [28]. Fakt o neplanárnosti kompletného grafu G_{21}^5 je zrejmá najskôr každému. Už dlhší čas je aj prostredníctvom Lua a Huanga [29] známa hodnota $\mathrm{cr}(G_{21}^5 + D_n)$, no výsledky pre cesty a kružnice sú zatiaľ len v tvare horných odhadov $\mathrm{cr}(G_{21}^5 + P_n) \leq Z(5,n) + 2n + \lfloor \frac{n}{2} \rfloor + 4$ a $\mathrm{cr}(G_{21}^5 + C_n) \leq Z(5,n) + 2n + \lfloor \frac{n}{2} \rfloor + 7$.

Tabuľka 2
Sumarizácia priesečníkových čísiel $G_i^5 + D_n$, $G_i^5 + P_n$ a $G_i^5 + C_n$

i	G_i^5	$\mathrm{cr}(G_i^5 + D_n)$, $n \geq 1$	$\mathrm{cr}(G_i^5 + P_n)$, $n \geq 2$	$\mathrm{cr}(G_i^5 + C_n)$, $n \geq 3$
1		$Z(5,n)$	$Z(5,n)$	$Z(5,n)+1$
2		$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor+2$
3		$Z(5,n)+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor+1$
4		$Z(5,n)+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor+1$
5		$Z(5,n)+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor+2$
6		$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor+2$
7		$Z(5,n)+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor+1$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor+2$
8		$Z(5,n)$	$Z(5,n)+1$	$Z(5,n)+2$
9		$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor+2$
10		$Z(5,n)+n$	$Z(5,n)+n+1$	$Z(5,n)+n+3$
11		$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor+1$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor+3$
12		$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor+2$
13		$Z(5,n)+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor+1$	$Z(5,n)+\lfloor\frac{n}{2}\rfloor+2$
14		$Z(5,n)+2\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor+1$	$Z(5,n)+2\lfloor\frac{n}{2}\rfloor+3$
15		$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor+2$	$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor+4$
16		$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor+1$ $(n \geq 3)$	$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor+3$
17		$Z(5,n)+n$	$Z(5,n)+n+1$	$Z(5,n)+n+3$
18		$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor+2$	$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor+4$
19		$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor$	$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor+1$	$Z(5,n)+n+\lfloor\frac{n}{2}\rfloor+4$
20		$Z(5,n)+2n$	$Z(5,n)+2n+2$	$Z(5,n)+2n+5$
21		$Z(5,n)+2n+\lfloor\frac{n}{2}\rfloor+1$	$?Z(5,n)+2n+\lfloor\frac{n}{2}\rfloor+4?$	$?Z(5,n)+2n+\lfloor\frac{n}{2}\rfloor+7?$

V Tabuľkách 1 a 2 si môžeme všimnúť štyri výnimky, ku ktorým dochádza práve v najmenších možných ohodnoteniach premennej n. V prvom prípade grafu H_{10}^5 je $\mathrm{cr}(H_{10}^5 + D_1) = 0$ kvôli problému so zápornou hodnotou pre $n = 1$. V zvyšných troch prípadoch sú výnimky zapríčinené riedkosťou optimálnych nakreslení uvažovaných grafov $H_{12}^5 + C_3$, $H_{13}^5 + P_2$ a $G_{16}^5 + P_2$. Chýbajúce hodnoty sú v tvare: $\mathrm{cr}(H_{12}^5 + C_3) = 7$, $\mathrm{cr}(H_{13}^5 + P_2) = 3$ a $\mathrm{cr}(G_{16}^5 + P_2) = 3$.

LITERATÚRA

[1] O. Aichholzer, R. Fabila-Monroy, A. Fuchs, C. Hidalgo-Toscano, I. Parada, B. Vogtenhuber, and F. Zaragoza, "On the 2-Colored Crossing Number," *In Proceedings of the 27th International Symposium on Graph Drawing and Network Visualization*, p. 87–100, 2019.
[2] E. Gethner, L. Hogben, B. Lidicky, F. Pfender, A. Ruiz, and M. Young, "On Crossing Numbers of Complete Tripartite and Balanced Complete Multipartite Graphs," *J. Graph Theory*, vol. 84, pp. 552–565, 2017.
[3] B. Jenny, D. M. Stephen, I. Muehlenhaus, B. E. Marston, R. Sharma, E. Zhang, and H. Jenny, "Design principles for origin-destination flow maps," *Cartogr. Geogr. Inf. Sci.*, vol. 45, pp. 62–75, 2018.
[4] K. Zarankiewicz, "On a problem of P. Turan concerning graphs," *Fundamenta Mathematicae*, vol. 41, pp. 137–145, 1955.
[5] D. J. Kleitman, "The crossing number of $K_{5,n}$," *J. Combinatorial Theory*, vol. 9, pp. 315–323, 1970.
[6] M. Klešč, "The join of graphs and crossing numbers," *Electron. Notes in Discrete Math.*, vol. 28, pp. 349–355, 2007.
[7] ——, "The crossing numbers of join of cycles with graphs of order four," *Proc. Aplimat 2019:* 18^{th} *Conference on Applied Mathematics*, pp. 634–641, 2019.
[8] M. Klešč and Š. Schrötter, "The crossing numbers of join products of paths with graphs of order four," *Discuss. Math. Graph Theory*, vol. 31, pp. 321–331, 2011.
[9] M. Staš and M. Švecová, "Disconnected spanning subgraphs of paths in the join products with cycles," *Art Discrete Appl. Math.*, vol. 6, p. 3.06, 2023.
[10] M. Staš, "Parity Properties of Configurations," *Mathematics*, vol. 10, p. 1998, 2022.
[11] M. Klešč and M. Staš, "Cyclic permutations in determining crossing numbers," *Discuss. Math. Graph Theory*, vol. 42, pp. 1163–1183, 2022.
[12] L. Li, "The Crossing Numbers of Join of a 5-Vertex Graph with Path and Cycle," *Math. Prac. Th.*, vol. 44, pp. 203–211, 2014.
[13] M. Klešč, M. Staš, and J. Petrillová, "The crossing numbers of join of special disconnected graph on five vertices with discrete graphs," *Graphs and Combinatorics*, vol. 38, p. 35, 2022.
[14] M. Staš and M. Timková, "The crossing numbers of join products of four graphs of order five with paths and cycles," *Opuscula Math.*, vol. 43, pp. 865–883, 2023.
[15] M. Staš, "On the crossing number of join product of the discrete graph with special graphs of order five," *Electron. J. Graph Theory Appl.*, vol. 8, pp. 339–351, 2020.
[16] M. Staš and M. Švecová, "The crossing numbers of join products of paths with three graphs of order five," *Opuscula Math.*, vol. 42, pp. 635–651, 2022.
[17] M. Staš and M. Timková, "The crossing numbers of join products of $K_4 \cup K_1$ with cycles," *Communications in Combinatorics and Optimization*, [to appear].
[18] M. Klešč and Š. Schrötter, "The crossing numbers of join of paths and cycles with two graphs of order five," *Combinatorial Algorithms*, vol. 7125, pp. 160–167, 2012.
[19] Y. Huang and T. Zhao, "The crossing number of $K_{1,4,n}$," *Discret. Math.*, vol. 308, pp. 1634–1638, 2008.
[20] M. Staš, "The Crossing Numbers of Join Products of Paths and Cycles with Four Graphs of Order Five," *Mathematics*, vol. 9, p. 1277, 2021.
[21] K. Asano, "The crossing number of $K_{1,3,n}$ and $K_{2,3,n}$," *J. Graph Theory*, vol. 10, pp. 1–8, 1986.
[22] Z. Su and Y. Huang, "Crossing numbers of $K_{2,3} \vee P_n$," *Appl. Math. China Ser. A*, vol. 27, p. 488–492, 2012.
[23] M. Staš, "Join Products $K_{2,3} + C_n$," *Mathematics*, vol. 8, p. 925, 2020.
[24] P. Ho, "The crossing number of $K_{1,1,3,n}$," *Ars Combinatoria*, vol. 99, pp. 461–471, 2011.
[25] Z. Ding and Y. Huang, "The crossing numbers of join of some graphs with n isolated vertices," *Discuss. Math. Graph Theory*, vol. 38, pp. 899–909, 2018.
[26] M. Staš and J. Valiska, "On the crossing numbers of join products of $W_4 + P_n$ and $W_4 + C_n$," *Opuscula Math.*, vol. 41, pp. 95–112, 2021.
[27] P. Ho, "On the crossing number of some complete multipartite graphs," *Utilitas Math.*, vol. 79, pp. 125–143, 2009.
[28] M. Staš and M. Timková, "The influence of separating cycles in drawings of $K_5 \setminus e$ in the join product with paths and cycles," *Mathematica Slovaca*, [to appear].
[29] S. Lu and Y. Huang, "The crossing number of $K_5 \times S_n$," *J. Math. Res. Exp.*, vol. 28, pp. 445–459, 2008.

Návrh zariadenia pre výrobu vodíka

Vladimír SZOMOSI, Dušan MEDVEĎ

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

vladimir.szomosi@student.tuke.sk, dusan.medved@tuke.sk

***Abstrakt* — Tento článok sa zaoberá počiatočnou fázou návrhu elektrolyzéra s cieľom efektívnej výroby vodíka elektrochemickým procesom. Analyzované sú rôzne parametre, ako sú teplota, koncentrácia elektrolytu a typy elektród na základe získaných údajov z predošlého výskumu zaoberajúceho sa možnosťami využitia elektrolyzéra pre elektroenergetické účely, s cieľom dosiahnuť optimálne podmienky pre elektrolýzu. Víziou je vytvorenie efektívneho elektrolyzéra, ktorý by mohol slúžiť na získavanie čistého vodíka určeného na akumuláciu nadbytočnej elektrickej energie.**

***Kľúčové slová* — vodík, elektrolýza, kyslík, elektrochemická premena, elektrolyzér, elektródy, niklová pena**

I. Úvod

Predošlá publikácia, o ktorú sa budeme opierať, sa zaoberala výrobou alkalického elektrolyzéra slúžiaceho na výrobu Brownovho plynu (HHO). Využitie HHO je avšak veľmi limitné v praktickej aplikácii. Z hľadiska využitia a bezpečnosti je oveľa zaujímavejšie vyprodukovať čistý vodík priamo elektrochemickou premenou. V tejto počiatočnej teoretickej fáze je cieľom správny výber materiálov a zvolenie efektívneho a reálne uskutočniteľného návrhu konkrétneho elektrolyzéra. Konečným cieľom výskumu je produkcia vodíka v navrhovanom elektrolyzéry [8].

II. Alkalická elektrolýza

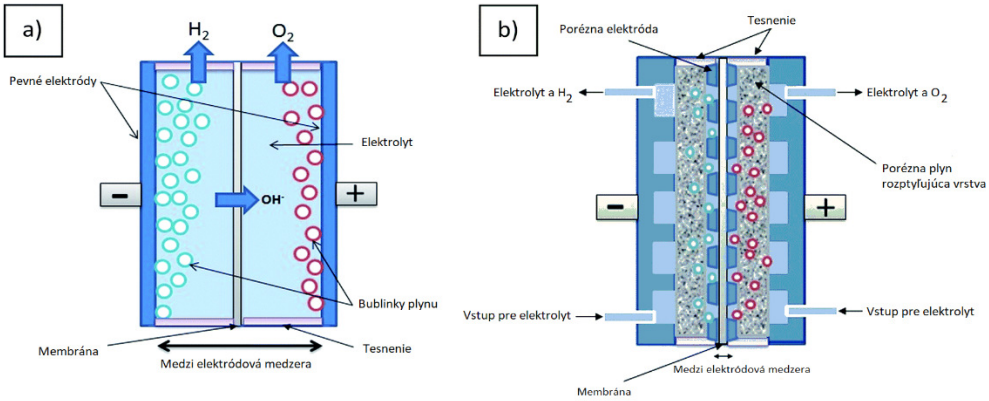
Alkalická elektrolýza je elektrochemický proces, ktorý je aktívny pri pôsobení stáleho jednosmerného prúdu na elektródové platne oddelené kvapalným alkalickým elektrolytom. Jednosmerný prúd spôsobí rozklad molekúl alkalického vodného roztoku procesom oxidačným alebo redukčným, s účinnosťou štandardne nepresahujúcou 60 %. Proces prebieha v uzavretej sústave elektrolyzéra zloženej z anódy, katódy a bipolárnych elektród. Proces redukcie prebieha na katóde, kde atómy elektrolytu spájajú na povrchu elektródy s elektrónmi, vzniká tak vodík. Oxidácia (odovzdávanie valenčných elektrónov) prebieha na anóde za vzniku kyslíka. Pri HHO elektrolyzéri je vodík a kyslík vylučovaný do jednej komory, vzniká tak zmes plynov v pomere 2:1. HHO má vyšší energetický potenciál ako vodík, problémom je ale jeho nestabilita. Jedná sa o vysoko výbušnú zmes, ktorá pri kontakte zo zápalným médiom imploduje.

Relatívne nízka účinnosť alkalickej elektrolýzy viedla aj k vývoju elektrolýzy typu PEM (Proton Exchange Membrane) a taktiež elektrolýzy pevných oxidov. PEM elektrolyzér je schopný pracovať s účinnosťou až 72 % pri vysokej prúdovej zaťažiteľnosti článku, približne 2 $A \cdot cm^{-2}$ oproti alkalickej elektrodýze 0,25 $A \cdot cm^{-2}$, kde ide o signifikantný nárast. Problémom je ale skutočnosť, že tieto vlastnosti sú kompenzované nákladmi na ušľachtilé katalyzátory membrán a elektród (platina, irídium). Vysokoteplotná elektrolýza pevných oxidov si vyžaduje dodatočné teplo (teda zvýšenie teploty) pre svoju funkciu, takže sa zvyšuje jej energetická náročnosť a je obťažná na konštrukciu a návrh.

PEM elektrolyzér pracuje na systéme článku s nulovou medzerou medzi membránou a elektródami. Pri implementovaní tohto riešenia do prostredia alkalického elektrolyzéra, ktorého výhodu je použite lacnejších materiálov, je možné dosiahnuť vysokú hustotu prúdu a účinnosť s nízkymi nákladmi.

Zloženie alkalického elektrolyzéra s nulovou medzerou je založené na poréznej vlastnosti elektród (anódy a katódy), ktorými bude obložená vodivá membrána prepúšťajúca hydroxidové ióny. Dosiahne sa tak medzera hrúbky membrány a zníži sa tak odpor elektrolytu. Redukcia

vodíka tak prebieha vo vnútri poréznej elektródy a núti plynové bublinky vznikajúce v poréznej štruktúre vystupovať zo zadnej časti elektród skrz prestupy elektrolyzéra [5], [8].



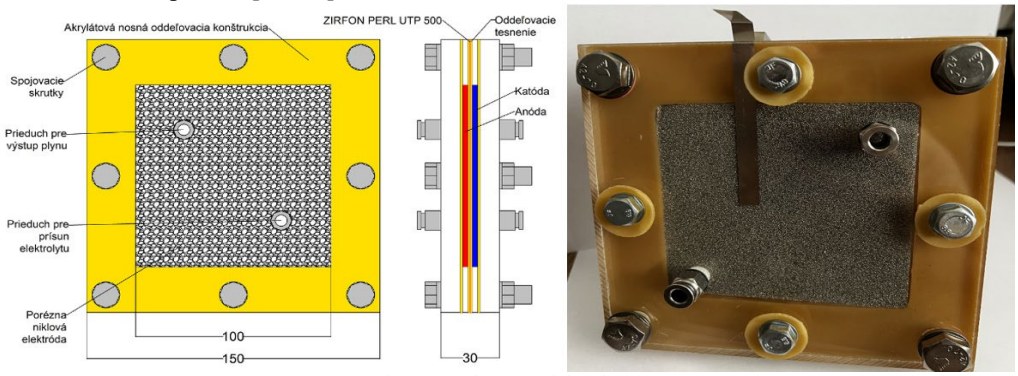
Obr. 1 a) Klasické usporiadanie alkalického elektrolyzéra, b) Usporiadanie elektrolyzéra s nulovou medzerou [5]

III. KONCEPT ELEKTROLYZÉRA

Primárnym zariadením v sústave slúžiacej na výrobu vodíka je elektrolyzér. Elektrolyzér bol navrhovaný na základe skúseností zo spomínaného výskumu. Princíp spočíva na suchom vyhotovení zloženom z jedného elektródového páru, anódy a katódy vyrobených z poréznej niklovej peny s mierou porozity 75 PPI (pores per inch) udávanej výrobcom. Elektródy sú oddelené membránou navrhnutou pre alkalickú elektrolýzu ZIFRON PERL UTP 500, ktorú poskytla spoločnosť Agfa-Gevaert NV na vedecké účely [1]. Hlavné komponenty, membránu a elektródy, spája priehľadná akrylátová konštrukcia navrhnutá tak, aby bolo možné sledovať elektrochemický proces tvorby bubliniek plynu na elektródach. Aparát elektrolyzéra produkuje separovane vodík a kyslík, ktoré je potrebné odviesť zo zariadenia; to zabezpečujú prestupy umiestené na bočných akrylátových paneloch. Set prestupov zároveň zabezpečuje prítok elektrolytu z rezervovaných nádrží na elektrolyt tak, aby sa zabezpečila cirkulácia elektrolytu a plynu. Rezervoáre patria k obslužným zariadeniam. Výstupom z nádrže na strane katódy je vodík v plynnej forme bez elektrolytu, ktorý zostáva v nádržke. Na strane anódy sa vylučuje vedľajší produkt – kyslík, ktorý bude voľne vylúčený do priestoru [1, 5, 6, 8].

A. Elektrolyzér

Pre správne fungovanie elektrolyzéra je potrebné priviesť na elektródy jednosmerný prúd, ten vyvolá elektrochemickú reakciu vo vnútri poréznych elektród stlačených k sebe. Elektródy sú ponorené vo vodnom roztoku hydroxidu draselného KOH. Silné korozívne účinky elektrolytu si vyžadujú použiť materiál odolný zásadám. Pre overenie tejto schopnosti bola niklová pena vystavená kúpeľu v 30 % roztoku KOH, kedy neboli pozorované negatívne účinky na vlastnosti materiálu. Antikorózne vlastnosti musia mať taktiež aj všetky prvky sústavy, ktoré prídu do kontaktu s elektrolytom. Preto sú ako hlavný typ materiálu použité polyméry a oceľ s antikoróznou úpravou [5, 6, 8].



Obr. 2 Navrhovaný elektrolyzér

B. Elektródy

Pri návrhu HHO elektrolyzéra boli použité oceľové elektródy s antikoróznou úpravou. Problémom je, že pri produkcii plynu sa pri elektrolýze vylučujú oxidy chrómu a oxidu železitého, spôsobené pochrómovaním oceľových elektród. Znečisteniu sa tak vyhneme použitím niklu, ktorý sa využíva v rôznych formách pri elektrolýze. Nízka cena niklu je jednou z hlavných

výhod alkalických elektrolyzérov. Náklady na poréznu niklovú elektródu rozmerov 100 × 100 × 5 mm, budú približne 15 €, čo je lacnejšie ako elektródy z ušľachtilých kovov irídia alebo platiny.

Primárna časť elektrolyzéra pozostávajúca z dvoch poréznych elektród anódy a katódy, ktoré sú v kontakte s membránou, sa nazýva MEA (Membrane Electrode Assembly). Výrobu primárnej časti MEA možno rozdeliť do dvoch základných kategórií, kedy rozlišujeme aplikáciu katalyzátora na buď membránu alebo elektródu [2, 5, 6, 8].

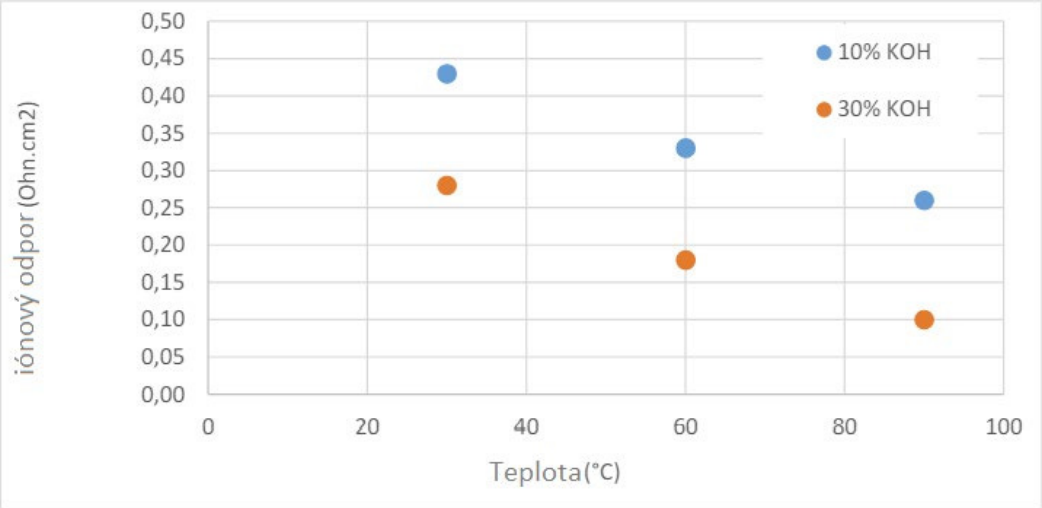
Aplikáciou katalyzátora na poréznu niklovú elektródu sa môže výrazne zvýšiť produkcia vodíka. V praxi využívanou metódou na túto aplikáciu je metóda elektrodepozície. Elektrokatalyzátor, teda vodný roztok Ni/Fe, sa nanáša na elektródu v elektrolytickej vani pod vplyvom konštantného jednosmerného prúdu o hustote 300 $mA \cdot cm^{-2}$ počas 30 s pri 25 °C. V navrhovanom elektrolyzéri sa táto úprava elektród plánuje vykonať až v nadchádzajúcom výskume [4].

Membrána pokrytá katalyzátorom sa zväčša využíva v PEM elektrolyzéroch. Katalyzátormi sú v tomto prípade ušľachtilé kovy (Ir_{02} pre anódu, Pt pre katódu) vo forme polymérovej pasty, ktorá je nanesená tenkou vrstvou na obe strany membrány. Poskytnutá membrána ZIFRON PERL UTP 500 je priamo určená pre alkalickú elektrolýzu, takže táto úprava poréznej membrány nebude potrebná [1, 5].

Prahové napätie elektrochemickej premeny sa nazýva reverzibilné napätie $E_{rev} = 1{,}23$ V. Problémom je však udržanie prahového napätia medzi elektródami z dôvodu premenlivého odporu elektrolytu spôsobeného teplotou a cirkuláciou elektrolyzérom. Tento problém je možné čiastočne vyriešiť dobrým prítokom elektrolytu využitím čerpadiel. Teplotný koeficient elektrolytu narastá vplyvom zvyšujúcej sa teploty. Ide o približne 2 % nárast vodivosti o jeden teplotný stupeň, preto sa v praxi alkalické elektrolyzéry prevádzkujú pri relatívne vysokých teplotách. Ďalším faktorom je medzera medzi elektródami, tento problém je zväčša vyriešený, keďže vyhotovenie elektrolyzéra je založené na nulovej medzere. V praxi elektrolyzéry pracujú pri vyššom prahovom napätí z dôvodu použitého typu elektrolytu, ktorého pH taktiež vplýva na E_{rev}. Pri vodnom roztoku NaOH je $E_{rev} = 1{,}69$ V a pri KOH je $E_{rev} = 1{,}67$ V. Ak bude napätie medzi elektródami prevyšovať tieto hodnoty, budú vznikať tepelné straty a dochádza k poklesu účinnosti. V našom prípade používame článok pozostávajúci z anódy a katódy. Napätie na svorkách elektrolyzéra by sa malo pohybovať pri mierne vyšších hodnotách s uvažovaním strát. Pri meraní napätia už funkčného zariadenia si tento predpoklad overíme [5, 7, 8].

C. Membrána

Oddeľovacia membrána pre alkalickú elektrolýzu ZIRFON PERL UTP 500 je zložená z polyfenylénsulfidovej tkaniny s otvorenou sieťovinou, ktorá je symetricky potiahnutá zmesou polyméru a oxidu zirkoničitého. Membrána môže byť použitá v atmosférických elektrolyzéroch – teda aj v predmetom navrhovanom elektrolyzéri, ale aj v elektrolyzéroch pracujúcich pod tlakom. Maximálna prevádzková teplota membrány je 110 °C. V aparáte je plánované maximálne zaťaženie teplotou do 40 °C, aby nedošlo k vyparovaniu elektrolytu. Štruktúra membrány je porézna, zabezpečuje sa tak iónová výmena, pri separácií plynov, ktoré vďaka veľkosti molekúl nepreniknú cez štruktúru membrány. Treba poznamenať, že iónový odpor membrány klesá so zvyšujúcou sa teplotou. Pri priemyselných elektrolyzéroch je výhodné pracovať pri vysokých teplotách elektrolytu. Klesajúci odpor je znázornený na obr. 3, t.j. závislosti iónového odporu od teploty elektrolytu, pri rôznych koncentráciách elektrolytu od 10 % do 30 % [1, 5, 8].



Obr. 3 Závislosť iónového odporu na teplote elektrolytu [1]

IV. Výpočet objemu produkcie vodíka a kyslíka

Približným výpočtom určíme produkciu vodíka a kyslíka podľa Faradayových vzťahov. Problémom je zistenie elektrického náboja Q ,respektíve prúdu I, ktorým budeme napájať proces elektrolýzy. Ak by bol aparát elektrolyzéra v prevádzke, bolo by jednouché odčítať priemernú veľkosť prúdu priamo z jednosmerného zdroja. Pri výpočte budeme uvažovať, že prúdová zaťažiteľnosť bežného alkalického elektrolyzéra je 0,25 A·cm^{-2},podľa geometrických rozmerov plochy elektródy odhadneme veľkosť prúdu na 2,5 A; túto veľkosť dosadíme do výpočtu elektrického náboja. Výsledné množstvo vyprodukovaného vodíka V_{H} následne overíme experimentálnym meraním, kde odmeriame množstvo vyrobeného vodíka pri nastavenom prúde 2,5 A za čas 60 s [3, 7].

$$V_{\mathrm{H}} = \frac{R \cdot T \cdot Q}{F \cdot p \cdot z_{\mathrm{H}}} = \frac{8{,}314 \cdot 293 \cdot (60 \cdot 2{,}5)}{9{,}648 \cdot 10^4 \cdot 101{,}325 \cdot 2} = 0{,}0187\ \mathrm{l/\,min} \quad (1)$$

$$V_{\mathrm{O}} = \frac{R \cdot T \cdot Q}{F \cdot p \cdot z_{\mathrm{O}}} = \frac{8{,}314 \cdot 293 \cdot (60 \cdot 2{,}5)}{9{,}648 \cdot 10^4 \cdot 101{,}325 \cdot 4} = 0{,}00934\ \mathrm{l/\,min} \quad (2)$$

kde: R plynová konštanta, 8,314 J·K^{-1}·mol^{-1}

T teplota okolia

Q elektrický náboj

z_H počet voľných elektrónov vodíka

z_O počet voľných elektrónov kyslíka

F Faradayova konštanta

p atmosférický tlak

Podľa predošlej skúsenosti, pri výpočte produkcie HHO plynu z predošlého výskumu, boli vypočítané hodnoty nepresné s viac ako 50 % odchýlkou. Cieľom je zistiť, do akej miery sa budú líšiť aj pri tomto type elektrolyzéra a vyvodiť záver, či ide o správny výpočet. Problémom pri určovaní bude zistiť čistotu produkovaného vodíka [7, 8].

V. Príprava merania

Skonštruovaný elektrolyzér so všetkými obslužnými aparatúrami bude podrobený sérií meraní ako pri spomínanom predošlom výskume. Cieľom meraní bude pomocou meracej aparatúry odmerať produkciu vodíka pri konštantách alebo časovo závislých parametroch a porovnať ich s predpokladom. Časovo závislé meranie bude určené na skúmanie, aký vplyv má doba prevádzky na správanie elektrolyzéra. Z predošlej skúsenosti vieme, že pri dlhšej prevádzke sa teplota elektrolytu zvyšuje a dochádza k zvýšenej produkcii plynu z dôvodu zvýšenia odoberaného elektrického prúdu zo zdroja. Meranie bude sledovať, či dôjde k porovnateľnému nárastu teploty aj pri elektrolyzéri s menším prúdovým zaťažením [7, 8].

VI. Záver

Cieľom príspevku bolo prestaviť počiatočnú fázu návrhu zariadenia na výrobu vodíka, ktorá nepriamo nadväzuje na predchádzajúcu prácu [7] zaoberajúcu sa možnosťami využitia elektrolyzéra na elektroenergetické účely. Ďalším krokom je spustenie už skonštruovaného

elektrolyzéra do prevádzky a vykonanie potrebných funkčných skúšok. Následne zrealizujeme merania, overíme hypotézy a uvedieme závery.

Poďakovanie

Tento príspevok vznikol vďaka Agentúre na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576 a APVV-21-0312, ako aj Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporená grantom VEGA 1/0627/24.

Referencie

[1] Agfa-Gevaert N.V., „*ZIRFON PERL UTP 500*“, [online] [cit. 2024-05-22]. Dostupné na internete: <https://www.agfa.com/specialty-products/wp-content/uploads/sites/8/2020/06/TDS_ZIRFON_PERL_UTP_500_20200525.pdf>.

[2] M. Bodner, et al. „*H2 generation from alkaline electrolyzer*“, In: WIREs Energy and Environment. 2015. Vol. 4, no. 4, s. 365–381. doi: 10.1002/wene.150.

[3] A. K. E. Soly et al., „*Comparative experimental investigation of oxyhydrogen (HHO) production rate using dry and wet cells*“. In: International Journal of Hydrogen Energy. 2021. Vol. 46, no. 24, s. 12639–12653. doi: 10.1016/j.ijhydene.2021.01.110.

[4] F. J. Pérez-Alonso, „*Ni/Fe electrodes prepared by electrodeposition method over different substrates for oxygen evolution reaction in alkaline medium*“, In: International Journal of Hydrogen Energy. 2014. Vol. 39, no. 10, s. 5204–5212. doi: 10.1016/j.ijhydene.2013.12.186.

[5] R. Phillips, C. Dunnill, „*Zero gap alkaline electrolysis cell design for renewable energy storage as hydrogen gas*“, In: RSC Advances. 2016. Vol. 6, no. 102, s. 100643–100651. doi: 10.1039/C6RA22242K.

[6] B. Sudarmanta, S. Darsopuspito, D. Sungkono, “*Application of dry cell HHO gas generator with pulse width modulation on Sinjai spark ignition engine performance*”, In: International Journal of Research in Engineering and Technology, Vol. 5, No. 2, pp. 105-112, 2016. ISSN: 2319-1163.

[7] V Szomosi, „*Možnosti využitia elektrolyzéra pre elektroenergetické účely*“, Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2023.

[8] V. Szomosi, D. Medveď, „*Možnosti využitia elektrolyzéra pre elektroenergetické účely*“, In: Electrical Engineering and Informatics 14: Proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. Košice: TUKE, 2023, s. 242 - 247. ISBN 978-80-553-4407-2.

Možnosti regulácie výkonu v ostrovnej prevádzke

Daniel ČORŇÁK, Dušan MEDVEĎ

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

daniel.cornak@student.tuke.sk, dusan.medved@tuke.sk

***Abstrakt* — Tento článok sa zaoberá možnosťami regulácie výkonu v ostrovnej prevádzke pre energetické účely. Jedná sa riešenie vybraných javov v elektrickej sieti. Pre tento účel bol navrhnutý model elektrickej siete, v ktorej sa bude vyrábať a regulovať výkon. Na konkrétnom modeli sa zrealizujú rôzne experimentálne simulácie za účelom získania výsledkov o zmenách v sieti. Tento článok je rozdelený do niekoľkých častí, ktoré sa zaoberajú návrhom a modelovaním konkrétnej siete a meraním jednotlivých zmien.**

***Kľúčové slová* — regulácia výkonu, ostrovná prevádzka, distribučná sústava, ATPDraw, napätie, vyššie harmonické, výkon**

I. Úvod

V posledných rokoch je ľudstvo čoraz viac závislejšie od decentralizovaných a udržateľných zdrojov energie. Preto sa tematika, zaoberajúca sa ostrovnou prevádzkou, stáva stredobodom pozornosti v oblasti elektroenergetiky. S narastajúcim dopytom po elektrine sa hľadajú nové alternatívne zdroje energie, ktoré dokážu efektívne využiť a tiež aj skladovať energiu. Preto sa aj vďaka technologickému pokroku viaceré podniky venujú problematike ostrovnej prevádzky. Výhoda ostrovnej prevádzky spočíva v ovládaní, stabilite a regulácii elektrickej energie. Nie všetko je však ideálne, a tak aj ostrovná prevádzka má niekoľko nevýhod, ktoré môžu spôsobovať nestabilitu siete.

II. Ostrovná prevádzka

Systém, kde zdroj na výrobu elektriny je priamo pripojený do izolovaného odberného elektrického zariadenia na konkrétnom odbernom mieste sa nazýva ostrovná prevádzka. Tento spôsob výroby elektrickej energie spočíva v tom, že dané odberné elektrické zariadenie je napájané vlastným zdrojom elektrickej energie. Dôležité je podotknúť, že v ostrovnej prevádzke musí byť dané odberné elektrické zariadenie technicky, ale aj elektricky oddelené od elektrickej prípojky a od zvyšku časti distribučnej sústavy. Prostredníctvom tohto bodu sa ostrovná prevádzka delí na 2 hlavné typy:

1. ***Riadna*** (trvalá) ***ostrovná prevádzka*** – táto prevádzka spočíva v tom, že odberné elektrické zariadenie je napájané vlastným zdrojom elektrickej energie a je natrvalo technicky, ale aj elektricky oddelené od elektrickej prípojky a od zvyšku časti distribučnej sústavy.
2. ***Núdzová ostrovná prevádzka*** – táto prevádzka naopak spočíva v tom, že odberné elektrické zariadenie je síce napájané vlastným zdrojom elektrickej energie, ale je dočasne technicky a elektricky oddelené od elektrickej prípojky a zvyšnej časti distribučnej sústavy pri bezprúdovom stave v distribučnej sústave.

Vzhľadom na to, že ostrovná prevádzka má niekoľko dobrých vlastností, ako sú spoľahlivosť, stabilita a využitie obnoviteľných zdrojov energie, tak stále sú tu aj riziká, ktoré treba zohľadniť, pretože môžu mať za následok nestabilitu alebo až úplný výpadok siete [6].

III. Možnosti regulácie výkonu

Ostrovná prevádzka má aj tú vlastnosť, že má niekoľko spôsobov alebo možností ako sa dá regulovať výkon. Každá z týchto možností má svoje kladné, ale aj záporné stránky:

1. ***Regulácia pomocou obnoviteľných zdrojov energie*** – OZE sa síce považujú za čistú (zelenú) energiu, ale aj napriek tomu môžu mať veľký vplyv na stabilitu a teda aj bezpečnosť siete. Existuje niekoľko typov obnoviteľných zdrojov energie, ktoré slúžia

na výrobu elektrickej energie, ale aj tepla. Najpopulárnejšie sú fotovoltické panely a veterné elektrárne. Hlavnou nevýhodou pri ich integrácii do siete je ich premenlivosť, teda náchylnosť na rôzne typy počasia. Preto je potrebná predikcia počasia, aby bolo riadenie OZE čo najefektívnejšie.

- *Fotovoltické systémy* – tieto systémy sú vybavené riadiacimi prvkami, ktorých hlavnou úlohou je riadiť a optimalizovať výrobu elektrickej energie. Striedače (invertory) v týchto systémoch sú nastavené tak, aby upravovali vyrábaný výkon podľa aktuálnej spotreby.
- *Veterné elektrárne* – v týchto elektrárňach sa reguluje výkon pomocou lopatiek turbíny (natáčanie lopatiek) alebo pomocou riadenia rýchlosti otáčok turbíny. Najmodernejšie systémy týchto elektrární dokážu predpovedať zmeny v poveternostných podmienkach a na základe toho prispôsobiť výrobu elektrickej energie [5].

2. ***Energetické úložiská*** – hlavnými prvkami energetických úložísk sú batérie. Vďaka ich okamžitej odozve pri zmenách dopytu a ponuke elektrickej energie hrajú hlavnú úlohu v ostrovných prevádzkach. Najpoužívanejšími typmi týchto úložísk sú chemické batérie:
   - *Li-ion batérie* (lítiovo-iónové batérie) – tieto batérie sú najpoužívanejším typom energetických úložísk. Ich hlavnou úlohou je vyrovnávať krátkodobé výkyvy vo výrobe a spotrebe elektrickej energie; vďaka ich schopnosti rýchleho nabitia a vybitia.
   - *Flow batérie* (prietočné batérie) – tento typ batérií sa používa na dlhodobé aplikácie, pretože sa vyznačujú ich dlhou životnosťou [10].

3. ***Riadenie dopytu*** – ďalšou dôležitou vlastnosťou pre reguláciu výkonu ostrovnej prevádzky je riadenie dopytu. Hlavnou vlastnosťou riadenia dopytu je, že prispôsobuje spotrebu elektrickej energie podľa aktuálnych podmienok v sieti. To sa dosiahne pomocou inteligentných sietí (inteligentné domácnosti) a moderných meracích prístrojov. Riadenie sa tak delí do 2 základných skupín:
   - *Inteligentné siete* (smart grids) – v týchto sieťach prebieha obojsmerná komunikácia medzi výrobcom a spotrebiteľom elektrickej energie. Aj vďaka inteligentným meracím prístrojom, ktoré poskytujú údaje v reálnom čase, zlepšuje sa riadenie spotreby.
   - *Dynamické cenové tarify* – v súčasnej dobe sa táto metóda používa na zníženie spotreby. To sa dosiahne tým, že sa počas dennej špičky zaťaženia umelo navýšia ceny, a tým spotrebitelia znižujú svoju spotrebu. To vedie k úsporám elektrickej energie a taktiež to napomáha k vyrovnávaniu zaťaženia [4].

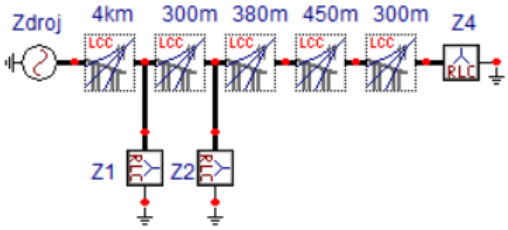
4. ***Hybridné systémy*** – sú to systémy, ktoré sú kombinované rôznymi zdrojmi a úložiskami elektrickej energie. Týmto spôsobom sa zvyšuje stabilita a spoľahlivosť v ostrovnej prevádzke. Rôznorodou kombináciou sa hybridné systémy delia na:
   - *Dieselové generátory a OZE* – kombináciou dieselových generátorov a OZE sa vytvorí spoľahlivý záložný zdroj elektrickej energie počas výpadku alebo nízkej výroby elektriny z OZE.
   - *Integrované energetické systémy* (IES) – hlavnou úlohou IES je zvýšiť efektivitu a spoľahlivosť vďaka kombinácii rôznych technológií. Využívajú pri tom rôzne moderné systémy pre skladovanie elektrickej energie a optimalizáciu [6].

Aj napriek rôznorodosti regulácie výkonu v ostrovnej prevádzke je tu niekoľko nedostatkov. Medzi hlavné nedostatky sa zaraďuje spolupráca medzi jednotlivými technológiami a systémami, ďalej medzi rôznymi zdrojmi a obnoviteľnými zdrojmi energie a iné. Ďalším problémom je variabilita obnoviteľných zdrojov energie, teda ich nestabilita v sieti. Tento problém sa však dá vyriešiť lepšími predikciami alebo systémami s rýchlymi odozvami na zmeny vo výrobe elektrickej energie.

IV. Modelovanie v programe ATPDraw

ATPDraw je grafický výpočtový program na výpočet ustálených ako aj prechodných dejov v elektrickej sieti. V tomto programe môže používateľ namodelovať rôzne elektrické obvody, kde si môže vybrať zo širokej ponuky elektrických komponentov. V tomto programe sa dá taktiež pomocou funkcie PlotXY vykresliť graf, z ktorého sa následne môžu odčítať údaje pre potrebné výpočty a porovnania výsledkov. V tomto prípade sa bude pracovať v najnovšej verzii programu,

kde sa namodeluje sieť (ostrovná prevádzka), v ktorej sa pomocou prifázovaného motora a k nemu striedača, budú sledovať zmeny v ostrovnej prevádzke pri rôznych podmienkach [2].



Obr. 1 Namodelovaná zjednodušená schéma siete

Na Obr. 1 je vyobrazená zjednodušená schéma namodelovaná v programe ATPDraw. Samotnú sieť je možné vyskladať z viacerých častí, ktoré sú napríklad motor SM59, namodelovaný striedač, vedenia, záťaže a meracie prístroje. Schéma na Obr. 1 je ilustratívna a slúži len ako príklad riešenej problematiky. V tejto sieti je trojfázový zdroj pripojený k vedeniu o rôznej dĺžke, na ktoré sú pripojené záťaže. Tieto záťaže reprezentujú jednotlivých spotrebiteľov/odberateľov elektrickej energie. Po namodelovaní siete sa spustí simulácia, ktorá pomocou funkcie PlotXY vykreslí priebehy napätia a vyšších harmonických.

A. Motor SM59

Pri tomto konkrétnom prípade sa uvažuje s použitím motora typu SM59. Je to typ synchrónneho stroja s možnosťou riadenia pomocou TACS. Iné ideálne trojfázové zdroje by vytvorili tzv. „tvrdú sieť“. Pri takejto sieti by nebola možná správna regulácia výkonu siete a tak by boli výsledky veľmi nepresné. Pri tomto elektrickom točivom stroji je možné nastaviť rôzne elektrické veličiny ako napríklad frekvenciu stroja, zdanlivý výkon, napätie, rôzne reaktancie, budiaci prúd statora, počet pólov stroja a mnoho ďalších [8].

B. Vedenie

Vedenie je možné namodelovať rôznymi spôsobmi a na výber je z viacerých druhov vedení. Pri modelovaní vedenia sa dá uvažovať ako s modelmi vedení so sústredenými, tak aj s rozloženými parametrami. Pri výbere modelu je na výber z hlavných 3 modelov:

a) Overhead Line – vonkajšie vzdušné vedenie
b) Single Core Cable – jednožilový kábel
c) Enclosing Pipe – viacžilový kábel

V tomto prípade sa zvolí ako vedenie jednožilový alebo zemný kábel, kde za postupne zadajú jednotlivé parametre, ako napríklad frekvencia, merná rezistivita zeme a dĺžka vedenia. Následne sa v záložke Data zadávajú rozmery vedenia a materiálové konštanty vodičov a izolantov [7].

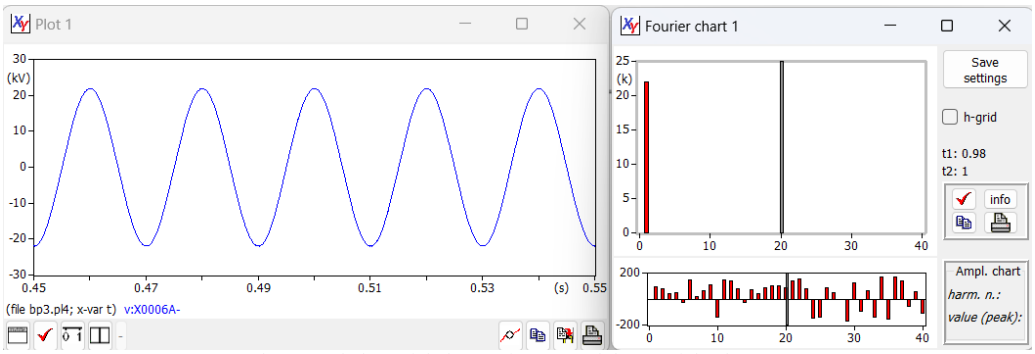
C. Switch (prepínač)

Do siete sa zakomponuje aj prepínač, vďaka ktorému je v sieti možné simulovať rôzne výpadky alebo poruchy. Týmto spôsobom sa daný zdroj alebo motor v určitom čase vypne a po vopred definovanom intervale opäť zapne. Takto sa napríklad odmerajú rôzne hranice stability ostrovnej prevádzky a tým sa určia výhody a nevýhody tejto prevádzky pri rôznych výpadkoch. V tomto príspevku bol použitý časovo kontrolovaný prepínač (Switch time controlled). Pre tento prípad je tento prepínač ideálny, pretože má dva režimy (trojfázový a jednofázový) [9].

D. Striedač

Striedač (alebo invertor) je elektrické zariadenie, ktoré premieňa jednosmerný prúd z jednosmerného zdroja na striedavý prúd. Používa sa hlavne pri pripájaní fotovoltických panelov na sieť, ktoré napájajú rôzne bežné spotrebiče. V tomto prípade, kde sa jedná o ostrovnú prevádzku, bude napájaná elektrickým motorom spolu s prepínačom, ktorý bude slúžiť ako fotovoltický panel. Daný striedač sa musí navrhnúť pomocou vopred stanoveného zapojenia a modelov a následne namodelovať v programe ATPDraw [3].

Pomocou týchto elektrických komponentov sa vytvorí zapojenie, ktoré bude charakterizovať ostrovnú prevádzku elektrickej siete, v ktorej budú prebiehať simulácie. Výsledky z týchto simulácií budú tvoriť priebehy/grafy napätia a vyšších harmonických:



Obr. 2 Výsledný priebeh napätia a vyšších harmonických

V. MERANIE A VYHODNOTENIE KVALITATÍVNYCH UKAZOVATEĽOV ELEKTRICKEJ ENERGIE

V prostredí ATPDraw sa pomocou simulácií vyhodnocuje niekoľko ukazovateľov riešenej problematiky. V tomto prípade sa budeme zameriavať na výpadky výkonu v ostrovnej prevádzke, ako napríklad zatienenie fotovoltického panelu alebo opačný prípad a to pripojenie viacerých panelov. Daným spôsobom sa odsledujú a porovnajú jednotlivé režimy, a to pred výpadkom a po výpadku. Výpadky v sieti sa dosiahnu vloženým vypínacím prvkom (switch), ktorý sa časovo nastaví na rôzne časové intervaly, ktoré budú simulovať zatienenie fotovoltického panelu. To spôsobí zníženie výkonu dodávaného do siete, na ktoré by mala ostrovná prevádzka zareagovať reguláciou iného zdroja elektrickej energie alebo priamo vykompenzovaním u spotrebiteľa.

Ďalej je možné sledovať napríklad vplyv vyšších harmonických, ktoré rozdeľujeme na harmonické napätia a prúdu. Sú to napätia a prúdy vyššej frekvencie, teda sú celočíselným násobkom základnej harmonickej (2×50, 3×50 až 50×50 Hz) okrem základnej harmonickej, ktorá je na Slovensku 50 Hz. Podľa normy STN EN 50160 majú vyššie harmonické danú limitnú hodnotu, ktorá sa nesmie prekročiť [1].

IEC 1000-2-2 Verträglichkeitspegel für öffentliche NS Netze IEC 1000-2-4 Verträglichkeitspegel für industrielle Netze				IEC 1000-2-2 Verträglichkeitspegel für öffentliche NS Netze IEC 1000-2-4 Verträglichkeitspegel für industrielle Netze			
n	Klasse1	Klasse2	Klasse3	n	Klasse1	Klasse2	Klasse3
1				26	0,20	0,20	1,00
2	2,00	2,00	3,00	27	0,20	0,20	1,00
3	5,00	5,00	6,00	28	0,20	0,20	1,00
4	1,00	1,00	1,50	**29**	**0,63**	**0,63**	**3,08**
5	**3,00**	**6,00**	**8,00**	30	0,20	0,20	1,00
6	0,50	0,50	1,00	**31**	**0,60**	**0,60**	**2,98**
7	**3,00**	**5,00**	**7,00**	32	0,20	0,20	1,00
8	0,50	0,50	1,00	33	0,20	0,20	1,00
9	1,50	1,50	2,50	34	0,20	0,20	1,00
10	0,50	0,50	1,00	**35**	**0,56**	**0,56**	**2,80**
11	**3,00**	**3,50**	**5,00**	36	0,20	0,20	1,00
12	0,20	0,20	1,00	**37**	**0,54**	**0,54**	**2,73**
13	**3,00**	**3,00**	**4,50**	38	0,20	0,20	1,00
14	0,20	0,20	1,00	39	0,20	0,20	1,00
15	0,30	0,30	2,00	40	0,20	0,20	1,00
16	0,20	0,20	1,00	**41**	**0,50**	**0,50**	**2,59**
17	**2,00**	**2,00**	**4,00**	42	0,20	0,20	1,00
18	0,20	0,20	1,00	**43**	**0,49**	**0,49**	**2,53**
19	**1,50**	**1,50**	**4,00**	44	0,20	0,20	1,00
20	0,20	0,20	1,00	45	0,20	0,20	1,00
21	0,20	0,20	1,75	46	0,20	0,20	1,00
22	0,20	0,20	1,00	**47**	**0,46**	**0,46**	**2,42**
23	**1,50**	**1,50**	**3,50**	48	0,20	0,20	1,00
24	0,20	0,20	1,00	**49**	**0,45**	**0,45**	**2,37**
25	**1,50**	**1,50**	**3,50**	50	0,20	0,20	1,00

Obr. 3 Tabuľka s povolenými hodnotami vyšších harmonických [1]

Výsledky simulácie z programu ATPDraw sa porovnajú pomocou tejto tabuľky na Obr. 3, kde sú určené hodnoty pre jednotlivé triedy. Trieda 1 je pre citlivé prostredia (nemocnice), trieda 2 je pre základné siete a taktiež pre verejné energetické siete. Napokon trieda 3 je pre priemyselné siete s vysoko-výkonnými frekvenčnými meničmi. Tieto hodnoty sa po odčítaní z grafu musia prepočítať na percentá pomocou vzorca na parameter THD_U a THD_I (celkové harmonické skreslenie napätia a prúdu) [1]:

$$THD = \frac{\sqrt{\sum_{k=2} C_k^2}}{C_1}$$

kde: C_1 amplitúda základnej harmonickej zložky
C_k amplitúda k-tej harmonickej zložky

Pre THD_U, norma STN EN 50160 uvádza maximálnu dovolenú hodnotu 8 %, avšak nesmie byť prekročená žiadna harmonická zložka napätia. Pri THD_I však norma neuvádza žiadnu maximálnu dovolenú hodnotu. Pre správne zistenie vyšších harmonických je potrebná dôkladná analýza tejto elektrickej siete [1].

VI. ZÁVER

Cieľom tohto príspevku bolo predstaviť možnosti regulácie výkonu v ostrovnej prevádzke na vybranom modeli elektrickej siete v prostredí ATPDraw. Podľa návrhu sa namodeluje sieť v ostrovnej prevádzke, pomocou ktorej sa overia nedostatky a problémy ostrovnej prevádzky pri výpadkoch, ako napríklad výkyvy vo výkone alebo zmeny a dodržanie vyšších harmonických.

POĎAKOVANIE

Tento príspevok vznikol vďaka Agentúre na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576 a APVV-21-0312, ako aj Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporená grantom VEGA 1/0627/24.

REFERENCIE

[1] P. Gardian, „*Kompenzácia vyšších harmonických v elektrických sieťach*“ [online] [cit. 2024-05-11]. Dostupné na internete: <http://www.power-grid.eu/prednaskaharmonickeBA08.pdf>.
[2] H. K. Hoidalen, „*ATPDraw*“, [online]. [cit. 2024-05-18]. Dostupné na internete: <https://www.atpdraw.net/>.
[3] Buderus, „*Čo je invertor? – slovník pojmov*“ [online]. [cit. 2024-05-12]. Dostupné na internete: <https://www.buderus.com/sk/sk/informacie/slovnik-pojmov/invertor.html>.
[4] S. Česaná, „*Riadenie dopytu*“, [online]. 2024. [cit. 2024-05-14]. Dostupné na internete: <https://www.ekonomicka.sk/riadenie-dopytu/ >.
[5] S. Štroffeková, „*Obnoviteľné zdroje energie*“, [online]. [cit. 2024-05-19]. Dostupné na internete: <https://www.enviroportal.sk/indicator/detail?id=2601>.
[6] P. Šimon, „*Ostrov, hybrid a distribučky*“, [online]. 2016. [cit. 2024-05-12]. Dostupné na internete: <https://www.energiaweb.energy/2016/01/17/ostrov-hybrid-a-distribucky-ii/>.
[7] ATPDraw 7.0 Help, „*The graphical user interface, Lines/Cables*“, [online]. [cit. 2024-05-01]. Dostupné na internete: <https://www.atpdraw.net/help7/lines_cables.html>.
[8] ATPDraw 7.0 Help, „*The graphical user interface, Machines*“, [online]. [cit. 2024-05-01]. Dostupné na internete: <https://www.atpdraw.net/help7/machines.html>.
[9] ATPDraw 7.0 Help, „*The graphical user interface, Swithces*“, [online]. [cit. 2024-05-01]. Dostupné na internete: <https://www.atpdraw.net/help7/switches.html>.
[10] P. Šperka, I. Vonkomer, „*Zdroje špičkovej energie a energetické úložiská*“, [online]. [cit. 2024-05-04]. Dostupné na internete: <https://www.atpjournal.sk/budovy/rubriky/nove-trendy/zdroje-spickovej-energie-aenergetickeuloziska.html?page_id=13880>.

Digitalizácia analógových meračov

[1]Jozef KOMPAN, [2]Anton BALÁŽ

[1]Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky Technická univerzita košice, Slovenská republika

[1]jozef.kompan@student.tuke.sk, [2]anton.balaz@tuke.sk

Abstrakt **– Tento článok sa zameriava na proces digitalizácie analógového vodomeru. V úvode detailne analyzujeme súčasné možnosti v tejto oblasti a navrhujeme architektúru pre hardvérové a softvérové riešenie. Základom hardvéru je Raspberry Pi Zero s POE adaptérom a kamerou. Vyvíjame púzdro, ktoré je navrhnuté priamo na vybraný hardvér a vytlačené pomocou 3D tlačiarne. Následným naprogramovaním a ladením docielime požadovanú funkcionalitu, teda prevedenie analógovej hodnoty na počítadle vodomera do digitálnej podoby. Tento problém sme riešili za pomoci Document Intelligence poskytovaným Microsoft Azure a knižnice PyTesseract. S cieľom zabezpečiť jednoduchší prístup k dátam a overenie ich presnosti, navrhujeme a implementujeme cloudovú aplikáciu v prostredí Microsoft Power Apps. Záver článku sa venuje analýze funkčnosti riešenia a jeho možným slabinám. Taktiež popisujeme možné rozšírenia pre budúce využitie z komerčného ale aj individuálneho.**

Kľúčové slová **– Databáza, Fusion360, Linux, Neurónová sieť, Power Apps, Raspberry**

I. Úvod

Ľudská práca sa stále viac nahradzuje automatizovanými procesmi, ktoré sú schopné efektívne využívať údaje a zabezpečiť ich správne vyhodnotenie. Tento trend sa nevyhnutne dotýka aj oblasti vodohospodárstva, kde sa stretávame s rastúcou potrebou monitorovať a riadiť spotrebu vody. Analógové meracie prístroje pre odpisovanie spotreby vody postupne ustupujú pred novými digitálnymi technológiami, ktoré prinášajú výhody nielen v rýchlosti a efektivite, ale aj v spoľahlivosti a presnosti meraní. Zároveň sa sústredíme na vývoj inteligentných systémov, ktoré dokážu nielen monitorovať spotrebu vody, ale aj včas identifikovať prípadné úniky a poruchy vodovodných systémov. V tomto článku sa budeme venovať práve tejto problematike, pričom sa zameriame na vývoj zariadenia pre automatickú kontrolu stavu vodomera. Cieľom je vytvoriť produkt, ktorý umožní užívateľom jednoducho sledovať ich spotrebu vody s možnosťou emailových notifikácií. Pozrieme na proces digitalizácie údajov z analógových zariadení a na výber vhodných hardvérových komponentov s dôrazom na ich dostupnosť a jednoduchú manipuláciu. Celkovým cieľom je vytvoriť produkt, ktorý bude nielen efektívny a spoľahlivý, ale aj cenovo dostupný a jednoducho ovládateľný pre užívateľov rôznych vekových skupín. V konečnom dôsledku veríme, že takýto produkt prinesie významné zlepšenie v kontrole a manažmente spotreby vody, čo je kľúčové pre udržateľné využívanie tohto dôležitého zdroja.

II. Návrh

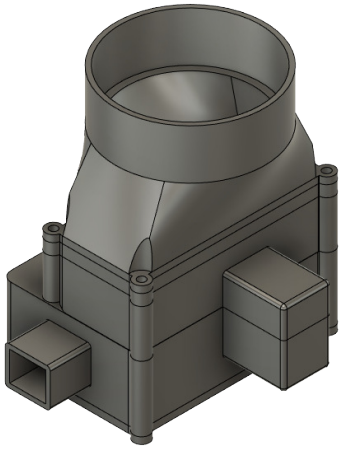
A. Výber hardvérovej a softvérovej výbavy

Digitalizácia sa stáva stále dôležitejšia pre využívanie moderných technologických nástrojov, ktoré umožňujú automatizáciu procesov a efektívne spravovanie dát. Jedným z takýchto nástrojov je Microsoft Power Platform, ktorá zahŕňa viacero aplikácií a služieb určených na tvorbu obchodných aplikácií a automatizáciu procesov. V tomto článku sa pozrieme na dve z týchto aplikácií - Microsoft Power Apps a Microsoft Power Automate. Microsoft Power Apps je webová platforma integrovaná priamo v Azure ekosystéme, ktorá umožňuje vývoj obchodných aplikácií a správu veľkého množstva dát. Jednou z jej hlavných výhod je jednoduchá rozšíriteľnosť a integrácia s rôznymi ďalšími procesmi v rámci Azure. Ďalšou výhodou je možnosť využiť ju aj ako zobrazovací systém pre dáta. Na druhej strane, Microsoft Power Automate je platforma určená na automatizáciu procesov, nazývaných Flow-y, ktoré spolupracujú s aplikáciami v rámci Power Platform. Ponúka širokú škálu spúšťačov a umožňuje reagovať na rôzne udalosti a dopyty. V rámci nášho projektu sme vybrali Linux Debian s nadstavbou pre Raspberry Pi, konkrétne Raspberry Pi Zero, ako vhodnú platformu pre naše účely. Tento výber bol motivovaný nielen jeho

výkonom a možnosťami pripojenia, ale aj dostupnosťou a kompatibilitou s rôznymi modulmi. Ďalším krokom bolo zvolenie externých modulov a nástrojov, ktoré budeme používať. Napájanie sme zvolili pomocou POE adaptéra, ktorý zároveň zaručuje aj internet. Vytvorenie fotografií je zabezpečované kamerovým modulom k Raspberry Pi. Pre spracovávanie fotografie sme použili PyTesseract OCR, teda knižnicu podporujúcu optické rozpoznávanie znakov. SQLite je jednoduchá a rýchla relačná databáza vhodná pre naše účely. Okrem toho sme sa rozhodli využiť aj cloudové služby, konkrétne Azure AI services a Microsoft Dataverse, pre lepšiu správu dát a možnosť integrácie s ďalšími aplikáciami v rámci ekosystému Microsoft. Spomínané nástroje a platformy nám umožnia vytvoriť kvalitný základ pre vytvorenie zariadenia na automatizovanú kontrolu a správu spotreby vody. Ich kombinácia nám poskytne potrebné nástroje na získanie, spracovanie a analýzu dát, čím prinesieme významné zlepšenie monitoringu [1] [2] [3] [4] [5] [6] [7] [8].

B. Modelovanie, tlačenie púzdra a montáž zariadenia

Uchytenie zariadenia k vodomeru je vyriešené pomocou manžety. Vzhľadom na to, že rôzne vodomery od rôznych výrobcov sa líšia krytom, táto manžeta nie je unifikovaná pre všetky, teda je potrebné ju prispôsobiť na konkrétny vodomer nachádzajúci sa v danej domácnosti. Hlavnou požiadavkou je presné priliehanie krytu vodomera, preto je potrebné zmerať jeho priemer a výšku uchytenia. Kónický tvar manžety je zvolený pre lepšie rozloženie svetla a jednoduchšiu tlač bez nutnosti tlače a odstraňovania podpôr. Na ukotvenie manžety k zvyšku zariadenia sme navrhli štyri montážne 'uši, ktorými sú prevlečené závitové tyče. Pri výrobe sme tiež zohľadnili kameru a prísvit, pre ktoré sme vytvorili dieru, aby nedochádzalo k obmedzeniu svetla. Vyosenie diery pre prísvit minimalizuje odrazy a zabezpečuje čo najkvalitnejšiu viditeľnosť pre vytvorenie fotografie. Manžeta je vyrobená z TPU filamentu vhodného pre jeho ohybnosť a schopnosť dobre znášať zmeny teploty. Zariadenie je uložené v kryte, ktorý je následne primontovaný na manžetu. Kryt pre zariadenie je navrhnutý tak, aby spĺňal potrebu vodotesnosti a unifikovanosti, umožňujúcej použitie s rôznymi manžetami. Potreba úpravy krytu môže nastať pri zmene kamery s odlišnou ohniskovou vzdialenosťou, čo vyžaduje prispôsobenie vzdialenosti od vodomera a vyosenie LED diódy. Kryt sa skladá z dvoch častí: pre PoE adaptér a pre Raspberry Pi Zero s kamerou a LED prísvitom. Toto delenie umožňuje jednoduchšiu výmenu pri použití rôznych PoE s inými rozmermi. Pre Raspberry Pi využívame externý model, čo zjednodušuje návrh montážnych stĺpikov. Kvôli úspore miesta vylučujeme pájkované piny, ktoré by zväčšili profil. Flex kábel medzi kamerou a Raspberry Pi vyžaduje pridanie montážnych stĺpikov pre kameru na kryte. Rozmery kamery čerpáme z datasheetu výrobcu. Vložky s metrickým závitom vsádzame do stĺpikov spájkovačkou a upevňujeme skrutkami. Pre LED diódu vŕtame otvor s priemerom 5 mm. Pri vŕtaní do PETG dbáme na prehrievanie materiálu. Kryt pre PoE je najvyšší segment zariadenia kvôli prívodu internetu a napájaniu cez RJ45 kábel, ktorý odpojujeme ako prvý. Tunel okolo otvoru pre napájanie zabezpečuje vodotesnosť. Pre spojenie krytov používame „uši“ s otvormi na závitovú tyč a drážky pre silikón alebo lepidlo. Tlačíme z PETG filamentu pri teplote 235°C, čo zaručuje dostatočnú pevnosť bez praskania. Po vytlačení a očistení krytov skladáme zariadenie v troch fázach. Najprv testujeme funkčnosť pred montážou, najmä spájkovanie LED na GPIO17, aby sme predišli problémom. Spoj testujeme pomocou python skriptu, pričom pohybujeme prívodnými drôtmi. Potom manuálne zaostrujeme kameru, pomaly otáčame objektívom a používame meter a text na odhad vzdialenosti. Montujeme kameru, LED, PoE adaptér a Raspberry Pi s SD kartou, predinštalovaným Raspbianom, povoleným VNC a statickou IP. Komponenty spájame pomocou 3mm závitových tyčí a matíc. Montáž zariadenia ukončujeme nasadením na vodomer [9] [10].

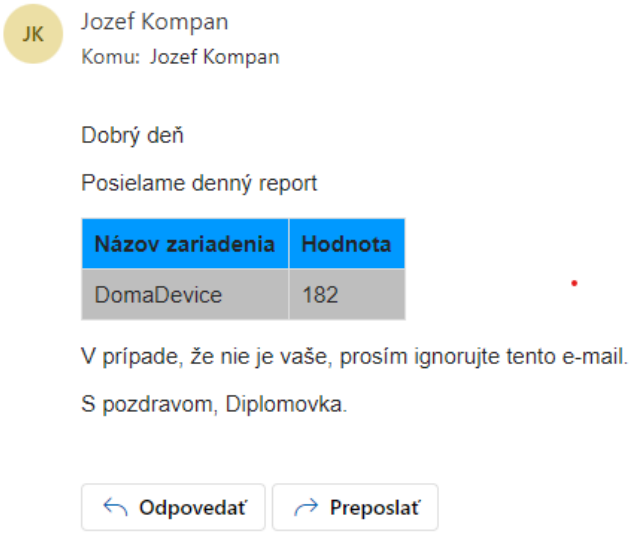


Obr. 1 Kompletovaný model v programe Autodesk Fusion360

C. Vytvorenie cloudovej aplikácie

V rámci prípravy a implementácie cloudových častí diplomovej práce pracujeme s Microsoft Azure ekosystémom, konkrétne s Power Automate, Microsoft Dataverse a Power Apps. Po prihlásení sa do Azure vyhľadáme Azure AI services, kde vyberáme Document Intelligence. Po kliknutí na tlačidlo 'nový' vyplníme formulár, kde zadáme 'Subscription' (používame študentskú licenciu), vyberieme 'Resource group' s názvom 'Vodomer' a nastavíme región na 'West EU'. Po zadaní mena a výbere 'Pricing tier' (zadarmo pre študentov) aplikáciu vytvoríme. Pre back-end prístup skopírujeme prístupové kľúče a endpoint. Pre návrh a vytvorenie databázy otvoríme PowerApps platformu, kde v navigačnom menu zvolíme 'Tables'. V existujúcich tabuľkách využívame 'Contact' a vytvárame tabuľku 'Device' so stĺpcami 'DeviceName', 'Address', 'ID device' a referenciou na 'Contact' v 'Relationships'. Túto reláciu nastavujeme na 'LookUp'. Využívame aj predgenerovaný stĺpec 'Created On'. Taktiež vytvárame tabuľku 'DataExtracted' pre ukladanie dát so stĺpcami 'ValueText', 'ValueNumber', 'IDDevice', 'extractionTime' a referenciou na 'Device', tiež typu 'LookUp'. Všetky stĺpce nastavujeme ako 'business required'. Pre vytvorenie Power Apps aplikácie v navigačnom menu klikneme na 'Create' a vyberieme 'Blank app based on Dataverse'. Po zadaní mena a popisu upravíme zdroje, z ktorých aplikácia čerpá dáta, a vyberieme tabuľky popísané vyššie. Upravením 'view' používateľa zabezpečíme prehľad jemu priradených zariadení, ktoré môže pridávať, odoberať, aktivovať a deaktivovať. V prehľade zobrazujeme hodnoty 'ValueNumber', 'CreatedOn', 'Device' a 'extractionTime' z tabuľky 'DataExtracted'. Aplikácia funguje s flowmi v záložke 'Flows', kde používame tri typy spúšťačov: časový pre plánované odosielanie notifikačných emailov, HTTP request received, ktorý sa spustí po prijatí HTTP požiadavky a kontroluje hlavičku requestu, a Dataverse aktivované flowy, ktoré sa spúšťajú pri zmene v Dataverse tabuľke. Pre ukladanie dát, párovanie a posielanie notifikácií vytvárame flow RegisterDevice, ktorý sa spustí po prijatí HTTP requestu s preddefinovaným JSON formátom dát. Tento flow vyberie používateľa a zariadenie z tabuliek 'Contacts' a 'Devices' podľa emailu a ID, čím zamedzí duplicite zariadení. Ak zariadenie neexistuje, vloží ho do tabuľky s informáciami z requestu. Flow Pridanie zariadenia Notification reaguje na pridanie nového riadku v tabuľke 'Devices', vyberie používateľa z 'Contacts', inicializuje premennú s HTML tabuľkou a odosiela email s informáciami o zariadení. Flow DataExtractedNew pracuje na princípe HTTP requestu s očakávaným JSON pre vyčítané dáta, zistí párovanie lookup stĺpca a pridá nový riadok do tabuľky 'DataExtracted' bez kontroly existencie, nakoľko odosielanie je časované na 2-hodinové intervaly. Flow Odosielanie Aktuálnych údajov je časovo riadený a vykonáva sa raz denne o 18:00. Načíta všetky kontakty z tabuľky 'Contacts', inicializuje premennú pre email, cyklicky pridáva do premennej HTML tagy s údajmi o zariadeniach a odosiela denné informácie používateľom. Na konci flowu premennú vymaže. Flow Zistenie GUID spúšťame v testovacom režime, vyberie všetky záznamy z 'Contacts', 'Devices' a 'DataExtracted' a použije 'Compose' pre zistenie názvu 'temporary' stĺpca na lookup stĺpcoch.

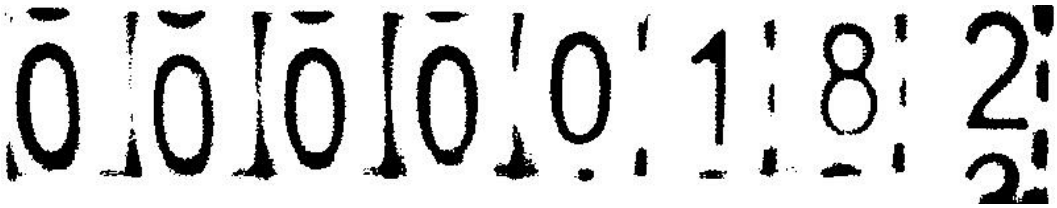
Po vytvorení všetkých flowov je prostredie pripravené pre reakcie na požiadavky.



Obr. 2 Emailová notifikácia o stave zariadení

D. Vytvorenie a príprava fotografie

Na správne fungovanie zariadenia programujeme ovládacie BASH skripty a programy v jazyku Python s využitím rôznych knižníc. Skripty sú krátke a obsluhujú špecifické funkcie, čo z nich robí jednotlivé komponenty. Využívame knižnice na ovládanie podprocesov, spracovanie obrázkov, systémový dátum a JSON na čítanie konfiguračného súboru. Najprv vytvárame fotografiu pomocou kamery a kvôli svetelným podmienkam zapíname a vypíname prísvit, čo zaručuje stabilné podmienky pre OCR a preprocessing fotky. Nástroj raspistill ukladá dáta na disk ako fotografiu, preto potrebujeme program, ktorý tieto dáta uloží do databázy nespracované. Následne program 'cropper' oreže a otočí obrázok, pripraví ho pre OCR. Potom vytvárame skript pre tesseract, ktorý zmenší a binarizuje obrázok, čím zlepšuje prácu tesseractu s textom o veľkosti 30 pixelov a kvalite aspoň 600 DPI. Tento skript vyčíta dáta z fotky, validuje ich a uloží do databázy. Ďalší skript odosiela dáta na Azure OCR, vyberá najlepšieho kandidáta, keďže Azure OCR nie je zadarmo. Používame pôvodne orezanú fotografiu optimalizovanú pre Azure OCR, ktoré funguje optimálne pri 600 DPI a veľkosti 800 x 600. Po vyčítaní a uložení dát do tabuliek odošleme tieto dáta na PowerApps aplikáciu, ktorá poskytuje jednoduchý prehľad a posiela notifikácie. Všetky potrebné nastavenia pre skripty a programy sú uložené v konfiguračnom súbore. Dáta sú odoslané na PowerApps aplikáciu pre prehľad a notifikácie [11].



Obr. 3 Binarizovaná fotografia pre pyTesseract skript

Aj napriek preprocesingu sa môže stať, že OCR vyčíta čísice 1 v mieste, kde sa nachádzajú medzery medzi číslicami číselníku. Čo najlepšie eliminovanie tejto chybovosti riešime druhým orezom, ktorý rozdelí jednotlivé číslice do samostatných obrázkov, ktoré sú následne vyčítávané po jednom. Nakoľko počet číslic sa môže líšiť, je to parameter programu uložený v konfiguračnom súbore. Výhodou tohoto riešenia je, že v prípade zle prečítanej číslice dokážeme vziať konkrétnu číslicu z predošlého vyčítavania.



Obr. 4 Vyrezané konkrétne číslo

E. Vyčítanie dát, validácie a odoslanie na cloud

Skript pre vyčítanie dát z obrázkov, validáciu údajov a ich uloženie do databázy má parametre načítané z config súboru, ako názov tesseract tabuľky, azure tabuľky, názov databázy, názov obrázka a maximálny počet litrov, ktorý slúži ako základný validačný krok. Po vyčítaní údajov z predspracovaného obrázka pomocou pytesseract, ktorý ignoruje nečíselné hodnoty, prebieha validácia údajov. Ak rozdiel voči posledne odoslanej hodnote prekročí maximálny počet litrov, nasleduje druhá validácia, kde skript prečíta číslice jednotlivo, čo je časovo náročné, preto sa spúšťa iba pri zlyhaní prvej validácie. Údaje sa potom uložia do tesseract tabuľky, pričom skript vyčíta maximálne ID z tabuľky photos a vloží dáta s časom extrahovania, photoid, číselnými hodnotami, reťazcom a validačným príznakom. Pre Azure OCR skript vyberá kandidáta na odoslanie podľa validovaných údajov tesseractu. Po pripojení na databázu vyberá fotku z poslednej polhodiny, ktorá nebola odoslaná na Azure, pomocou LEFT JOIN a filtračnej podmienky. Vybrané dáta predspracuje cropper na optimálnu veľkosť pre Azure OCR. Dáta odošle na Azure a čaká na ich vyčítanie, ktoré skompletizuje do požadovaného reťazca. Skript porovná hodnoty s tesseractom, a ak sú Azure dáta menšie, uložia sa do dataAzure pričom je potrebná dĺžka dát rovná počtu číslic na číselníku. Na záver skript vymaže dočasný súbor. Skript na odosielanie dát do PowerApps aplikácie definuje základné nastavenia a pripraví posledné dáta z dataAzure tabuľky do JSON formátu pre HTTP request. Tento skript posiela extrahované dáta, ktoré následne spracuje cloudová aplikácia [6] [12].

F. Ostatné skripty a programy

Pre mazanie obrázkov z databázy vytvárame skript, ktorý záznamy staršie ako určitý počet dní maže, aby sa zabránilo preplneniu SD karty. Tento skript je naplánovaný v crontab súbore na 1:00 ráno. Pre správne sekvenčné spúšťanie programov vytvárame 'localScript.sh' na lokálne spracovanie a 'azScript.sh' na odosielanie dát na Azure OCR a PowerApps. Tieto skripty obsahujú funkciu 'cd' pre správne umiestnenie súborov a sú automaticky spúšťané cez crontab. Registrácia zariadenia na PowerApps zahŕňa vytvorenie používateľa a registráciu zariadenia pomocou 'testCreateUserPowerapps.py' a 'registerDevice.py', ktoré posielajú HTTP requesty na PowerApps. Úspešné priradenie zariadenia je potvrdené e-mailom s informáciami o zariadení.

G. Vyhodnotenie riešenia

Zameriame sa na silné a slabé stránky zariadenia a riešenie problémov počas jeho výroby. Hlavným cieľom pri modelovaní krytu bola vodotesnosť, čo viedlo k integrácii drážok pre silikón vo všetkých spojoch, spôsobujúcich zúženie štruktúry. Tento problém sme riešili ich zosilnením, najmä v oblasti tunela pre kábel spájajúci Raspberry s POE adaptérom. Pri tlači prototypu sme zistili, že flex kábel kamery nie je možné správne ohnúť, preto sme rozšírili vnútorný priestor. V priebehu tlače sa vyskytol prehyb kvôli zvýšenej teplote tlačovej hlavy, čo sme riešili oddelením časti s POE adaptérom a Raspberry, čím sme eliminovali veľké premostenia a znížili spotrebu filamentu. Pri implementácii Power Apps sme narazili na problémy kvôli pokroku systému, najmä v rozhraní, čo spôsobilo zdĺhavejšie vytváranie aplikácie. Pri Power Automate sme riešili problémy s odosielaním údajov kvôli štruktúre databázy a LookUp stĺpcom, čo sme vyriešili pomocným flowom. Rozdielne formáty dátumov v Microsoft Dataverse a zariadeniach sme riešili nastavením rovnakého ISO formátu. Problémom bolo pripojenie zariadenia k internetu cez WiFi, čo sme riešili priamym pripojením cez USB a úpravou skriptu. WiFi slúžilo ako

záložné pripojenie v prípade preseknutia kábla medzi smerovačom a POE adaptérom. Najväčším problémom bolo upravenie obrázka pre Tesseract. Riešili sme to binarizáciou s thresholdom 25% a nastavením OCR enginu na LSTM sieť, čím sme čiastočne eliminovali problém s medzerami medzi číslicami. Validáciu sme riešili tak, že prvú hodnotu sme nevalidovali a osoba musela skontrolovať inicializačné hodnoty, prípadne spustit priložený program pre ich manuálny zápis do databázy. Pri ukladaní do databázy sme zapisovali menšie číslo, ak Azure vrátilo menšie, aby sme riešili chybne vyčítané medzery. Databázu nadmerne zapĺňali veľké fotografie, čo viedlo k zníženiu výkonu zariadenia a spotrebovaniu úložného priestoru do dvoch mesiacov. Preto sme uchovávali údaje za posledných 7 dní a automatický skript staršie dáta ma, pričom toto nastavenie je možné upravovať v konfiguračnom súbore [13] [7].

H. Záver

Cieľom bolo navrhnúť, vytvoriť a naprogramovať zariadenie na digitalizáciu analógových meracích prístrojov pomocou neurónových sietí. Analyzovali sme existujúce riešenia a procesy, manuálne aj automatizované. Zariadenie splnilo požiadavky na modulárnosť a vodotesnosť. Eliminovali sme vplyv rôznych svetelných podmienok použitím LED prísvitu a zakrytovaním. Problém s neúplným otočením cifry sme čiastočne riešili pomocou Azure neurónovej siete a čítaním jednotlivých číslic. Aplikácia v Power Apps zabezpečila prístup k dátam cez internet. Rozšírenie funkcionalít zahŕňa detekciu únikov vody a integráciu s inteligentnou domácnosťou. Výsledky práce naznačujú potenciál pre efektívne riadenie a monitorovanie zdrojov, čo je v dnešnej dobe žiadané pre súkromné aj komerčné využitie.

Poďakovanie

Článok bol podporený v rámci projektu - 015TUKE-4/2024 Moderné metódy a formy vzdelávania v oblasti kyberbezpečnosti.

Literatúra

[1] “What is Raspberry Pi?” https://www.spiceworks.com/tech/networking/articles/what-is-raspberry-pi/, accessed: 2024-04-19.
[2] *Overview of Microsoft Power Apps*, https://learn.microsoft.com/sk-sk/power-apps/powerapps-overview, 2024, accessed: 2024-04-19.
[3] Vectra. (2024) Learning about Microsoft Power Automate. https://www.vectra.ai/learning/power-automate. Accessed: 2024-04-19.
[4] A. AuthorLastName, “Handwritten digit recognition using cnn,” *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 11, no. 1, January 2023, an ISO 9001:2008 Certified Journal. [Online]. Available: https://www.ijircce.com
[5] IBM. (2024) Understanding Convolutional Neural Networks. https://www.ibm.com/topics/convolutional-neural-networks. Accessed: 2024-04-19.
[6] *AI Document Intelligence on Microsoft Azure*, https://azure.microsoft.com/en-us/products/ai-services/ai-document-intelligence#overview, 2024, accessed: 2024-04-19.
[7] “Pytesseract python-tesseract is an optical character recognition (ocr) tool for python,” https://pypi.org/project/pytesseract/, accessed: 2024-04-19.
[8] “Introduction to microsoft dataverse,” https://learn.microsoft.com/en-us/power-apps/maker/data-platform/data-platform-intro, accessed: 2024-04-19.
[9] “All about petg 3d printing filament: Materials, properties, definition,” https://www.xometry.com/resources/3d-printing/petg-3d-printing-filament/, accessed: 2024-04-19.
[10] “Camera Module 2 Mechanical Drawing,” https://datasheets.raspberrypi.com/camera/camera-module-2-mechanical-drawing.pdf, accessed: 2024-04-18.
[11] “Style guide for writing bash scripts,” https://lug.fh-swf.de/vim/vim-bash/StyleGuideShell.en.pdf, accessed: 2024-04-19.
[12] “5 advantages and disadvantages of mysql,” https://www.hitechwhizz.com/2022/10/5-advantages-and-disadvantages-limitations-benefits-of-mysql1.html, accessed: 2024-04-19.
[13] “Introduction to Long Short-Term Memory (LSTM),” https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/, 2021, accessed: 2024-04-19.

Systémy IDS cez proaktívne vyhľadávanie zraniteľnosti v sieťovom prostredí

[1]Tomáš SIVÁK, [2]Anton BALÁŽ

Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky Technická Univerzita v Košiciach, Slovenská Republika

[1]tomas.sivak@student.tuke.sk, [2]anton.balaz@tuke.sk

Abstrakt **– Cieľom tejto práce je znižovanie falošne pozitivných výstrah vznikajúcich pri detekcií všetkých výstrah systémami detekcie narušenia. Predstavujeme návrh, architektúru a implementáciu riešenia. Práca pozostáva z opisu týchto systémov a ich fungovania. Na základe získaných vedomostí zvolíme príslušný typ a metódu detekcie systému Snort. Prvotná obrana vzniká v podobe skeneru zraniteľnosti Nessus. Celková implementácia prebieha v podobe korelačného algoritmu. Pre získanie údajov sme po dosiahnutí znalostí vytvárali útoky a falošne pozitívne výstrahy. Tie sa stali súčasťou prvého súboru údajov. Tento súbor je získaný systémom Wireshark pre funkciu filtrovania výstrah Snort. Pre otestovanie riešenia sme použili súbor DARPA2000 LLDOS1.0. Obidve súbory slúžia ako zdroj údajov v našej implementácií pomocou programovacieho jazyka Python. Výsledkom implementácie sú vizualizácie a získané hodnoty, ktoré prešli analýzou. Analyzované počty preukazujú konkrétne hodnoty redukcií, ktoré je možné vidieť v podobe grafov na porovnávanie.**

Kľúčové slová **– Snort, falošné pozitívne výstrahy, paket, korelácia, zraniteľnosť**

I. Úvod

Každá doba so sebou priniesla a aj prinesie výrazný technologický pokrok, ktorý si dokázal získať ľudí a to ovplyvnilo ich každodenný život. Niet pochýb, že existuje len málokto bez určitej skúsenosti so slovom internet. Ale aj napriek pozitívam, ktoré so sebou priniesla, vznikli nástrahy vytvárajúce dôsledky, ktoré môžu mať dopad práve na ľudí. Každým dňom pribúda počet útokov rôzneho typu, ktoré sa dokážu vyhnúť ich odstráneniu a zabráneniu. Kvôli tomu sa vyvíjajú systémy schopné vyriešiť takéto každodenné záležitosti. Jedným z takýchto systémov sú systémy detekcie narušenia.

Systémy, ktoré sú adekvátnou metódou ochrany siete pred rôznymi nebezpečnými útokmi. Ich využiteľnosť je podstatná. Ale aj voči ich neustálemu vývoju prichádza chyba, ktorú dokážu zneužiť ľudia produkujúci problémy kvôli ktorým tieto systémy vznikajú. Takýmto problémom sú aj falošne pozitívne alarmy. Odkazujú na útoky, ktoré si privlastňujú nebezpečenstvo, ale v skutočnosti žiadne nevzniklo. Takto musia zistiť, kde a či vôbec daný útok prebehol. To ich stojí čas a aj peniaze, kedže aj pri nesprávnej klasifikácií môže dôjsť k výrazným finančným škodám.

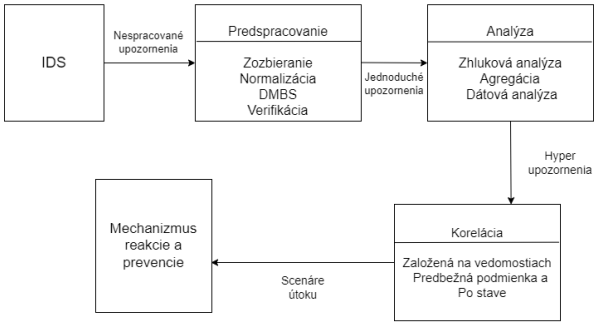
V dnešnom svete dokážu tieto výstrahy pomáhať práve útočníkom, ktorí spolu s nimi vytvárajú aj reálne útoky. Cieľom práce je podeliť sa o nadobudnuté vedomosti, poznatky a riešenie, ktorým je práve detekcia a zníženie falošne pozitivných výstrah. Práve vedomosti sú podkladom pre vytvorenie samotného riešenia. Navrhuté riešenie by pomohlo bezpečnostným administrátorom vyriešiť ich každodenný problém.

II. Návrh riešenia

A. Korelácia medzi výstrahami

Návrh sa zameriava na zníženie false positives upozornení, ktoré sú často generované viacerými IDS založené na podpisoch bez maninpulácie s predvolebnou sadou [1] podpisov. Sľubnou metódou na automatizáciu analýzy výstrah detekcie narušenia je nájdenie korelácie medzi výstrahami a takýto systém na obr. 1 je známy ako Alert Correlation System (ACS). Jedna z hlavných aplikácií korelácie výstrah (AC) je diagnóza útoku. Aj keď sa proces korelácie často prezentuje ako jeden krok, v skutočnosti analýzu vykonáva niekoľko komponentov, z ktorých každý má špecifický cieľ.

Cieľom korelácie [1] by bolo prezentovať informácie generované IDS zmysluplným spôsobom, t.j spôsob, ktorý môže bezpečnostný analytik ľahko pochopiť a spracovať a ktorý mu pomôže odhaliť útoky a incidenty. Toto zobrazenie môže odhaliť scenáre útoku a zámery útočníka.



Obr. 1 Vystražný korelačný systém(ACS)

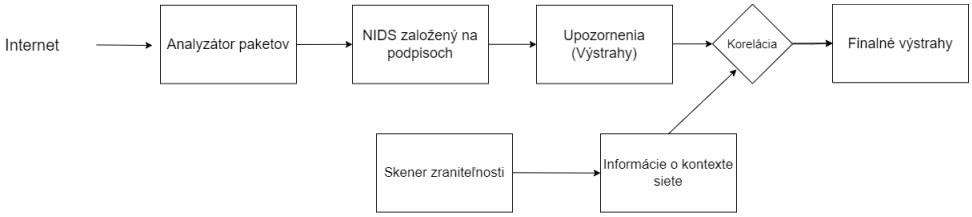
Tento proces pozostáva zo zbierky komponentov, ktoré transformujú senzorové signály detekcie narušenia na správy o vniknutí. Keďže upozornenia sa môžu vzťahovať [2] na rôzne druhy útokov na rôznych úrovniach granularity, proces korelácie nemôže zaobchádzať so všetkými upozorneniami rovnako. Namiesto toho je potrebné poskytnúť súbor komponentov, ktoré sa zameriavajú na rôzne aspekty celkovej korelačnej úlohy.

Rámec ACS pozostáva z niekoľkých operácií. Podľa obr. 1 sa vo fáze predbežného spracovania [1] zhromažďujú nespracované výstrahy z jedného alebo mnohých IDS a prevedú sa do štandardného formátu (alebo známeho ako normalizácia), ktorý sa nazýva Intrusion Detection Message Exchange Format (IDMEF). Hlavným cieľom fázy analýzy je zvyčajne zbaviť sa väčšiny falošných upozornení, nie nezabudnúť zlúčiť nadbytočné výstrahy. Zatiaľ čo výskumník zoskupuje výstrahy na základe rovnakých hlavných príčin, budúce zaťaženie výstrah by sa mohlo znížiť odstránením najviac prevládajúcich a pretrvávajúcich základných príčin.

V prípade korelácie je [1] bežným prístupom definovanie pravidiel alebo znalostí na určenie vzťahu medzi dvoma výstrahami. Overovanie sa vykonáva pomocou pasívnych aj aktívnych techník a každý prístup si vyžaduje [3] inú podporu zo strany narušiteľa infraštruktúry na detekciu narušenia. Rozdiel medzi aktívnymi a pasívnymi mechanizmami je založený výlučne na časovom bode, kedy sa [4] konfigurácia alebo údaje zhromažďujú. Jeden z mechanizmov aktívneho overovania so vzdialeným prístupom je založený na použití skenerov zraniteľností. Keď sa zistí útok, skener sa môže použiť [4] na kontrolu na zraniteľnosť, ktorú sa tento útok pokúša zneužiť. Každý incident je pripútaný k iným [5] incidentom sieťou kombinácií spojenia a disjunkcie a generuje možnú sieť útokov.

B. Zostavenie systému

Využitá bude metóda aktívného overovanie výstrah, ktorá je integrovaná ako modul Snort a využíva nástroj Nessus skener zraniteľností, ktorý poskytuje IDS kontextové informácie o monitorovanej sieti. Pomocou tejto konfigurácie demonštrujeme nástroj Snort, ktorý bude upravený tak, aby využíval informácie poskytované populárnym skenerom zraniteľností Nessus, čím by sa výrazne mala zlepšiť presnosť detekcie Snortu a zníženie falošných pozitívných poplachov. Na realizovanie tohto návrhu na obr. 2 sa použijú nástroje medzi ktoré patria Snort a Nessus. Je to systém detekcie narušenia siete a skener zraniteľnosti siete, ďalej programovací jazyk Python bude použitý na implementáciu a vyhodnotenie úlohy.



Obr. 2 Návrh systému

IDS Snort je vybraný na základe hodnotenia a porovnania s inými open source nástrojmi, ktorými sú Suricata alebo Zeek. Snort má jednoduché nasadenie a bohatú sadu pravidiel, ktoré sú

aktualizované často komunitou. Nessus bol vybraný ako overovací nástroj kvôli vysokej kvalite jeho kontrol zraniteľností, minimálnemu vplyvu na jednoduchosti, s akou by sa dal integrovať do systému Snort. Pre každé upozornenie, ktoré je spracované overovacím vláknom, sa extrahuje zodpovedajúci identifikátor Common Vulnerabilities and Exposures [CVE] a použije sa ako index do systému Nessus. Výsledok každého overenia je uložený vo vyrovnávacej pamäti na určitý čas, aby sa znížilo zaťaženie siete. Filtrovanie ako výsledok údajov budú naprogramované v programovaciom jazyku Pythone. Zvolený bol pretože poskytuje bohatú sadu knižníc potrebných pre čítanie a analýzu údajov súvisiacich so sieťou. Celková schéma zostavenia takéhoto systému v podstate koreluje zraniteľnosti siete s identifikátormi alarmov a nakoniec filtruje falošne pozitívne upozornenia.

C. Zápis v podobe algoritmov

Znížiť počet nadbytočných výstrah a falošne pozitívnych poplachov, odhaliť skutočné jednokrokové alebo viackrokové útoky a vykonávať [6] včasné varovanie, je naliehavo potrebné používať algoritmy na analýzu a spracúvať výstrahy. Ide o zlepšenie IDS, ktorý môže získať presnejšie pravidlá porovnávania a zameriavať sa na analýzu a filtrovanie údajov o výstrahách [6] s cieľom oddeliť správne a nesprávne výstrahy. Aby sa dosiahol tento cieľ využije sa spomínaná korelácia pomocou algoritmu založenom na korelačnej analýze. Zistíme charakteristiky prostredníctvom analýzy výstrah IDS a rozdielom medzi nimi odstránime falošné poplachy.

Podľa štúdie sa v určitom časovom [6] intervale počet susedných súvisiacich výstrah falošných poplachov vygenerovalo oveľa menej ako polovica skutočných poplachov. Štatistiky ukazujú, že pravdivé pozitívne výstrahy často prinášajú rozsiahle výstrahy v priebehu 0 - 1 sekundy a falošne pozitívne výstrahy v rozsahu 1 až 5 sekúnd, čo je oveľa menej [6] ako pravé pozitívne výstrahy. Prostredníctvom rôznej hustoty výstrah možno tiež rozlíšiť pravé a falošné pozitívne výstrahy. Pomocou charakteristík pravdivých a falošne pozitívnych hlásení môžeme preveriť výstrahy generované systémami na detekciu narušenia, čím sa zníži počet zbytočných poplachov a zvýši sa účinnosť korelačnej analýzy.

Po prvom kroku zoskupíme podobné výstrahy výpočtom funkcie podobnosti každého alarmu podľa zdrojovej IP a cieľovej IP, zdrojového a cieľového portu a potom ďalej znížime počet alarmov. IDS vyvolá upozornenia [1] nezávisle (pre každý útok), hoci medzi nimi môžu existovať logické súvislosti. Tieto upozornenia sa spracujú a analyzujú pomocou spomínaného systému Alert Correlation System (ACS). Viaceré dvojice pravidiel môžu byť vysoko korelované. Vysoká korelácia týchto pravidiel sa môže použiť na kombináciu výstrah, ktoré predstavujú tú istú udalosť. S týmto poznatkom sa počet generovaných výstrah na jednu udalosť [7] môže znížiť zlúčením dvoch výstrah a tak vznikne jedna super výstraha, ktorá obsahuje obe pravidlá. Ak je výstraha vygenerovaná z rovnakého zdroja do rovnakého cieľa vo vopred definovanom časovom intervale.

Práve tento krok sa týka hľadania vzťahov medzi výstrahami generovanými jedným alebo viacerými zdrojmi údajov a spájanie týchto informácií s ďalšími znalosťami. Celkovo to poskytuje stručnejší a vysokoúrovňový pohľad na pokusy o narušenie. Cieľom korelácie by v skutočnosti malo byť prezentovať informácie generované IDS zmysluplným spôsobom, t. j. spôsobom, ktorý by bol ľahko pochopiteľný a spracovateľný ľudským analytikom a pomáhal mu odhaliť útoky a incidenty. Je dôležité poznamenať, že cieľom úloh korelácie je pomôcť čo najviac ľudským operátorom, ale nie ich nahradiť.

III. Overenie

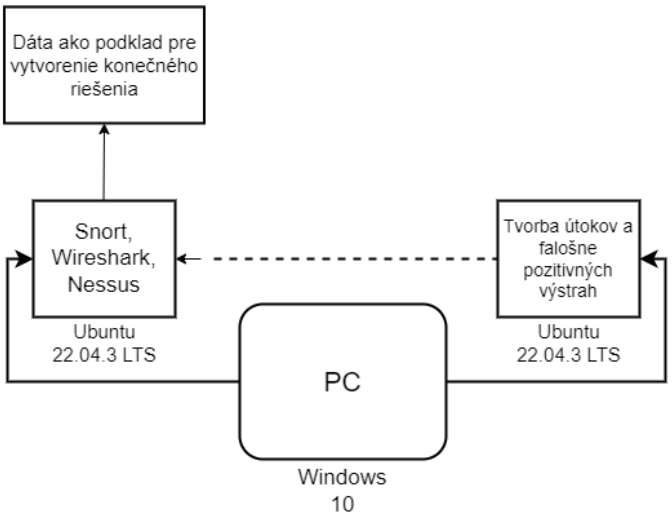
Predovšetkým rozoberáme hĺbkovú analýzu každého kroku a jeho značný vplyv na konečný výstup s niektorými potrebnými argumentmi založenými na overených tvrdeniach a metodológií. Uvádzame tiež rôzne algoritmy pre každý hlavný nezávislý modul. Súčasná implementácia nášho nástroja na overovanie výstrah bola vyhodnotená na experimentálnom testovacom zariadení alebo experimentálnom testovaní s ohľadom na jeho účinnosť pri znižovaní miery falošných výstrah systému Snort.

Na obr. 3 môžte vidieť, že na jednom termináli v systéme Ubuntu boli zavedené rôzne známe zraniteľnosti a na senzorovom Ubuntu boli povolené príslušné signatúry na detekciu útokov využívajúcich tieto zraniteľnosti v podobe systému Snort. Daľej bol spustený program Wireshark spolu so skenerom zraniteľnosti Nessus. Na cieľ sa potom spustila široká škála útokov. Hoci Snort poskytuje vstavané súbory pravidiel na detekciu známych útokov zároveň tento detekčný systém nám umožnil napísať vlastné pravidlá. Na tento účel sme použili rôzne pravidlá. Použili sme pravidlo na detekciu icmpevent a tiež sme napísali jednoduché pravidlo ping na testovanie Snort. Potom sme definovali ďalšie pravidlá na zhromažďovanie škodlivých údajov alebo údajov

o vniknutí jedním z nich je aj generovanie útoku DoS. Zo siete je možné zhromažďovať a označovať údaje. Údaje obsahujú tzv. 'normálnu' prevádzku spolu s niektorými útokmi.

V našich experimentoch sa útoky vykonávali najmä pomocou skrípt na zásadne zraniteľnosti. Na získanie výsledkov sa testy opakujú deväťkrát. Sada testov pre tento experiment sa nazýva testy Snort 1 so systémom Ubuntu a sadou pravidiel, ktoré boli vytvorené a bude obsahovať nasledujúce položky:

- Vytvorené útoky
- Vytvorené falošné pozitívne útoky (výstrahy)
- Nessus: Súbor údajov, ktorý bude obsahovať údaje zo siete



Obr. 3 Proces fungovania architektúry

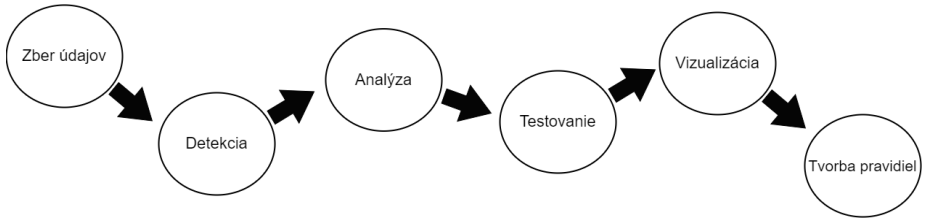
A. Architektúra systému

Každá udalosť pri príchode v IDS do centra zabezpečenia sa analyzuje, aby sa zistilo, či súvisí s konkrétnou zraniteľnosťou a či je zraniteľnosť prítomná v cieľovom systéme a na cieľovom porte. Toto je veľmi presné na detekciu útokov, pretože súbor údajov zo systému Nessus môže odhaliť vopred možné zraniteľnosti ktoré môže utočník využiť. Snort nasadený v predvolenom režime počas tejto práce prijímal všetky pakety, ktoré prešli firewallom ako aj všetky odchádzajúce požiadavky. Súbory pravidiel, ktoré boli použité, sú dostupné open-source sady pravidiel pre vznikajúce hrozby a vytvorené sady pravidiel.

Útočná prevádzka bola generovaná po vytvorení a spustení útokov. Skripty pre falošne pozitívne útoky sú naprogramované v jazyku Python konkrétne pomocou Scapy. V krátkosti je Scapy výkonný interaktívny program a knižnica na manipuláciu s paketmi v jazyku Python a dokáže falšovať alebo dekódovať pakety veľkého počtu protokolov, posielať ich, zachytávať ich, ukladať alebo čítať a mnoho ďalšieho. Údaje boli zozbierané. V určitých hodinách boli vygenerované útoky, aby boli k dispozícii pravdivé údaje. Všetka ostatná pozorovaná prevádzka je zahrnutá počas tohto procesu a považuje sa za pravidelnú.

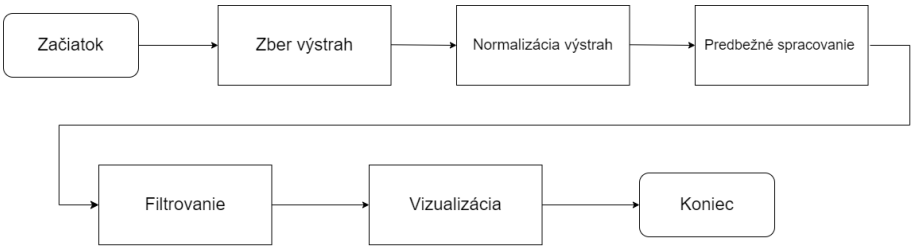
Predovšetkým sa zohľadňujú všetky alarmy týkajúce sa práve jedného hostiteľa počas časového okna, t. j. vybrané korelačné parametre sú čas, cieľová a zdrojová IP adresa, cieľový a zdrojový port a protokol. Tie nám poskytuje systém Wireshark. Používajú sa veľmi presné údaje a údaje o zraniteľnostiach v reálnom čase. Podotýkame, že korelácia nie je závislá od údajov o zraniteľnostiach z programu Nessus. Konštatujeme, že v tomto prípade je korelácia výstrah s kontextovými informáciami získanými prostredníctvom skenovania siete a zraniteľností veľmi obmedzená. V špecifických prípadoch by táto metóda mohla byť veľmi užitočná, ale vo všeobecnosti tento druh korelácie nevedie k jasnej a presnej indikácii či je upozornenie pravdivé alebo nepravdivé.

Súbor údajov, ktorý sa používa na tento výskum, obsahuje tento základný súbor znakov: time, source, sport, destination, dport, protocol, length, info. Tieto vlastnosti poskytujú informácie o prevádzke, ktorá je zistená v sieti. Na základe výsledkov je možné použiť iba podmnožinu znakov, ktorá bola vybraná na ďalšie spracovanie. Súbor údajov bol zredukovaný vynechaním prvkov, ktoré neposkytovali dodatočné informácie o správaní snímača. Najmä je potrebné poskytnúť súbor komponentov, ktoré sa zameriavajú na rôzne aspekty celkovej úlohy korelácie. Šesťstupňová metodika projektu je zhrnutá na obr. 4 a zahŕňa zber údajov, detekciu, analýzu, testovanie, vizualizáciu a tvorbu pravidiel. Korelácia výstrah analyzuje výstrahy a snaží sa prepojiť rôzne



Obr. 4 Fázy celého procesu

výstrahy, aby sa vytvoril celkový obraz o útoku, a tým sa získal prehľad o stave bezpečnosti na vysokej úrovni. Navrhovaný model využíva korelačný algoritmus na obr. 5.



Obr. 5 Diagram korelačného algoritmu

Práve každý z týchto jednotlivých modulov má pozitívny vplyv na zníženie počtu falošne pozitívnych výstrah. Na obr. 5 je uvedené grafické znázornenie korelačného algoritmu, ktorý sme implementovali pomocou Pythonu. Prvé dve úlohy sa vykonávajú pre všetky výstrahy. V prebeiehajúcej fáze normalizačná zložka konvertuje každý výstražný záznam, ktorý je prijatý do štandardizovaného formátu, ktorému rozumia všetky korelačné komponenty. Ostatné zložky korelácie nášho rámca pracujú s jednotlivými alebo úzko súvisiacimi udalosťami. Spájajú výstrahy založené na sieti s výstrahami založenými na hostiteľoch, ktoré súvisia s tým istým útokom. A tak dávajú upozornenia do kontextu s ohľadom na konkrétny cieľ siete.

B. Overenie účinnosti riešenia

S cieľom overenia účinnosti systému riešenia sa v tomto dokumente vyberá štandardný súbor údajov z odvetvia DARPA ako experimentálne údaje. Používame DARPA 2000 LLDOS1.0, ktorý obsahuje viacstupňový útok. Na tento súbor je použité riešenie korelačného algoritmu.

Navrhovaný model sa postupne snaží minimalizovať počet spracovaných výstrah na každom komponente. Takto sme vykonali podrobnú analýzu prevádzky. Naše pravidlá korelácie pre túto úroveň sú založené na základe zdrojovej IP a časovej pečiatky. V prvotných fázach korelácie odstraňuje nerelevantné, nereálne a falošné výstrahy. Takto tento model dokáže predstavovať jednotlivé 'vzory' útokov z masívnych záznamov alarmov, čo pomáha znížiť tvorenie falošných pozitívnych výstrah. Na dosiahnutie konečného cieľa je práve vizualizácia výstrah, ktorá pomáha správcovi siete pochopiť celkový obraz stavov siete. Narušitelia väčšinou spúšťajú svoj útok v niekoľkých krokoch. Tradičné IDS nie sú schopné odhaliť tieto komplexné útoky série a riešia každý malý útok ako individuálny útok bez korelácie s inými útokmi. Výsledkom je, že IDS generuje obrovské množstvo výstrah a falošne pozitívnych hlásení čo sťažuje proces analýzy pre správcu siete.

C. Analýza riešenia

Keďže výstrahy nie sú významné, ak sú izolované, preto je dôležitou fázou hľadanie vzťahov medzi nimi. Zoskupovanie a zhlukovanie výstrah v podobe analýzy na základe podobnosti ich vlastností môže v skutočnosti odhaliť kroky útoku, ktoré útočníci spustili. Okrem toho možno ľahko odhaliť a zlúčiť nadbytočné výstrahy. Vrámci tohto výskumu sú zhromaždené dva súbory údajov, ktoré sa počas tohto výskumu použili. Aby bolo možné riešiť informácie, ktoré sú často prezentované na funkciách výstrah, je potrebné hľadať dodatočné informácie. Údaje o zraniteľnosti poskytujú podporné primerané základné znalosti o monitorovaných sieťach, na základe ktorých možno vypočítať ďalšie funkcie výstrah.

Pred vygenerovaním našej sieťovej premávky sme pre zabezpečenie údajov o zraniteľnostiach použili skener zraniteľností Nessus. Údaje o zraniteľnosti sa používajú pri overovaní platnosti

výstrah a zároveň na zabezpečenie potecialných novovzniknutých hrozieb, ktoré môžu byť odstránené nami zabezpečovacími meódami. Účelom je zhromaždiť informácie o súvisiacich útokov a predložiť jedinú inštanciu upozornenia, ktorá zhrnie všetky relevantné informácie pre analytika. Proces zberu a prebieh zachytávania paketov(výstrah), získavame pomocou systému Snort spolu so systémom Wireshark.

D. Výsledné hodnoty

Ako môžte vidieť na na obr. 6 v tabuľke sú zaznamenané celkové hodnoty pre obidve datasety. Práve pomocou systému Wireshark sme odfiltrovali pakety(výstrahy) a získali tie, ktoré úspešne detekoval systém Snort. Tento údaj prezentuje hodnota 41763 uvedená na obr. 6, ktorá sa rovná 48,015 podielu. V prípade datasetu Darpa2000 bol celkový počet paketov(výstrah) 649787. Tento údaj na obr. 6 bol overený systémom Snort a zároveň aj systémom Wireshark kvôli pravosti tohto údaju. Hodnoty pre odelenie paketov a získanie paketov systému Snort pomocou funkcie filtrovania systému Wireshark nie sú uvedené. Dôvodom nezískania týchto údajov spočíva v tom, že je to cudzí dataset a v tomto prípade nám systém Wireshark bránil získanie týchto hodnôt. Možným dôvodom je aj to, že premávka tohto datasetu prebiehala na cudzom zariadení.

Pre posun spomínaného spracovania cieľa sme obidve súbory údajov transformovali do csv súborov. Následne sa výstrahy spájajú na základe funkcií podobnosti v komponente spájania výstrah. Tie sa následne agregujú a potom klasifikujú práve pomocou podobnosti príznakov. Okrem toho sú využité techniky založené na znalostiach spolu s zhromažďovaním informácií na základe čoho sa vyhodnocujú s cieľom eliminovať falošne pozitívne výsledky.

Dataset	Počet paketov (výstrah)	Počet paketov (výstrah) zachytených systémom Snort	Percentuálne zastúpenie systémom Snort
Snort 1	86979	41763	48,015 %
Darpa2000	649787		

Obr. 6 Celkové hodnoty

Časová charakteristika sa zistuje, keď jeden výstražný signál sa porovnáva so susedným výstražným signálom z čoho získavame ich časovú odchýlku. Práve tieto vypočítané prvky výstrah sa používajú na výpočet vzdialeností medzi výstrahami pri určovaní podobnosti výstrah. Cieľom je kombinovať výstrahy, ktoré predstavujú detekciu toho istého útoku. Potom nastáva rozhodnutie o zlúčení dvoch výstrah, ktoré je založené na časovom údaji medzi týmito výstrahami a informácií, ktoré obsahujú. Na to sme sa rozhodli použiť jupyter notebook Python. Jupyter notebook obsahuje kolekciu vizualizačných nástrojov, podporuje analýzu údajov a niekoľko štandardných úloh dolovania údajov, výber príznakov a to všetko pomocou programovacieho jazyka Python.

Všetky údaje sú agregované v rámci daného časového okna. Adekvátnym prístupom pri vymazávaní údajov zo súboru údajov je hlavne potrebné postupovať veľmi opatrne, malo by ísť o posledný z postupov, ktoré sa vykonajú pri zistení nezrovnalostí, pretože nemôžeme riskovať stratu užitočných údajov z iných stĺpcov len tým, že tieto nulové hodnoty odstránime. Pri ďalšom kroku sme pristúpili k vyčisteniu údajov na účely kvalitnej analýzy. Chýbajúce hodnoty sa spracujú a zároveň sa tak údaje štandardizujú. Po nahradení alebo odstránení nulových, neexistujúcich údajov pokračujeme v overovaní existencie duplicitných riadkov. V oboch testovacích súborov existovalo množstvo hodnôt spojených s protokolmi, ktoré pre náš prípad bolo v najlepšom prípade ich odstránenie, keďže neboli považované práve za falošne pozitívne údaje. Analýza sa vzťahuje na kritický proces vykonávania počiatočného výskumu údajov s cieľom odhaliť vzory, otestovať metódy a overiť tak predpoklady pomocou súhrnných štatistík a grafických znázornení.

Pri analýze útokov v čase sa predbežne zameriame na najviac súbežných cieľových ip adries pre obidve súbory údajov. V súbore údajov Snort 1 je najviac krát zaregistrovaný cieľ s adresou 192.168.0.108 zatiaľ čo pre Darpa2000 to je 172.16.115.20. Pri týchto cieľových adresách je dôležité brať do úvahy aj ostatné, ktoré môžu byť cieľom práve falošne pozitívnych výstrah. Preto sa získavajú všetky spomenuté vlastnosti pre minimalizáciu a zároveň zvyšovanie pravdepodobnosti pre konkrétny typ cieľovej adresy.

To, čo sa deje s vybranou cieľovou ip adresou, v tomto prípade najviac zastúpenou, zistíme práve vytvorením grafického zobrazenia, kde každému útoku zodpovedá bod súvisiaci s cieľovým portom a protokolom. Spravíme to pomocou funkcie scatterplot(). Pre predpoklad vzorov je ešte

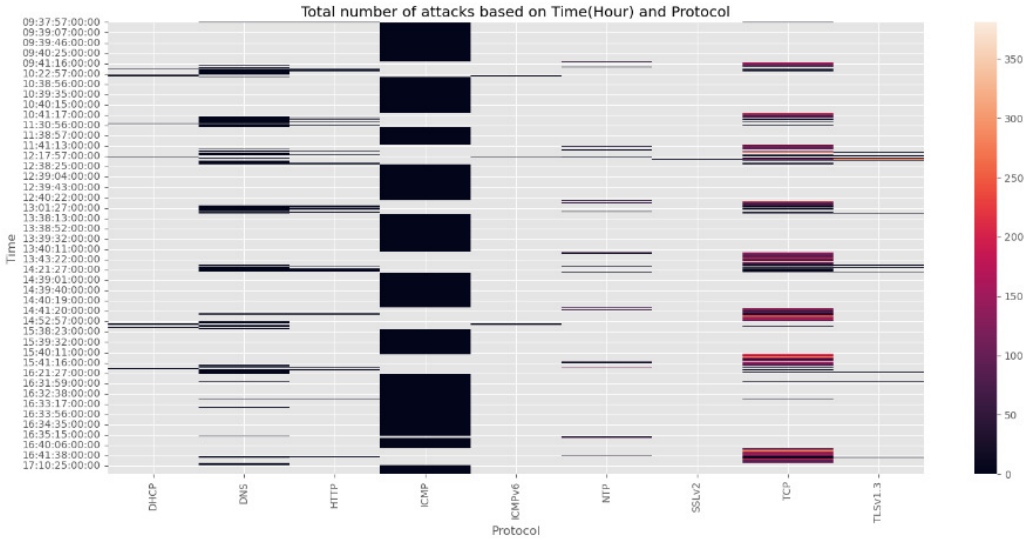
Destination	
192.168.0.108	40562
192.168.0.107	1156
255.255.255.255	21
ff02::16	14
ff02::1:ff97:1aa8	5

Obr. 7 Cieľové adresy Snort 1

Destination	
172.16.115.20	107808
131.84.1.31	73725
172.16.112.50	72937
172.16.112.100	35997
172.16.113.84	22749

Obr. 8 Cieľové adresy Darpa2000

potrebné identifikovať ich správanie v čase. Aby sme to ukázali, vykreslili sme mapu, ktorá súvisí s počtom útokov, protokolmi, cieľovými adresami a časom útoku.



Obr. 9 Odhaľujúce zistenie Snort 1

Je vidieť, že v tomto prípade trvali útoky vo väčšom počte v opakujúcom rozsahu, pričom s rastúcim počtom, útoky v menšom počte má trvanie udalostí klesajúcu tendenciu. V taktom prípade sme potrebovali a pristupovali k získaniu vzťahu medzi časom vykonania útokov a protokolov útoku a samotnej najviac zasiahnutej cieľovej ip adresy. To možno vykonať hľadaním údajov súvisiacich s časom, v ktorej boli útoky vykonané. Práve na základe časovej odchýlky, ktorú sme vytvorili v podobe nového stĺpca a ktorá nám tak prezrádza časový interval od posledného útoku.

Dataset	Vstupný počet výstrah	Výstupný počet výstrah	Celková redukcia (%)	FP redukcia (%)
Snort 1	41763	34889	16,45	5,272
Darpa2000	649787	542239	16,55	14,609

Obr. 10 Finálne výsledky

Takto sme implementáciou prechádzali datasety a kombinovali cieľové adresy s protokolmi a časovými odchýlkami aby sme dosiahli zistenie false positive výstrah. Jednotlivými korelačnými prístupmi sme odkrývali zistenia. Výpočty sa realizovali pre celkovú redukciu a falošne pozitívnu redukciu pomocou vzorcov.

Pri použití súboru údajov DARPA 2000 naša stratégia zaznamenala zníženie počtu výstrah o 16,55 %. Zároveň toto zníženie na obr. 10 obsahuje práve 14,609 % redukcie false positive výstrah. V súbore údajov Snort 1 naša stratégia zaznamenala zníženie počtu výstrah o 16,45 % z čoho 5,272 % predstavovalo práve všetky zistené false positive výstrahy. V oboch experimentoch navrhované riešenie výrazne znížilo objem výstrah a false positive výstrahy.

IV. Záver

Vo svete existuje mnoho podôb hrozieb. Útočníci využívajú každú možnú príležitosť, zraniteľnosť a sú schopní vytvárať ďaleko sofistikovanejšie metódy pre spôsobenie škôd. Jedným z nich sú aj falošne pozitívne výstrahy. Práve na nás je, aby sme tieto výstrahy v podobe útokov dokázali odhaliť a zredukovať. To bolo aj hlavným cieľom tejto práce zníženie počtu falošne pozitívnych výstrah. Pre získanie takéhoto cieľa bolo potrebné zvoliť kompletný prístup. Na základe poznatkov sme postupne formovali návrh riešenia a postupným získavaním informácií sme sa posúvali k nášmu cieľu. Počas toho sme potrebovali vyriešiť problémy ako získavať informácie o údajoch a ako z týchto informácií vytvoriť základ riešenia nášho cieľa. Preto sme si zvolili systém Wireshark a korelačný prístup v podobe nášho korelačného algoritmu.

Hlavným zdrojom boli pre nás súbory údajov. Jeden z nich sme si vytvorili pomocou útokov a skript na tvorbu falošne pozitívnych výstrah a druhý nám slúžil práve na otestovanie a overenie nášho riešenia. Korelačný algoritmus bol aplikovaný ako proces, podľa ktorého sme postupovali pri našej implementácií. Táto implementácia pozostávala zo série znalostí a komponentov, ktoré v nej boli použité ako postup pre uskutočnenie nášho cieľa. Zároveň sme dospeli k vizualizácií informácií a tak získavali konkrétne hodnoty, čo viedlo k úspešnému dosiahnutiu nášho hlavného cieľa. Jeho výsledky sme analyzovali a získali tak konkrétne počty jednotlivých redukcií.

Náš prístup dosiahol zlepšenie kvality súborov údajov a to znížením množstva výstrah navrhnutím lepšieho mechanizmu, ktorý získava podobnosť výstrah na základe definovaných podmienok. Okrem tohto cieľa sme dodatočne získali aj celkovú redukciu súboru údajov. Zároveň sa počas tohto výskumu zodpovedali aj výskumné otázky, ako zabezpečiť siete a aké nástroje je potrebné používať. Kombinácia výhod nástrojov Snort, Nessus, Wireshark a Python v tejto štúdii viedla k maximálnemu prínosu a zvýšeniu úrovne bezpečnosti sietí.

Poďakovanie

Článok bol podporený v rámci projektu - 015TUKE-4/2024 Moderné metódy a formy vzdelávania v oblasti kyberbezpečnosti.

Literatúra

[1] S. Z. M. H. Maheyzah Md Siraj, "Network instrustion alert corellation challenges and techinques," *Faculty of Computer Science*, vol. 4, no. 15, p. 24, 12 2008, Žiadne. [Online]. Available: https://core.ac.uk/download/pdf/11784911.pdf

[2] C. K. Fredik Valeur, Giovanni Vigna and R. a Kemmerer, "A comprehensive approach to intrusion detection alert correlation," *TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING*, vol. 4, no. 11, p. 23, 7 2004, Žiadne. [Online]. Available: https://escholarship.org/uc/item/3r33v62n

[3] ——, "A comprehensive approach to intrusion detection alert correlation," *TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING*, vol. 4, no. 11, p. 23, 7 2004, Žiadne. [Online]. Available: https://sites.cs.ucsb.edu/~chris/research/doc/tdsc04_correlation.pdf

[4] C. Kruegel and W. Robertson, "Using alert verification to identify successful intrusion attempts," *Reliable Software Group University of California, Santa Barbara*, vol. 4, no. 23, p. 15, 3 2004, Žiadne. [Online]. Available: https://sites.cs.ucsb.edu/~vigna/publications/2004_kruegel_robertson_vigna_PIK_AlertVerification.pdf

[5] S. A. a. R. J. Seyed Ali Mirheidari, "Alert correlation algorithms: A survey and taxonomy," *Springer Link*, vol. 4, no. 35, p. 15, 8 2020, Žiadne. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-03584-0_14

[6] L. S. L. Jianyi and Zhang, "Algorithm of reducing the false positives in ids based on correlation analysis," *IOP Conference Series: Materials Science and Engineering*, vol. 4, no. 12, p. 6, 6 2016, Žiadne. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/322/6/062016/pdf

[7] M. Simidžioski, *An Unsupervised Approach for False Alarm Filtering in Rule-based NIDS*, 3rd ed., ser. 10. The address: Technische Universiteit Delft, 6 2020, vol. 4, an optional note.

Procedurálne generovanie modelu prostredia pre Unreal engine

[1]Dominik KNIEŽA, [2]Branislav MADOŠ

[1,2]Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]dominik.knieza@student.tuke.sk, [2]branislav.mados@tuke.sk

Abstrakt **– Článok opisuje návrh a implementáciu programu pre procedurálne generovanie terénu pomocou vstupných vektorových dát. Vstupné dáta reprezentujú jednoduchý tvar terénu, podľa ktorého generátor vygeneruje detailný terén. Program disponuje grafickým používateľským rozhraním pre jednoduchšiu kontrolu generátora. V úvodnej časti je článok venovaný popisom niektorých druhov generátorov. V tejto časti sa nachádza aj analýza existujúcich programov pre generovanie terénu a porovnanie ich vlastností. V d'alšej časti je popísaný formát vstupných dát a detailný návrh generátora a použitých algoritmov pre generovanie. Článok je ukončený príkladom využitia generátora pri procedurálnom modelovaní prostredia.**

Kľúčové slová **– Terén, Procedurálne generovanie, Unreal engine**

I. Úvod

S rastúcou komplexnosťou počítačových hier rastú aj nároky na veľké, realistické a detailné herné prostredia. Jeden zo spôsobov ako vytvoriť takéto prostredie je použitie realistického terénu. Vytvoriť rozsiahly a realistický terén je ale časovo veľmi náročné a preto sa neustále hľadajú nové spôsoby ako tento proces urýchliť.

V dnešnej dobe už existuje množstvo nástrojov na zjednodušenie a zrýchlenie tvorby terénu ako sú napríklad rôzne generátory terénu. Generátory umožňujú vytvárať obrovské detailné a často realistické terény za veľmi krátky čas. Namiesto toho aby musel vývojár manuálne vytvoriť svoj vlastný terén, tak iba zadá do generátora očakávané vlastnosti terénu a generátor podľa nich automaticky takýto terén vytvorí. Nevýhodou procedurálnych generátorov je, že vo väčšine prípadov používateľ buď stráca kontrolu nad tým ako bude výsledný terén vyzerať, alebo je dodatočná úprava terénov príliš komplikovaná.

V práci sú analyzované výhody a nevýhody existujúcich programov pre procedurálne generovanie terénu. Podľa výsledkov analýzy bol navrhnutý nový program pre procedurálne generovanie terénu, ktorého cieľom je vyriešiť niektoré nevýhody ostatných procedurálnych generátorov.

II. Generátory terénu

Najjednoduchší spôsob, ako vytvoriť terén je využiť generátor terénu. Generátory terénu sú efektívne nástroje pre tvorbu rozsiahlych terénov.

A. Procedurálne generátory

Tieto generátory vytvárajú terén na základe matematických pravidiel, procedúr a algoritmov. Ich výstup obsahuje rôzne náhodné prvky, ktoré je možné parametrizovať. Parametrizovaním procedurálnych generátorov vieme docieliť konzistentný a predvídateľný výstup (pri rovnakých vstupoch je vždy rovnaký výstup). Generátory je možné použiť v počítačových hrách alebo pri matematických a fyzikálnych simuláciách alebo analýzach. Ich výstupy bývajú často monotónne a predvídateľné, a preto sa najčastejšie používajú v kombinácií s inými druhmi generátorov. Okrem monotónnosti, vygenerované terény bývajú veľmi hladké, a chýbajú im malé rozdiely vo výškach. Silnou stránkou procedurálnych generátorov je rýchlosť generovania a preto sa často používajú pri generovaní terénov v reálnom čase.

B. Fraktálne generátory

Tieto generátory používajú pre tvorbu terénov špecializovanú časť geometrie – fraktály. Ide o matematické procedúry, ktoré majú zaujímavé vlastnosti ako napríklad rekurziu, ktorá umožňuje vytváranie nekonečných detailov. Grafické znázornenie fraktálov pripomína geometrické tvary, ktoré je možné nájsť v prírode, ako napríklad ulity alebo koruny stromov, a preto je ich použitie pri tvorbe digitálnych terénov veľmi zaujímavé. Fraktály, ktoré sa používajú pri generovaní terénov bývajú znázornené geometrickými tvarmi ako sú trojuholníky alebo štvorce, ktoré sa kombinujú s deformačným algoritmom aby vytvorili realistické terény [1]. Používajú sa najmä pri generovaní hôr a kopcov. Tieto generátory bývajú často zaraďované medzi procedurálne generátory kvôli ich podobným výstupom aj technike generovania. Príkladom fraktálneho generátora je algoritmus diamantových štvorcov [2].

C. Rastrové generátory

Tieto generátory vytvárajú výstupný terén na základe vstupného rastra (matice) buniek s vopred definovanou výškou. Tento postup sa najčastejšie používa pri tvorbe terénov v geografických informačných systémoch (GIS), kedy je takýto model schopný na základe niekoľkých vstupných dát, ktoré boli získané napríklad digitalizáciou, vytvoriť model s vysokou presnosťou. Rastrové generátory generujú terény zo skutočných geografických dát. Takýto terén presne zodpovedá jeho skutočnému ekvivalentu. Využívajú sa skôr pre simulácie, architektonické vizualizácie alebo v spomínaných geografických informačných systémoch. Ich nevýhodou je, že z jedného vstupného rastra je možné vygenerovať vždy iba rovnaký terén, čo obmedzuje ich širšie využitie. Sú dobrou voľbou pre video hry, ktorých svet má byť založený na skutočnom mieste.

D. Geomorfologické generátory

Ide o špecializované generátory, ktoré pre tvorbu využívajú vopred naprogramované algoritmy, simulujúce skutočné prírodné a biologické procesy. Najčastejšie sa používajú pre simuláciu erózie ale môžu sa použiť aj pri simulácii sedimentácie alebo iných prírodných javov. Ich použitie je dôležité pre dosiahnutie realistických terénov. Nevýhoda geomorfologických generátorov je ich časová náročnosť a ich použitie na generovanie terénu v reálnom čase je obmedzené. Táto nevýhoda sa však dá s dostatočnou optimalizáciou algoritmov vyriešiť [3].

E. Špeciálne generátory

Patria sem ďalšie špecializované generátory, ktoré sa zameriavajú na konkrétne časti krajiny. Ide napríklad o generátory vodných tokov a hladiny morí. Tieto generátory väčšinou upravujú terény, ktoré boli vygenerované inými generátormi. Pod špeciálne generátory je možné zaradiť aj generátory, ktoré využívajú strojové učenie pre generovanie terénu. Pri týchto generátoroch je však potrebné ich model natrénovať na veľkej vzorke predpripravených terénov. Získať veľkú vzorku môže byť komplikované, a preto sa väčšinou využívajú iné druhy generátorov. Ich ďalšou nevýhodou je, že vlastnosti vygenerovaných terénov určuje druh vzoriek. Pre iný druh terénu je potrebné znova natrénovať model na novej vzorke.

III. Programy na generovanie terénu

Na generovanie terénu využívajú kombinované generátory a v niektorých prípadoch môžu výstupy generátorov kombinovať s rôznymi matematickými funkciami a operáciami. Generátory sú ovládané parametrami, ktoré nastavuje používateľ. Počet dostupných parametrov sa líši od komplexnosti konkrétneho programu. Okrem zmeny parametrov pred generovaním terénu, niektoré programy ponúkajú aj nástroje na úpravu výsledných terénov. Niektoré programy využívajú pre nastavovanie parametrov uzlové grafy. Funkcionalita týchto generátorov je rozdelená do samostatných uzlov, ktoré si používateľ môže ľubovoľne vybrať. Tieto uzly potom navzájom prepojí a vznikne jeden veľký graf. Každý uzol má vlastnú úlohu pre generovanie terénu. Niektoré uzly zabezpečujú generovanie terénu a iné upravujú výstup predošlých uzlov ale kombinujú výstupy viacerých uzlov. Analyzovaných bolo niekoľko programov na generovanie terénu a ich porovnanie sa nachádza v Tab. 1.

IV. Formát vstupných dát navrhnutého procedurálneho generátora

Bolo potrebné zabezpečiť aby vstupné dáta boli jednoduché na vytvorenie, ale zároveň boli dostatočne komplexné na to, aby mohli slúžiť ako dobrý základ pre vygenerovaný terén. Vstupné dáta obsahujú prvky, ktoré určujú základný tvar terénu, podľa ktorého sa vygeneruje terén. V rámci tejto práce navrhnutý procedurálny generátor podporuje dva druhy prvkov. Polygón, ktorý

Tabuľka 1
Porovnanie vlastností generátorov terénu

	Blender	**L3DT**	**TerraForge 3D**	**Gaea**	**World Machine**	**Instant Terra**
Kontrola nad terénom	malá	stredná	stredná	veľká	veľká	veľká
Procedurálne generátory	áno	áno	áno	áno	áno	áno
Fraktálne generátory	áno	áno	áno	áno	áno	áno
Simulácia erózie	nie	nie	áno	áno	áno	áno
uzly na generovanie terénov	-	-	-	48	17	21
uzly na úpravu terénov	-	-	-	29	7	46
uzly na generovanie textúr	-	-	-	22	28	40
export do Unreal enginu	áno	áno	áno	áno	áno	áno

znázorňuje jednotlivé vrstevnice terénu a línia, ktorá znázorňuje rieky a cesty. Tieto prvky obsahujú parametre, ktoré bližšie určujú tvar terénu. Vstupné dáta sa vytvárajú pomocou programu Quantum Geographic Information System (QGIS). Parametre prvkov je možné nastaviť v QGIS alebo aj priamo v programe generátora cez editor. V editore programu je možné okrem zmeny parametrov zmeniť aj veľkosť alebo pozíciu prvkov.

V. Generátor

Výstup generátora je deterministický a nezávislý od rozlíšenia výslednej výškovej mapy. Okrem výškových máp ponúka možnosť vygenerovať aj rôzne špeciálne mapy, ktoré môžu slúžiť pre označenie rôznych regiónov alebo oblastí pre konkrétnu vegetáciu.

Prvým krokom generovania je vyhladzovanie vrstevníc vo vektorových dátach. Vyhladzovanie prebieha tak, že pre každé dva prvky, ktoré sa nachádzajú v rovnakej vrstve sa vytvorí nový prvok, ktorého body sú výsledkom lineárnej interpolácie dvoch najbližších bodov susedných prvkov. Výška nového prvku je výsledkom lineárnej interpolácie ich výšiek. Silu vyhladzovania je možné meniť pomocou počtu iterácií.

V druhom kroku sú vektorové dáta konvertované na rastrové dáta. Postup konverzie je jednoduchý. Pre každý prvok sa vytvorí štvoruholník, ktorý daný prvok ohraničuje. Následne sa pre každý bod vo vnútri tohto štvoruholníka priradí hodnota výšky podľa toho, či sa daný bod nachádza vo vnútri prvku alebo nie. Výsledkom tejto konverzie je dvojrozmerné pole 16 bitových čísel typu short, na ktorom prebiehajú ďalšie úpravy generátorom.

Ostatné funkcie generátora sú rozdelené na samostatné časti nazývané segmenty. Každý segment reprezentuje konkrétny algoritmus alebo funkciu generátora. Každý segment obsahuje niekoľko parametrov, ktoré ovplyvňujú ich výstupy. Segmenty je možné ľubovoľne kombinovať a pridávať do generátora podľa potreby. Vďaka voľnosti pri výbere segmentov sa dá vygenerovať množstvo rozdielnych terénov s rozdielnymi detailmi z rovnakých vstupných dát ako je vidno na Obr. 1.

A. Segment procedurálneho šumu

Vytvára rôzne nerovnosti v teréne, čím vytvára detaily na inak rovnom teréne. Pre generovanie nerovností využíva generátor procedurálneho šumu. Typ procedurálneho šumu je možné nastaviť.

Generátor Perlinovho šumu

Perlinov šum bol vyvinutý okolo roku 1981 Kennethom Perlinom a dodnes sa v rôznych modifikovaných verziách používa na generovanie prirodzene vyzerajúcich textúr, terénov alebo štruktúr. Generovanie Perlinovho šumu začína generovaním matice ohraničujúcich bodov a im príslušným gradientovým vektorom. Gradientový vektor pre každý bod matice so súradnicami x,y označíme ako g(i,j). V prvom kroku samotného výpočtu sa pre každý bod mriežky T(x,y) určia body, ktoré ho ohraničujú. Ďalej je potrebné určiť nové „vzdialenostné“ vektory medzi ohraničujúcimi bodmi a pôvodným bodom. Nasleduje výpočet skalárneho súčinu medzi gradientovým a vzdialenostným vektorom ohraničujúceho bodu:

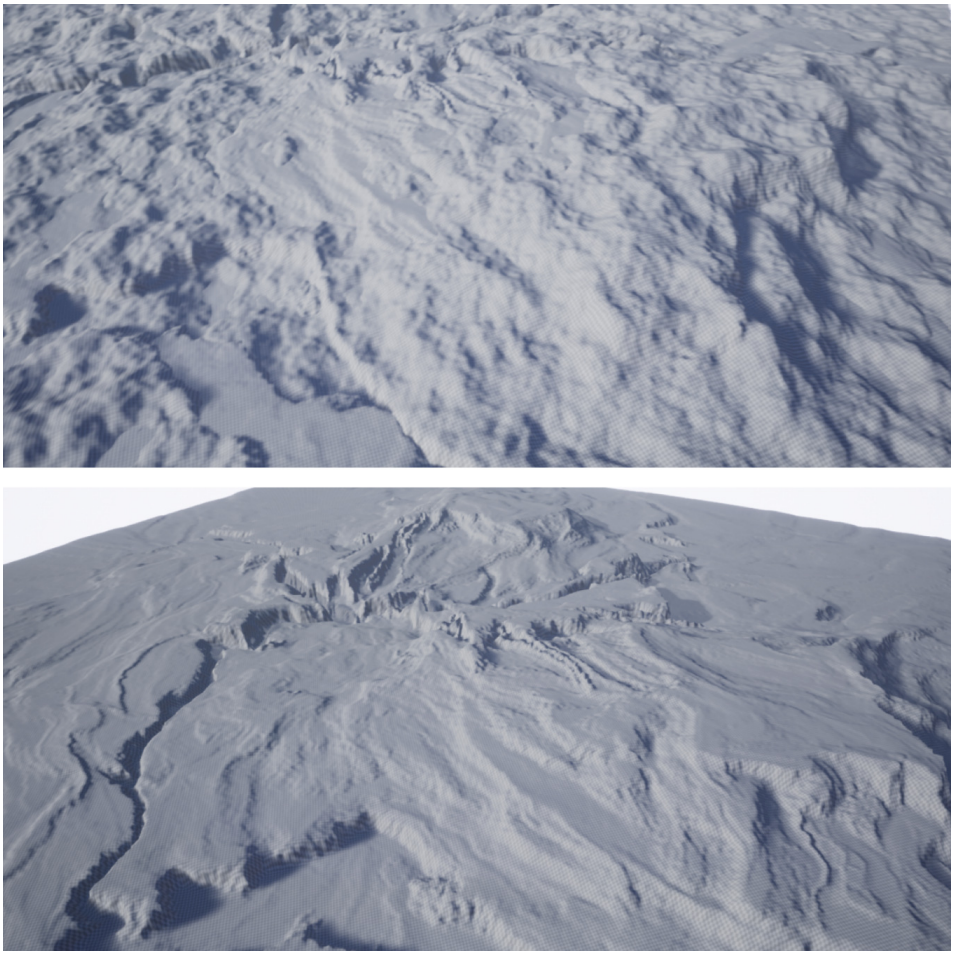
$$\begin{gathered} \boldsymbol{dx = x - i, dy = y - j} \\ \boldsymbol{N(i,j) = dx \cdot g(i,j) \cdot x + dy \cdot g(i,j) \cdot y} \end{gathered} \tag{1}$$

Ďalším krokom je interpolácia takto získaných hodnôt. Použije sa bilineárna interpolácia všetkých hodnôt, ktorá najprv lineárne interpoluje x súradnice a potom y. Tu by sa mohol algoritmus

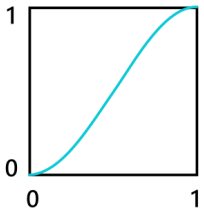
skončiť, avšak takto získaný výsledok po lineárnej interpolácii vyzerá neprirodzene. Na ďalšie zjemnenie prechodov medzi gradientami sa ešte použije funkcia na princípe S-krivky, ktorej tvar je znázornený na Obr. 2.

Generátor OpenSimplex šumu
Je to vylepšená verzia originálneho Perlinovho šumu, ktorá vyriešila niektoré jeho nedostatky, ako napríklad vznik rôznych smerových artefaktov, alebo veľkú výpočtovú náročnosť pri mriežkach s väčšími rozmermi. Princíp algoritmu je podobný ako pri Perlinovom šume až na niekoľko rozdielov. Pôvodný Perlinov šum používa pravidelnú mriežku, kde najmenšou štruktúrou je štvoruholník, tak OpenSimplex používa jednoduchú hexagonálnu mriežku, kde najmenšou štruktúrou je trojuholník [4]. Pre každý vstupný bod P(x,y) sa určia hraničné body príslušného trojuholníka. Súradnice vrcholov trojuholníkov mriežky sa najskôr transformujú po hlavnej diagonále tak, až je vstupná mriežka normalizovaná do siete dvojíc trojuholníkov, ktoré tvoria štvorce. Transformácia je znázornená na Obr. 3. Určovanie príslušného trojuholníka pre každý bod sa tak výpočtovo zmení z matematicky komplikovanej operácie na jednoduché, celočíselné porovnávanie súradníc x a y. Simplex algoritmus namiesto výpočtovo náročnej interpolácie používa jednoduchý súčet hodnôt hraničných bodov. Zjednodušene vplyv každého vrcholu zároveň výrazne klesá od jeho vzdialenosti od bodu P(x,y). Zisťuje sa to tak, že okolo každého vrcholu trojuholníka je opísaná kružnica s polomerom, ktorý nepresiahne hranice vedľajšieho trojuholníka. Týmto sa zabezpečí, že do výpočtu vstupujú naozaj len tri ohraničujúce body.

Generátor Bunkového šumu
Na rozdiel od predchádzajúcich generátorov je tento jednoduchší a nepotrebuje komplikované výpočty. Popísal ho v roku 1996 Steven Worley [5]. Základný princíp výpočtu Worleyho šumu spočíva v náhodnom rozmiestnení určitého počtu bodov v priestore a zisťovaní vzdialenostných vektorov ku každému tomuto bodu z každého skúmaného bodu, pre ktorý sa určuje výsledná hodnota [6]. Vzdialenostné vektory sú potom zoradené podľa veľkosti a najkratší z nich určuje



Obr. 1 Dva rozdielne terény vygenerované z rovnakých vstupných dát.



Obr. 2 Znázornenie S-krivky Perlin generátora

hodnotu príslušného pixelu. Pre urýchlenie výpočtu sa priestor v ktorom sa počíta šum rozdelí do mriežky. Na začiatku je každý náhodne vytvorený bod pridelený konkrétnej bunke mriežky. Výpočet vektorov sa potom realizuje celočíselným porovnávaním súradníc, čím sa výrazne skráti potrebný čas. Worleyho algoritmus nepredpisuje presnú metódu akou sa vypočíta výsledná hodnota skúmaného bodu. Závisí to vždy od konkrétnej implementácie. Niekedy sa pre zlepšenie výstupu môžu pridať ďalšie parametre ako napríklad výpočet farebných prechodov odvodených od vzdialeností. V implementácii sa používa Euklidovská vzdialenosť. Tá je pre body v n-rozmernom priestore zo vzorca (2) daná vzťahom vo vzorci (3) [7]:

$$x = [x_1, ..., x_n], y = [y_1, ..., y_n] \tag{2}$$

$$p(x, y) = \sqrt{(x_1 - y_1)^2 + ... + (x_n - y_n)^2} \tag{3}$$

Tieto vzdialenosti sa počítajú pre každý jeden skúmaný bod a náhodne vygenerované body. Ako bolo spomenuté vyššie, tieto hodnoty sa usporiadajú a najmenšia z nich potom určuje hodnotu skúmaného bodu P. Finálny výpočet je možné vyjadriť vzťahom:

$$P = min\|S_i - P\| \tag{4}$$

Kde, ‖ ‖ vyjadruje Euklidovskú vzdialenosť, S_i je náhodne vygenerovaný bod z celkového počtu **i**. **P** je skúmaný bod a **min** predstavuje výber najmenšej hodnoty z množiny vygenerovaných Euklidovských vzdialeností. Vďaka možnosti použiť vlastnú metódu na výpočet výslednej hodnoty skúmaného bodu je tento algoritmus vysoko modifikovateľný a obľúbený. Výsledkom je často štruktúra podobná bunkám a preto sa tento generátor šumu v niektorých zdrojoch uvádza aj ako „bunkový“ generátor šumu.

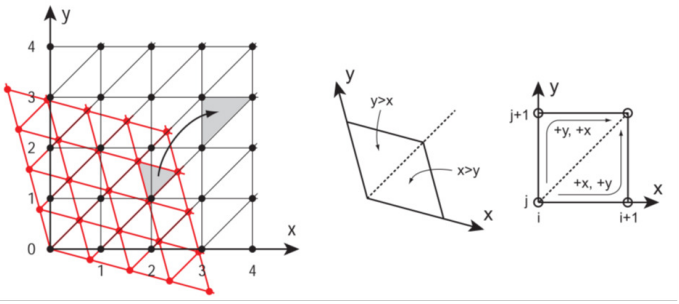
B. Segment modifikátora deformácie domény

Deformuje priame hrany vrstevníc, čím vytvára prirodzenejší vzhľad terénu. Pre deformáciu využíva rovnaké procedurálne generátory ako segment procedurálneho šumu.

Algoritmus deformácie funguje na jednoduchom princípe. Pre každý bod P(x,y) vypočíta nové súradnice následovne.

$$\begin{aligned} dx &= noise(x, y) \cdot s \\ dy &= noise(y, x) \cdot s, \end{aligned} \tag{5}$$

Kde, **s** je veľkosť deformácie a **noise** je funkcia generátora šumu, ktorá vráti hodnoty v rozmedzí $\langle -1, 1 \rangle$.



Obr. 3 Transformácia miežky v Simplex generátore

C. Segment manipulácie výšky

Upravuje výšku terénu podľa mapy, ktorú tvorí vybraná vrstva terénu, a zadanej konštanty. Zmena výšky časti terénu umožňuje vyriešiť menšie nedostatky, ako napríklad vysoký rozdiel vo výške prvkov. Vrstva terénu sa vyberá osobitne od importovaných vrstiev. Výšku upravuje pomocou jednoduchých matematických operácií ako je sčítanie, odčítanie alebo násobenie. Ponúka aj prepísanie výšky zadanou konštantou, čím sa zarovná daná časť terénu. Zarovnanie terénu slúži pre vytvorenie oblasti, na ktorej sa dajú ľahšie položiť veľké dekoračné alebo funkčné prvky sveta, ako napríklad vyhliadková veža alebo mesto.

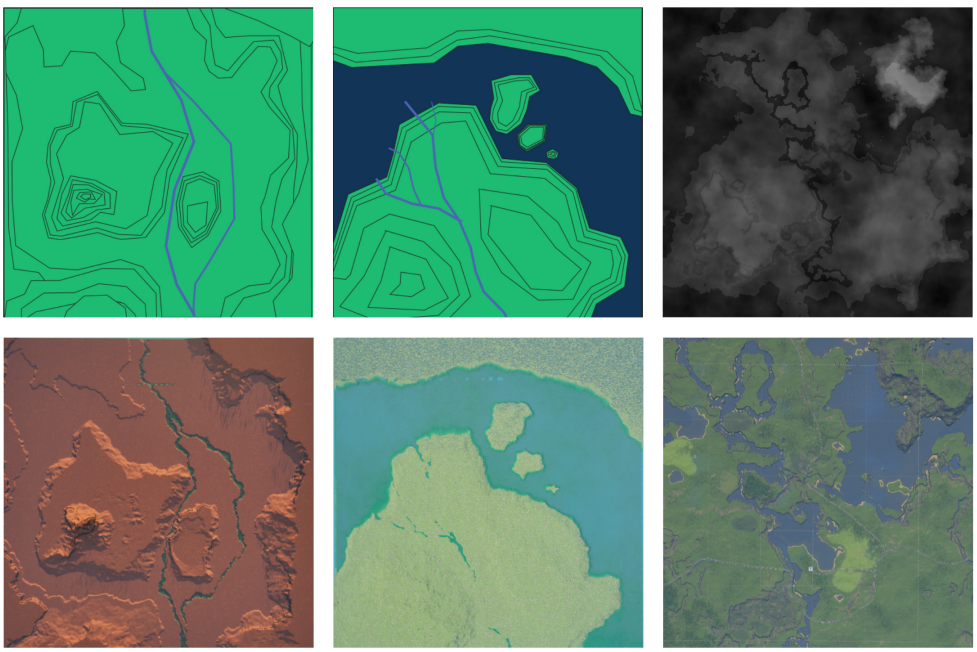
D. Segmenty vyhladzovania mapy

Aj po vyhladzovaní vrstiev sa na mape nachádzajú viditeľné prechody medzi vrstevnicami terénu. Tieto segmenty viditeľné prechody vyhladzujú. Fungujú na rovnakom princípe ako algoritmy pre rozmazávanie fotiek a obrázkov. K dispozícii sú dva segmenty vyhladzovania. Jeden segment využíva na vyhladzovanie mapy Gaussovské rozostrenie a druhý segment funguje na jednoduchom princípe, kde pre každý bod na mape priradí priemernú výšku susedných bodov.

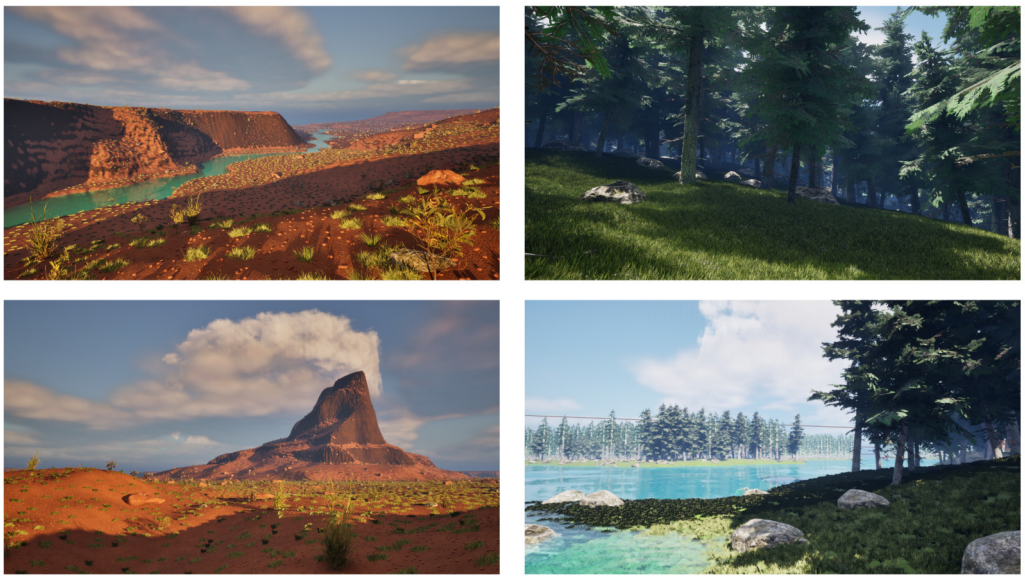
VI. Funkčnosť a využitie generátora

Vygenerované výškové mapy sú primárne určené pre importovanie do Unreal enginu. V Unreal engine je možné pomocou výškových a špeciálnych máp vytvárať detailné prostredia. Unreal engine ponúka možnosť vytvoriť procedurálne nástroje pre náhodné pokladanie rôznych objektov ako sú stromy, rastliny alebo kamene. Pre určenie oblasti pokladania je možné použiť vygenerované špeciálne masky, čo umožňuje vizuálne rozdeliť terén na viaceré regióny ako sú lúky, lesy alebo púšte.

Tým, že je formát výškových máp všeobecne využívaný, je možné použiť výškové mapy aj v iných programoch ako napríklad Unity3D alebo Blender. Niektoré hry ponúkajú možnosť importovania výškovej mapy, ktorá sa použije ako terén v hre. Jedna z týchto hier je Cities Skylines, v ktorej bola výšková mapa otestovaná. Počas hrania nebola nájdená žiadna chyba v teréne, ktorá by obmedzila hrateľnosť hry. Na Obr. 4 sú zobrazené vygenerované terény spolu s ich vektorovou reprezentáciou a terén v hre Cities skylines s použitou výškovou mapou. Na Obr. 5 sú detailné zábery na prostredia vytvorené v Unreal engine pomocou vygenerovaného terénu a procedurálne poukladanou vegetáciou.



Obr. 4 Zobrazenie výstupu generátora a jeho využitie



Obr. 5 Detailné zábery na vytvorené prostredia pomocou vygenerovaných výškových máp

VII. Záver

Cieľom práce bolo navrhnúť a implementovať procedurálny generátor terénu. Procedurálny generátor mal umožňovať rýchle a efektívne generovanie realistických terénov. Používateľ má mať vysokú kontrolu nad výslednou formou no zároveň nemusí nad tvorbou terénu stráviť veľa času aby bol terén detailný. Pre dosiahnutie cieľa boli analyzované a porovnané existujúce generátory terénu. Podľa výsledkov porovnania bol navrhnutý procedurálny generátor, ktorý pre dosiahnutie veľkej kontroly nad výsledným terénom využíva vektorové vstupné dáta. Vstupné dáta sú vektorovou reprezentáciou základného tvaru požadovaného terénu. Implementované riešenie môže slúžiť ako dobrý základ pre budúce rozšírenie. Tým, že su jednotlivé funkcie a algoritmy rozdelené, umožňujú jednoduché rozšírenie programu o nové funkcie.

Poďakovanie

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

Literatúra

[1] K. Piispa, "Methods for procedural terrain generation," p. 61, 2022, dostupné online: https://urn.fi/URN:NBN:fi-fe2022041929481.
[2] "Perlin noise vs diamond square," dostupné online: https://dknazar.wordpress.com/portfolio/perlin-noise-vs-diamond-square/; Navštívené: 21.2. 2024.
[3] X. Mei, P. Decaudin, and B.-G. Hu, "Fast hydraulic erosion simulation and visualization on gpu," 10 2007.
[4] Uniblock, "Noise!" 09 2014, navštívené: 10.2. 2024. [Online]. Available: https://uniblock.tumblr.com/post/97868843242/noise
[5] S. Worley, "A cellular texture basis function," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: Association for Computing Machinery, 1996, p. 291–294. [Online]. Available: https://doi.org/10.1145/237170.237267
[6] C.-J. Rosén, "Cell noise and processing," 2006.
[7] E. Beliana, "Vzdialenosť dvoch bodov," 12 2020, navštívené: 7.5. 2024. [Online]. Available: https://beliana.sav.sk/heslo/vzdialenost-dvoch-bodov

Výber polovodičov pre usmerňovač mostíkového meniča s fázovým posunom

Daniel GORDAN, Marek PASTOR, Tadeáš KMECIK, Tomáš BASARIK

Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach

daniel.gordan@tuke.sk, marek.pastor@tuke.sk, tadeas.kmecik@tuke.sk,
tomas.basarik@tuke.sk

***Abstrakt* — V článku je opísaný výber polovodičov pre synchrónny usmerňovač nepriameho jednosmerného mostíkového meniča s fázovým posunom, pričom matematické výpočty strát sú overené simuláciou v programe LTspice.**

***Kľúčové slová* — mostíkový menič s fázovým posunom, synchrónny usmerňovač, výber polovodičov, SI MOSFET**

I. Úvod

Topológia nepriameho jednosmerného mostíkového meniča s fázovým posunom (Phase-Shifted Full-Bridge PSFB) je známa desiatky rokov[1] a je veľmi rozšírená všade, kde je vyžadovaný vysoký výstupný výkon, galvanické oddelenie a vysoká účinnosť. Medzi takéto aplikácie môžeme zaradiť meniče pre fotovoltiku[2], batériové úložiská[3], či nabíjačky pre elektrické autá[4].

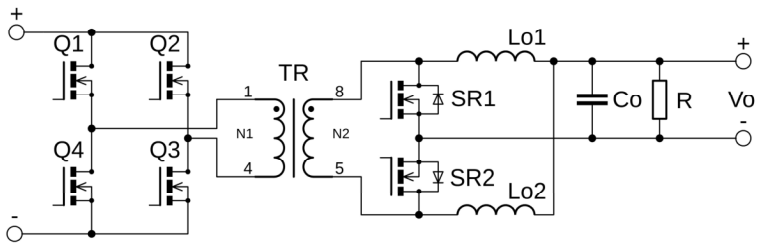
Menič je zložený z troch základných častí, striedača, transformátora a usmerňovača. V súčasnosti je striedač tvorený tranzistorom typu MOSFET, pričom moderné technológie typu GaN a SiC umožňujú zvýšiť frekvenciu pri zachovaní rovnakej účinnosti, čím dochádza k zmenšeniu rozmerov a hmotnosti výsledných meničov[5]. Takáto redukcia je výhodná nie len v domácich, ale aj v automobilových, či vesmírnych aplikáciách, kde je na hustotu výkonu kladený veľký dôraz. Striedač vytvára striedavé napätie, ktorým je napájaný vysokofrekvenčný transformátor. Transformátor je navrhnutý tak, aby bol schopný preniesť energiu z primárnej strany na sekundárnu stranu, pri zachovaní vysokej účinnosti, najmä pre vysoké frekvencie je vhodné použiť planárny transformátor[6].

Za transformátorom nasleduje usmerňovač, ktorý je tvorený diódami, alebo tranzistormi typu MOSFET. V tomto druhom prípade hovoríme o synchrónnom usmerňovači, ktorý je pri snahe o dosiahnutie vysokej účinnosti nevyhnutný. [7,8]

Článok opisuje výber tranzistorov od troch výrobcov, pričom sú pre každý tranzistor vypočítané vodivostné, spínacie a celkové straty. Správnosť matematických výpočtov je overená simuláciou v programe LTspice.

II. Topológia meniča

Jednosmerný menič pracuje v topológií mostíkového meniča s fázovým posunom, usmerňovač je typu zdvojovač prúdu, Obr. 1. Parametre meniča sú zobrazené v Tab. 1, komponenty boli určené na základe [8], a sú zobrazené v Tab. 2.



Obr. 1 Mostíkový menič s usmerňovačom typu zdvojovač prúdu

Tabuľka 1
Parametre meniča

Parameter	Hodnota
Napájacie napätie	390 V
Výstupný prúd	25 A
Výstupné napätie	24 V
Výstupný výkon	600 W
Spínacia frekvencia	100 kHz

Tabuľka 2
Komponenty meniča

Parameter	Hodnota
Prevodový pomer transformátora	6:1
Filtračná indukdnosť L_{o1}, L_{o2}	30 µH
Filtračný kondenzátor C_O	56 µF

Na primárnej strane sú použité GaN tranzistory typu **GS-065-030-2-L**, ktoré sa vyznačujú dobrým pomerom medzi výstupnou kapacitou C_{oss} a sériovým odporom v zopnutom stave Rds_{on}. Voľba sekundárnych tranzistorov, ktorá je témou tohto článku bude opísaná v ďalších kapitolách.

Primárne tranzistory na Obr. 1, Q_1 – Q_4 spínajú so striedou 50 % a malou bezpečnostnou medzerou, aby sa zabránilo skratu vo vetve. Výkon je regulovaný pomocou zmeny fázového posunu, čiže prekrytia medzi zapnutím tranzistora Q_1 a Q_3, čo analogický platí pre druhú vetvu. Takýmto spínaním je vytvorené striedavé napätie pre transformátor, na ktorého sekundárnej strane je usmerňovač typu zdvojovač prúdu. Tento typ usmerňovača vyžaduje dve filtračné indukčnosti, na úkor jednoduchšieho transformátora s jedným vinutím. Posledným prvkom v zjednodušenej schéme je výstupný filtračný kondenzátor, ktorý zabezpečuje želané zvlnenie.

III. VÝBER POLOVODIČOV PRE USMERŇOVAČ

A. Matematický opis strát

Matematický opis strát je uvedený v literatúre [8].

Vodivostné straty:

$$P_{con} = R_{DSon} I_s^2, \tag{1}$$

kde R_{DSon} je odpor tranzistora v zopnutom stave a I_s je efektívna hodnota prúdu na sekundárnej strane.

Straty od výstupnej kapacity tranzistora:

$$P_{oss} = 0{,}5 Q_{oss} \frac{V_o}{ph_{eff}} f, \tag{2}$$

kde Q_{oss} je výstupný náboj tranzistora, V_o je výstupné napätie, ph_{eff} je efektívna doba zapnutia a f je spínacia frekvencia.

Straty od budenia:

$$P_{gate} = V_g Q_g f, \tag{3}$$

kde V_g je napätie na hradle, Q_g je náboj hradla a f je spínacia frekvencia.

Hodnoty potrebné pre výpočet sú uvedené v Tab. 3. Pre výpočet strát je potrebné poznať aj hodnoty Rds_{on}, Q_{oss} a Q_g, tieto sú špecifické pre každý tranzistor. Tab. 4 – Tab. 6 zobrazuje parametre a im zodpovedajúce straty určené na základe vzorcov (1) – (3). Tab. 4 zobrazuje súčiastky od firmy Infineon, Tab. 5 od firmy Vishay a Tab. 6 od firmy Toshiba. Celkové straty sú dané súčtom jednotlivých strát definovaných v (1) – (3).

Tabuľka 3
Parametre meniča potrebné pre výpočet strát

Parameter	Hodnota
Efektívna hodnota sekundárneho prúdu I_S	16,48 A
Výstupné napätie V_o	24 V
Napätie hradla V_g	15 V
Spínacia frekvencia	100 kHz
Efektívna doba ph_{eff}	0,3692

Tabuľka 4
Parametre a straty súčiastok od firmy Infineon

Súčiastka	Vds_{max} (V)	I_D[1] (A)	Rds_{on}[2] (mΩ)	Q_{oss}[3] (nC)	Q_g (nC)	P_{con} (W)	P_{oss} (W)	P_{gate} (W)	P_{total} (W)
IPT039N15N5	150	190	3,9	219	78	1,06	0,712	0,117	1,89
IPT044N15N5	150	179	4,4	188	67	1,19	0,611	0,101	1,91
IPT054N15N5	150	143	5,4	155	55	1,47	0,504	0,083	2,05
IPT063N15N5	150	122	6,4	131	47	1,71	0,426	0,071	2,21

1) I_D pri V_{GS} = 10 V a T_C = 25 °C
2) Rds_{on} pri V_{GS} = 10 V a I_D = 50 A
3) Q_{oss} pri V_{DS} = 75 V a V_{GS} = 0 V

Tabuľka 5
Parametre a straty súčiastok od firmy Vishay

Súčiastka	Vds_{max} (V)	I_D[1] (A)	Rds_{on}[2] (mΩ)	Q_{oss}[3] (nC)	Q_g (nC)	P_{con} (W)	P_{oss} (W)	P_{gate} (W)	P_{total} (W)
SiR570DP	150	144	7,9	111	35	2,15	0,114	0,053	2,31
SiR578DP	150	70,2	8,8	105	24	2,39	0,079	0,037	2,51
SiR572DP	150	59,7	10,8	84	27	2,93	0,087	0,040	3,06
SiR574DP	150	47,1	13,5	68	23	3,67	0,077	0,035	3,78

1) I_D pri T_C = 25 °C
2) Rds_{on} pri V_{GS} = 10 V a I_D = 20 A
3) Q_{oss} pri V_{DS} = 75 V a V_{GS} = 0 V

Tabuľka 6
Parametre a straty súčiastok od firmy Toshiba

Súčiastka	Vds_{max} (V)	I_D[1] (A)	Rds_{on}[2] (mΩ)	Q_{oss}[3] (nC)	Q_g (nC)	P_{con} (W)	P_{oss} (W)	P_{gate} (W)	P_{total} (W)
TPH9R00CQ5	150	64	7,9	87	44	2,15	0,143	0,066	2,35
TPH1100CQ5	150	49	9	70	38	2,44	0,124	0,057	2,62
TPH1400CQ5	150	32	11,4	56	31	3,1	0,101	0,066	3,24

1) I_D pri T_C = 25 °C
2) Rds_{on} pri V_{GS} = 10 V a I_D = 22 A
3) Q_{oss} pri V_{DS} = 75 V a V_{GS} = 0 V

Pri výbere súčiastok bol dôraz kladený na nízky odpor v zopnutom stave, podobnosť púzdra a dostupnosť potrebných údajov v katalógových listoch. V neposlednom rade musia súčiastky zvládnuť prúdové a napäťové namáhanie. V danej topológií vznikajú na sekundárnej strane prirodzené prepätia, ktoré je možné pri výstupnom napätí 24 V odhadnúť približne na 70 V. Z tohto dôvodu sú tranzistory napäťovo predimenzované. [8]

Súčiastky od firmy Infineon sú v púzdre TOLL, od Vishay a Toshiba vo výkonovom púzdre SO-8. V tomto prípade z Tab. 4 – Tab. 6 vyplýva, že najmenšie straty mali súčiastky od firmy Infineon. Je tomu tak preto, lebo väčšina strát je tvorená vodivostnými stratami a Infineon v danom púzdre ponúka radu súčiastok s najnižším Rds_{on}, ktorý vplýva práve na vodivostné straty. Pre férové porovnanie je však vhodné uviesť pomer medzi veľkosťou Rds_{on} a Q_{oss}, a Rds_{on} a Q_g. Čím menší je násobok parametrov, tým má polovodič lepšiu kvalitu v daných parametroch. Tieto hodnoty sú uvedené v Tab. 7.

Z porovnania uvedeného v Tab. 7 vyplýva, že najlepší pomer medzi Rds_{on} a Q_{oss} má kremíkový tranzistor od firmy Toshiba. Súčiastky v tabuľke boli vybrané na základe podobného Rds_{on}, aj keď rozdiely v rovnakej rade súčiastok budú takmer zanedbateľné. Je vhodné doplniť, že polovodič od firmy Infineon ponúkal najvyššiu prúdovú zaťažiteľnosť, čo sa prejaví väčšou robustnosťou polovodiča a tým aj väčšími parazitnými kapacitami. Vzhľadom k tomu je možné konštatovať, že rozdiely v kvalite medzi výrobcami pri daných súčiastkach sú minimálne.

Na základe matematických výpočtov je simulačné overenie realizované so súčiastkami od firmy Infineon, pretože v danom púzdre ponúkajú nízky Rds_{on}, pričom si zachovávajú dobrý pomer medzi odporom v zopnutom stave a veľkosťou parazitných kapacít.

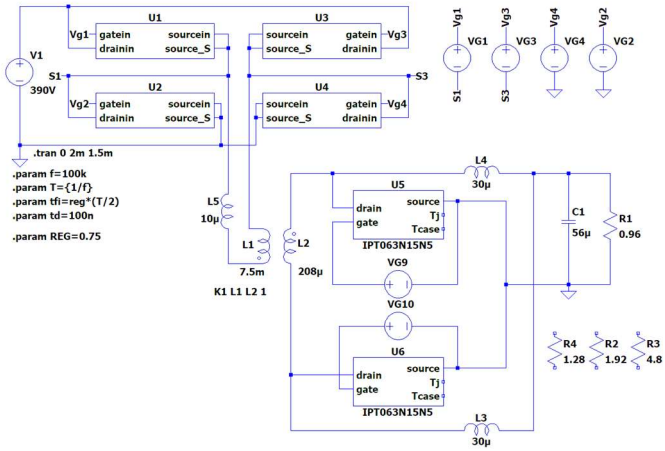
Tabuľka 7
Porovnanie kvality polovodiča pri sledovaných parametroch

Súčiastka	Výrobca	$Rds_{on}Q_{oss}$ (mΩnC)	$Rds_{on}Q_g$ (mΩnC)
IPTG063N15NM5	Infineon	875,7	315
SiR570DP	Vishay	876,9	277,29
TPH9R00CQ5	Toshiba	687,3	347,6

B. Simulačné overenie

Simulácia je vyhotovená v programe LTspice, pre štyri záťaže. Modely súčiastok sú stiahnuté priamo zo stránky výrobcu. Výsledné straty sú zobrazené v Tab. 8. Zelenou farbou je zvýraznená súčiastka, ktorá dosiahla najnižšie straty pri danej záťaži. Pri väčšej záťaži sa viac prejavia vodivostné straty. Pri menšej naopak spínacie straty od výstupnej kapacity. Z tohto tvrdenia vyplýva, že tranzistor s nižším Rds_{on} je vhodnejší pre väčšie záťaže, a naopak tranzistor s menšou výstupnou kapacitou pre menšie záťaže. Výsledné straty sú približne rovnaké, ako v prípade matematických výpočtov, kde bolo počítané s maximálnym výkonom. Straty sú prehľadne zobrazené v grafe na Obr. 3.

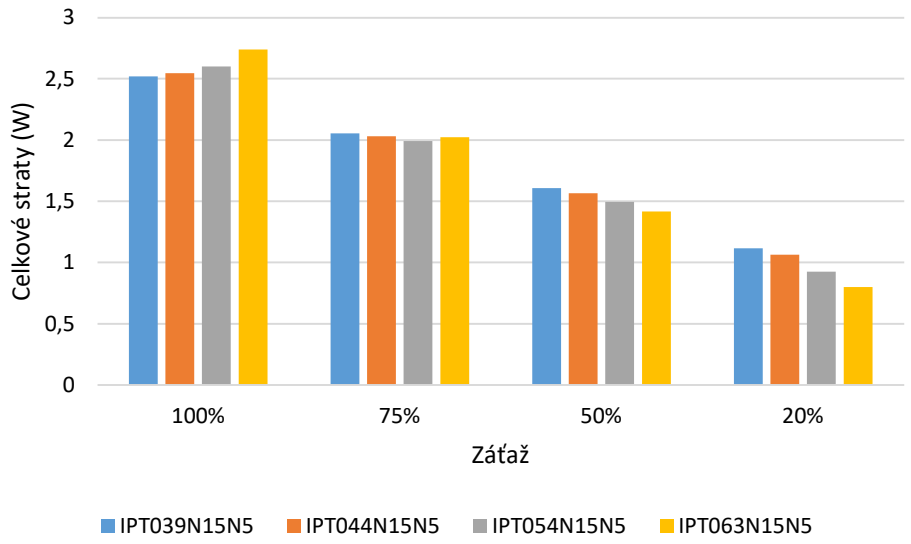
Na základe grafu na Obr. 3 je možné konštatovať, že tranzistor **IPT044N15N5** by mal byť vhodnou voľbou pre usmerňovač v danom mostíkovom meniči s fázovým posunom. Pri záťaži 75% dosiahol nižšie celkové straty, ako tranzistor IPT039N15N5, zatiaľ čo pri nominálnej záťaži 100% boli jeho straty len o trochu vyššie ako pri tranzistore s nižším odporom v zopnutom stave. Zároveň je dobré myslieť aj nato, že meniče prevažne pracujú pri trochu menšej ako nominálnej záťaži.



Obr. 2 Simulačná schéma v programe LTspice

Tabuľka 8
Simuláciou zistené straty súčiastok od firmy Infineon

Súčiastka	Záťaž 100%	Záťaž 75%	Záťaž 50%	Záťaž 20%
IPT039N15N5	2,5194	2,0533	1,6065	1,1146
IPT044N15N5	2,5446	2,0298	1,5661	1,0623
IPT054N15N5	2,5997	1,9925	1,4939	0,9243
IPT063N15N5	2,7396	2,0221	1,4172	0,7994



Obr. 3 Výkonová strata na polovodičoch pri rôznej záťaži

IV. Zhodnotenie

Článok opisuje topológiu mostíkového meniča s fázovým posunom a usmerňovačom typu zdvojovač prúdu. Pre zadané parametre meniča sú počítané spínacie a vodivostné straty pre viaceré polovodiče od troch výrobcov. Výsledkom matematického výpočtu je výber štyroch polovodičov od firmy Infineon, ktorých straty sú simulované v programe LTspice.

Výsledkom simulácie je výber jednej súčiastky IPT044N15N5, ktorá by mala v danom type meniča a pri zadaných parametroch pracovať najlepšie.

Poďakovanie

Tento príspevok vznikol s podporou Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied v rámci projektu VEGA 1/0584/24.

Referencie

[1] L. H. Mweene, C. A. Wright and M. F. Schlecht, "A 1 kW, 500 kHz front-end converter for a distributed power supply system," Proceedings, Fourth Annual IEEE Applied Power Electronics Conference and Exposition, Baltimore, MD, USA, 1989, pp. 423-432, doi: 10.1109/APEC.1989.36994.

[2] S. Mukherjee, S. S. Saha and S. Chowdhury, "Battery Integrated Three-Port Soft-Switched DC–DC PSFB Converter for SPV Applications," in IEEE Access, vol. 11, pp. 62472-62483, 2023, doi: 10.1109/ACCESS.2023.3287149.

[3] A. Brandis, K. Knol, D. Pelin and D. Topić, "Prototype Proposal of an 18 kW Non-Isolated Bidirectional Converter for Battery Energy Storage System," 2023 22nd International Symposium on Power Electronics (Ee), Novi Sad, Serbia, 2023, pp. 01-06, doi: 10.1109/Ee59906.2023.10346198.

[4] A. Ghatak, T. Pandit, S. R. Kavitha and C. A, "Performance Analysis of Charge Controller for Electric Vehicle with Vehicle to Grid System," 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia, 2021, pp. 1-7, doi: 10.1109/i-PACT52855.2021.9696489.

[5] M. Heintze and I. Stefan Butzmann, "A GaN 500 kHz High Current Active Clamp Phase-Shifted Full-Bridge Converter With Zero-Voltage Switching Over the Entire Line and Load Range," 2018 20th European Conference on Power Electronics and Applications (EPE'18 ECCE Europe), Riga, Latvia, 2018, pp. P.1-P.9.

[6] L. Dixon, "Designing Planar Magnetics," Texas Instruments, Dostupné na internete : <https://www.ti.com/download/trng/docs/seminar/Topic4LD.pdf>

[7] Texas Instruments, “Phase-Shifted Full Bridge DC/DC Power Converter Design Guide,” Application Note, 2014, Texas Instruments

[8] S. Abdel-Rahman, “Design of Phase Shifted Full-Bridge Converter with Current Doubler Rectifier,” Design Note, 2013, Infineon Technologies North America

Porovnanie metód riadenia BLDC motora s využitím umelej inteligencie

[1]*Peter GIROVSKÝ,* [2]*Tadeáš KMECIK,* [3]*Daniel GORDAN,* [4]*Tomáš BASARIK*

Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky
Technická Univerzita v Košiciach, Slovenská Republika

[1]peter.girovsky@tuke.sk, [2]tadeas.kmecik@tuke.sk, [3]daniel.gordan@tuke.sk,
[4]tomas.basarik@tuke.sk

***Abstrakt* — Článok prezentuje metódy riadenia trojfázového BLDC motora s využitím umelej inteligencie. V úvodnej časti si rozoberieme model motora a riadiacu schému použitú pri simuláciách. Následne v ďalšej časti porovnáme dve metódy umelej inteligencie s klasickým PI regulátorom. Jednotlivé typy metód riadenia boli porovnané na základe stanovených parametrov a analýzy odsimulovaných priebehov. V závere na základe týchto priebehov boli opísané kvality jednotlivých typov regulátorov.**

***Kľúčové slová* — BLDC motor, umelá inteligencia, neurónová sieť, fuzzy logika, riadenie**

I. ÚVOD

V súčasnosti sa stretávame s pojmom umelá inteligencia čoraz viac a to nielen medzi odborníkmi z oblasti, ale aj medzi laickou verejnosťou. Umelá inteligencia v jednoduchosti využíva algoritmy, ktorými určuje správanie sa nejakému systému na základe definovaných pravidiel [1],[2],[5]. V praxi existuje niekoľko prístupov, ktorými vieme vytvoriť umelú inteligenciu, pričom najčastejšie sa využíva strojové učenie, neurónové siete a hĺbkové učenie. Tieto prístupy majú jedno spoločné. Pristupujú k vytváraniu umelej inteligencie na základe podobnosti s fungovaním ľudského mozgu [3],[4],[6],[7],[8],[9]. Podstatou fungovanie ľudského mozgu je to, že sa dokáže učiť a rozvíjať svoje schopnosti. Spojenie týchto vlastností s rozvojom výpočtovej techniky spôsobí nárast nasadzovania umelej inteligencie pri riešení problémov a úloh, pri ktorých klasické metódy zlyhávajú, alebo sú neefektívne.

Využitie umelej inteligencie [10] v technickej oblasti vedie k optimalizácií riadiacich procesov, najmä pri vysoko nelineárnych systémoch vyššieho radu. V tejto oblasti sa umelá inteligencia používa na modelovanie nelineárnych systémov [18], predikciu a identifikáciu [16],[17] parametrov alebo na riadenie [13],[14],[15],[19],[20],[21]. Konkrétne pre oblasť riadenia elektrických pohonov [11],[12],[22] umelú inteligenciu využívame buď ako pozorovatele potrebných stavov alebo ako regulátory.

II. BLDC MOTOR

BLDC je bezkefový motor s permanentnými magnetmi. Môže byť vyhotovený s vnútorným alebo vonkajším rotorom. Vinutie je umiestnené na statore a magnety sú na rotore. Statorové vinutie býva najčastejšie trojfázové. Z vyšším počtom fáz zabezpečíme plynulejší chod motora, vyšší výkon a tichšiu prevádzku.

V porovnaní s komutátorovým motorom neobsahuje kefové uhlíky a komutátor. Preto sa vyznačuje vyššou spoľahlivosťou, životnosťou, efektívnosťou, výkonnosťou a robustnosťou. Okrem spomínaných výhod ponúka aj menšie nároky na údržbu a možnosť prevádzky v nebezpečnom prostredí.

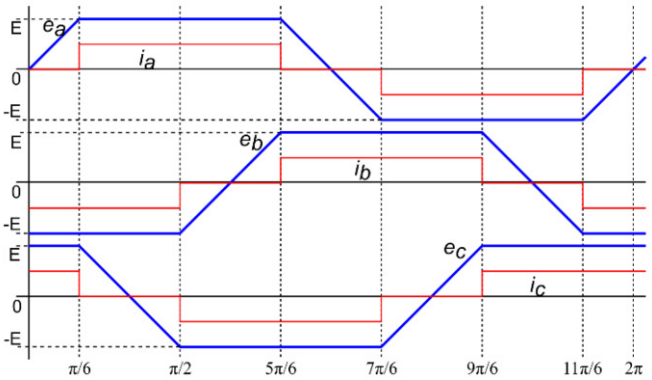
K nevýhodám môžeme zaradiť vyššie obstarávacie náklady a zložitejšie riadenie. V praxi je široko používaný v rôznych oblastiach priemyselnej automatizácie a takisto automobilovom priemysle.

A. Matematický model BLDC motora

Ak uvažujeme motor so symetricky rozloženým vinutím odpor, indukčnosť a vzájomná indukčnosť je rovnaká pre každé jednotlivé vinutie motora. Trojfázový BLDC motor môžeme matematicky popísať nasledovne (1) [23]:

$$V_a = i_a R + L\frac{di_a}{d_t} + M\frac{di_b}{d_t} + M\frac{d_{ic}}{d_t} + e_a$$
$$V_b = i_b R + L\frac{di_b}{d_t} + M\frac{di_a}{d_t} + M\frac{d_{ic}}{d_t} + e_b \quad (1)$$
$$V_c = i_c R + L\frac{di_c}{d_t} + M\frac{di_b}{d_t} + M\frac{d_{ia}}{d_t} + e_c$$

kde R – je fázový elektrický odpor vinutia, L – je fázová indukčnosť vinutia, M – vzájomná indukčnosť medzi vinutiami, i_a, i_b, i_c – fázové prúdy, V_a, V_b, V_c – fázové napätia, e_a, e_b, e_c – spätné indukované napätie. Na Obr. 1 môžeme vidieť priebeh indukovaného napätia pre jedno elektrické otočenie.



Obr. 1 Priebeh spätného indukovaného napätia

Hodnota spätného indukovaného napätia(EMF) závisí od pozície rotora a od amplitúdy E. Amplitúdu vypočítame podľa (2), v ktorej vystupuje aj konštanta K_e čiže konštanta spätného EMF.

$$E = K_e \omega \quad (2)$$

Elektromagnetický moment BLDC motora vypočítame podľa (3):

$$T_e = \frac{e_a i_a + e_b i_b + e_c i_c}{\omega} \quad (3)$$

B. Riadenie a komutácia

Komutáciu v BLDC motore zabezpečuje algoritmus, podľa ktorého sa spínajú tranzistory v meniči. Na základe polohy rotora riadený menič prepína prúdy do jednotlivých fáz motora. Na určenie polohy rotora sa používajú enkodéry, rezolvery alebo Hallové snímače. Použitie týchto snímačov umožňuje zvýšiť presnosť riadenia rýchlosti a momentu motora. Pri dynamicky nenáročných aplikáciách môžeme využiť aj bezsnímačové riadenie. Tento typ riadenia využíva meranie indukovaného napätia.

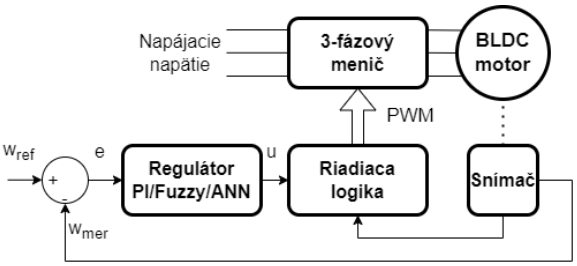
V praxi sa väčšinou využíva všeobecne známy 6-krokový komutačný algoritmus. Tento algoritmus generuje počas jednej elektrickej otáčky, 6 rôznych vektorov napätia. Napäťové vektory sú generované takým spôsobom, že sa udržiava vzájomná kolmosť medzi magnetickým poľom statora a rotora. Uhlová rýchlosť motora sa riadi pomocou zmeny strednej hodnoty napätia, ktorým napájame motor, čiže pomocou modulácie (PWM).

III. NÁVRH RIADENIA BLDC MOTORA S VYUŽITÍM UMELEJ INTELIGENCIE

Ako bolo spomenuté, v tejto práci sa analyzujú vybrané metódy umelej inteligencie, ktoré sa používajú na riadenie elektrických pohonov. Vybrané metódy umelej inteligencie porovnáme s klasickým PI regulátorom a vyhodnotíme kvalitu a vlastnosti skúmaných typov regulátorov.

A. Simulačný model

Testovanie kvality jednotlivých regulátorov bolo vykonané na základe simulačnej schémy uvedenej na Obr. 2. Parametre použitého motora v simuláciu sú uvedené v Tabuľke I.



Obr. 2 Principiálna bloková schéma riadenia BLDC motora

Vstupným signálom je žiadaná referenčná uhlová rýchlosť w_{ref}, od ktorej sa odčíta meraná skutočná uhlová rýchlosť. Rozdielom je regulačná odchýlka, ktorá vstupuje do regulátora. Výstupom z regulátora je akčný zásah u, ktorý vstupuje do riadiacej logiky. Riadiaca logika obsluhuje menič, čiže určuje, ktorý tranzistor v meniči má byť spínaný.

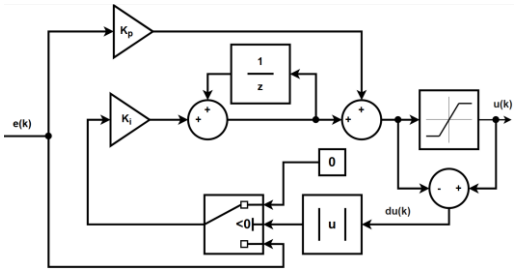
Pozícia rotora sa vyhodnocuje na základe signálov z Hallových snímačov umiestnených v motore. Tieto signály takisto vstupujú do riadiacej logiky. Spínanie tranzistorov sa riadi na základe spomínaného 6-krokového algoritmu. Ten zabezpečí vytvorenie točivého magnetického poľa.

Tabuľka I
Parametre simulovaného BLDC motora

Parameter	Hodnota
p – počet párov pólov	4
U_N – nominálne statorové napätie [V]	48
n_N – nominálne otáčky motora [RPM]	4000
T_N – nominálny moment motora [Nm]	0,24
I_N – nominálny prúd motora [A]	2,43
k_e – elektrická konštanta [Vrms/Krpm]	6,88
k_t – momentová konštanta [Nm/Arms]	0,099
J – moment zotrvačnosti [kg.m^2]	0,38.10^{-4}
R – statorový odpor [Ω]	0,71
L – statorová indukčnosť [mH]	1,35

B. PI regulátor

Principiálne zapojenie PI regulátora s obvodom blokovania integrátora môžeme vidieť na Obr. 3.

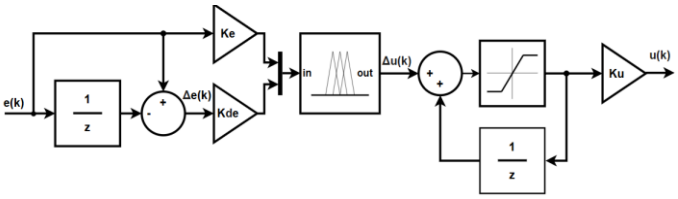


Obr. 3 Bloková schéma PI regulátora

Parametre regulátora boli určené pomocou Ziegler-Nichols metódy a následným expertným doladením. Proporcionálne zosilnenie $K_p=0.0084$ a integračné zosilnenie $K_i=3.9626$.

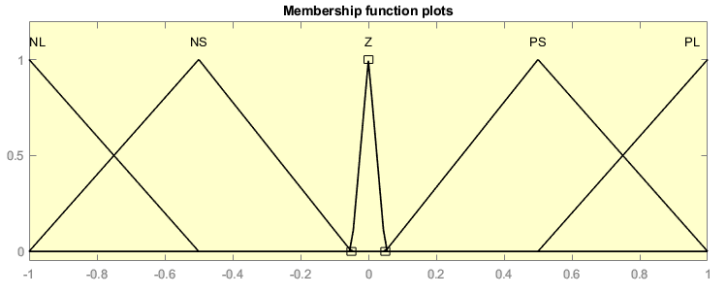
C. Fuzzy regulátor

Principiálne zapojenie Fuzzy regulátora môžeme vidieť na Obr. 4. Vstupom do regulátora je regulačná odchýlka e ale takisto aj zmena regulačnej odchýlky Δe. Zmena sa vypočíta ako rozdiel medzi aktuálnou regulačnou odchýlkou a regulačnou odchýlkou v minulom kroku. Oba vstupy sú normované na rozsah {-1,1}. Pre jednoduchšie ladenie regulátora sme pridali aj normovacie a denormovacie bloky. Pomocou zmeny zosilnení v týchto blokoch ladíme Fuzzy regulátor. Hodnoty zosilnení sú nasledovné: $K_e=0.05$, $K_{\Delta e}=0.3$ a $K_u=3$.



Obr. 4 Bloková schéma Fuzzy regulátora

Pre oba vstupy sme zvolili päť funkcií príslušnosti s trojuholníkovým tvarom Obr. 5. Taký istý tvar a počet bol zvolený aj pre výstup. Zvolený počet a tvar funkcií príslušnosti nám zabezpečil požadované riadenie. S vyšším počtom funkcií príslušnosti, vieme zabezpečiť jemnejšiu a presnejšiu reguláciu. Označenie jednotlivých funkcií príslušnosti: NM – záporná veľká, NS – záporná malá, Z – nulová, PS – kladná malá, PM – kladná veľká. V Tabuľke II sú uvedené pravidlá fuzzy riadenia.



Obr. 5 Rozloženie funkcií príslušnosti vstupu a výstupu Fuzzy regulátora

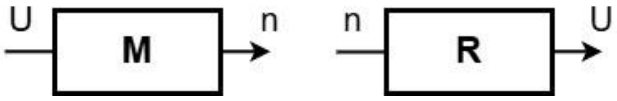
Tabuľka II
Tabuľka fuzzy riadiacich pravidiel

e(k)/ Δe(k)	**NM**	**NS**	**Z**	**PS**	**PM**
NM	NM	NM	NM	NS	Z
NS	NM	NM	NS	Z	PS
Z	NM	NS	Z	PS	PM
PS	NS	Z	PS	PM	PM
PM	Z	PS	PM	PM	PM

D. Neurónový regulátor

Ako neurónový regulátor nám poslúži dopredná sieť s jednou skrytou vrstvou. Použitá neurónová sieť pozostáva vo skrytej vrstve z desiatich neurónov s nelineárnou tansigmoidálnou aktivačnou funkciou. Výstup zo siete je reprezentovaný jedným neurónom s lineárnou aktivačnou funkciou. Neurónový regulátor v tomto prípade predstavuje inverzný model BLDC motora. Principiálna bloková schéma je uvedená na Obr. 6. Pri neurónovom regulátore je vstupnou veličinou požadovaná uhlová rýchlosť a výstupnou veličinou je napájacie napätie.

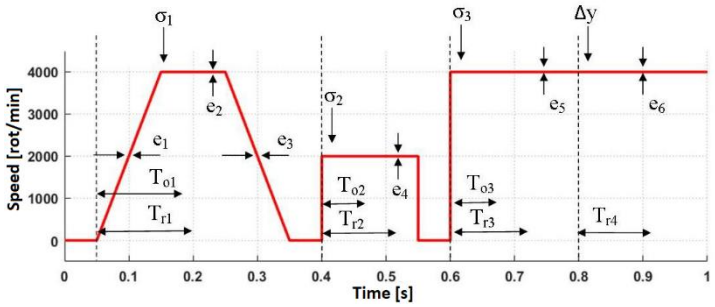
Prvým krokom v procese učenia neurónovej siete je získanie potrebného súboru trénovacich vzorov. Tento súbor vzorov bol získaný simulovaním BLDC motora pri rôznych napájacích napätiach. Dĺžka simulácie bola 20 s. Simulovaním bola získaná vstupno-výstupná množina s 400 000 vzorkami. Z celkového súboru vzorov sa vhodne vybrali určité vzorky, ktoré boli použité na učenie, validáciu a testovanie neurónovej siete. Pomer vybraných vzoriek bol 70-15-15 %. Trénovacie dáta slúžia na trénovanie modelu, na základe nich sa model naučí predpovedať budúce stavy. Validačné dáta slúžia na sledovanie výkonnosti siete a na výber jej najlepšej verzie. Testovacie slúžia na overenie konečnej verzie modelu, čiže na vyhodnotenie presnosti a úspešnosti natrénovaného modelu. Na učenie siete sa použil Levenberg-Marquardtov algoritmus.



Obr. 6 Neurónový regulátor ako inverzný model BLDC motora

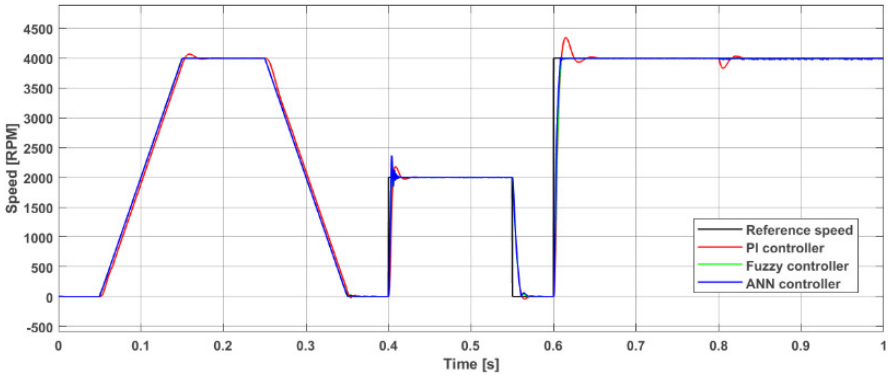
IV. TESTOVANIE REGULÁTOROV BLDC MOTORA

Pre vyhodnotenie kvality jednotlivých regulátorov, bolo zvolených niekoľko typických zmien v žiadanej rýchlosti. Priebeh žiadanej rýchlosti môžeme vidieť na Obr. 7.

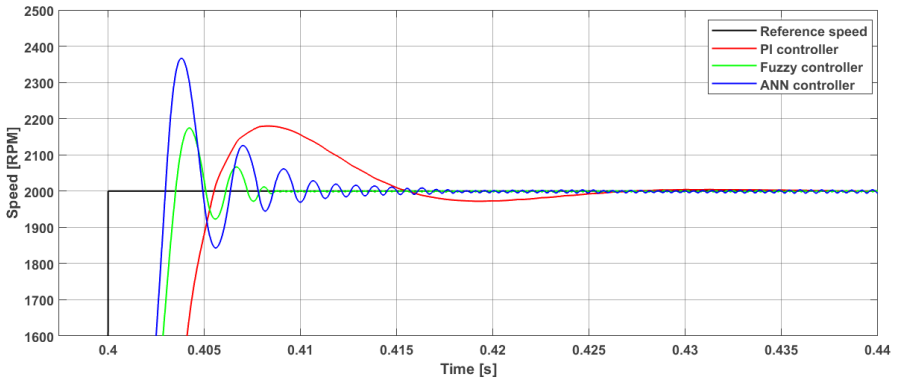


Obr. 7 Priebeh žiadanej rýchlosti pre porovnanie kvalít jednotlivých regulátorov

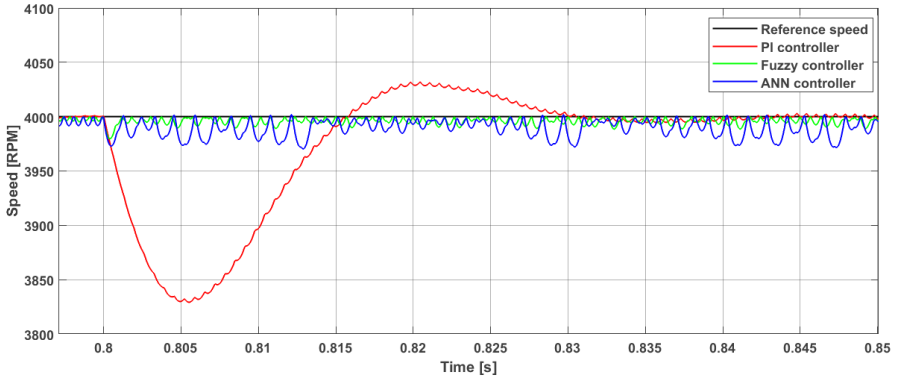
σ_1, σ_2, σ_3 – maximálna hodnota prekmitu aktuálnej rýchlosti po dosiahnutí žiadanej, Δy – podkmit rýchlosti po momentovom zaťažení motora v čase 0.8 s, T_{o1}, T_{o2}, T_{o3} – čas za ktorý skutočná hodnota prvýkrát dosiahne žiadanú, T_{r1}, T_{r2}, T_{r3}, T_{r4} – čas za ktorý sa skutočná hodnota ustáli v rozsahu ±5% žiadanej hodnoty, e_1, e_2, e_3, e_4, e_5, e_6 – veľkosť regulačnej odchýlky v čase 0.1 s, stredná hodnota regulačnej odchýlky v čase od 0.2 do 0.25 s, veľkosť regulačnej odchýlky v čase 0.3s, stredná hodnota regulačnej odchýlky v čase od 0.5 do 0.55 s, stredná hodnota regulačnej odchýlky v čase od 0.7 do 0.8 s a stredná hodnota regulačnej odchýlky v čase od 0.9 do 1 s. Porovnanie odsimulovaných výsledkov je ukázané na nasledujúcich obrázkoch.



Obr. 8 Priebeh žiadanej a skutočnej rýchlosti pri riadení PI, fuzzy a neurónovým regulátorom



Obr. 9 Detail na priebeh rýchlosti po dosiahnutí žiadanej hodnoty 2000 rpm

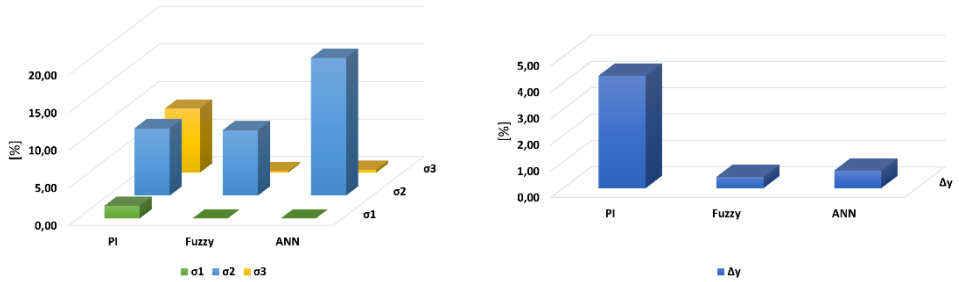


Obr. 10 Detail na priebeh rýchlosti po momentovom zaťažení motora v čase 0.8 s

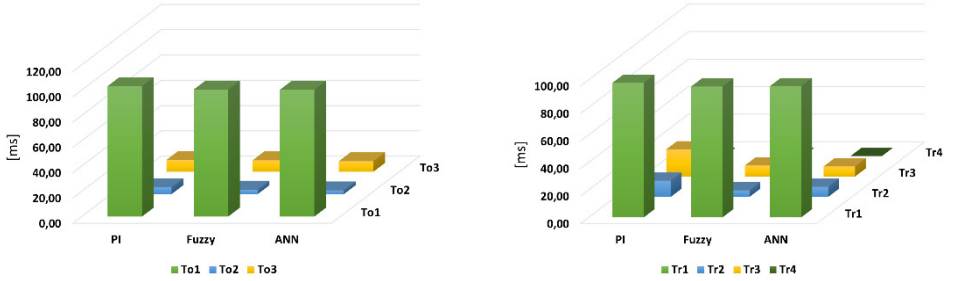
V Tabuľke III sú uvedené odsimulované hodnoty, na základe ktorých sa vyhodnotí kvalita spomínaných regulátorov.

Tabuľka III
Odsimulované hodnoty jednotlivých typov regulátorov

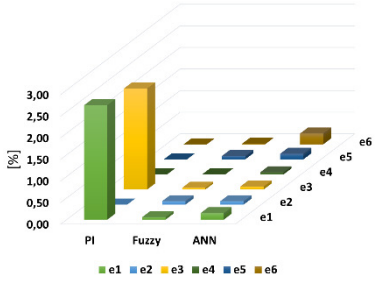
	PI regulátor	Fuzzy regulátor	Neurónový regulátor
σ_1 [%]	1,68	0,00	0,00
σ_2 [%]	8,95	8,70	18,32
σ_3 [%]	8,52	0,20	0,38
Δy [%]	4,27	0,43	0,67
T_{o1} [ms]	103,14	100,45	100,30
T_{o2} [ms]	5,51	3,50	2,99
T_{o3} [ms]	9,40	9,12	8,40
T_{r1} [ms]	97,73	95,04	95,24
T_{r2} [ms]	11,80	4,72	7,36
T_{r3} [ms]	19,70	8,04	7,33
T_{r4} [ms]	0,00	0,00	0,00
e_1 [%]	2,65	0,06	0,15
e_2 [%]	0,00	0,08	0,08
e_3 [%]	2,34	0,04	0,07
e_4 [%]	0,00	0,00	0,05
e_5 [%]	0,00	0,08	0,13
e_6 [%]	0,00	0,01	0,25



Obr. 11 Maximálna hodnota prekmitu aktuálnej rýchlosti po dosiahnutí žiadanej hodnoty (vľavo), percentuálny podkmit po momentovom zaťažení motora (vpravo)



Obr. 12 Porovnanie časov za, ktorých skutočná rýchlosť prvýkrát dosiahne žiadanú hodnotu (vľavo), porovnanie časov za ktorých skutočná rýchlosť sa ustáli v danom pásme žiadanej hodnoty ±5 % (vpravo)



Obr. 13 Porovnanie veľkosti regulačných odchýlok v daných časových pásmach

V. Záver

Na základe parametrov uvedených v Tabuľke II, môžeme povedať, že každý zo simulovaných regulátorov vykazoval lepšie vlastnosti v inej oblasti. Preto nie je jednoznačne možné rozhodnúť, ktorý zo simulovaných regulátorov je najlepší. Testované regulátory splnili základné pravidlá

regulácie, čo znamená, regulovanie bez výraznej regulačnej odchýlky, rýchle dosiahnutie žiadanej hodnoty a dobrú dynamiku vyregulovania vstupujúcej poruchy do systému.

Pre dosiahnutie optimálnych výsledkov regulácie, je potrebné vhodne vybrať najoptimálnejší regulátor, s ohľadom na konkrétnu špecifickú oblasť aplikácie. Takisto okrem samotného výberu typu regulátora je potrebné zvolený regulátor naladiť tak, aby sa zabezpečilo také riadenie, ktorým dosiahneme žiadaný priebeh regulovanej veličiny.

Aj keď PI regulátor dosiahol v porovnaní s ostatnými regulátormi výrazne horšie výsledky v určitých častiach regulácie, jeho celkovú kvalitu, pokiaľ ide o reguláciu rýchlosti BLDC motora, hodnotíme ako veľmi dobrú. Tento typ regulátora nám zabezpečil stabilnú a spoľahlivú reguláciu rýchlosti a ako jediný zo všetkých regulátorov odstránil trvalú regulačnú odchýlku. Implentácia takéhoto regulátora je veľmi jednoduchá a rýchla. Nevýhodou je, že sa jedná o lineárny regulátor, preto ním vieme spoľahlivo regulovať iba lineárne systémy.

Pri návrhu fuzzy regulátora nebolo potrebné poznať presnú štruktúru a podrobné parametre systému. To predstavuje jeho najväčšiu výhodu. Okrem toho tento regulátor dosiahol vo väčšine jednotlivých častiach regulácie najlepšie výsledky. Fuzzy regulátor pri riadení rýchlosti BLDC motora, nám zabezpečil rýchlu a stabilnú reguláciu. V porovnaní s PI regulátorom vykazoval menšiu presnosť v ustálených dejoch.

Vytvorenie a nastavenie fuzzy regulátora bolo najzložitejšie a časovo najnáročnejšie zo všetkých troch typov regulátorov. Na dosiahnutie optimálnej regulácie boli potrebné empirické poznatky resp. odborné znalosti. Nesprávne nastavenie regulátora viedlo k zlým výsledkom a neschopnosti regulátora riadiť daný systém.

Neurónový regulátor vykazoval podobné vlastnosti ako fuzzy, avšak v niektorých častiach regulácie za ním mierne zaostával. Ani pri vytváraní a ladení tohto regulátora sa nevyžadovala presná štruktúra a údaje o regulovanom systéme. Pre dosiahnutie optimálneho naladenia tohto typu regulátora je potrebné získať dostatok vstupných trénovacich údajov. Bez nich by riadenie viedlo k zlým výsledkom a neschopnosti regulátora riadiť systém.

Poďakovanie

Táto práca bola podporená Kultúrno a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky v rámci projektu KEGA 032TUKE-4/2024.

Zdroje

[1] Priddy, L. Kevin; Keller, E. Paul. Artificial Neural Networks – An Introduction. SPIE Bellingham, Washington USA, 2005. ISBN 0-8194-5987-9
[2] G.E. Hinton and J.A. Anderson. Parallel Models of Associative Memory. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1989.
[3] M. Minsky and S. Papert. Perceptrons. An Introduction to Computational Geometry. The MIT Press, Cambridge, MA, 1969.
[4] C.L. Giles and T. Maxwell. Learning, invariance, and generalization in high-order neural networks. Applied Optics 26:4972-4978, 1987.
[5] J. Hertz, A. Krogh, and R.G. Palmer. Introduction to The Theory of Neural Computation. Addison-Wesley Publishing Company, Redwood City, California, 1991.
[6] J.L.Elman and D. Zipser. Learning the hidden structure of speech. Journal of the Acoustical Society of America, 83: 1615-1626, 1988.
[7] M.W. Hirsh. Saturation at High Gain in Discrete Time Recurrent Networks. Neural Networks, 3: 449-453, 1994.
[8] H.-U. Bauer and K.R. Pawelzik. Quantifying the neighborhood preservation of selforganizing feature maps. IEEE Transactions on Neural Networks, 3(4):570-579, 1992.
[9] C. Bouton, M. Cottrell, J.C. Fort, and G. Pagés. Self-organization and convergence of the Kohonen algorithm. In: Probabilités Numériques (Eds. N. Bouleau and D. Talay), INRIA, Paris, France, 163-180, 1991.
[10] P. Vas. Artificial-Intelligence-based electrical machines and drives, Oxford University Press, 1999.
[11] W. Leonhard. Control of Electrical Drives. Springer Berlin Heidelberg, 2001. ISBN 3-540-41820-2
[12] W. Piotr. Dynamics and Control of Electrical Drives. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-20221-6
[13] M. Lešo, J. Žilková, J. and P. Girovský, " Development of a simple fuzzy logic controller for DC-DC converter, " PEMC 2018, 18th IEEE International Conference on Power Electronics and Motion Control, Budapest, Hungary, pp. DOI:10.1109/EPEPEMC.2018.8521896
[14] M. Lešo, J. Žilková, M. Pástor and J. Dudrik, " Fuzzy logic control of soft-switching DC-DC converter," Elektronika ir Elektrotechnika, vol. 22, no 5, pp.3-7, 2016. DOI:10.5755/j01.eie.22.5.16334/WOS 10/
[15] Girovský, P., Žilková, J., Kaňuch, J.: Optimization of Vehicle Braking Distance Using a Fuzzy Controller. Energies, Vol.13,no. 11, pp. 3022, 2020. DOI:10.3390/en13113022
[16] Girovský, P; Timko, J; Žilková J., Fedák, V: Neural Estimators for Shaft Sensorless FOC Control of Induction Motor ,IEEE International Conference on Power Electronics and Motion Control, EPE PEMC 2010, Ohrid, MACEDONIA DOI:10.1109/EPEPEMC.2010.5606907
[17] Balara D., Timko J., Žilková J. Application of neural network model for parameters identification of non-linear dynamic system. Neural Network World. 2013, pp.103-116. DOI: 10.14311/NNW.2013.23.008.
[18] Perduková D., Fedor P., Lacko M.: DC Motor Fuzzy Model Based Optimal Controller. MM Science Journal, Vol. 2021, pp. 4879-4885, 2021. ISSN: 1803-1269. DOI: 10.17973/MMSJ.2021_10_2021033

[19] Perduková D., Fedor P., Bačík J., Herčko J., Rofár J.: Multi-motor drive optimal control using a fuzzy model based approach. Journal on Ambient Intelligence and Smart Environments, Thematic Issue on Intelligent Systems, Applications and Environments for the Industry of the Future, Vol. 9, no. 3, 2017, pp. 329-344. ISSN: 1876-1364. DOI: 10.3233/AIS-17043
[20] Bose B.K.Expert system, fuzzy logic and neural network applications in power electronics and motion control. Proceedings of the IEEE, 1994,82(8), pp. 1303-1323,DOI:10.1109/5.301690.
[21] Narendra, K. S.: Neural networks for control: Theory and practice. IEEE Trans.on Neural Networks, 1996.
[22] J. Kaňuch, P. Girovský.: Motor for Direct Drive of Electric Wheelchair. International journal of engineering research in Africa, Vol. 31, 2017, pp. 94-103. ISSN: 1663-3571. DOI: 10.4028/wwwscientific.net/JERA.31.94
[23] Immaneni, H. Mathematical Modelling And Position Control Of Brushless DC (BLDC). In: International Journal of Engineering Research and Applications (IJERA), 2013, Vol.3, Issue 3, ISSN: 2248-9622.

Určenie parametrov planárneho transformátora z nameraných priebehov

[1]Tomáš BASARIK, [2]Milan LACKO,
[3]Tadeáš KMECIK, [4]Daniel GORDAN

[1,2,3,4]Katedra Elektrotechniky a Mechatroniky, Fakulta Elektrotechniky a Informatiky,
Technická Univerzita v Košiciach, Slovenksá Republika

[1]tomas.basarik@tuke.sk, [2]milan.lacko@tuke.sk,
[3]tadeas.kmecik@tuke.sk, [4]daniel.gordan@tuke.sk

***Abstrakt* – Tento príspevok sa zaoberá určovaním magnetizačnej a rozptylovej indukčnosti planárneho transformátora z nameraných priebehov napätí a prúdov primárnej a sekundárnej strany transformátora. V práci je opísaný postup merania ako aj celý postup výpočtu. Výsledky sú porovnané s nameranými hodnotami oboch indukčností pomocou komerčného RLC metra. Záver práce je venovaný diskusii o dosiahnutých výsledkoch.**

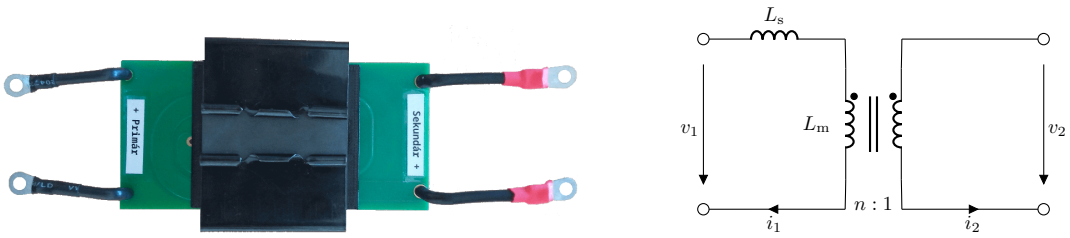
***Kľúčové slová* – planárny transformátor, rozptylová indukčnosť, magnetizačná indukčnosť**

I. Úvod

Transformátor je netočivý elektrický stroj, ktorý premieňa elektrickú energiu znova na elektrickú. Nachádza uplatnenie takmer vo všetkých odvetviach elektrotechniky, a to od elektroenergetiky a rozvodných sietí, až po najmenšiu spotrebnú elektroniku. Vo výkonovej elektronike čoraz častejšie nachádza uplatnenie takzvaný planárny transformátor. Tento transformátor je možné nájsť v nepriamych blokujúcich meničoch [1], jednosmerných meničoch [2], polomostikových meničoch [3], mostikových meničoch [4], rezonančných meničoch [5], v meničoch v uzlovom zapojení [6], izolovaných DC-DC meničoch [7] a ďalších. Jedná sa o transformátor, ktorý namiesto drôtových vinutí používa tenké a ploché vinutie, ktoré je často vyrobené na doske plošného spoja. Pri tejto konštrukcií sa dosahujú nízke profily transformátorov. Takéto transformátory dokážu pracovať s vysokou frekvenciou, čo taktiež znižuje počet závitov vinutí a tým aj komplexnosť návrhu a rozmery. Kvôli plochým vinutiam majú lepšie tepelné vlastnosti a efektívnejšie odvádzajú teplo. Tento príspevok je zameraný na určenie hlavných parametrov transformátora, a to magnetizačnej indukčnosti L_{m} a rozptylovej indukčnosti L_{s}, ktoré sú kritické pre návrh meničov a regulácii.

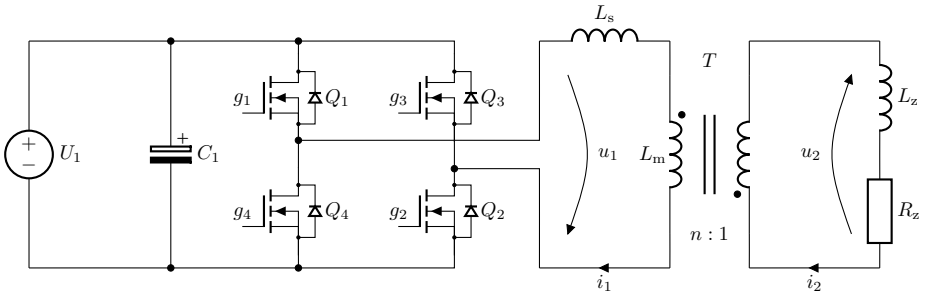
II. Meranie transformátora

Na Obr. 1 (vľavo) je zobrazený planárny transformátor. Na Obr. 1 (vpravo) je zobrazená náhradná schéma transformátora, ktorá sa používa pri simuláciách výkonových meničov. Táto schéma zanedbáva odpory vinutí na primárnej a sekundárnej strane, a taktiež straty v jadre transformátora. Tieto hodnoty sú zanedbané, nakoľko dosahujú veľmi malé hodnoty. Indukčnosť L_{s} predstavuje rozptylovú indukčnosť a indukčnosť L_{m} predstavuje magnetizačnú indukčnosť v náhradnej schéme transformátora.



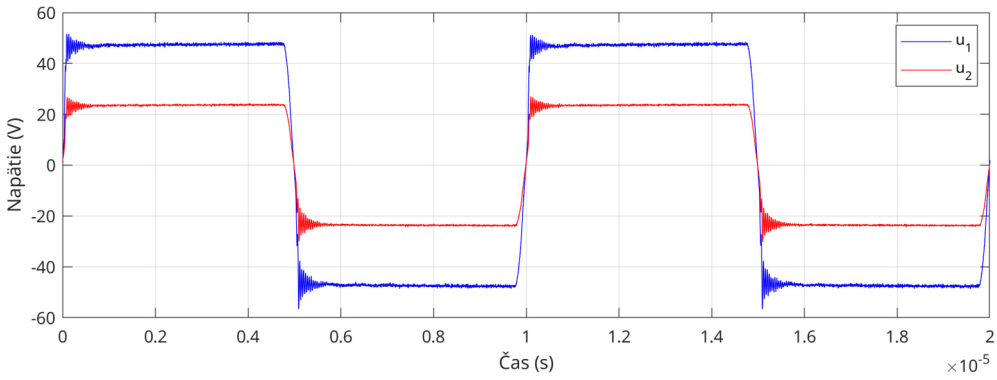
Obr. 1 Planárny transformátor (vľavo), náhradná schéma transformátora (vpravo)

Na Obr. 2 je zobrazená schéma, pri ktorej bol transformátor meraný. Primárna strana transformátora bola pripojená na jednofázový striedač, ktorý bol napájaný jednosmerným napätím. Na sekundárnu stranu bola pripojená RL záťaž v podobe výkonového odporu.



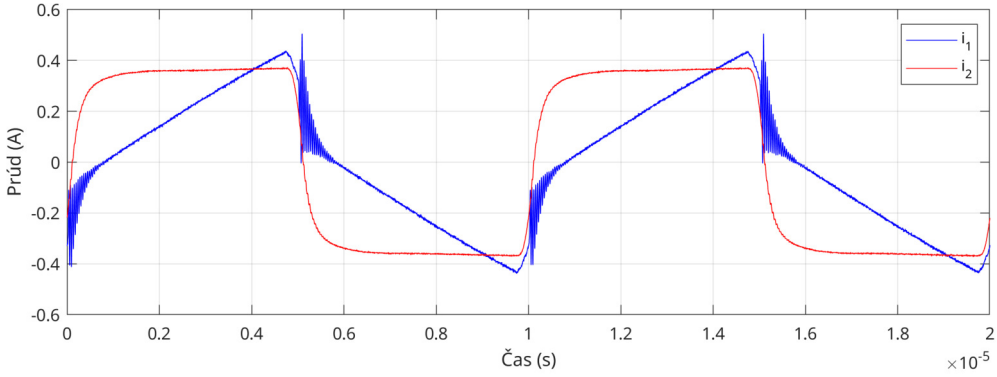
Obr. 2 Schéma merania planárneho transformátora s RL záťažou

Na Obr. 3 sú zobrazené priebehy napätí na primárnej a sekundárnej strane. Primárne napätie u_1 je označené modrou farbou a sekundárne napätie u_2 červenou farbou. Primárne napätie je dané veľkosťou napätia jednosmerného zdroja a striedačom, ktorý strieda napätie s 50 % šírkou impulzu. Nakoľko napätie na sekundárnej strane je menšie ako na primárnej, je možné usúdiť, že transformátor je znižovací.



Obr. 3 Priebehy primárneho a sekundátrneho napätia u_1 a u_2

Na Obr. 4 sú zobrazené priebehy prúdov na primárnej a sekundárnej strane planárneho transformátora. Veľkosť prúdu na sekundárnej strane, znázornená červenou farbou, je závislá od veľkosti záťaže a rozptylovej indukčnosti L_s. Na primárnej strane tečie okrem prepočítaného záťažového prúdu aj magnetizačný prúd, ktorý je daný magnetizačnou indukčnosťou L_m. Magnetizačný prúd má tvar lineárneho nárastu. Keďže je primárny prúd značne zašumený vplyvom spínania, je lepšie ho pozorovať na sekundárnej strane.



Obr. 4 Priebehy primárneho a sekundátrneho prúdu i_1 a i_2

III. Určenie indukčností transformátora

Nasledujúca časť je venovaná určeniu indukčností planárneho transformátora z nameraných priebehov. V Tabuľke 1 sú uvedené namerané alebo určené hodnoty z grafov. Tieto hodnoty sú

potrebné pre nasledujúce výpočty.

Tabuľka 1
Hodnoty potrebné pre výpočty

Parameter	Označenie	Hodnota	Jednotka
Čas potrebný pre integrál a diferenciu	t_1	1.000	µs
Čas potrebný pre integrál a diferenciu	t_2	4.000	µs
Prúd i_1 v čase t_1	$i_1(t_1)$	27.30	mA
Prúd i_1 v čase t_2	$i_1(t_2)$	365.0	mA
Časová konštanta zotrvačnosti prvého rádu	τ	204.0	ns
Indukčnosť záťaže	L_z	12.70	µH
Odpor záťaže	R_z	68.38	Ω

A. Prevod transformátora

Prevod transformátora je možné určiť z podielu maximálnych hodnôt primárneho napätia u_1 a sekundárneho napätia u_2. Z priebehov zobrazených na Obr. 3 je zrejmé, že maximálne hodnoty napätia u_1 a u_2 je možné odčítať približne v čase od t_1 do t_2. Pre výpočet je vhodné použiť vzorec pre strednú hodnotu nasledovne:

$$u_{1\max} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} v_1(t)dt = 47.3844\,V \tag{1}$$

$$u_{2\max} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} u_2(t)dt = 23.6099\,V \tag{2}$$

Prevod transformátora je potom daný nasledovnou rovnicou:

$$n = \frac{u_{1\max}}{u_{2\max}} = 2.007 \tag{3}$$

B. Výpočet magnetizačnej indukčnosti

Na primárnej strane transformátora tečie okrem primárneho prúdu i_1 aj magnetizačný prúd i_m. Magnetizačná indukčnosť L_m je pripojená paralelne k zdroju napätia u_1. Pretože magnetizačná indukčnosť transformátora je omnoho väčšia ako rozptylová indukčnosť, jej nárast prúdu bude pomalší a pri vysokých frekvenciách je možné tento priebeh linearizovať. Pri výpočte budeme uvažovať s časovou odozvou indukčnosti, ktorá vraví, že napätie $u(t)$ na indukčnosti je úmerné indukčnosti L a časovej zmene prúdu [8]:

$$u(t) = L\frac{di(t)}{dt} \tag{4}$$

Z praktického hľadiska to znamená, že pri pripojení konštantného napätia na indukčnosť bude prúd lineárne rásť. Nárast prúdu bude závisieť od veľkosti indukčnosti. Ak priebeh nárastu prúdu je linearizovaný, potom je možné deriváciu nahradiť diferenciou, vďaka čomu sa výpočet zjednoduší. Z rovnice (4) je potom možné odvodiť vzťah pre výpočet magnetizačnej indukčnosti:

$$L_\mathrm{m} = U_{1\max}\frac{t_2 - t_1}{i_1(t_2) - i_1(t_1)} = 421.07\,\mu H \tag{5}$$

Úbytok napätia na rozptylovej indukčnosti L_s je možné zanedbať, nakoľko jej hodnota je malá a teda aj úbytok bude malý.

C. Výpočet rozptylovej indukčnosti

Na sekundárnej strane transformátora tečie iba sekundárny prúd i_2, ktorý je v sérii so zdrojom napätia U_1 a RL záťažou. Nárast sekundárneho prúdu i_2 má tvar odozvy zotrvačnosti prvého rádu. Časová konštanta zotrvačnosti prvého rádu v sériovom zapojení indukčnosti a odporu je daná nasledovnou rovnicou [8]:

$$\tau = \frac{L}{R} \tag{6}$$

Časová konštanta τ predstavuje čas, za ktorý sekundárny prúd i_2 dosiahne hodnotu 63,2 % z ustálenej hodnoty. V sérii sa okrem rozptylovej indukčnosti L_s a odporovej záťaže R_z nachádza aj indukčnosť záťaže L_z. Indukčnosť záťaže, taktiež ovplyvňuje nárast prúdu a teda aj časovú

konštantu. Preto je potrebné po uskutočnení výpočtu odčítať hodnotu záťažovej indukčnosti z výsledku. Nakoľko výpočet je realizovaný na sekundárnej strane transformátora, hodnotu rozptylovej indukčnosti je potrebné prepočítať na primárnu stranu cez kvadrát prevodu transformátora. Výsledný vzorec pre výpočet vyzerá nasledovne:

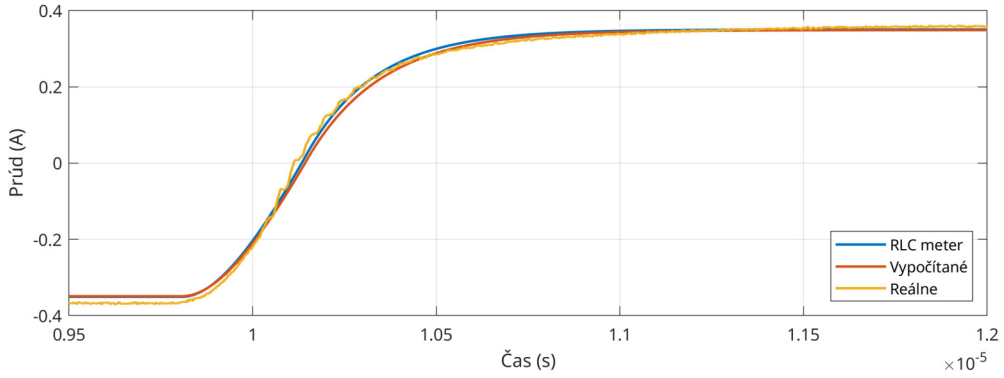
$$L_{\mathrm{s}} = (R_{\mathrm{z}} \cdot \tau - L_{\mathrm{z}}) \cdot n^2 = 5.0325\,\mu H \tag{7}$$

IV. Porovnanie výsledkov a diskusia

V Tabuľke 2 sú uvedené hodnoty indukčností planárneho transformátora, ktoré boli namerané pomocou RLC metra a taktiež vypočítané hodnoty, ktoré boli určené z nameraných priebehov napätia a prúdu. Určenie rozptylovej indukčnosti L_{s} transformátora pomocou RLC metra bolo vykonané pri skratovanej sekundárnej strane, v tzv. stave nakrátko. Určenie magnetizačnej indukčnosti L_{m} transformátora pomocou RLC metra bolo vykonané pri rozpojenej sekundárnej strane, v tzv. stave naprázdno. Ako vidieť v tabuľke, hodnoty magnetizačnej indukčnosti sú podobné aj keď s menším rozdielom. Rozdiel môže byť zapríčinený tým, že RLC meter meria indukčnosť pri napájacom napätí $0.6\,\mathrm{V}$ a prúde $2\,\mathrm{mA}$. Magnetický obvod preto nebol ideálne sýtený. Prúdová sonda osciloskopu má taktiež danú presnosť, ktorá sa mohla podujať na rozdiele výsledkov. Rozdiel hodnôt rozptylovej indukčnosti L_{s} bol značne vyšší. Pri meraniach týchto hodnôt bol rozdiel v tom, že pri meraní RLC metrom bol transformátor odpojený od striedača a celkovo od celého obvodu meniča. Pri meraní s RL záťažou bol menič pripojený a tým pádom pôsobili na transformátor parazitné odpory a indukčnosti vodivých ciest DPS a vodičov. Nárast a pokles napätia pri spínaní tranzistorov nebol skokový ale trval istý čas. To ako aj parazitné prvky obvodu meniča mohli spôsobiť skreslenie časovej konštanty RL obvodu. Každopádne pri návrhu regulácie je potrebné poznať reálne hodnoty transformátora v zapojenom meniči a jeho odozvy. Na Obr. 5 je zobrazené porovnanie simulovaných prúdov a nameraného prúdu. Výsledné odozvy prúdov sú si veľmi podobné, čo znamená, že pre návrh regulácie to postačuje.

Tabuľka 2
Namerané a vypočítané parametre transformátora

Parameter	Označenie	Vypočítané	RLC meter	Jednotka
Rozptylová indukčnosť	L_{s}	5.0325	0.971	µH
Magnetizačná indukčnosť	L_{m}	421.07	347.6	µH



Obr. 5 Porovnanie simulovaných prúdov a nameraného prúdu

V. Záver

V tomto príspevku je uvedený postup pre určenie rozptylovej a magnetizačnej indukčnosti planárneho transformátora na základe nameraných priebehov napätí a prúdov. Parametre transformátora sa podarilo určiť. Meranie parametrov sa vykonalo aj pomocou komerčného RLC metra. Vypočítané a namerané hodnoty sú porovnané v poslednej kapitole tohoto príspevku. Medzi hodnotami vznikli malé rozdiely, ktoré sú taktiež bližšie opísané v poslednej kapitole. Postup určenia rozptylovej a magnetizačnej indukčnosti planárneho transformátora na základe nameraných priebehov sa teda javí ako efektívny a je možné ho použiť v praxi pri návrhu regulácií výkonových meničov, ktoré obsahujú planárny transformátor.

POĎAKOVANIE

Tento príspevok vznikol s podporou Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied v rámci projektu VEGA 1/0363/23. Príspevok bol taktiež podporovaný pomocou Agentúry na podporu výskumu a vývoja v rámci projektu APVV-18-0436.

LITERATÚRA

[1] M. Ursino, S. Saggini, R. Specogna, A. Bianco, F. Ciappa, and G. Scappatura, "A novel mixed planar litz transformer for high frequency active clamp flyback converters," in *2020 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2020, pp. 1327–1332.

[2] X. Zhang, A. Ferencz, T. Takken, B. Nguyen, and P. Coteus, "A 12v-to-0.9v active-clamp forward converter power block with planar transformer, standing slab inductor and direct edge solder to motherboard," in *2017 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2017, pp. 237–240.

[3] B. Li, X. Huang, and J. Zhang, "A high density 400 w dc/dc power module with integrated planar transformer and half bridge gan ic," in *2024 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2024, pp. 94–100.

[4] D.-W. Lee, H.-S. Youn, and J.-K. Kim, "Development of phase-shift full-bridge converter with integrated winding planar two-transformer for ldc," *IEEE Transactions on Transportation Electrification*, vol. 9, no. 1, pp. 1215–1226, 2023.

[5] S. Ann, W.-J. Son, J. H. Lee, J. Byun, and B. K. Lee, "Design of a sic-based llc resonant converter for on-board chargers considering parasitic capacitance of planar transformer," in *2020 23rd International Conference on Electrical Machines and Systems (ICEMS)*, 2020, pp. 315–319.

[6] F. Pereira, A. Ferreira, V. Sobrado, A. Martins, and A. Carvalho, "Push-pull dc-dc converter with planar transformer for pv applications," in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, 2013, pp. 1180–1184.

[7] A. Marcinek, Ž. Ferková, and M. Pástor, "Design and simulation verification of planar transformer for multiport power converters," in *2022 IEEE 20th International Power Electronics and Motion Control Conference (PEMC)*, 2022, pp. 70–74.

[8] N. S. Nise, *Control Systems Engineering*. John Wiley & Sons, Inc., 2011.

The Technological Aspects of Flexible Solar Cells

[1]*Barbora VISKI,* [2]*Peter LUKACS*

[1,2] Department of Technologies in Electronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]barbora.viski@student.tuke.sk, [2]peter.lukacs@tuke.sk

***Abstract* — This paper presents an extensive review of the technological advancements in flexible solar cells, with particular focus on a variety of thin-film photovoltaic technologies. Edmond Becquerel discovered the photovoltaic effect in 1839, and Bell Laboratories made significant progress in solar cell efficiency by developing silicon-based cells in the 1950s. The necessity for energy solutions that were portable, lightweight, and easily manufactured led to the pursuit of flexible solar cells. The study analyzes the structures, manufacturing processes, efficiency rates, and advantages of various thin-film solar cell types, such as amorphous silicon, CIGS, CdTe, perovskite, and organic materials. The work focuses on the ecological and economic concerns that are associated with flexible solar cell technologies. We have identified current challenges and prospective research directions, as well as the potential of flexible solar cells to revolutionize energy accessibility and sustainability, through this review.**

***Keywords* — photovoltaic cells, solar cells, solar energy, perovskites, organic cells**

I. Introduction

The photovoltaic effect was discovered in 1839 by Edmond Becquerel, who noted that certain materials generate voltage when exposed to sunlight. The first solar cell was created in1883 by Charles Fritts, by coating selenium with a thin layer of gold, achieving an energy conversion rate of 1-2%.

In the 1950s, the researchers at Bell Laboratories have discovered that silicon is more efficient than selenium. This led to creating the first practical solar cell that was 6% efficient. Satellite Vanguard 1, the first satellite to use solar energy, was launched in 1958. This paved the way for research to reduce the costs of solar energy.

Over the years, solar panels have gained popularity, and now they are presented as a greener and carbon-neutral way of generating electricity. The research has led not only towards more efficient photovoltaic cells but also towards more flexible and portable ones [1].

A. Short Theoretical Background

The photovoltaic effect in solar panels happens because of two layers of semiconductors; p-type and n-type sandwiched together to form a p-n junction.

When solar radiation hits a solar cell, some of the photons are absorbed by the cell which results in electron-hole pair production in the cell. With forming of an external circuit, this difference drives electrons from the n-side to the p-side of the junction producing electric current [2].

The quality and efficiency of solar cells are calculated by calculating the Fill factor and solar cell efficiency. Fill factor measures quality of solar cell and is defined as the ratio of maximum power point P_{max} and open-circuit voltage a short circuit current. This relation is described in formula (1).

$$FF = \frac{P_{max}}{V_{OC} * I_{SC}} \tag{1}$$

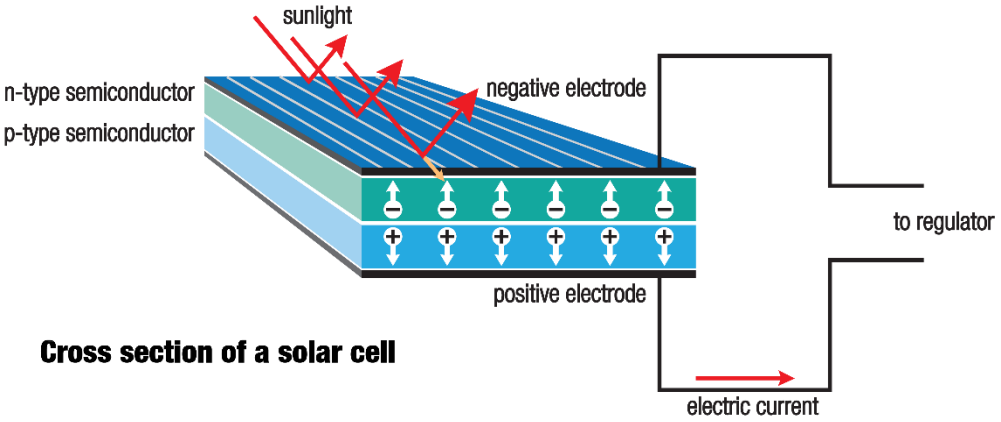


Fig. 1 Description of the photovoltaic effect in photovoltaic cell

IV curve describes combinations of current and voltage output in photovoltaic devices. The photovoltaic device produces maximum current when there is a short circuit between positive and negative terminals, this current is abbreviated as I_{SC}. Maximum voltage occurs on break in the circuit, it's called open-circuit voltage V_{OC}. At V_{OC}, resistance is infinitely high, and current is zero because of an incomplete circuit [2], [3].

Solar cell efficiency (power conversion efficiency, PCE) is a ratio, where electrical power output at the maximum point of the IV curve is divided by the incident light power.

$$\eta = \frac{V_{OC} * I_{SC} * FF}{P_{inc}} \tag{2}$$

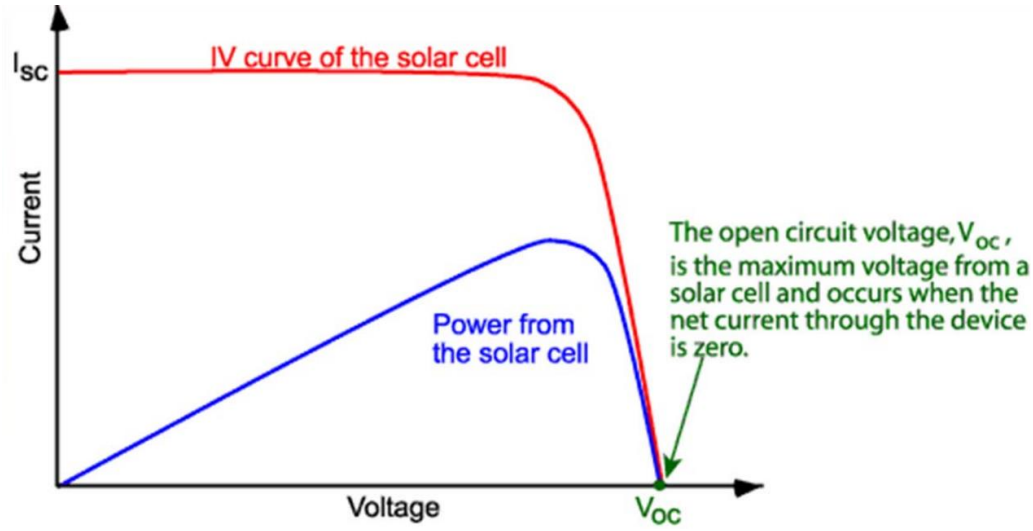


Fig. 2 IV curve of solar cell [3]

II. Thin Film Photovoltaic Cells

The flexibility of photovoltaic cells is achieved primarily by depositing thin-film layers on flexible materials, and while efficiency is better in conventional, rigid solar panels, their flexible thin-film counterparts are becoming prominent. A big advantage of thin-film solar panels is beside their flexibility, portability, and lightweight, they are also easier to manufacture, as illustrated in Fig. 3.

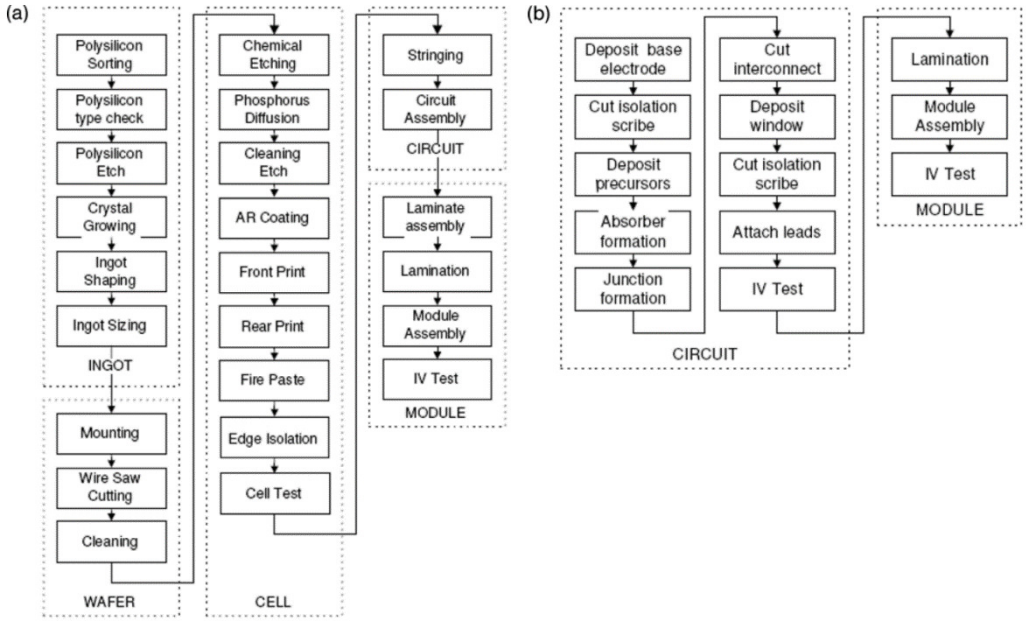


Fig. 3 Process sequence for manufacturing of crystalline silicon modules (a) and thin-film modules (b) [4]

There are various thin-film technologies, some examples can be seen listed below:

- Amorphous Silicon thin film
- CIGS thin film on metal foil
- CdTe thin film
- Perovskite thin film
- Organic thin film

In Fig. 4 we can see the growth in efficiency for various solar cell technologies. CIGS, CdTe, a-Si, perovskite, and organic solar panels are also shown on the graph and their comparison to other solar cell technologies [4].

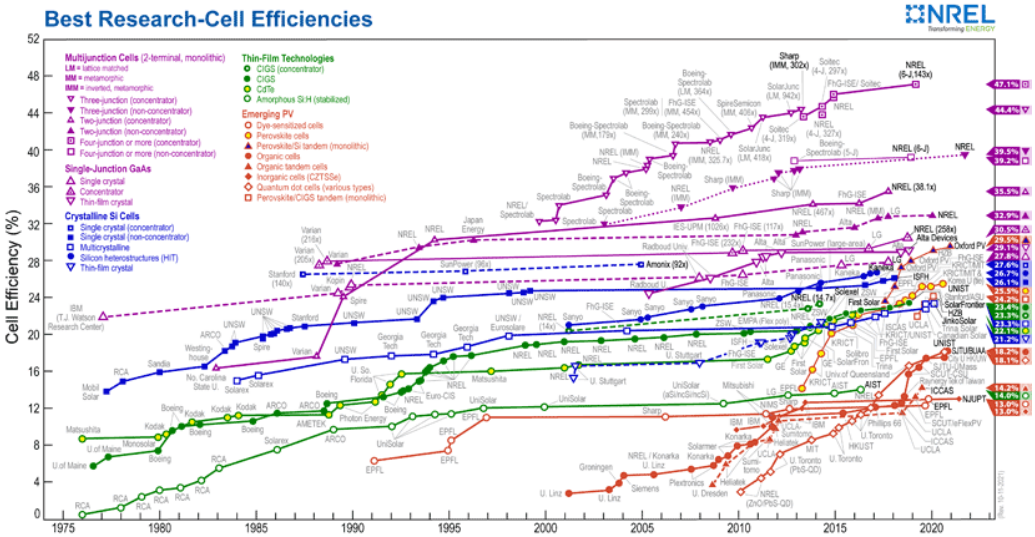


Fig. 4 Comparison of different solar cells by type [6]

A. Amorphous Silicon Thin Films

Word amorphous means lacking a crystalline structure. Which is a difference between amorphous silicon (a-Si) thin-film solar cells and conventional silicon-based solar cells, that are crystallized.

An important step in the development of this type of cell was the introduction of triple junction modules (Fig. 5). Those were able to provide higher efficiency and stability. The triple-junction cell stacks cells of different bandgaps together. These layers separate photons of different optical gaps, which allows capturing a greater percentage of incident light energy, improving efficiency and energy output under diffused light.

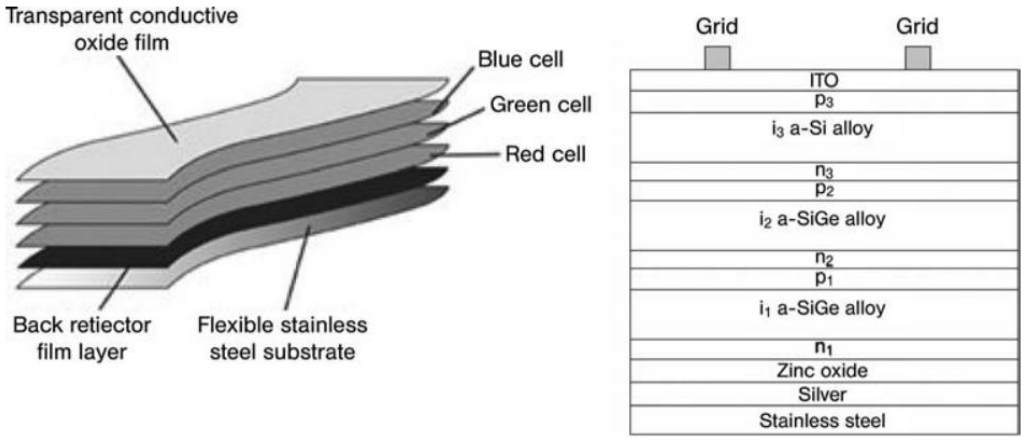


Fig. 5 Triple-junction structure in the a-Si solar cell [4]

Advantages of a-Si compared to other technologies are, that they use less silicone and making substrate for them is inexpensive.

When it comes to manufacturing processes, Swiss company Flexcell is using a four-step manufacturing process. In this process, a metal coating of the plastic roll is followed by reposition of chemical vapors of a-Si layer and TCO (transparent conductive oxide) through structuring layers, and lastly encapsulating modules with plastic foils [4].

According to NREL, the best research-cell efficiency for the a-Si solar cells was 14% in 2016 [6].

B. CIGS Thin Films on metal foil

CIGS or Copper Indium Gallium Diselenide with chemical formula $CuIn_{(1-x)}Ga_{(x)}Se_2$ are solar cells with the highest performance among thin-film solar cells, with the best CIGS cell having efficiency comparable to polycrystalline-silicon cells.

Big advantages of these cells are their resistance to high temperatures, good response to low light, and longevity. Modules that contain these cells are very stable, and their performance improves during the operation due to self-healing mechanisms caused by defect relaxation.

Initially, a disadvantage of these cells was the complexity of the materials as well as the manufacturing process. This process was simplified by the development of nanostructured CIGS ink by the company Nanosolar. As the CIGS ink was homogenously mixed, the atomic ratios were preserved after the ink was printed even across large areas. The development of this ink has industrialized roll-to-roll printing by minimizing the use of expensive and ineffective high vacuum manufacturing equipment [4].

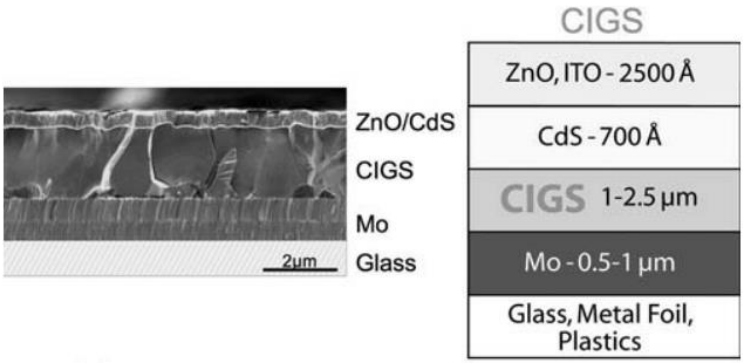


Fig. 6 Structure of CIGS solar cell [4]

Reading the cell efficiency graph from Fig. 4, CIGS has achieved the best efficiency at 23.4% in 2018 [6].

C. CdTe Thin Films

Cadmium Telluride (CdTe) solar cells are capable to convert solar energy into electricity at efficiency rates that are comparable to silicon-based technologies while requiring fewer materials.

CdTe has the advantage of being less susceptible to cell temperature increases, which allows them to produce more electricity under higher temperatures compared to traditional solar cells. These kinds of solar cells are also converting low and diffuse light better, making them more efficient under cloudy weather, at dawn or dusk. Due to this, the CdTe solar modules will produce more electricity compared to other solar modules with similar power ratings [4].

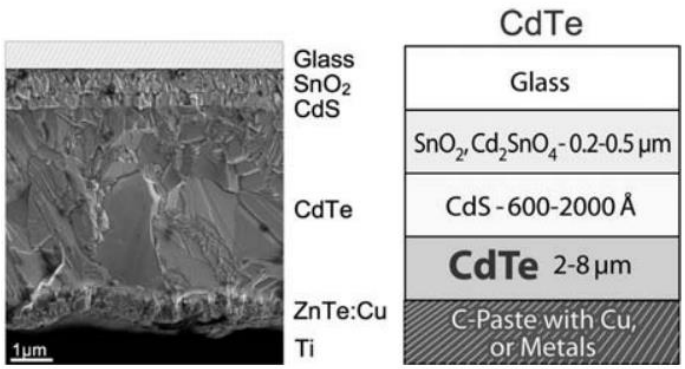


Fig. 7 Structure of CdTe solar cell [4]

The most common deposition method for the CdTe devices involves acquiring SnO_2-coated glass, that has deposited CdS using the chemical bath. The CdTe absorber is then applied by close-spaced sublimation, vapor-transport deposition, or electrodeposition. This process is finalized by $CdCl_2$ treatment [4].

The best research cell efficiency for the CdTe solar cells was 22.1% according to NREL [6].

D. Perovskite Thin Films

Perovskites are organic-inorganic hybrid with structure ABX_3, where A is an organic cation, B is a divalent cation and X is a halide anion.

Perovskite solar cells are gaining research attention due to their advantages like high electron and hole mobility, bandgap tunability, the low binding energy of exciton, and long carrier diffusion length. They can also be produced using low-cost processes and materials.

As this is a very new technology, there are drawbacks to it. The cells have very poor stability while also having very high toxicity, which prompted research into eliminating these issues, namely the stability.

If the issues with perovskite solar cells were solved, this technology has the potential to become one of the highest efficiency and lowest cost photovoltaic technology.

Two technologies have roots in perovskites, perovskite/Si and perovskite/CIGS. These new technologies have good efficiencies, perovskite/Si having around 29.5% and perovskite/CIGS having around 24.2% [8]. Perovskites have achieved recently research-cell efficiency of 25.5%, based on data from NREL [6].

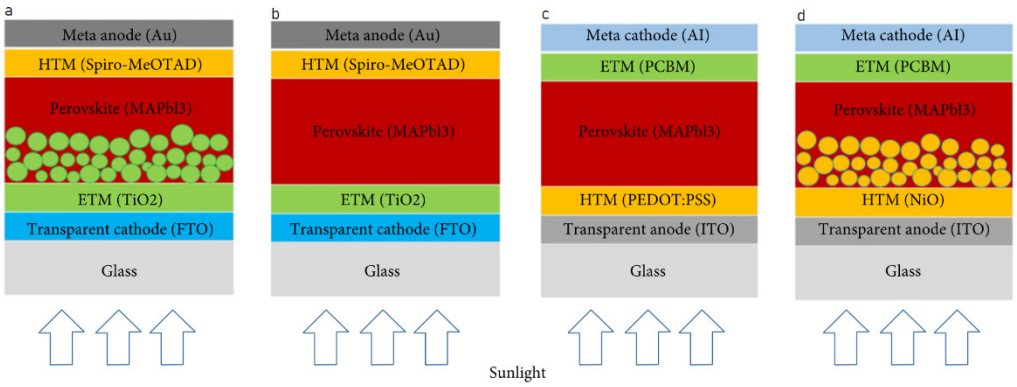


Fig. 8 Cell structure of perovskites in n-i-p mesoscopic (a), n-i-p planar (b), p-i-n planar (c) and p-i-n mesoscopic (d) structures

E. Organic Thin Films

Organic solar cells use organic molecules printed in a thin layer onto a substrate, compared to crystalline silicon, with the semiconducting layer being what makes the most notable difference between the two cell types.

One of the greatest disadvantages is poor reproductivity, making manufacturing them less scalable. Other disadvantages are longevity issues, as being made from organic materials makes them prone to degradation.

The great advantages of organic solar cells are light, environmentally friendly, and have low material and manufacturing costs. Another advantage of organic solar cells is, that we can tailor

the molecule properties accordingly to the application [7], [8].

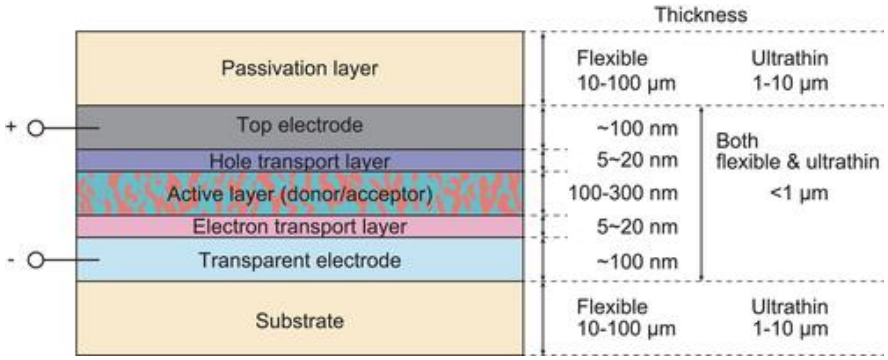


Fig. 9 Structure of organic solar cells [7]

When it comes to manufacturing processes, organic solar cells were made by spin-coating or thermal evaporations. But in modern laboratories, the organic solar cells are being printed onto the substrate via inkjet printing [5].

The research-cell efficiency of organic solar cells is 18.2% according to NREL [6].

III. Ecological and economic concerns

Unlike other types of energy plants, solar energy doesn't release contaminants under normal conditions. The ecological burden mostly lies in manufacturing these panels and their recycling. Another ecological problem comes in a form of batteries, which are mostly made from toxic materials.

There are however ecological and economical concerns when it comes to flexible solar cells as well. Cadmium (CdTe) is a toxic heavy metal and indium (CIGS) is in short supply. Additionally, the prices for materials like In, Se, and Te are rising steadily with their demand [4].

While manufacturing solar cells can be cheap or expensive, depending on the type of cell, access to solar power has changed lives for many people who would otherwise not have access to electricity.

IV. Conclusion

Perovskite solar cells have promising efficiency, and they are rising as a possible solar technology of the future, once their stability issues will be resolved.

Organic solar cells, on the other hand, could have great uses in powering wearable electronics, which could give them a place in the medical and fashion field equally.

Future of the solar energy is bright. Technologies are constantly improving and beating the previous generations and iterations.

Acknowledgment

This paper was supported by project KEGA no. 011TUKE-4/2023, project no. APVV 22-0289 and projects VEGA no. 1/0049/24 and no. 1/0278/24, and by the project ITMS kód 313011T557 („Podpora výskumno-vývojového potenciálu v oblasti dopravných prostriedkov").

References

[1] Solar Power Authority Staff, "A History of Solar Cells: How Technology Has Evolved," [Online]. Available: https://www.solarpowerauthority.com/a-history-of-solar-cells/
[2] M. Pagliaro, G. Palmisano, R. Ciriminna, "Photovoltaics", Flexible Solar Cells, WILEY-VCH, 2008, pp 31-53
[3] R. J. Ellingson, "Fundamental Properties Of Solar Cells," [Online]. Available: http://astro1.panet.utoledo.edu/~relling2/teach/archives/6980.4400.2012/20120131_PHYS_6980_4400_FundamentalProperties.pdf
[4] M. Pagliaro, G. Palmisano, R. Ciriminna, "Inorganic Thin Films", Flexible Solar Cells, WILEY-VCH, 2008, pp 55-81
[5] M. Pagliaro, G. Palmisano, R. Ciriminna, "Organic Thin Films", Flexible Solar Cells, WILEY-VCH, 2008, pp 85-105
[6] NREL, "Best Research-Cell Efficiency Chart," [Online]. Available: https://www.nrel.gov/pv/cell-efficiency.html
[7] Fukuda, K., Yu, K., Someya, T., The Future of Flexible Organic Solar Cells. Adv. Energy Mater. 2020, 10, 2000765. https://doi.org/10.1002/aenm.202000765
[8] M. Giannouli, "Current Status of Emerging PV Technologies: A Comparative Study of Dye-Sensitized, Organic, and Perovskite Solar Cells", International Journal of Photoenergy, vol. 2021, Article ID 6692858, 2021. https://www.hindawi.com/journals/ijp/2021/6692858/

Image generation using diffusion deep models

[1]*Yana KRASNOVA*, [2]*Marián MACH*

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

[1]yana.krasnova@student.tuke.sk, [2]marian.mach@tuke.sk

Abstract **– The focus of the paper is generation of images using diffusion deep models. Particularly, the focus is on two models - Denoising Diffusion Probabilistic Model and Denoising Diffusion Implicit Model. The paper concentrates on training these models as well as on their fine-tuning. Experiments with unconditional image generation using these two models were designed and conducted on two datasets using Hugging Face platform.**

Keywords **– deep learning, diffusion models, generative models, image generation**

I. Introduction

The field of image generation has witnessed a significant evolution over the past years, driven by advancements in computational capabilities, data availability, and breakthroughs in machine learning methodologies.

Generative models aim to create new data instances that resemble the training data. These models can be broadly classified into two main categories [1]:

- Implicit models learn a function that can transform a simple noise distribution into the target data distribution. This transformation process is learned through a neural network, which can approximate complex functions. Examples of implicit models are Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs).
- Explicit models directly learn the probability distribution of the training data. Once learned, new data instances can be generated by sampling from this distribution. Examples of explicit models are Gaussian Mixture Model (GMM) and Diffusion Models (DMs).

II. Diffusion deep models

A. Principle of diffusion models

Diffusion models are named for their similarity to the process of diffusion, since these models describe how data can be transformed into a predefined noise distribution (usually Gaussian) through a sequence of small diffusion steps [2].

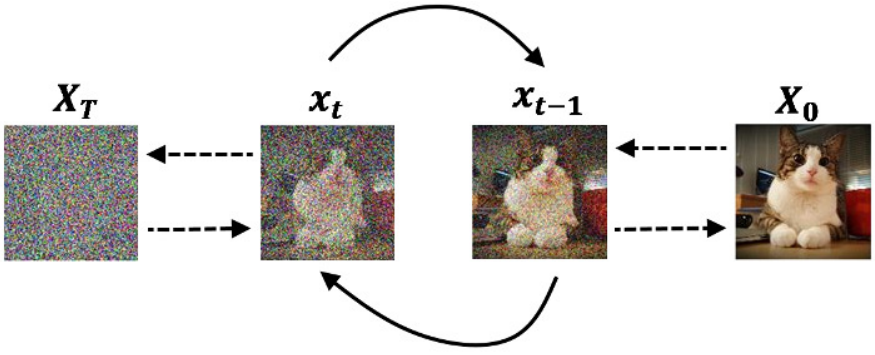


Fig. 1 Forward and backward process of diffusion models.

The diffusion process is illustrated in Fig. 1, where rightmost image represents real (or generated) image and leftmost image represents complete noise.

The fundamental idea behind diffusion models is to reverse the diffusion process to generate new data. In a typical setting, the diffusion process starts with the real data and gradually adds

noise until reaching a noise distribution after a number of steps. This forward process can be thought of as 'destroying' the data.

To generate new image, reverse process is used. The model attempts to reverse this process, starting from the noise and step by step, recovering the original data. Achieving it is possible by training a neural network, often a CNN or a variant of it, to predict the reverse of each diffusion step given the data from the previous step.

Each diffusion step is small, which makes the reverse prediction task easier for the network. After training, the network can generate new data by starting with samples from the noise distribution and applying the reversed diffusion steps. The result is a sample from the model's learned data distribution. The generation process is typically slower than other deep generative models due to the sequential nature of the diffusion steps [3].

III. Types of diffusion models

Diffusion models are a versatile family of generative models, and several variations have been proposed, each with its own unique features and advantages. The presented paper focuses on two diffusion model types:

Denoising Diffusion Probabilistic Models (DDPM): Introduced by [2], DDPMs are among the most influential in the diffusion model family. DDPMs approach data generation as a reverse diffusion process, where noise is gradually removed from a sample from a simple noise distribution until a sample from the target distribution is formed. DDPMs are known for their high-quality image generation and relatively stable training dynamics.

Denoising Diffusion Implicit Models (DDIMs): DDIMs represent a type of generative model that intended to accelerate the process of sampling from denoising diffusion probabilistic models. DDIMs are a more efficient iteration of these models. While generative process in DDPMs is defined as the reverse of a particular Markov diffusion process, DDIMs generalize this by using a class of non-Markov diffusion processes that lead to the same training objective [4].

In addition to these two basic models, there are other models as Score-Based Generative Models via Stochastic Differential Equations (Score SDEs), Latent Diffusion Models (LDMs), or Stable Diffusion Model (SDM).

IV. Preparation of experiments with deep models

A. Datasets overview

Two datasets were chosen in this study – dataset of cats and dogs for training and Van Gogh paintings dataset for fine-tuning. Both datasets are accessible from the Hugging face platform.

The Animal Faces HQ (AFHQ) dataset provides a collection of 15000 high-quality 512x512 pixels resolution images labeled by 3 classes (cats, dogs and wild). This dataset was ideal for the limitations imposed by our computational resources. The distinctive features present in both cats and dogs, such as eyes, ears, and fur patterns, allow us to make visual assessments of the model's output. In particular, these distinctive features serve as clear markers of successful generation. Any generated image should ideally retain these elements, enabling an easier evaluation of the model's performance based on whether these animal-specific characteristics are accurately reproduced.

The Van Gogh paintings dataset comprises digital reproductions of a wide range of paintings from the renowned artist and initially intended for training a model to transform Van Gogh paintings to realistic photos. That is why it has 6290 rows, where each row represents two images – one painting and one photo. In our work only the first column will be used. The diversity within the dataset in terms of subject matter (portraits, landscapes, still life images, etc.) allows for a wide range of generated images.

B. Data preparation for training

After downloading the datasets, we initiated the data preparation procedures. Firstly, we used the torchvision transforms library to define data augmentations and transformations.

The transformations included resizing the images to the target dimensions – 64x64 pixels for animal images and 256x256 pixels for paintings. The selection of these dimensions directly influences the computational resources required for the training process. Lower dimensions are preferred due to their lesser demand on memory and processing power, hence, allowing faster training times.

Augmentation consisted of randomly flipping the images horizontally and converting the images to tensors, Random horizontal flipping is a commonly employed data augmentation technique that helps in increasing the dataset's diversity, effectively providing the model with more varied examples to learn from, thereby improving its generalization capabilities.

Lastly, normalization was applied, converting tensor values to within the range of -1 to 1. Without normalization, pixel intensity values in images usually range from 0 to 255. Normalization was used to ensure that all input features (pixel intensity values in this case) have a similar data distribution, making the training process more manageable and improving model performance.

C. Structure of the used deep network

We used a variant of U-Net architecture. The model begins with a series of downsampling steps – the image is processed with a series of convolutional and max pooling layers. The purpose of the downsampling process is to reduce the dimensionality of the image. Following downsampling, the model goes through an upsampling phase. Here, the dimensionality of the feature maps is increased through a series of transposed convolutions, restoring the image to its original resolution. Skip connections allow the model to use detailed information from the downsampling path in the upsampling path, improving the quality of the output. The model consists of:

- two regular downsampling blocks,
- two downsampling blocks with spatial self-attention,
- two upsampling blocks with spatial self-attention,
- two regular upsampling blocks.

Each block comprises two ResNet layers. which are used to effectively learn features from the input by implementing shortcut connections that allow to bypass layers during training.

Since we worked with standard color images, there were three input channels as well as output channels (corresponding to red, green and blue). Downsampling blocks produced 64, 128, 128 and 256 channels in succession while upsampling blocks consumed these channels in reverse order.

D. Metrics to asses image generation quality

For our experiments with image generation, the Frechet Inception Distance (FID) will serve as a vital metric. The FID [5] offers a measurement of similarity between two sets of images. The rationale behind using FID lies in its ability to capture the nuances between the distributions of real and generated images in the feature space of a pretrained model, in this case, the Inception model.

$$FID = \sqrt{|| \mu_x - \mu_y ||^2 + Tr(C_x + C_y - 2\sqrt{C_x C_y})} \tag{1}$$

where x represents real data distribution, y represents generated data distribution, C is covariance matrix and μ is mean feature vector. Tr denotes trace of a matrix (sum of elements on its main diagonal).

The computation process for the FID involves certain steps. Initially, we obtain the Inception activations for the real and the generated images. These activations are the outputs from the Inception model preceding the final classification layer. Subsequently, a multivariate Gaussian distribution is fitted to these activations. Ultimately, the FID is computed as a measure of similarity between these two Gaussian distributions. In our study, we'll use the PyTorch-FID library to perform these computations efficiently.

V. Results of the selected experiments

A. Model training from scratch on AFHQ dataset

To evaluate the performance of different models and scheduling strategies, we carried out a systematic analysis using the AFHQ dataset (64x64 pixels size of images), with a focus on the Frechet Inception Distance (FID) as a measure of image quality [6]. In order to establish a reference point for interpreting FID scores, a 'worst case' scenario was set up by generating 300 Gaussian noise tensors, yielding a high FID score of approximately 480.71. This provided a benchmark against which the performance of the trained models could be assessed.

DDPM Pipeline with DDPM Scheduler: The first model under investigation utilized the Denoising Diffusion Probabilistic Model (DDPM) coupled with the original DDPM scheduler. Although the training process was time-consuming, taking approximately 17 minutes and 12 seconds for the generation of 300 images, the model demonstrated a promising performance. Over the course of 50 training epochs, the loss function indicated rapid initial improvement, followed by more gradual enhancements. Using a total of 1000 inference steps to generate each image, the model achieved an FID score of 211.32, considerably below our noise benchmark.

DDPM Pipeline with DDIM Scheduler: Next, we applied the same pretrained model, but employed the DDIM scheduler instead. Despite the process being faster (39 seconds for all images), the quality of the images and the FID score (224.86) slightly deteriorated compared to the previous model. This suggests a trade-off between processing speed and image quality.

Table 1
Comparison of FID values of each model

Pipeline type	Scheduler type	Number of epochs	FID score
DDPM pileline	DDPM scheduler	50	211.32
DDPM pipeline	DDIM scheduler	50	224.86
DDIM pipeline	DDPM scheduler	50	414.29
DDIM pipeline	DDIM scheduler	50	415.88
DDIM pipeline	DDPM scheduler	20	238.39
DDIM pipeline	DDIM scheduler	20	238.79

DDIM Pipeline with 50 epochs: Subsequent experiments were carried out with the DDIM model trained for 50 epochs using both the DDIM and DDPM schedulers. In both cases the loss function showed a significant increase after the 20th epoch. This sudden surge in loss can potentially be attributed to numerical instabilities within the weights or gradients, leading to results that were qualitatively poor, which closely approached our noise benchmark, indicating a marked decrease in the quality of the generated images.

DDIM Pipeline with DDIM Scheduler on 20 epochs: Aiming to mitigate the sudden increase in loss observed in the 50-epoch model, the DDIM model was then trained for only 20 epochs. Loss continued decreasing over training time. The speed remained the same (49 seconds for all images), but there was a considerable improvement in the accuracy and a corresponding reduction in the FID score to 238.79. This demonstrates the importance of the appropriate selection of training duration.

DDIM Pipeline with DDPM Scheduler on 20 epochs: The 20-epoch DDIM model, when paired with the DDPM scheduler, yielded a slightly more accurate result with an FID score of 238.39. Although the difference was minor, it implies that the DDPM scheduler may offer slight performance enhancements.

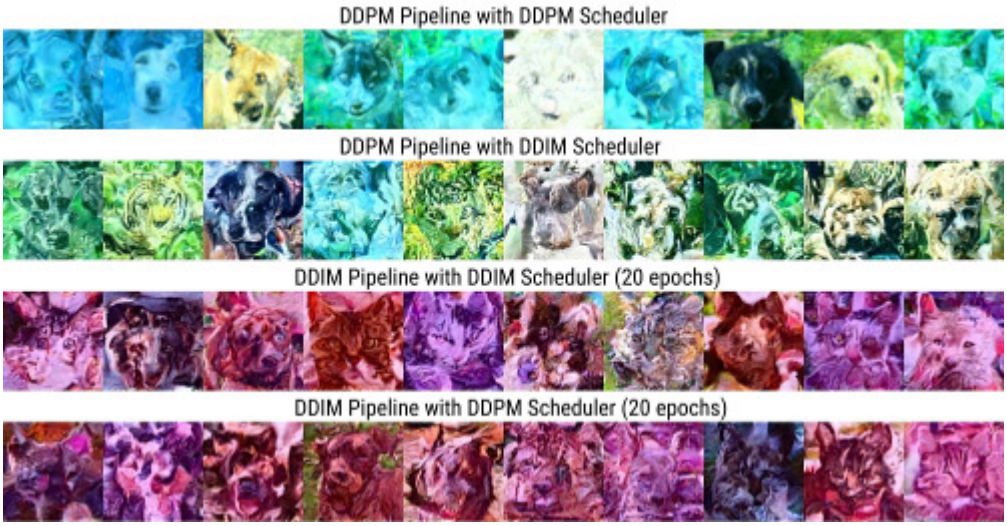


Fig. 2 Images generated by models trained from scratch on AFHQ dataset.

A visual examination of the images generated by various models, as seen in Fig. 2, reveals significant variation in quality and realism. The initial model under consideration, DDPM with the original scheduler, demonstrates impressive performance by producing well-defined images that accurately depict animals. Notably, distinguishing features like ears and eyes are rendered with clarity, indicating the model's ability to faithfully replicate intricate details. Among the models examined, it exhibited the most natural color rendition, with clear differentiation between different animal types, especially cats and dogs. The DDIM pipeline, trained for a mere 20 epochs, also generated recognizable images. However, these images displayed an overly red tint, lending them an unnatural appearance. In contrast, the DDPM model did not exhibit this anomaly, highlighting an area for improvement in the DDIM pipeline.

B. Fine-tuning of the pretrained models on the Van-Gogh dataset

As an extension of our previous analysis, we undertook the task of fine-tuning the pretrained models using the Van-Gogh dataset [6]. Since we did not want to waste significant computational

resources to reiterate experiments with learning from scratch to obtain a pretrained model, we decided to use a ready-made pretrained model. We select a DDPM model trained on the Church dataset, consisting of 126,227 images with a resolution of 256x256 pixels.

Observations reveal that the DDPM model with the DDPM scheduler took 9 minutes and 35 seconds to generate 5 images of 256 pixels each for the default 1000 inference steps. Conversely, when the DDPM pipeline was paired with the DDIM scheduler, the generation time significantly reduced to 22 seconds for 40 inference steps. The DDPM pipeline with its original scheduler performs superior due to the lower occurrence of image corruption. The DDIM pipeline, while slightly falling behind in quality, compensates with a markedly faster generation speed.

We limited the fine-tuning process to two epochs for each model to focus on rapid adaptation while also observing how well the models handle the new data. Despite the short span of two epochs, the model displayed great adaptability, mirroring Van Goghs signature style in the generated images (Fig. 3)

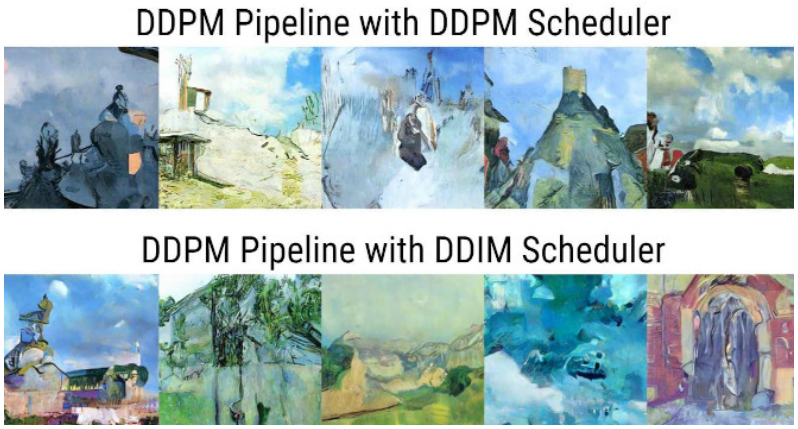


Fig. 3 Images generated by models fine-tuned on the Van Gogh dataset.

Interestingly, the generation time remained consistent with the non-fine-tuned model, implying that the fine-tuning process did not overly burden the model's computational efficiency. Further, the training loss for fine-tuning remained quite low, hovering around 0.029, signaling the model's ability to adjust effectively to the new dataset without significant errors in image reconstruction trained on new data.

VI. Conclusion

In conclusion, the performed experiments demonstrated that the choice of the scheduler and the duration of training significantly influence the performance of the diffusion models. The best quality of generated images was delivered by the DDPM model combined with its own scheduler. Unfortunately, with unbearably slow generation speed. The DDIM model combined with the DDPM scheduler, particularly when trained for an optimized number of epochs, achieved the best balance between image quality and generation speed from those compared.

Our experiments underscores the importance of model selection and optimization, emphasizes the necessity for models to be robust and adaptable, not just in terms of handling new datasets but also in the context of computational efficiency.

Acknowledgment

This paper is supported by Slovak Science Grant Agency project "Basic Research of Deep Learning for Image processing - DL4VISION" supported 2022-2025 under registration code VEGA 1/0394/22.

References

[1] Q. Wu, R. Gao, and H. Zha, "Bridging explicit and implicit deep generative models via neural stein estimators," in *Proc. of the 35th Conf. on Neural Information Processing Systems*, ser. (neurIPS), 2021, [arXiv:1909.13035].
[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," [arXiv:2006.11239], 2020.
[3] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," [arXiv:2105.05233], 2021.
[4] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," [arXiv:2010.02502], 2020.
[5] M. Heusel and et al., "Gans trained by a two time-scale update rule converge to a local nash equilibrium," [arXiv:1706.08500], 2017.
[6] Y. Krasnova, *Image generation using diffusion deep models*. Faculty of Electrical Engineering and Informatics, Technical University of Košice, 2023, bachelor thesis.

Procedurálny generátor grafickej reprezentácie rozsiahlych modelov riek

[1]*Alžbeta SKLENÁRIKOVÁ,* [2]*Branislav MADOŠ*

[1,2] Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]alzbeta.sklenarikova@student.tuke.sk, [2]branislav.mados@tuke.sk

***Abstrakt* — Táto práca sa zaoberá procedurálnym generovaním so zameraním na metódy využívané pri tvorbe modelov riek, pričom v úvodnej časti práce analyzuje metódy generovania terénu a riečnej hladiny. Následne je v práci predstavený vlastný konceptuálny návrh procedurálneho generátora modelu rieky založeného na princípe kontrolných dlaždíc. Navrhnuté riešenie je v závere testované a vyhodnotené.**

***Kľúčové slová* — počítačová grafika, 3D model rieky, procedurálne generovanie obsahu, LWJGL**

I. Úvod

Problematika procedurálneho generovania riek v oblasti počítačovej grafiky je pomerne rozsiahla oblasť, ktorá v sebe skĺbuje hneď niekoľko disciplín. Existujú v nej všeobecne zaužívané poznatky, ako sú napríklad postupy pri tvorbe vodnej hladiny alebo frekventovane využívané funkcie šumu na úpravu povrchu okolitého terénu. Na druhej strane existuje určitý deficit týkajúci sa nedostatočnej klasifikácie a kategorizácie. Príčinou je samotný charakter simulácie, skombinovaný s rôznorodými aspektmi, ktoré možno pri uvažovaní o rieke ako o simulačnej predlohe pozorovať. Charakteristickou vlastnosťou simulácie je výber a extrahovanie konkrétnych zákonitostí z pozorovanej reality, pričom táto selekcia je ovplyvnená zámerom, pre ktorý bol simulačný proces vytvorený. Simulovaný model nikdy nebude odrážať všetky aspekty napodobňovanej reality, pretože aj za malým kúskom skutočnosti sa skrýva príliš veľa stavebných prvkov, pôsobiacich faktorov, a z toho vyplývajúcich uhlov pohľadu, akými na ňu možno nazerať.

Dôvody, pre ktoré sa vytvárajú modely riek, či ich častí, siahajú od zábavného priemyslu až po vedecké predikcie. Správanie rieky možno simulovať z hľadiska pohybu tekutín, či formovania povrchu brehu v závislosti od rôznych faktorov. Môžu sa týkať výskumu, či sa sústrediť na sprostredkovanie estetického zážitku v hre, ideálne za čo najkratší čas a spôsobom šetrným k výpočtovým zdrojom počítača. Pri pohľade na fascinujúci rozsah možností, niet pochýb, že sa ambiciózna vízia o „technológii, ktorá spája umelcov a vedcov“ [1], zdieľaná inžiniermi, ktorí v 60-tych rokoch stáli za vytvorením počítačovej grafiky, naplnila.

Súčasťou tejto práce je nielen skúmanie algoritmických metód využívaných pri tvorbe riek, ale aj návrh vlastného procedurálneho generátora rozsiahleho polygonálneho modelu rieky. Cieľom generátora je poskytnúť nástroj tvorby modelov riek zakomponovateľných do prostredia 3D počítačových hier.

Vo všeobecnosti, procedurálne generovanie je systém, ktorý využíva algoritmy pre produkciu obsahu na základe určených pravidiel [2]. Pomocou procedurálneho generovania obsahu (PCG), sa taktiež často vytvárajú zvuky a grafické efekty odpovedajúce v reálnom čase, napríklad na hráčovu interakciu s konkrétnymi objektmi. Čas vývoja PCG algoritmov, ktoré generujú majoritnú časť herného obsahu, nemožno podceňovať. Akákoľvek zmena v smerovaní projektu môže mať za následok nutnosť obsiahlych reforiem už napísaného kódu [3]. Procedurálne generovacie algoritmy nemusia byť vždy súčasťou zverejnenej hry. Môžu sa využiť aj ako pomoc vo fáze dizajnovania herného obsahu. Napríklad pri tvorbe mesta v hre Skyrim, sa procedurálne vygenerovalo veľké množstvo objektov, ktoré dizajnéri ďalej upravovali do finálnej podoby [4]. Niektoré hry sa nemusia spoliehať na PCG algoritmy, ale ponúkajú svojím hráčom takzvaný „infinity modul“ [5], ktorý generuje napr. nekonečne veľkú krajinu.

Medzi hlavné výhody využitia spoľahlivých PCG algoritmov patrí ušetrenie času generovaním obsahu oproti jeho manuálnemu vkladaniu do herného sveta, ďalej možnosť rozšírenia využitia

generátora na tvorbu rôznych elementov, nie len tých na ktoré bol prvotne určený. Výhodou je tiež istota, že keď sa raz ustanovia pravidlá generovania, PCG algoritmus má schopnosť ich dodržať lepšie ako keby mal byť daný obsah vyprodukovaný človekom. V hrách využitie PCG predlžuje čas hrateľnosti daného produktu. Môže zvýšiť aj schopnosť prispôsobiť herný obsah na základe rôznych aspektov, ako je napríklad úroveň schopností postavy, za ktorú hráč hrá [6]. Jednou z hlavných nevýhod procedurálneho generovania v hrách je príliš veľké riziko straty kontroly nad tým, čo sa v hernom svete deje. Nie je možné, aby bol vo fáze testovania skontrolovaný každý komponent procedurálne generovaného obsahu. Preto sa môže stať, že chyby ostanú ukryté, až kým sa hra nezverejní a dostatočné množstvo hráčov hru nevyskúša [7].

Algoritmy a postupy pri vytváraní riek v počítačovej grafike a simulácii sa líšia vzhľadom na očakávaný cieľ, aký má ich vygenerovanie spĺňať. Teda je dôležité determinovať, či je cieľom len, aby rieka pôsobila dôveryhodne, alebo či jej model má odrážať reálne fyzikálne zákonitosti. Procedurálne algoritmy umožňujú implementovanie širokého spektra výstupov, založeného len na malej obmene vstupných parametrov. Fyzikálne-orientované algoritmy však ako jediné poskytujú výsledky odzrkadľujúce matematickú zákonitosť simulovaného javu v realite. Ich slabinou je nedostatočná možnosť kontroly nad generovaným výstupom [8].

II. Súvisiace Práce

Rieka je prirodzený jednosmerný prúd vody ústiaci v smere spádu vplyvom gravitácie od svojho prameňa do mora, jazera, či inej rieky [9]. Veľkosť toku rieky závisí najmä od hĺbky rieky, teda vzdialenosti vodnej hladiny od pôdorysu povodia, a jej prietoku, t. j. objemu vody, ktorý pretiekol priečnym profilom koryta toku [10]. Pri procedurálnej tvorbe prírodných povrchov sa kladie otázka, ako pravidlá rôznych prírodných vzorov vyjadriť algoritmom.

Prvým takýmto algoritmom na generovanie realistických textúr bol Perlinov šum [11]. Šum sú stochastické fluktuácie v očakávanom signále, pričom rôzna koncentrácia a intenzita týchto fluktuácií vytvára odlišné vzory, ktoré môžu byť reprezentované v n-dimenzionálnom priestore. Funkcie šumu sú využívané ako pseudo-náhodné sekvenčné generátory na tvorbu prírodných povrchov, textúr či efektov. Využívajú sa na počítačovú simuláciu mrakov, ohňa, či dreva. V procedurálnych generátoroch zabezpečujú, napríklad, aby textúry, či krivky neboli neprirodzene pravidelné. Možno teda konštatovať, že v rámci počítačovej grafiky je generátor šumu generátorom čísiel, ktorý vytvára náhodný a neštruktúrovaný vzor používaný na pridanie realistických detailov už vytvoreným rozsiahlym štruktúram [12].

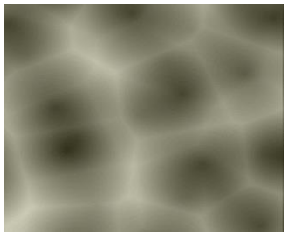
Perlinov šum je bežne využívaný v procedurálnon generovaní organického herného obsahu. Algoritmus sa používa pre mnohé typy nerovnomerných materiálov a textúr, ktoré môžu byť využité na generovanie terénu, či efektov pripomínajúcich pohyby ohňa, vody, alebo mrakov [13]. Algoritmus je najčastejšie reprezentovaný v 2D (Obr. 1) a v 3D dimenzii, ale je ho možné využiť aj pre vyššie dimenzie. Pre každý vstup sa vytvorí modulus 1.0 za účelom nájdenia jednotkovej kocky. Pre bod v 3D priestore vznikne kocka, v 2D priestore vznikne štvorec, v 4D teserakt; spolu patria do rodiny hyperkociek.

Worleyov šum, taktiež známy ako celulárny šum, je algoritmus využívaný na generovanie rozličných vzorov. Tento šum je výpočtovo drahší ako Perlinov, keďže zahŕňa výpočet vzdialenosti medzi všetkými bodmi. Jeho výhodou je však skutočnosť, že zatiaľ, čo iné zámerné funkcie šumu často potrebujú niekoľko iterácií, aby dokázali vytvoriť zaujímavé vzory, Worleyov algoritmus šumu dokáže vytvoriť pútavý obraz už po jednej iterácii [14].

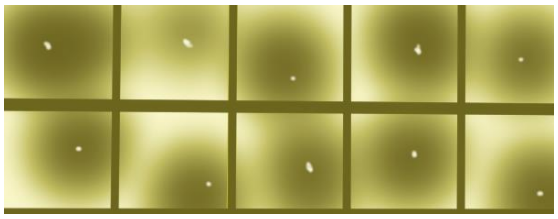
Základný algoritmus funguje tak, že na vstup sa vyberie náhodný bod v 2D alebo 3D priestore a pre tento bod sa vypočíta vzdialenosť n-najbližších bodov. To vytvorí zaujímavý vzor, napríklad ako ten, ktorý možno vidieť na Obr. 2.



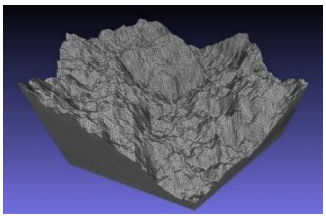
Obr. 1 2D Perlinov šum. Najvyššie hodnoty sú reprezentované ako biele body, najnižšie ako čierne.



Obr. 2 2D Worleyov šum.



Obr. 3 Poissonova distribučná mriežka využitá pri generovaní Worleyovho šumu.



Obr. 4 Fraktálny šum môže byť dosiahnutý aj niekoľkými iteráciami jedného typu šumu.

Pre Worleyov šum je vhodná izotropická distribúcia. Príkladom takejto distribúcie je Poissonova distribúcia, ktorá priestor rozdelí do mriežky, a teda vzniknú identické kocky, pričom v jednej kocke je jeden bod a pre ten sa budú počítať vzdialenosti od bodov zo susedných kociek [14], tak, ako to možno vidieť na Obr. 3.

Fraktálny šum sa získa zvolením niekoľkých vygenerovaných šumov, buď rovnakého, alebo odlišného typu a ich pridaním k sebe. Pre každú za sebou idúcu oktávu sa upraví šum dvoma spôsobmi, prudkým znížením amplitúdy a zvýšením frekvencie [15]. Príklad takto vygenerovaného šumu možno vidieť na Obr. 4.

Pre pohybujúce sa kontinuum (teda tekutinu alebo deformovateľnú tuhú látku, príp. teleso), existujú dva prístupy na opis jeho pohybu: Lagrangenove hľadisko a Eulerove hľadisko [16].

Lagrangeove hľadisko pomenované po francúzskom matematikovi Lagrangeovi, uvažuje o kontinuu ako o častícovom systéme. Každý bod v tekutine alebo v tuhej látke je označený ako oddelená častica s pozíciou x a rýchlosťou u. O oddelených časticiach možno rozmýšľať ako o molekulách a tento spôsob simulácie sa využíva najmä pri tuhých látkach s diskrétnym počtom častíc [17][18].

Eulerove hľadisko pomenované po švajčiarskom matematikovi Eulerovi, využíva hľadisko, kde namiesto, aby bola pozorovaná každá častica v celkovom útvare, v centre záujmu stojí to, ako sa mení správanie vybraných faktorov (napr. hustota, rýchlosť, teplota) vo fixných bodoch v priestore [19].

Navier-Stokesove rovnice opisujú ako sú rýchlosť, tlak, teplota a hustota pohybujúcej sa tekutiny prepojené. Tieto rovnice pozostávajú z časovo závislej kontinuálnej rovnice pre zachovanie masy, troch časovo závislých rovníc pre zachovanie momentu a časovo závislej rovnice na zachovanie energie [20][21]. Navzdory ich širokému praktickému využitiu, je teoretické objasnenie týchto rovníc nekompletné - a preto dôkaz o ich existencii patrí medzi "sedem problémov tisícročia", ktoré boli vyhlásené Clayovym matematickým ústavom [22].

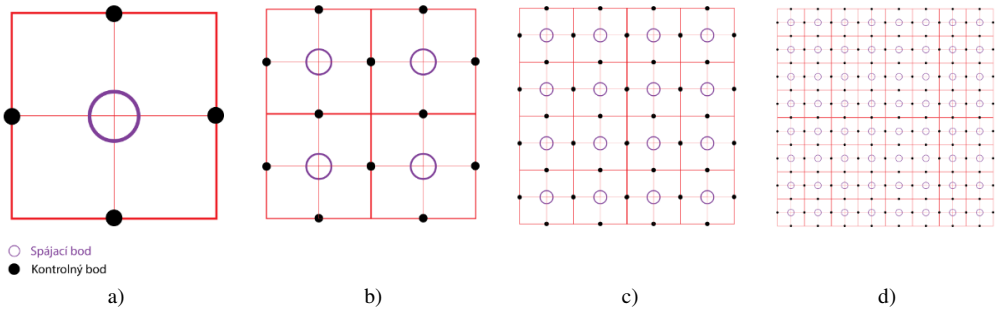
MAC mriežka (Marker And Cell grid) bola prvýkrát publikovaná v roku 1965 americkým teoretickým fyzikom Francisom H. Harlowom a J. Eddie Welchom [23]. Táto metóda sa stala jedným z najpopulárnejších spôsobov simulácie toku tekutín v počítačovej grafike. Ide o "stupňovitú" mriežku kombinujúcu niekoľko typov uzlových bodov umiestnených na odlišných geometrickým pozíciách. Tento na prvý pohľad zložitý typ mriežky umožňuje prehľadnú a presnú formuláciu výsledkov pre parciálne diferenciálne rovnice s konečnými rozdielmi, kam spadajú aj Navier-Stokesove rovnice [24].

Level set metóda na simulovanie pohybujúceho sa povrchu bola v roku 1987 formulovaná S. Osherom a J.A. Sethianom [25] a do dnešného dňa sa zaznamenáva v oblasti počítačovej grafiky široké spektrum jej využitia; od špeciálnych filmových efektov, vizualizáciu až po počítačové videnie či zachytávanie multifázových dynamických tokov kvapalín [26]. Ide o jednu z najvyužívanejších metód sledovania kvapalinového povrchu. Level set metóda umožňuje väčšiu kontrolu nad povrchom kvapaliny, a teda sa simulácia dokáže viac priblížiť k skutočnému správaniu hladiny riek - pri minimálnom pohybe a nižšej rýchlosti vyzerá hladina rieky hladko, a naopak pri rýchlejšom toku rieky sa hladina bude správať búrlivo až "trhane" [27].

Interpolácia a extrapolácia sú dva bežne využívané typy matematickej predikcie na odhad hypotetických hodnôt. Kým interpolácia predikuje hodnoty, ktoré sa už nachádzajú niekde v rámci existujúcej množiny hodnôt funkcie, extrapolácia sa používa na predikciu hodnôt spadajúcich mimo tejto množiny [28]. Obe metódy sa pri procedurálnom generovaní rieky používajú na doplnenie generovaných hodnôt.

III. Navrhnuté Riešenie

V nasledujúcej časti článku bude predstavené vlastné riešenie procedurálneho generátora riečnej cesty, navrhnuté v rámci tejto práce, ďalej bude vysvetlený jeho konceptuálny dizajn a taktiež budú vyzdvihnuté najdôležitejšie čiastkové problémy návrhu — spoločne s postupmi ich riešenia.



Obr. 5 Kontrolná dlaždica a) Typ 3, b) Typ 5, c) Typ 9, d) Typ 17.

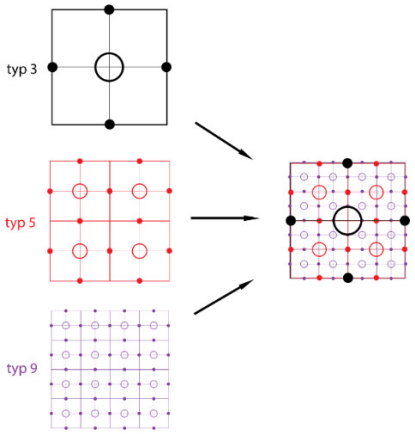
Koncept kontrolnej dlaždice. Úlohou kontrolnej dlaždice je vytvoriť pre rieku kontrolovateľné prostredie, v rámci ktorého sa jej cesta po 3D svete môže rozvíjať.

Koncept kontrolnej dlaždice pozostáva z (Obr. 5):

- **Kontrolných bodov** - Kontrolné body reprezentujú miesta, kam môže riečna cesta ďalej pokračovať vo svojom toku. V prípade, že v konkrétnom smere rieka pokračovať nemôže (napríklad z dôvodu, že sa dlaždica nachádza na konci mapy sveta), tak sa daný kontrolný bod z tejto dlaždice odstráni.
- **Spájacích bodov** - Prepája cestu z kontrolného bodu do všetkých zvyšných. Inými slovami je to miesto 'ohybu rieky' pri jej toku z jedného kontrolného bodu do druhého.

Typ dlaždice určuje veľkosť rieky, ktorá ňou preteká, a zároveň determinuje (podľa počtu voliteľných kontrolných bodov) koľko rozvodníc je v dlaždici možné vytvoriť. Na obr. 5 možno vidieť dlaždice rôznych typov.

Základnou vlastnosťou riadiacich dlaždíc je schopnosť rešpektovať pozície kontrolných a spájacích bodov ich predchodcov s nižšou hustotou tak, aby umiestnenie ich vlastných kontrolných bodov tie predošlé neprekrývalo, ale naopak dopĺňalo, tak ako to možno vidieť na Obr. 6, kde sú prekryté dlaždice typu 3, 5 a 9.



Obr. 6 Prekrytie kontrolných dlaždíc typu 3, 5 a 9, kde je možné vidieť, že sa spájacie a kontrolné body navzájom neprekrývajú ale vzájomne dopĺňajú.

Rieka nemôže prechádzať voľne medzi všetkými typmi dlaždíc. Aby jej výsledný obraz pôsobil prirodzene, zaviedli sa nasledujúce pravidlá prechodu:

1. Po kontrolnej dlaždici môže pokračovať iba dlaždica rovnakého alebo vyššieho typu. Keď je do sveta vložená kontrolná dlaždica, môže po nej nasledovať iba dlaždica rovnakej alebo vyššej hustoty. Je to z dôvodu, aby po rieke s úzkym korytom nenasledoval neprirodzene veľký tok.
2. Zamlčanie vybraných kontrolných bodov na základe strany, ktorou rieka do dlaždice vstúpila. Zamlčanie vybraných kontrolných bodov vzhľadom k umiestneniu prameňa rieky v dlaždici má zamedziť vytváraniu riečnych ramien smerujúcich proti pôvodnému prúdeniu toku.

Vytvorenie vizualizácie rieky pozostáva zo štyroch samostatných krokov:

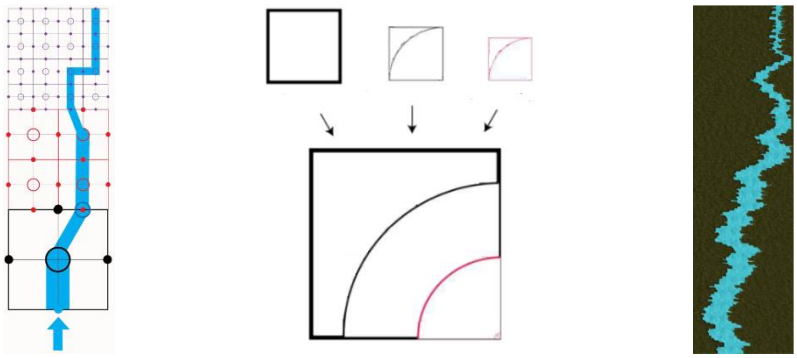
1. Fáza nastavenia parametrov koryta rieky
2. Fáza náčrtu koryta rieky
3. Fáza úpravy koryta rieky
4. Fáza vykresľovania koryta rieky

Fáza náčrtu koryta rieky je fázou, v ktorej sa vyberajú jednotlivé kontrolné dlaždice, ich typy a umiestnenie v priestore.

Fáza úpravy koryta rieky berie do úvahy zmeny smeru toku rieky a vytvára zákruty zodpovedajúce umiestneniu kontrolných a spájacích bodov v rámci kontrolných dlaždíc vybraných typov.

Fáza vykresľovania koryta rieky predstavuje finálnu fázu, v ktorej sa definitívne určí tvar brehu rieky a vykreslí sa v plných detailoch.

Jednotlivé fázy vytvorenia vizualizácie rieky možno vidieť na Obr. 7.



a) b) c)

Obr. 7 Fázy vytvorenia vizualizácie rieky, kde je a) fáza náčrtu koryta rieky, b) fáza úpravy koryta rieky c) fáza vykresľovania koryta rieky.

IV. VÝSLEDKY

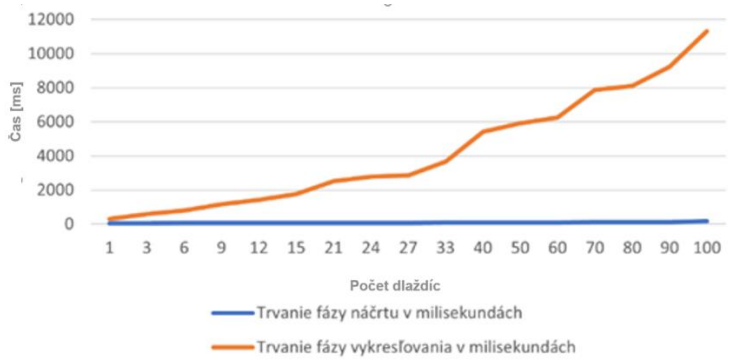
Testovanie navrhnutého generátora rieky sa zameralo na rôzne aspekty, od rýchlosti generovania rieky až po množstvo pamäte spotrebovanej pre reprezentáciu jej modelu.

Testovanie rýchlosti výpočtov prebehlo na 64-bitovom systéme s procesorom Intel Core i7-8550U s taktovacou frekvenciou 1.80GHz a 8GB pamäťou RAM. Simultánne sa uskutočnili tri typy meraní rýchlosti vykonávania jednotlivých fáz proceduráleho generátora:

- Testovanie rýchlosti fázy náčrtu — testovanie meralo trvanie výpočtov funkcií z fázy náčrtu aj požiadaviek.
- Testovanie rýchlosti fázy vykresľovania — testovanie meralo trvanie výpočtov funkcií z fázy vykresľovania.
- Testovanie celkovej rýchlosti generátora — testovanie meralo celkové trvanie výpočtov uskutočnených v rámci procedurálneho generátora.

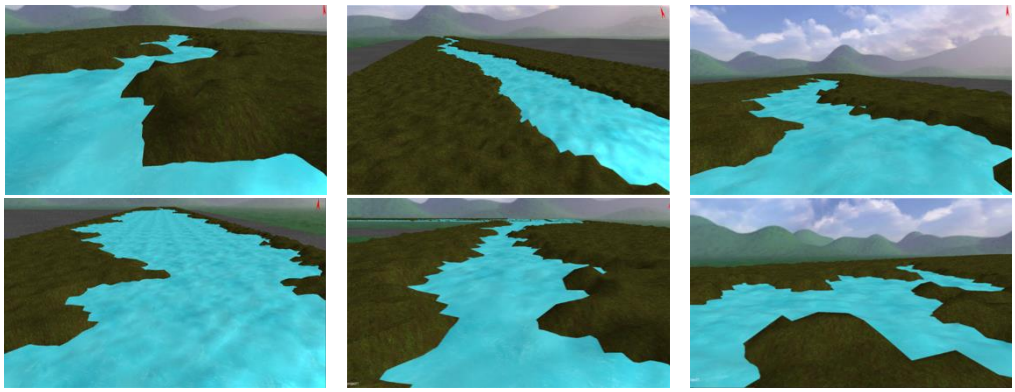
Testovanie v programe prebiehalo zaznamenávaním času začiatku a času konca vybraných výpočtov. Na zaznamenanie jednotlivých časov bola použitá funkcia currentTimeMillis() z java triedy System. Po ukončení výpočtov generátora boli zaznamenané časy pre jednotlivé merania od seba odčítané a hodnota rozdielu, indikujúca trvanie daného výpočtu, bola zapísaná. Testovalo sa trvanie výpočtov pre rôzne dĺžky rieky (resp. pre rôzny počet kontrolných dlaždíc predstavujúcich dĺžku rieky). Pre každú dĺžku rieky sa robilo niekoľko meraní a výsledné hodnoty sa následne spriemerovali.

Výsledky testovania rýchlosti výpočtov procedurálneho generátora potvrdili predpoklad o efektívnosti používania kontrolných dlaždíc z hľadiska rýchlosti výpočtov. Testy rýchlosti ukázali, že vytvorenie náčrtu rieky s dĺžkou 100 kontrolných dlaždíc trvá v priemere 187 milisekúnd, zatiaľ čo vygenerovanie finálneho modelu rieky dĺžky 1 dlaždice trvá 317 milisekúnd; teda výpočet fázy náčrtu pre najdlhšiu meranú rieku trvá v priemere o 41 % kratšie, ako výpočet fázy vykresľovania pre najkratšiu meranú rieku (Obr. 8).



Obr. 8 Trvanie jednotlivých fáz realizácie generovania rieky prostredníctvom navrhnutého algoritmu v milisekundách (ms) vzhľadom k veľkosti rieky uvádzanej v počte dlaždíc.

Pri hustote 65 vrcholov, bude jeden finálny objekt pevniny obsahovať 4225 vrcholov. Každý vrchol má tri pozičné hodnoty (x, y, z), určujúce jeho umiestnenie v zobrazovanom 3D priestore. Teda vrchol pri svojom zápise do poľa typu float zaberá 3 pozície. Každá hodnota uložená do primitívneho dátového typu float v jazyku java zaberá 4 bajty. Vizualizácie generovaných modelov riek možno vidieť na Obr. 9.



Obr. 9 Príklady vizualizácií riek vygenerovaných s využitím navrhnutého generátora.

V. ZÁVER

Hlavným zámerom pri vytváraní vlastného konceptuálneho návrhu riečneho generátora bolo docielenie spoľahlivej kontroly nad procesom procedurálnej tvorby riečnej cesty. Pre tento cieľ boli sformulované dve zásady návrhu softvérového generovania 3D obsahu. Prvou bola zásada o možnosti zásahu do výpočtových algoritmov generátora, ktorá sa veľmi podobá zaužívanému a uznávanému konceptu modularizácie. Druhou bola zásada o práci s čo najmenším možným množstvom dát, ktorá odporúča na začiatku implementačného návrhu vytvoriť a pracovať s riadiacimi dátami, na základe ktorých potom vzniknú zmeny v celom výstupnom dátovom sete. V súlade s predostrenými normami bol vytvorený a implementovaný vlastný koncept kontrolných dlaždíc, ktoré riadia celý proces generovania riečnej cesty.

POĎAKOVANIE

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

REFERENCIE

[1] KRANZ, S.; FOWLER, E.; HOLTON, M. Science & technology in the arts. New York: Van Nostrand Reinhold Co, 1974, ss. 335, ISBN 9780442245320.

[2] EBERT, David S; MUSGRAVE, F Kenton; PEACHEY, Darwyn; PERLIN, Ken; WORLEY, Steven. Texturing & modeling: a procedural approach. 3rd edition. The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling, Morgan Kaufmann, 2003, ss. 687, ISBN: 1–55860–848–6

[3] SHORT, Tanya; ADAMS, Tarn. Procedural generation in game design. CRC Press, Routledge, 2017. ss. 336, ISBN 9781498799195.

[4] FREIKNECHT, Jonas; EFFELSBERG, Wolfgang. A survey on the procedural generation of virtual worlds. Multimodal Technologies and Interaction. 2017, roč. 1, č. 4, s. 27, https://doi.org/10.3390/mti1040027.

[5] LANDIN, Per. THE ROAD TO DUNGEONS, Exploring the random art of procedural levels. Dostupné tiež z: https://www.minecraft.net/nl-nl/article/the-road-dungeons. Naposledy pristúpené: 20.6.2022.

[6] SMITH, Gillian; GAN, Elaine; OTHENIN-GIRARD, Alexei; WHITEHEAD, Jim. PCG-based game design: Enabling new play experiences through procedural content generation. 2011. Dostupné z doi: 10.1145/2000919.2000926.

[7] TOGELIUS, Julian; WHITEHEAD, Jim; BIDARRA, Rafael. Guest editorial: Procedural content generation in games. IEEE Transactions on Computational Intelligence and AI in Games. 2011, roč. 3, č. 03, s. 169–171.

[8] GÉNEVAUX, Jean-David; GALIN, Éric; GUÉRIN, Eric; PEYTAVIE, Adrien; BENES, Bedrich. Terrain generation using procedural models based on hydrology. ACM Transactions on Graphics (TOG). 2013, roč. 32, č. 4, s. 1–13.

[9] LEOPOLD, Luna Bergere. A View of the River. Harvard University Press, 2006, ss. 298, ISBN 9780674018457.

[10] VÝLETA, R.; KOHNOVÁ, S.; VALENT, P. Riešené úlohy z hydrológie I.: Povodie, zrážky, prietok a hydrologická bilancia. Edícia skrípt. Bratislava: Slovenskátechnická univerzita v Bratislave, 2011.

[11] GONZALEZ, P.; LOWE, J. The Book of Shaders. 2015. Dostupné tiež z: https://thebookofshaders.com/11/. Naposledy pristúpené: 20.8.2022.

[12] LAGAE, Ares; LEFEBVRE, Sylvain; COOK, Rob; DEROSE, Tony; DRETTAKIS, George; EBERT, David S; LEWIS, John P; PERLIN, Ken; ZWICKER, Matthias. A survey of procedural noise functions. In: Computer Graphics Forum. 2010, zv. 29, s. 2579–2600. Č. 8.

[13] DORAN, Jonathon; PARBERRY, Ian. Controlled procedural terrain generation using software agents. IEEE Transactions on Computational Intelligence and AI in Games. 2010, roč. 2, č. 2, pp. 111–119.

[14] WORLEY, Steven. A cellular texture basis function. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. 1996, s. 291–294.

[15] VIVO, Patricio Gonzalez; LOWE, Jen. The Book of Shaders: Fractal Brownian Motion. Patricio Gonzalez Vivo, https://thebookofshaders. com/13. 2015.

[16] BRDIČKA, M.; SAMEK, L.; SOPKO, B. Mechanika kontinua. Praha: Academia, 2011, ss. 880, ISBN 978-802-00-20-390

[17] BRIDSON, Robert; MÜLLER-FISCHER, Matthias. Fluid simulation: Siggraph 2007 course notes video files associated with this course are available from the citation page. In: ACM SIGGRAPH 2007 courses. 2007, s. 1–81.

[18] SMYTH, B.: Lagrangian and Eulerian descriptions, Oregon State University, Dostupné tiež z: https://eng.libretexts.org/Bookshelves/Civil_Engineering/Book%5C%3A_All_Things_Flow__Fluid_Mechanics_for_the_Natural_Sciences_(Smyth)/05%5C%3A_Fluid_Kinematics/5.01%5C%3A_Lagrangian_and_Eulerian_descriptions. Naposledy pristúpené: 05.12.2021.

[19] BRIDSON, Robert. Fluid simulation for computer graphics. AK Peters/CRC Press, 15 október, 2015, ss. 276, ISBN-10 9781482232837, ISBN-13 978-1482232837

[20] JANKUHN, Thomas; OLSHANSKII, Maxim A; REUSKEN, Arnold. Incompressible fluid problems on embedded surfaces: modeling and variational formulations. Interfaces and Free Boundaries. 2018, roč. 20, č. 3, s. 353–377.

[21] CENTER, Glenn Research. Navier-stokes equations. Dostupné tiež z: https://www.grc.nasa.gov/WWW/k-12/airplane/nseqs.htmll. Naposledy pristúpené: 12.12.2021.

[22] CARLSON, James A; JAFFE, Arthur; WILES, Andrew. The millennium prize problems. Clay Mathematics Institute, Cambridge, Massachusetts, ss. 161, ISBN 0-8218-3679-X.

[23] HARLOW, Francis H; WELCH, J Eddie. Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. The physics of fluids. 1965, roč. 8, č. 12, s. 2182–2189.

[24] DOMBROSKI, J.; GREGORY, T.; SHOTANDE, M.; RACKEAR, E. Marker-and-Cell Method (MAC). Dostupné tiež z: http://plaza.ufl.edu/ebrackear/. Naposledy pristúpené: 28.12.2021.

[25] Level Set Methods: An Overview and Some Recent Results. Journal of Computational Physics. 2001, roč. 169, č. 2, s. 463–502.

[26] OSHER, Stanley, Fedkiw Ronald: Level Set Methods and Dynamic Implicit Surfaces. Applied Mathematical Sciences, Vol. 153, Springer Verlag, New York, 2003, ss. 273, ISBN 0-387-95482-1.

[27] CLINE, David; CARDON, David; EGBERT, Parris K. Fluid flow for the rest of us: Tutorial of the marker and cell method in computer graphics. Brigham Young University. 2013.

[28] Interpolation and Extrapolation, Harper College, Dostupné tiež z: http://dept.harpercollege.edu/chemistry/chm/100/dgodambe/thedisk/labtech/data7.htm. Naposledy pristúpené: 13.12.2021.

Analýza spoľahlivosti OhmegaPly rezistorov rôznych tvarov

Tomáš *LENGER*

Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Košice, Slovenská republika

tomas.lenger@tuke.sk

***Abstrakt* — Tenkovrstvové rezistory predstavujú možnosť vnorenia súčiastok do dosiek plošných spojov. Elektronické súčiastky osadené technológiou povrchovej montáže vyžadujú priestor na povrchu dosky plošných spojov, čo sťažuje miniaturizáciu elektronických zariadení. Tenkovrstvové rezistory je možné jednoducho vnoriť do vnútorných vrstiev dosiek plošných spojov bez toho, aby sa zväčšila hrúbka dosiek plošných spojov. OhmegaPly rezistory predstavujú jednu z tenkovrstvových technológií. Tento článok analyzuje stabilitu OhmegaPly rezistorov dvoch rôznych tvarov, ako aj dvoch rôznych šírok. OhmegaPly rezistory tiež znižujú počet spájkovaných spojov. OhmegaPly rezistory boli realizované na povrchu substrátu, ako aj vnorené do viacvrstvovej dosky plošných spojov pozostávajúcej z OhmegaPly laminátu, prepreg vrstvy a FR-4 laminátu. Stabilita odporov OhmegaPly bola meraná počas tepelného šokovania. Tento článok porovnáva stabilitu OhmegaPly rezistorov na základe ich tvaru (obdĺžnik alebo meander), šírky (0,5 mm alebo 1 mm) a umiestnenia (na povrchu alebo vnorené).**

***Kľúčové slová* — OhmegaPly, vnorené rezistory, spoľahlivosť**

I. Úvod

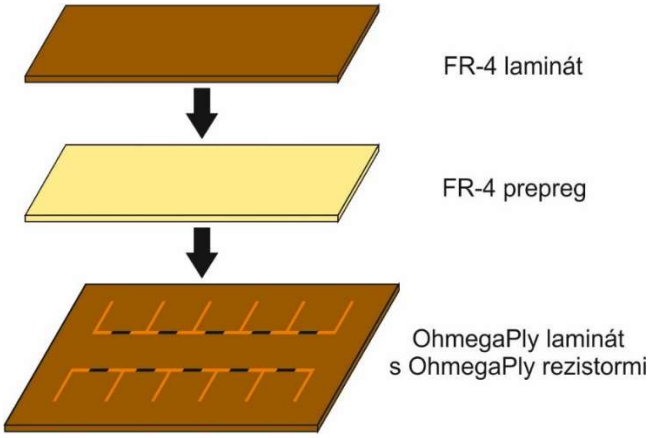
Dosky plošných spojov (DPS) predstavujú základnú súčasť všetkých elektronických zariadení. Rastúca zložitosť DPS si vyžaduje aj zvyšovanie množstva osadených elektronických súčiastok. V posledných rokoch miniaturizácia elektronických súčiastok pre povrchovú montáž (SMD) umožňuje zvyšovanie zložitosti DPS bez zväčšenia rozmerov DPS. Miniaturizácia SMD súčiastok však dosahuje svoje fyzikálne limity, čo si vyžaduje hľadanie iných spôsobov miniaturizácie DPS [1], [2], [3], [4], [5].

Vnáranie súčiastok do viacvrstvových DPS predstavuje odlišný spôsob miniaturizácie DPS. Vnorené súčiastky môžu zmenšiť rozmery DPS bez zníženia jej funkčnosti. Existuje mnoho rôznych technológií používaných na vnáranie elektronických súčiastok. Technológia OhmegaPly je jednou z možností vnárania súčiastok. Technológia OhmegaPly je založená na tenkých vrstvách s hrúbkou menšou ako 1 μm. OhmegaPly rezistory vďaka svojej hrúbke môžu byť vnorené do viacvrstvovej DPS a zmenšiť rozmery DPS bez zväčšenia jej hrúbky [6], [7], [8], [9].

OhmegaPly rezistory sú založené na iných materiáloch ako viacvrstvové DPS. Nesúlad tepelných vlastností materiálov môže spôsobiť iniciáciu mechanického napätia vo vnútri DPS [10], [11]. Cieľom tohto článku je analyzovať stabilitu OhmegaPly rezistorov, v závislosti od ich ich tvaru, šírky a umiestnenia. OhmegaPly rezistory nie je možné počas životnosti elektronického zariadenia vymeniť, z tohto dôvodu je stabilita ich vlastností kľúčovou súčasťou životnosti elektronických zariadení.

II. Návrh a príprava vzoriek

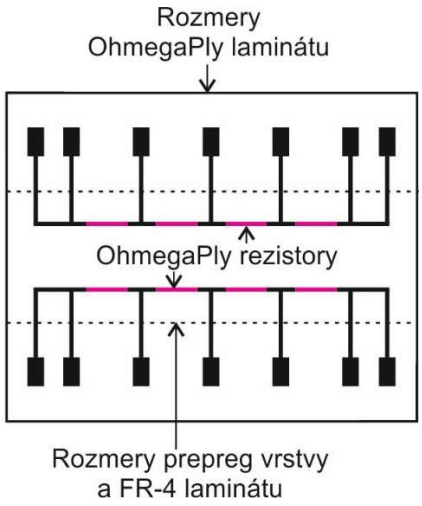
Tento článok je zameraný na stabilitu OhmegaPly rezistorov. OhmegaPly rezistory boli realizované na povrchu substrátu, ako aj vnorené do viacvrstvovej štruktúry. Viacvrstvová štruktúra pozostáva z OhmegaPly laminátu, prepreg vrstvy a FR-4 laminátu (Obr. 1).



Obr. 1 Štruktúra viacvrstvovej DPS

Viacvrstvové štruktúry boli laminované prepreg materiálom Isola DE104. Laminovanie viacvrstvových štruktúr bolo realizované konštantným tlakom (2 MPa) počas definovaného tepelného profilu. Maximálna teplota laminácie bola 170 °C počas 50 minút. Laminovaním bola vytvorená nerozdeliteľná viacvrstvová štruktúra s vnorenými OhmegaPly rezistormi.

Rozmery prepreg vrstvy a FR-4 laminátu boli menšie ako rozmery OhmegaPly laminátu. Menšie rozmery týchto vrstiev umožňujú meranie elektrického odporu vnorených OhmegaPly rezistorov bez medzivrstvových prepojení. Návrh DPS bol optimalizovaný pre meranie každého vnoreného rezistoru samostatne (Obr. 2). Elektrický odpor bol meraný Milliohmmetrom Agilent 4338B.



Obr. 2 Dizajn a ilustrácia rozmerov viacvrstvovej štruktúry s vnorenými OhmegaPly rezistormi

Elektrický odpor OhmegaPly rezistorov je definovaný ich plošným odporom. Existujú rôzne typy OhmegaPly laminátov, ktoré sa líšia hodnotou plošného odporu. Všetky testované OhmegaPly rezistory boli realizované na OhmegaPly lamináte s plošným odporom 100 Ω/□. Elektrický odpor OhmegaPly rezistorov obdĺžnikového tvaru je definovaný rovnicou:

$$R = R_s * N \tag{1}$$

v ktorom R_s je plošný odpor OhmegaPly laminátu a N je počet štvorcov

$$N = \frac{l}{w} \tag{2}$$

V ktorom l je dĺžka rezistora a w je šírka rezistora.

Počet štvorcov rezistorov v tvare meandra sa počíta analogicky ako pri rezistoroch v tvare obdĺžnika, avšak plošný odpor rohových štvorcov sa násobí 0,56. Vzhľadom na vplyv rohových štvorcov je elektrický odpor OhmegaPly rezistorov v tvare meandra definovaný rovnicou:

$$R = R_s * \left(N + (C_s * 0.56)\right) \tag{3}$$

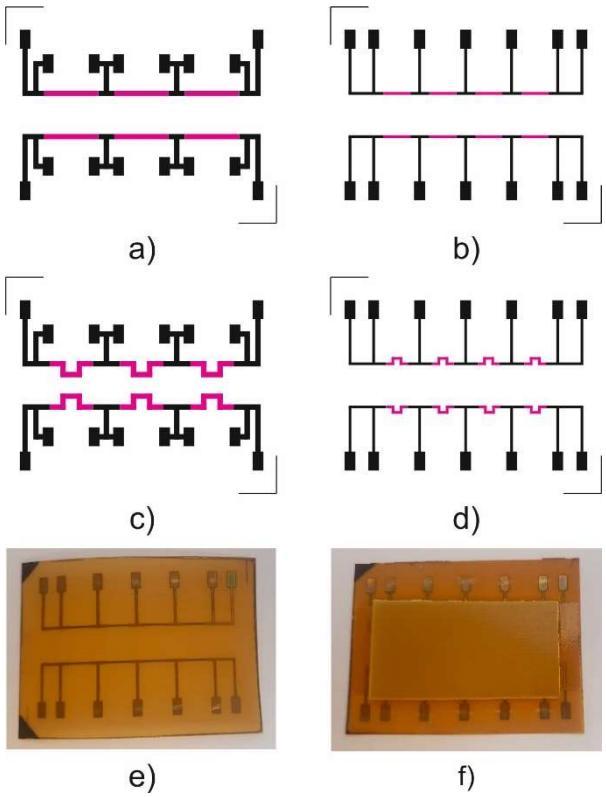
v ktorom R_s je plošný odpor OhmegaPly laminátu, N je počet štvorcov a C_s je počet rohových štvorcov.

Stabilita vnorených rezistorov bola analyzovaná na základe ich šírky a tvaru. Realizovaných bolo osem typov OhmegaPly rezistorov (Tabuľka 1). Pre každý typ OhmegaPly rezistorov bolo realizovaných 24 vzoriek.

Tabuľka 1
Typy a označenie realizovaných vzoriek

Vzorka	Popis
SR0,5	Obdĺžnikový tvar, 0,5 mm šírka, na povrchu
SM0,5	Meandrový tvar, 0,5 mm šírka, na povrchu
SR1	Obdĺžnikový tvar, 1 mm šírka, na povrchu
SM1	Meandrový tvar, 1 mm šírka, na povrchu
ER0,5	Obdĺžnikový tvar, 0,5 mm šírka, vnorený
EM0,5	Meandrový tvar, 0,5 mm šírka, vnorený
ER1	Obdĺžnikový tvar, 1 mm šírka, vnorený
EM1	Meandrový tvar, 1 mm šírka, vnorený

OhmegaPly rezistory v tvare obdĺžnika pozostávajú z 10 štvorcov a boli realizované so šírkou 0,5 mm (obr. 3b) a 1 mm (obr. 3a). Elektrický odpor OhmegaPly rezistorov v tvare obdĺžnika vypočítaný podľa (1) bol 1000 Ω. OhmegaPly rezistory s meandrovým tvarom boli navrhnuté tak, aby zodpovedali elektrickému odporu obdĺžnikových rezistorov. Meandrové rezistory boli realizované so šírkou 0,5 mm (obr. 3d) a 1 mm (obr. 3c). Elektrický odpor OhmegaPly rezistorov v tvare obdĺžnika vypočítaný podľa (3) bol 1024 Ω.



Obr. 3 Dizajn vzoriek a) SR1/Er1, b) SR0,5/ER0,5, c) SM1/EM/, d) SM0,5/EM0,5, a realizované vzorky e) SR0,5, f) ER0,5

Tento článok je zameraný na analýzu zmien elektrického odporu rezistorov OhmegaPly s rôznymi tvarmi. Viacvrstvové štruktúry s vnorenými OhmegaPly rezistormi pozostávajú z viacerých materiálov s rôznymi vlastnosťami. Rôzne vlastnosti použitých materiálov môžu spôsobiť mechanické namáhanie vo vnútri viacvrstvovej DPS počas životnosti elektronických zariadení. Vnorené OhmegaPly rezistory boli vystavené vysokej teplote aj počas procesu lamináce viacvrstvovej DPS.

Na analýzu životnosti elektronických zariadení sa používajú testy zrýchleného starnutia. Teplotné cyklovanie je jedným z testov zrýchleného starnutia. Podmienky tepelného cyklovania sú definované normou IPC-9701. Podľa normy IPC-9701 teplotné cyklovanie realizované

v rozsahu teplôt 125°C a -45°C. Jeden teplotný cyklus trvá 40 minút a realizovaných bolo 1000 teplotných cyklov. Tepelné cyklovanie bolo realizované v komore Clima Temperatur System TSS – 70.

III. Dosiahnuté výsledky a diskusia

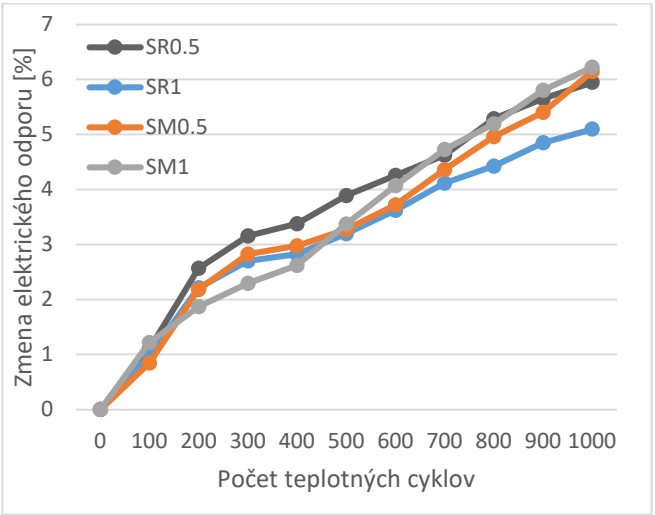
Cieľom tohto článku bolo analyzovať vplyv tvaru OhmegaPly rezistorov na stabilitu ich elektrického odporu. Analyzovaný bol aj vplyv šírky OhmegaPly rezistorov na ich stabilitu. Aj keď vypočítaná hodnota rezistorov bola 1000 Ω a 1024 Ω, namerané hodnoty vyleptaných OhmegaPly rezistorov boli približne o 20% vyššie (Tabuľka 2). Tabuľka 2 zobrazuje priemerné hodnoty pre každý typ realizovaných OhmegaPly rezistorov pred teplotným cyklovaním.

Tabuľka 2
Elektrický odpor OhmegaPly rezistorov

Vzorka	Elektrický odpor [Ω]	
	Vypočítaný	Nameraný
SR0,5	1000	1124,4
SR1	1000	1151,7
SM0,5	1024	1230,1
SM1	1024	1234,7
ER0,5	1000	1168,1
ER1	1000	1093,6
EM0,5	1024	1220,9
EM1	1024	1207,2

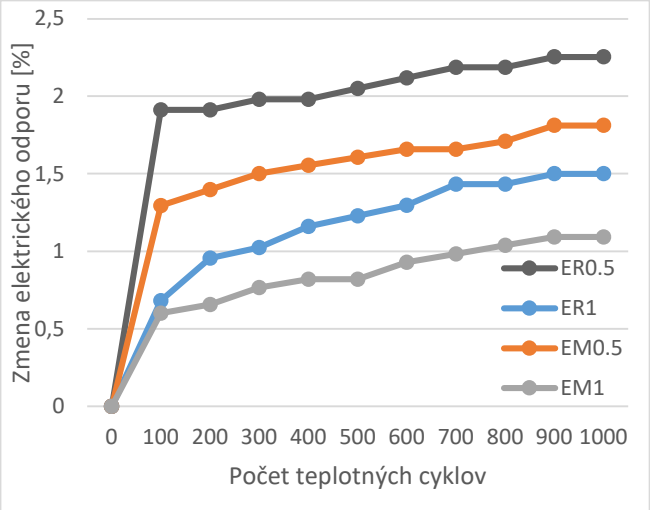
Hodnota elektrického odporu OhmegaPly rezistorov je vysoko závislá od procesu leptania rezistorov, ako vyplýva z Tabuľky 2. Podľa nameraného elektrického odporu rezistorov OhmegaPly by bolo vhodnejšie dizajn týchto rezistorov navrhnutý na nižšie hodnoty elektrického odporu a požadovaná hodnota elektrického odporu by mala byť dosiahnutá následným laserovým trimovaním.

Obr. 4 zobrazuje priebeh zmeny elektrický odpor OhmegaPly rezistorov umiestnených na povrchu. Obr. 4 znázorňuje reprezentatívnu vzorku pre každý typ OhmegaPly rezistora. Medzi analyzovanými vzorkami sú len zanedbateľné rozdiely, ktoré mohli byť spôsobené aj chybou merania. Zmena elektrického odporu pre každý typ vzorky bola približne 5-6%. OhmegaPly rezistory so šírkou 1 mm vykazujú nevýznamne menší nárast elektrického odporu. Nebol pozorovaný žiadny vplyv tvaru rezistoru na stabilitu jeho elektrického odporu. Avšak, pri 3 vzorkách so šírkou 0,5 mm, elektrický odpor po 700-800 tepelných cykloch výrazne vzrástol (desiatky percent), na základe čoho je možné konštatovať, že vyššiu odolnosť voči teplotným zmenám vykazujú OhmegaPly rezistory so šírkou 1 mm.



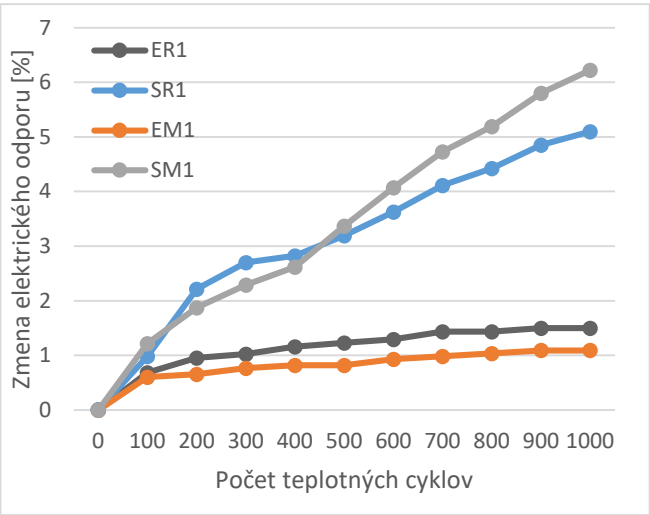
Obr. 4 Závislosť zmeny elektrického odporu od počtu teplotných cyklov OhmegaPly rezistorov umiestnených na povrchu substrátu

Obr. 5 zobrazuje závislosť elektrického odporu vnorených OhmegaPly rezistorov. Vnorené rezistory vykazujú najvýraznejšiu zmenu elektrického odporu medzi 0 a 100 teplotnými cyklami. Počiatočné hodnoty týchto rezistorov boli namerané pred laminációu. Zvýšenie elektrického odporu vnorených rezistorov je pravdepodobne spôsobené teplotou lamínácie. Zmeny elektrického odporu medzi 100 a 1000 teplotnými cyklami boli menšie ako 1%.



Obr. 5 Závislosť zmeny elektrického odporu od počtu teplotných cyklov vnorených OhmegaPly rezistorov

Obr. 6 zobrazuje porovnanie zmeny elektrického odporu medzi vnorenými OhmegaPly rezistormi a OhmegaPly rezistormi umiestnenými na povrchu. Aj keď viacvrstvová štruktúra s vnorenými OhmegaPly rezistormi pozostáva z viacerých rôznych materiálov s rôznymi vlastnosťami, zmena elektrického odporu OhmegaPly rezistorov umiestnenými na povrchu bola výrazne vyššia ako u vnorených OhmegaPly rezistorov. Prepreg vrstva a FR-4 laminát slúžili ako ochranná vrstva pre vnorené OhmegaPly rezistory. OhmegaPly rezistory umiestnené na povrchu neboli pokryté žiadnou ochrannou vrstvou. Vnorenie OhmegaPly rezistorov zvyšuje stabilitu ich elektrického odporu.



Obr. 6 Závislosť zmeny elektrického odporu OhmegaPly rezistorov umiestnených na povrchu a vnorených do viacvrstvovej DPS

IV. ZHRNUTIE

Tento článok bol zameraný na vplyv tvaru OhmegaPly rezistorov na stabilitu ich elektrického odporu. Stabilita OhmegaPly rezistorov bola analyzovaná na základe zmien ich elektrického odporu počas teplotného cyklovania. Na základe realizovaných experimentov nemá tvar OhmegaPly rezistorov žiadny preukázateľný vplyv na stabilitu ich elektrického odporu. OhmegaPly rezistory realizované na povrchu so šírkou 0,5 mm po 800 tepelných cykloch občas vykazovali výrazné zvýšenie ich elektrického odporu. Vnorenie OhmegaPly rezistorov do viacvrstvovej DPS zvýšilo stabilitu ich elektrického odporu. Zmena elektrického odporu vnorených OhmegaPly rezistorov bola pre každý typ rezistoru menšia ako 3 %. Prepreg vrstva FR-4 a laminát slúžili ako ochranná vrstva pre vnorené OhmegaPly rezistory.

Najvýraznejší nesúlad elektrického odporu realizovaných OhmegaPly rezistorov bol medzi vypočítanými a nameranými hodnotami po procese leptania. Proces leptania sa ukázal ako

najkritickejšia časť výroby OhmegaPly rezistorov.

Realizované experimenty ukázali, že vnorené OhmegaPly rezistory majú stabilný elektrický odpor. OhmegaPly rezistory vďaka svojej malej hrúbke (do 1 μm) mohli byť vnorené do viacvrstvovej DPS bez zväčšenia jej výslednej hrúbky DPS. Vnorené OhmegaPly rezistory môžu nahradiť SMD súčiastky osadené na povrchu a prispieť k miniaturizácii DPS.

POĎAKOVANIE

Tento článok bol podporený projektom KEGA č. 011TUKE-4/2023, projektom APVV 22-0289, projektmi VEGA č. 1/0049/24 a č. 1/0278/24 a projektom ITMS kód 313011T557 (názov projektu „Podpora výskumno-vývojového potenciálu v oblasti dopravných prostriedkov“).

ZOZNAM POUŽITEJ LITERATÚRY

[1] Z. Li et al., “Review of thin-film resistor sensors: Exploring materials, classification, and preparation techniques”, Chemical Engineering Journal, roč. 477, s. 147029, dec. 2023, doi: 10.1016/j.cej.2023.147029.

[2] A. Dziedzic, P. Osypiuk, a W. Steplewski, “Stability of electrical properties for mechanically exposed thick- and thin-film resistors on flexible substrates”, SSMT, roč. 29, č. 1, s. 54–58, feb. 2017, doi: 10.1108/SSMT-11-2016-0030.

[3] G. Tomaszewski et al., “Investigation of inkjet printed path resistance in the context of manufacture and flexible application”, CW, roč. 45, č. 1, s. 15–25, feb. 2019, doi: 10.1108/CW-11-2018-0089.

[4] S. Kardos, S. Slosarcik, a P. Balog, “Sensor Array for Evaluation of Gait Cycle”, AEEE, roč. 17, č. 4, s. 459–465, dec. 2019, doi: 10.15598/aeee.v17i4.3352.

[5] Y. Zhai, T. Xu, G. Xu, X. Cao, Z. Tao, a H. Li, “Temperature Characteristics of 4H-SiC Substrate and Thin-Film Resistor Applied in MEMS Piezoresistive Sensors”, IEEE Sensors J., roč. 23, č. 24, s. 30231–30240, dec. 2023, doi: 10.1109/JSEN.2023.3330486.

[6] T. Rovensky, A. Pietrikova, I. Vehec, a L. Livovsky, “Stability of miniaturized non-trimmed thick- and thin-film resistors”, Microelectronics Reliability, roč. 84, s. 88–94, máj. 2018, doi: 10.1016/j.microrel.2018.03.011.

[7] P. Lukacs, A. Pietrikova, I. Vehec, a P. Provazek, “Influence of Various Technologies on the Quality of Ultra-Wideband Antenna on a Polymeric Substrate”, Polymers, roč. 14, č. 3, s. 507, jan. 2022, doi: 10.3390/polym14030507.

[8] A. Pietrikova, T. Lenger, L. Livovsky, I. Vehec, a P. Lukacs, “Moisture Absorption of Glass-Epoxy Sandwich Structure”, v 2022 International Conference on Diagnostics in Electrical Engineering (Diagnostika), Pilsen, Czech Republic: IEEE, sep. 2022, s. 1–5. doi: 10.1109/Diagnostika55131.2022.9905216.

[9] P. Osypiuk, A. Dziedzic, a W. Stęplewski, “Influence of mechanical exposures on electrical properties of thin and thick-film flexible resistors and conductors”, Soldering & Surface Mount Technology, roč. 28, č. 1, s. 33–38, feb. 2016, doi: 10.1108/SSMT-10-2015-0035.

[10] G. Zhou et al., “Co-deposition of C and Ni P to fabricate high-ohmic resistor for board-level integration with on-line monitoring method”, Microelectronics Reliability, roč. 142, s. 114906, mar. 2023, doi: 10.1016/j.microrel.2023.114906.

[11] T. Lenger, A. Pietrikova, a L. Livovsky, “Reliability of Embedded SMD Resistors Realized by Face–down Technology”, v 2021 44th International Spring Seminar on Electronics Technology (ISSE), Bautzen, Germany: IEEE, máj. 2021, s. 1–5. doi: 10.1109/ISSE51996.2021.9467671.

Úvod do kompresného snímania

Štefan HRIČÁK, Ondrej KOVÁČ

Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická Univerzita v Košiciach, Slovenská Republika

stefan.hricak@student.tuke.sk; ondrej.kovac@tuke.sk

***Abstrakt* — V súčasnosti sa pri neustálom zvyšovaní množstva digitálnych údajov sa význam kompresie údajov stáva čoraz výraznejším. Nakoľko exponenciálne narastá objem denne generovaných a vymieňaných informácií, potreba efektívneho ukladania, prenosu a spracovania údajov sa stáva nevyhnutnou. Kompresia údajov zohráva kľúčovú úlohu pri riešení týchto výziev tým, že zmenšuje veľkosť súborov údajov bez toho, aby boli ohrozené ich základné informácie, čím sa uľahčuje úspornejšie využívanie úložného priestoru a rýchlejší prenos údajov. Relatívne novým prístupom do tejto oblasti kompresie údajov je takzvané kompresné snímanie. Štandardné kompresné techniky používané v multimédiách, ako napríklad JPEG alebo MP3, využívajú na dosiahnutie kompresie redundanciu v rámci údajov. Na rozdiel od týchto techník, kompresné snímanie ponúka revolučný prístup ku kompresii údajov tvrdením, že určité signály možno rekonštruovať v dostatočnej kvalite z výrazne redukovaného súboru meraní, čo poskytuje účinnejší spôsob kompresie. Význam kompresného snímania spočíva nielen v jeho schopnosti komprimovať údaje, ale aj v jeho potenciáli revolučne zmeniť získavanie signálov. Túto schopnosť možno uplatniť hlavne v situáciách, kde je zber veľkého množstva údajov nákladný alebo časovo náročný. Tieto situácie sa môžu odohrávať v oblastiach ako napríklad medicína, diaľkové snímanie, telekomunikácie a v oblastiach kde je efektívne zachytávanie a reprezentácia údajov prvoradá. V rámci tejto práce bude bližšie priblížená problematika kompresného snímania.**

***Kľúčové slová* — kompresia dát, kompresné snímanie, rekonštrukčné algoritmy**

I. Úvod

Kompresné snímanie predstavuje nový prístup snímania údajov, pri ktorom je potrebný oveľa menším počet vzoriek, ako vyžaduje Shannonova-Nyquistova veta o vzorkovaní. Nyquistova frekvencia je definovaná ako minimálna vzorkovacia frekvencia, ktorá musí byť väčšia než dvojnásobok maximálnej frekvencie signálu podľa Nyquistovho teorému. Je zrejmé, že problémy nastanú, keď má signál veľmi veľkú šírku pásma, čo si vyžaduje veľmi vysokú vzorkovaciu frekvenciu a čo má za následok veľké množstvo nameraných údajov [1]. Napriek tomu dokáže kompresné snímanie získať signál s menším počtom vzoriek a následne pomocou matematických optimalizačných techník zrekonštruovať pôvodný signál. V dôsledku toho prináša táto technika spracovania signálov niekoľko výhod do informačno-komunikačných systémov. Medzi tieto výhody patrí napríklad vyššia rýchlosť prenosu, menší počet potrebných senzorických zariadení, menšie nároky na ukladanie dát, lepšia energetická účinnosť na úrovni senzoriky.

Kompresné snímanie funguje na základe skutočnosti, že údaje, ktoré nesú informácie, majú riedku povahu a riedkosť údajov je mierou informačnej hodnoty. Preto sa vzorky údajov lineárne miešajú a snímajú ako oveľa menší počet vzoriek [2]. Pôvodné vzorky údajov s väčším objemom sa načítajú z kompresne snímaných údajov, pričom sa berie do úvahy skutočnosť, že medzi nekonečne možných riešení neúplného súboru lineárnych rovníc je to, ktoré má s najväčšou riedkosťou informačnú hodnotu a môže reprezentovať pôvodné údaje.

A. Aplikačné možnosti

V rámci aplikačných možností preukazuje kompresné snímanie obrovský potenciál. Jednou z oblasti v ktorej kompresné snímanie pritiahlo veľkú pozornosť výskumníkov je práve medicína. V posledných rokoch sa zvýšil záujem o ambulantné monitorovanie bioelektrických signálov. Toto monitorovanie má veľké výhody, ako je zvýšená mobilita pacienta, neustále pozorovanie

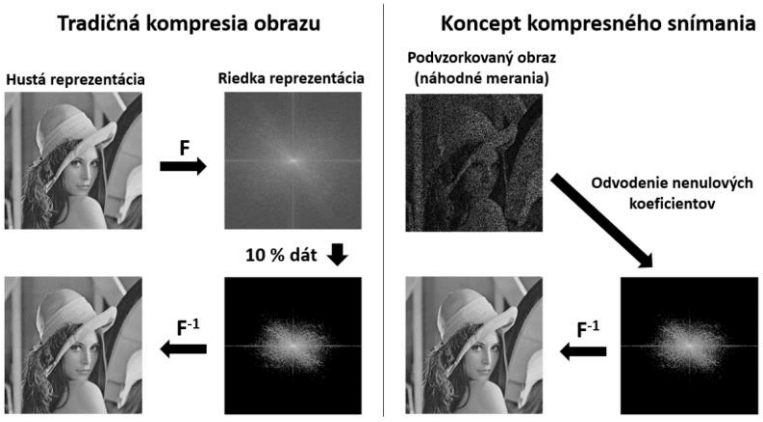
pacienta a zníženie nákladov na zdravotnú starostlivosť. Dôležitým klinickým nástrojom, ktorý sa používa na monitorovanie a rozšírenie poznatkov o mozgu, je elektroencefalogram (EEG). Toto zariadenie slúži na zaznamenávanie elektrickej aktivity mozgu a je schopné diagnostikovať neurologické ochorenia alebo poruchy [3]. Okrem elektroencefalogramu (EEG) si našlo kompresné snímanie uplatnenie aj napríklad v rýchlej magnetickej rezonancií (MRI), v zrýchlenej dynamickej MRI, rekonštrukcii počítačovej tomografie, ale aj v elektrokardiograme (EKG). EKG funguje tak, že zachytáva elektrické aktivity srdca. Každému pohybu kontrakcie alebo relaxácie tohto orgánu predchádza elektrický signál. Správne zachytenie týchto signálov umožňuje lekárovi skontrolovať, či srdce funguje správne [4].

Medzi ďalšiu aplikačnú oblasť patria radarové systémy. Po vyslaní určitého druhu impulzov sa v bežnom radarovom systéme používa prispôsobený filter, ktorý sa používa na koreláciu prijímaného signálu s vysielanými impulzmi. Následne sa na strane prijímača používa analógovo-digitálny prevodník, ktorého úlohou je komprimovať prijaté impulzy s vysokou rýchlosťou prenosu údajov [5]. To predstavuje problém nakoľko takáto technológia je veľmi náročná a nákladná. Hlavným konceptom takzvaného komprimovaného zobrazovania RADAR-u je reprezentácia časovo-frekvenčnej zložky do diskrétneho formátu a premietanie každej scény do matice. V prípade, ak bude detegovaných cieľov málo, matica bude riedka a preto bude možné ľahko použiť kompresné snímanie na rekonštrukciu cieľovej scény [5].

Okrem spomenutých aplikačných oblastí možno hľadať kompresné snímanie aj napríklad v lokalizačných službách, spracovaní signálov, telekomunikáciách, vo vylepšovaní obrazu či jeho adaptácii, spracovaní textových údajov, v poľnohospodárskych strojoch, v ťažbe dát, v systémoch verejnej dopravy, v plánovaní energetických systémoch a mnohých ďalších [6].

B. Porovnanie tradičnej kompresie oproti kompresnému snímaniu

Čoraz väčšie množstvo informácii obsiahnutých v obrazových súboroch si vyžaduje efektívnejšie spôsoby ukladania. Táto požiadavka viedla k vývoju celého radu techník na kompresiu obrazu. Napríklad Shannon-Fanovo kódovanie, Huffmanovo kódovanie, algoritmy LZ78 a LZW, JPEG a mnohé ďalšie. Niektoré techniky využívajú transformačné kódovanie nakoľko tento prístup umožňuje dosiahnuť vyššiu úroveň kompresie [7]. Pod pojmom transformácia si je možné predstaviť matematickú operáciu, ktorá zmapuje vybranú funkciu alebo sekvenciu do určitej transformačnej domény. Prostredníctvom transformácie je možné jednoduchšie vyriešiť daný problém, zistiť dodatočnú informáciu o danej funkcii, jednoduchšie pracovať s danou funkciou. Medzi najznámejšie transformácie patrí Diskrétna Kosínusová Transformácia (DCT), Diskrétna Fourierová Transformácia (DFT), Diskrétna Waveletová Transformácia (DWT).

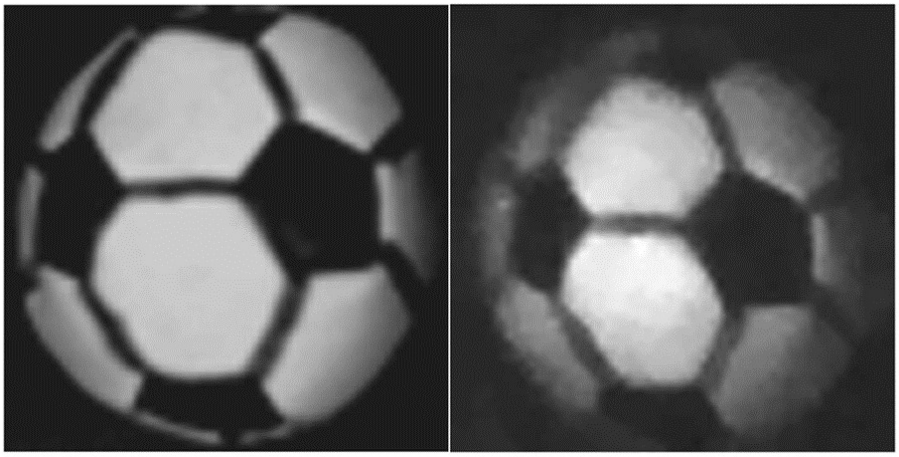


Obr. 1 Porovnanie tradičnej kompresie obrazu (vľavo) a konceptu kompresného snímania (vpravo)

Na Obr. 1 možno vidieť porovnanie tradičnej kompresie obrazu a konceptu kompresného snímania. V prípade tradičnej kompresii obrazu je postup nasledovný. V prvom rade je potrebné daný obraz zachytiť, resp. odfotiť. Tento obrázok predstavuje takzvanú hustú reprezentáciu nakoľko možno vidieť podrobne všetky jeho detaily.. Následne sa tento obraz transformuje, napríklad pomocou spomenutých transformácii, do domény v ktorej bude jeho reprezentácia riedka. V ďalšom kroku bude použitý proces prahovania koeficientov. V prípade, ak sa hodnota koeficientu nachádza pod nastavenou hranicou, tak dôjde k jej vynulovaniu. Následne je možné inverznou transformáciou dostať komprimovaný obraz. Tento proces kompresie sa radí do skupiny stratových kompresných metód nakoľko pri procese zahodenia transformačných koeficientov dochádza k strate informácie.

Kompresné snímanie predstavuje nový spôsob k samotnej kompresii. Hlavnou myšlienkou je zbytočnosť úplných meraní pre opis obrazu. Matematici z rôznych amerických univerzít sa

zamysleli nad otázkou, či je nutné uskutočňovať milióny meraní, pokiaľ je možné dosiahnuť adekvátny opis obrazu prostredníctvom oveľa menšieho počtu, napríklad len niekoľko desiatok tisíc meraní.[8]. V roku 2004 David Donoho, Justin Romberg, Terence Tao, Emmanuel Candes predstavili svetu koncept kompresného snímania. O pár rokov neskôr Richard Baraniuk a Kevin Kelly pre dokázanie konceptu kompresného snímania vyvinuli kameru s jedným pixelom, prostredníctvom ktorej realizovali experiment. Táto kamera realizovala proces zobrazovania osvetlenia scény svetelnými vzormi a proces snímania jedným svetelným detektorom, ktorý meral intenzitu svetla po interakcii so scénou [9].



Obr. 2 Pôvodná fotografia (vľavo) a zrekonštruovaná fotografia pomocou kompresného snímania (vpravo)

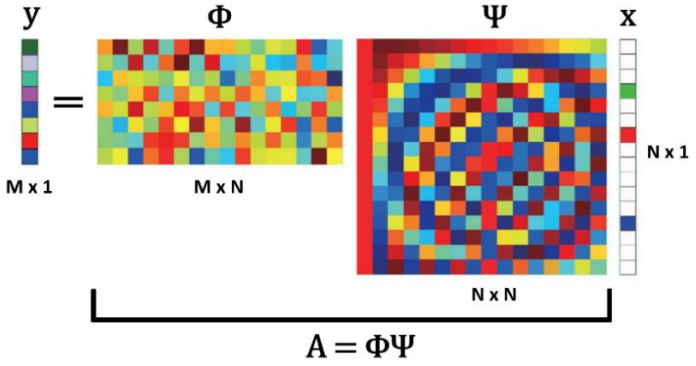
Na Obr. 2 vľavo možno vidieť pôvodnú fotografiu futbalovej lopty, ktorá mala rozmer 64x64 pixelov (spolu 4096 pixelov) a bola použitá v experimente. Túto fotografiu následne odfotili náhodne prostredníctvom jedno-pixelovej kamery 1600 krát. Následne prostredníctvom kompresného snímania sa im podarilo zrekonštruovať pôvodnú fotografiu, ako možno vidieť na Obr.2 vpravo a následne tak potvrdiť samotný koncept.

II. Matematický Model kompresného snímania

V rámci tejto kapitoly budú popísané všetky časti matematického modelu kompresného snímania. Kompresné snímanie možno v štandardnom prípade modelovať podľa vzťahu:

$$y = Ax = \Phi\Psi x \tag{1}$$

Z matematického hľadiska sa v kompresnom snímaní hľadajú riešenia nedeterminovaných lineárnych sústav rovníc pomocou optimalizačných techník. Cieľom kompresného snímania je teda obnovenie pôvodného signálu x [10]. Premenná y predstavuje vektor snímaných dát a písmeno M predstavuje počet prvkov daného vektora. Ako už bolo povedané neznáma x predstavuje pôvodný signál, o ktorého rekonštrukciu sa snaží kompresné snímanie a písmeno N predstavuje počet prvkov daného vektora. Premenná A sa nazýva matica kompresného snímania alebo taktiež snímacia matica. Táto matica definuje proces merania. Predstavuje spôsob pozorovania alebo snímania pôvodného signálu počas kompresnej fázy. Štruktúra a vlastnosti matice A sú rozhodujúce pre úspech kompresného snímania. Táto matica sa skladá z ďalších dvoch matíc. Z transformačnej matice Ψ a meracej matice Φ.



Obr. 3 Matematický model kompresného snímania

A. Transformačná matica Ψ

Aby bolo možné realizovať kompresné snímanie je nevyhnutné, aby daný signál mal riedku reprezentáciu v určitej transformačnej doméne, ako už bolo spomenuté. Nakoľko prirodzené signály nie sú v časovej oblasti dostatočne riedke, je potrebné ich transformovať [10]. Tento účel v kontexte kompresného snímania plní transformačná matica.

$$s = \Psi x \tag{2}$$

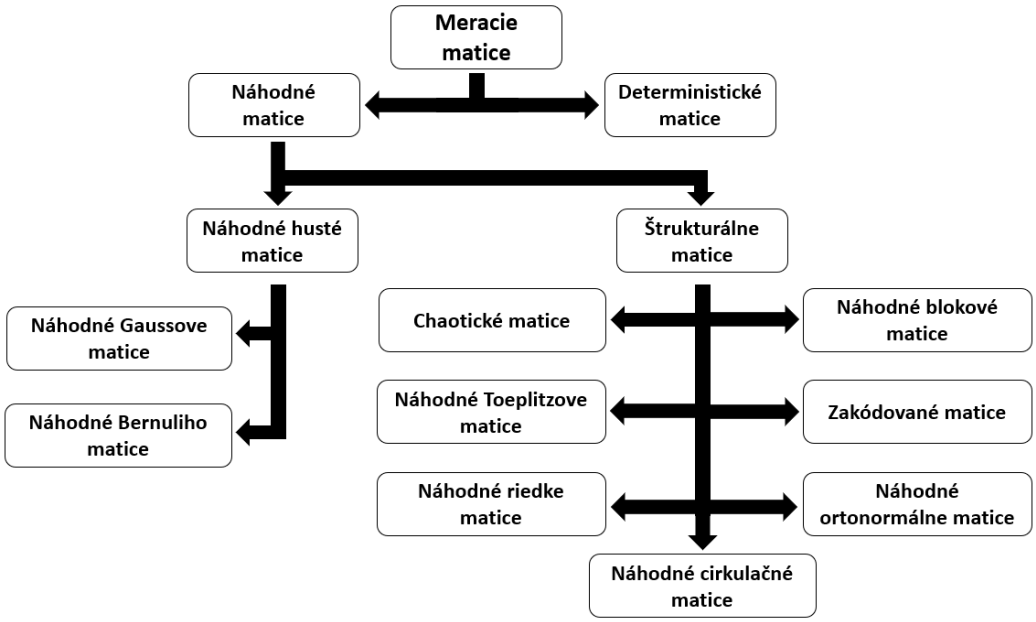
Ak uvažujeme signál x v časovej oblasti s veľkosťou N x 1, je možné ho transformovať do inej domény vynásobením ortogonálnou štvorcovou transformačnou maticou Ψ s veľkosťou N x N, čím sa získa riedky signál s, s rovnakou veľkosťou ako signál x. Veľkosť N je dĺžka signálu rovnako v časovej a aj transformovanej oblasti. Riedky signál s má iba K nenulových prvkov, pričom platí že K je oveľa menšie ako N.

B. Meracia matica Φ

Po transformovaní signálu do oblasti, kde má riedku transformáciu sa vykoná kompresia vynásobením riedkej reprezentácie signálu s vzájomne s meracou maticou Φ, ktorá ma veľkosť M x N. Cieľ je vytvoriť výsledný komprimovaný signál y s veľkosťou M x 1.

$$y = \Phi s \tag{3}$$

Veľkosť M predstavuje počet meraní výsledného signálu y, ktoré sa použijú na rekonštrukciu pôvodného signálu x. V dôsledku kompresie je veľkosť y menšia ako x, pričom platí K < M < N.



Obr. 4 Rozdelenie meracích matíc

Meracie matice sa rozdeľujú do dvoch hlavných skupín, deterministické a náhodne matice. Po vynásobení s riedkym signálom sa vytvorí komprimovaný signál spĺňajúci požiadavky rekonštrukčných algoritmov. Náhodné husté matice sú najčastejšie používaným typom matíc vrátane náhodných Gaussových a náhodných Bernoulliho matíc, pretože sú univerzálne nekoherentné s akýmkoľvek riedkym signálom. To znamená, že počet komprimovaných meraní potrebných na dostatočnú rekonštrukciu je takmer minimálny. Napriek svojim výhodám trpia potrebou obrovskej vyrovnávacej pamäte na ukladanie prvkov matice a vysokou výpočtovou zložitosťou v dôsledku ich neštruktúrovanej povahy. Spomenuté nedostatky náhodných matíc dokážu do istej miery napraviť štrukturálne náhodne matice. Je to možné hlavne vďaka ich organizovanej štruktúry, vďaka čomu umožňujú jednoduchú implementáciu a primerané nároky na pamäťové úložisko. Na Obr. 4 možno vidieť rozdelenie meracích matíc používaných v kompresnom snímaní.

C. Podmienky pre snímaciu maticu A

Snímacia matica predstavuje kľúčovú zložku pri procese rekonštrukcie pôvodného signálu x. Táto matica musí byť nezávislá od pôvodného signálu x, preto by mala byť navrhnutá v procese

neadaptívneho merania. To znamená, že súčasné merania nezávisia od predchádzajúcich. V prvých fázach vývoja kompresného snímania sa odborníci zaujímali o návrh vhodných snímacích matíc. Bolo to v dôsledku toho aby sa zabezpečilo, že sa signál nepoškodí alebo nestratí v dôsledku transformácie [10]. Preto boli naformulované dodatočné podmienky, ktoré viac menej zaručujú bezproblémovú rekonštrukciu signálu. Medzi tieto podmienky patrí napríklad vzájomná koherencia, vlastnosť nulového priestoru (z angl. Null space property) a vlastnosť obmedzenej izometrie (z angl. Restricted Isometry Property).

1) Vzájomná koherencia

Vzájomná koherencia snímacej matice používanej pri kompresnom snímaní je mierou podobnosti medzi stĺpcami matice. Je definovaná ako maximálny absolútny skalárny súčin medzi stĺpcami matice normalizovaný stĺpcovými normami. Menšia vzájomná koherencia znamená, že stĺpce matice sú takmer ortogonálne, čo je pre aplikáciu kompresného snímania nevyhnutnosťou.

$$\mu(A) = \max_{i \neq j, 1 \leq i,j \leq N} \left| \frac{\langle A_i, A_j \rangle}{\|A_i\|^2 \|A_j\|^2} \right| \tag{4}$$

Vzájomná koherencia zabezpečuje dobré záruky obnovy a ak je jej hodnota malá, môže sa zlepšiť výkonnosť algoritmov kompresného snímania [11].

2) Vlastnosť nulového priestoru

Vlastnosť nulového priestoru predstavuje jednu z nevyhnutných podmienok, ktorá zaručuje rekonštrukciu riedkych signálov prostredníctvom optimalizačných techník, ako napríklad prostredníctvom L1-minimalizácie [12]. Táto podmienka definuje, že pre ľubovoľný r-riedky vektor (najviac r nenulových prvkov) x je jediným riešením rovnice Ax = 0, triviálne riešenie x = 0.

$$\mathbb{N}(A) = \{x : Ax = 0\} \tag{5}$$

Vlastnosť nulového priestoru je dôležitá, pretože zaručuje, že riešenie problému, prostredníctvom napríklad L1-minimalizácie, je jedinečné a rovná sa pôvodnému riedkemu signálu. Avšak v prítomnosti šumu je potrebné silnejšiu podmienku.

3) Obmedzená vlastnosť izometrie

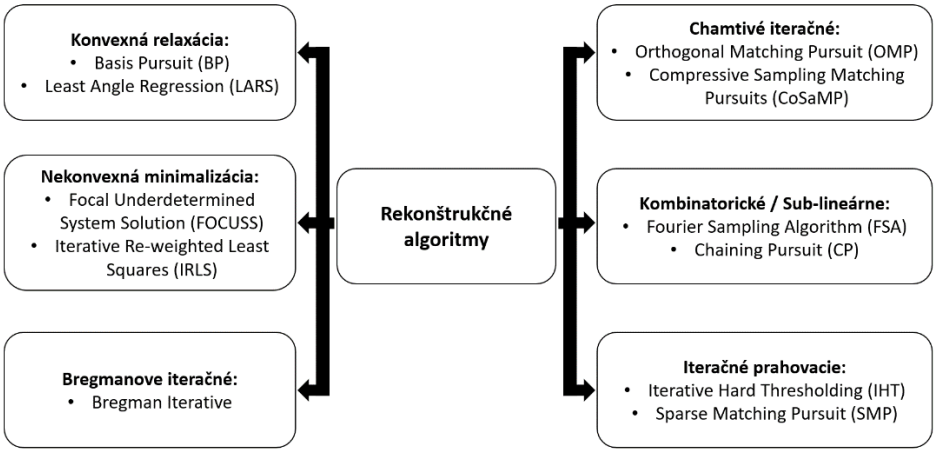
Obmedzená vlastnosť izometrie charakterizuje neštvorcové matice, ktoré sú takmer ortonormálne pri práci s riedkymi alebo stlačiteľnými signálmi [10]. Táto vlastnosť definuje ako dobre snímacia matica zachováva vzdialenosti medzi riedkymi vektormi. O snímacej matici možno povedať, že spĺňa podmienku obmedzenej vlastnosti izometrie, ak všetky možné výbery K stĺpcov matici A sú podobné ortonormálnej matici. Obmedzená izometrická konštanta δ_K hovorí, ako dobre snímacia matica A zachováva energiu komprimovaného signálu. Ak však snímacia matica nevyhovuje podmienke obmedzenej vlastnosti izometrie, potom nemožno zistiť, či signál sa dá obnoviť alebo nie.

$$(1 - \delta_K)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_K)\|x\|_2^2 \tag{6}$$

Vo všeobecnosti môže byť obmedzená vlastnosť izometrie výpočtovo ťažko riešiteľná podobne, ako v prípade vlastnosti nulového priestoru. Namiesto hľadania vhodných meracích matíc bolo dokázané, že náhodne meracie matice túto podmienku spĺňajú s vysokou pravdepodobnosťou. Gaussové náhodné matice preukazujú dobré izometrické konštanty, podobne aj ostatné meracie matice spomenuté na Obr. 4. Ak neexistuje špecifický typ presnej aplikácie kompresného snímania, ako napríklad operácia v konečných poliach, tieto matice predstavujú najjednoduchšiu a jednu z najlacnejších možností.

III. Rekonštrukčné algoritmy

Po úspešnom návrhu modelu kompresného snímania a dodržaní potrebných podmienok nasleduje posledný krok, samotná rekonštrukcia. Ako už bolo povedané samotná rekonštrukcia zahŕňa riešenie nedeterminovaných sústav lineárnych rovníc, ktoré majú nekonečne veľa riešení. Proces rekonštrukcie signálu spočíva v podstate vo výbere najlepšieho odhadu pôvodného signálu zo všetkých možných riešení [13]. To sa dá zrealizovať prostredníctvom konvexného optimalizačného algoritmu alebo mnohých ďalších, ktoré možno vidieť na Obr. 5.



Obr. 5 Rozdelenie rekonštrukčných algoritmov

Rozdelenie rekonštrukčných algoritmov do rôznych skupín je založené na základných princípoch a metodikách, ktoré používajú na obnovu riedkych signálov z redukovaného súboru meraní. Výber algoritmu často závisí od špecifických vlastností problému, úrovne riedkosti, prítomnosti šumu a výpočtových hľadísk. Každá skupina má svoje silné a slabé stránky, vďaka čomu sú vhodné pre rôzne scenáre a aplikácie.

IV. Záver

Cieľom tejto práce bolo poskytnúť prehľad do pomerne novej techniky kompresie údajov s názvom kompresné snímanie. Bližšie boli popísané začiatky samotného konceptu, taktiež bolo porovnané kompresné snímanie s tradičným spôsobom kompresie dát. Následne boli taktiež spomenuté aplikačné sféry danej techniky. Bližšie bola popísaná štruktúra matematického modelu kompresného snímania a taktiež boli popísané jej jednotlivé časti. Spomenuté boli taktiež podmienky, ktoré garantujú úspešnú rekonštrukciu pôvodného signálu prostredníctvom kompresného snímania. V poslednej časti boli taktiež popísané rekonštrukčné algoritmy.

Použitá literatúra

[1] D. Fonseca Resende, M. Khosravy, H. L. M. Monteiro, N. Gupta, N. Patel, a C. A. Duque, "Chapter 11 - Neural signal compressive sensing", v *Compressive Sensing in Healthcare*, M. Khosravy, N. Dey, a C. A. Duque, Ed., Academic Press, 2020, s. 201–221. doi: 10.1016/B978-0-12-821247-9.00016-0.

[2] M. Khosravy, N. Nitta, N. Gupta, N. Patel, a N. Babaguchi, "Chapter 3 - A descriptive review to sparsity measures", v *Compressive Sensing in Healthcare*, M. Khosravy, N. Dey, a C. A. Duque, Ed., Academic Press, 2020, s. 43–63. doi: 10.1016/B978-0-12-821247-9.00008-1.

[3] D. Gurve, D. Delisle-Rodriguez, T. Bastos-Filho, a S. Krishnan, "Trends in Compressive Sensing for EEG Signal Processing Applications.", *Sensors*, roč. 20, č. 13, júl. 2020, doi: 10.3390/s20133703.

[4] T. W. Cabral *et al.*, "Chapter 4 - Compressive sensing in medical signal processing and imaging systems", v *Sensors for Health Monitoring*, roč. 5, N. Dey, J. Chaki, a R. Kumar, Ed., Academic Press, 2019, s. 69–92. doi: 10.1016/B978-0-12-819361-7.00004-X.

[5] V. Upadhyaya a D. M. Salim, "Compressive Sensing: Methods, Techniques, and Applications", *IOP Conf. Ser. Mater. Sci. Eng.*, roč. 1099, č. 1, s. 012012, mar. 2021, doi: 10.1088/1757-899X/1099/1/012012.

[6] M. Khosravy, N. Gupta, N. Patel, a C. A. Duque, "Chapter 2 - Recovery in compressive sensing: a review", v *Compressive Sensing in Healthcare*, M. Khosravy, N. Dey, a C. A. Duque, Ed., Academic Press, 2020, s. 25–42. doi: 10.1016/B978-0-12-821247-9.00007-X.

[7] V. Bairagi, A. M. Sapkal, a M. S. Gaikwad, "The Role of Transforms in Image Compression", *J Inst Eng India Ser B*, roč. 94, s. 135–140, jún. 2013, doi: 10.1007/s40031-013-0049-9.

[8] Dana Mackenzie, "What's Happening in the Mathematical Sciences", v 1. , American Mathematical Society, 2008.

[9] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling", *IEEE Signal Process. Mag.*, roč. 25, č. 2, s. 83–91, 2008, doi: 10.1109/MSP.2007.914730.

[10] M. Taghouti, "Chapter 10 - Compressed sensing", v *Computing in Communication Networks*, F. H. P. Fitzek, F. Granelli, a P. Seeling, Ed., Academic Press, 2020, s. 197–215. doi: 10.1016/B978-0-12-820488-7.00023-2.

[11] Q. Xu, Z. Sheng, Y. Fang, a L. Zhang, "Measurement Matrix Optimization for Compressed Sensing System with Constructed Dictionary via Takenaka-Malmquist Functions.", *Sensors*, roč. 21, č. 4, feb. 2021, doi: 10.3390/s21041229.

[12] J. Cahill, X. Chen, a R. Wang, "The gap between the null space property and the restricted isometry property", *Linear Algebra Its Appl.*, roč. 501, s. 363–375, júl. 2016, doi: 10.1016/j.laa.2016.03.022.

[13] S. Sreeharitha, G. Sabarinath, a B. R. Jose, "Compressive Sensing Recovery Algorithms and Applications- A Survey", *IOP Conf. Ser. Mater. Sci. Eng.*, roč. 396, č. 1, s. 012037, aug. 2018, doi: 10.1088/1757-899X/396/1/012037.

Prehľad metód kalibrácie vnútorných parametrov kamery

Lukáš FERENC, Ondrej KOVÁČ

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

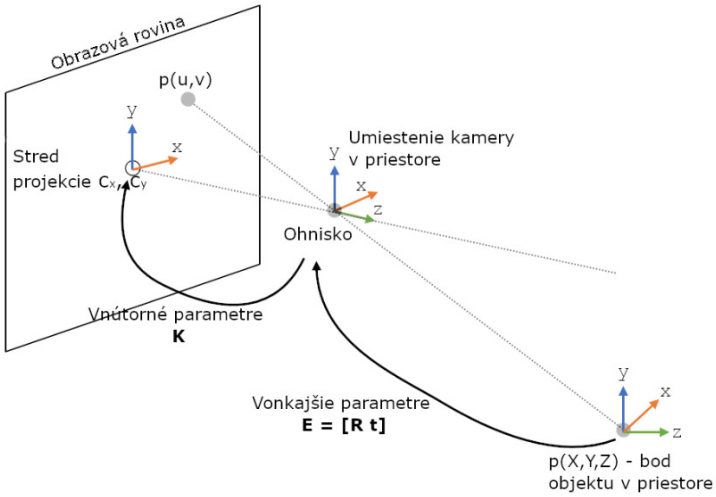
lukas.ferenc@student.tuke.sk; ondrej.kovac@tuke.sk

***Abstrakt* — Cieľom tejto práce je stručne popísať problematiku kalibrácie digitálnych kamier a oboznámiť čitateľa s rôznymi metódami kalibrácie kamery, ich algoritmom a výhodami či nevýhodami. Táto práca opisuje tri metódy odhadu vnútorných parametrov kamery. Prvé dve sú Zhangová a Tsaiová metóda. Obe metódy sú dlhodobo známe a v súčasnosti sú často používané v širokej škále oblastí od tzv. domácich, či školských experimentov spracovania obrazu až po masovú produkciu digitálnych kamier spracúvajúcich obraz. Zhangová metóda sa vyznačuje najmä svojou univerzálnosťou a Tsaiová metóda vyššou presnosťou odhadu vnútorných parametrov kamery. Okrem týchto dvoch metód boli časom vyvinuté ďalšie metódy pre odhad vnútorných parametrov, napríklad aj tzv. hybridné metódy, ktoré kombinujú algoritmy iných metód. Metóda odhadu vnútorných parametrov pomocou testovacieho poľa je práve jednou z hybridných metód kombinujúcich časti algoritmov Zhangovej a Tsaiovej metódy. Odhad vnútorných parametrov kamery touto metódou je veľmi presný, avšak táto metóda nie je univerzálna, keďže testovacie pole je väčšinou prispôsobené určitému typu kamier.**

***Kľúčové slová* — Vnútorné parametre kamery, vonkajšie parametre kamery, digitálna kamera, skreslenie obrazu,**

I. Úvod

Pojem kalibrácia digitálnej kamery označuje softvérový proces eliminácie tangenciálneho a radiálneho skreslenia. Je to nevyhnutnou požiadavkou pre získavanie metrických dát z obrazu, fotogrametrii - rekonštrukcie 3D objektu zo sérii snímok, spracovanie obrazu pre detekciu objektov v reálnom čase, aplikácie stereo videnia alebo rozšírenej reality [1].
Najväčšiu časť algoritmu kalibrácie kamery tvorí odhad parametrov kamery. Tieto parametre definujú vzťah medzi súradnicami bodov v trojrozmernom priestore a súradnicami týchto bodov v obrazovej rovine. Parametre kamery sa delia na vnútorné a vonkajšie. Na obr. 1 je graficky znázornený princíp perspektívnej projekcie bodov z 3D priestoru do 2D obrazovej roviny so znázornením vnútornej a vonkajšej oblasti kamery.



Obr. 1 Perspektívna projekcia bodu z 3D priestoru do 2D obrazovej roviny [2]

Vonkajšie parametre kamery definujú polohu a náklon kamery v priestore, teda transláciu a rotáciu kamery. Nech E je matica 3×4, potom vonkajšie parametre kamery sú definované maticou E podľa vzťahu *(1)*.

$$E = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \quad (1)$$

Vnútorné parametre kamery definujú, ako sa premietne bod z 3D priestoru prechádzajúci šošovkou na 2D obrazovú rovinu. Tieto parametre zahŕňajú ohniskovú vzdialenosť f_x a f_y v dvoch osiach, súradnice bodu c_x a c_y, v ktorom os šošovky pretína obrazovú rovinu a parametre skreslenia obrazu. Nech K je matica 3x3 vnútorných parametrov kamery, potom táto matica je definovaná vzťahom *(2)* [3].

$$K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

Matica vnútorných parametrov kamery pôvodne obsahovala aj parameter *S*, ktorý definoval prípadný strih výsledného obrazu spôsobeného tangenciálnym skreslením, v súčasnosti sa ale už nezvykne udávať [4].

K tangenciálnemu skresleniu dochádza vtedy, ak os šošovky nie je kolmá na rovinu snímacieho CCD alebo CMOS čipu a to buď dôsledkom nepresného osadenia šošovky, ak je šošovka lepená na PCB potom veľkú rolu zohráva aj proces vytvrdzovania tohto lepidla a jeho homogenita. Dôsledok tangenciálneho skreslenia je napríklad zmena tvaru výsledného obrazu z obdĺžnika napríklad na lichobežníkový tvar, avšak rovnosť čiar ostáva zachovaná. Naopak radiálne skreslenie je prirodzený jav, spôsobený samotnou geometriou šošovky a na rozdiel od tangenciálneho skreslenia, nie je ho možné eliminovať v procese výroby fotosnímacieho zariadenia. Tento typ skreslenia nezanecháva rovnosť čiar, čo je najlepšie viditeľné pri šošovkách typu rybie oko najmä na ich okrajoch [3]. Na obr. 2 je znázornené radiálne skreslenie a jeho korekcia jednou z metód kalibrácie kamery. Z obr. 2 je zrejmé, že obraz je deformovaný hlavne na jeho krajoch.



a) b)

Obr. 2 a) Skreslená snímka dôsledkom radiálneho skreslenia b) Aplikovaná korekcia skreslenia na pôvodnú snímku

II. ZHANGOVA METÓDA KALIBRÁCIE KAMERY

Kalibrácia kamery Zhangovou metódou je v súčasnosti najznámejšou a najpoužívanejšou metódou odhadu vnútorných parametrov kamery. Vyznačuje sa najmä svojou jednoduchosťou a nenáročnosťou, keďže pre kalibráciu postačí rovná doska so šachovnicovým vzorom, pričom pre odhad vnútorných parametrov stačí poznať rozmer strany jedného štvorca. Okrem toho, táto metóda je univerzálna a aplikovateľná pre každú digitálnu kameru [1].

Princíp odhadu vnútorných parametrov digitálnej kamery Zhangovou metódou spočíva v prvom kroku v detekcii šachovnicového vzoru a jeho pravého dolného rohu, ktorý je počiatkom trojrozmerného súradnicového systému, tak, ako to je znázornené na obr. 3. Detekcia kalibračného vzoru - šachovnice je dosiahnutá detektorom hrán ako napríklad ako známy Harrisov detektor rohov. Po úspešnej detekcii šachovnicového vzoru z obrazu nasleduje odhad vonkajších parametrov kamery, teda orientácia kamery v priestore daná transláciou a rotáciou kamery. Tieto parametre sú vypočítané na základe nadväznosti dvoch a viacerých snímok [1]. Vzťah medzi dvoma na seba nadväzujúcimi snímkami možno definovať použitím detektora významných bodov SIFT alebo jeho vylepšenej verzie SURF, ktoré každému významnému bodu priradia unikátny deksriptor charakterizujúci tento bod a vytvoria páry týchto bodov z oboch snímok [5]. Pre výpočet vzťahov medzi dvoma snímkami je potrebné nájsť aspoň 3 skutočné páry bodov. Tu môžu vznikať aj falošné páry bodov, ktoré je potrebné použitím algoritmu RANSAC eliminovať. Viac o týchto algoritmoch sa môže čitateľ dozvedieť napríklad v problematike spájania obrazov [6].

Po vyjadrení vzťahov medzi jednotlivými snímkami sú vypočítané vonkajšie parametre kamery použitím priamej lineárnej transformácie – DLT. Priama lineárna transformácia je metóda výpočtu vzťahu medzi bodom v 3D priestore a týmto istým bodom premietnutým do obrazovej

roviny. Mapovanie bodu z 3D priestoru do obrazovej roviny sa v oblasti spracovania obrazov nazýva aj ako homografia a toto mapovanie je definované rovnicou *(3)* [7].

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = H \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \tag{3}$$

Kde H je matica homografie 3×3, x,y sú súradnice bodu v obrazovej rovine a X,Y sú súradnice toho istého bodu v rovine šachovnice. Hodnoty matice homografie H sú vypočítané pomocou matematickej metódy singulárneho rozkladu – SVD [7].
Po výpočte matice homografie H nasleduje výpočet vnútorných parametrov podľa vzťahu *(4)*.

$$H_i = K R_i \tag{4}$$

Kde H_i je matica homografie 3x3 pre i-ty obraz, K je matica vnútorných parametrov a R_i je matica 4x3 opisujúca rotáciu a transláciu kamery. Ak $H_i = [\,h_1 \quad h_2 \quad h_3]$ a $R_i = [\,r_1 \quad r_2 \quad t]$, potom úpravou vznikajú rovnice *(5)* a *(6)*.

$$[\,r_1 \quad r_2 \quad t] = K^{-1}[\,h_1 \quad h_2 \quad h_3] \tag{5}$$

Z čoho vyplýva:

$$\begin{aligned} r_1 &= K^{-1}\, h_1 \\ r_2 &= K^{-1}\, h_2 \end{aligned} \tag{6}$$

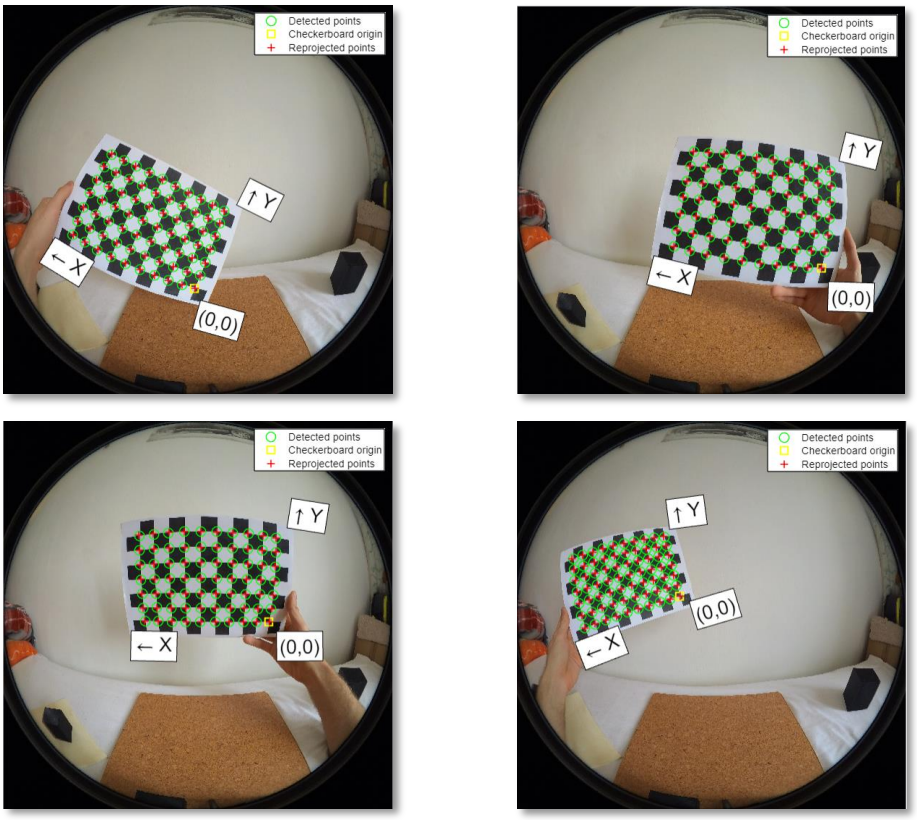
Z rovníc *(6)* sú neznáme r_1, r_2 a K. Na základe určitých vlastností rotácie pre vektory r_1, r_2 platí, že

$$\begin{aligned} {r_1}^T \cdot r_2 &= 0 \\ \|r_1\| &= \|r_2\| = 1 \end{aligned} \tag{7}$$

Dosadením a úpravou rovníc *(6)* do rovníc *(7)* vzniká:

$$\begin{aligned} {h_1}^T K^{-T} K^{-1}\, h_2 &= 0 \\ {h_1}^T K^{-T} K^{-1}\, h_1 - {h_2}^T K^{-T} K^{-1}\, h_2 &= 0 \end{aligned} \tag{8}$$

Keďže matica K má 5, resp. 6 stupňov voľnosti – neznáme, vzťahy *(8)* predstavujú sústavu dvoch rovníc o 5 neznámych. K vypočítaniu neznámych je potrebné získať ešte minimálne dve rovnaké sústavy z ďalších dvoch snímok šachovnice, pričom matica K je pre všetky snímky rovnaká [7].
Na obr. 3 sú znázornené štyri vstupné snímky a na obr. 2, je znázornená korekcia skresleného obrazu pomocou tejto metódy.



Obr. 3 Vstupné obrazy s detegovaným kalibračným vzorom pre kalibráciu kamery Zhangovou metódou

III. TSAIOVA METÓDA KALIBRÁCIE KAMERY

Tsaiova metóda na rozdiel od Zhangovej metódy odhaduje vnútorné parametre kamery pomocou trojrozmerných objektov, ktorých rozmery a poloha v priestore sú známe. Ku kalibrácii kamery touto metódou možno pristupovať dvoma spôsobmi. Pri prvom spôsobe nie je definovaná presná poloha kamery v priestore, preto je potrebné zhotoviť aspoň dve fotografie kalibračných objektov, na základe ktorých budú vonkajšie parametre kamery. Druhý spôsob spočíva v manuálnom definovaní vonkajších parametrov do algoritmu kalibrácie kamery [4].
Nasleduje výpočet vnútorných parametrov kamery, kde namiesto použitia priamej lineárnej transformácie, ako tomu bolo pri Zhangovej metóde, je použitý Lavenberg-Marquardtov algoritmus, inak nazývaný ako metóda najmenších štvorcov pre riešenie nelineárnych problémov. To spočíva v hrubom odhade vnútorných parametrov kamery a následne každou iteráciou riešením lineárnych rovníc je chyba odhadu menšia [8].
Matematicky to možno vyjadriť nasledovne. Nech *Xw, Yw a Zw* sú súradnice jedného bodu kalibračného objektu, *Xc, Yc a Zc* sú súradnice toho istého bodu ale vzhľadom ku kamere definované podľa vzťahu *(9)* [8].

$$(X_C \quad Y_C \quad Z_C) = R(X_W \quad Y_W \quad Z_W)^T + T \tag{9}$$

R je rotačná matica 3×3 a *T* je translačná matica 3×1. Nech *x, y* sú súradnice výsledného obrazu bez skreslenia, potom sú definované nasledovne.

$$\begin{aligned} x &= s \cdot \frac{X_C}{Z_C} \\ y &= s \cdot \frac{Y_C}{Z_C} \end{aligned} \tag{10}$$

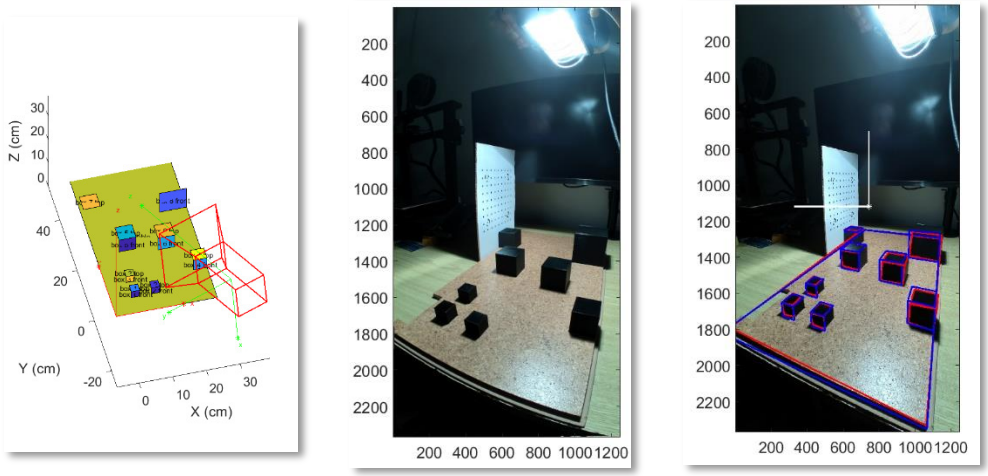
Kde *s* je škála daná súčinom ohniskovej vzdialenosti a rozmerom jedného pixelu. Uvažujúc aj samotné skreslenie, potom výsledné súradnice obrazu a x_D, y_D so zahrnutím skreslenia sú dané vzťahom *(11)*.

$$\begin{aligned} x_D &= x \cdot (1 + K_1 \cdot r^2 + K_2 \cdot r^4) \\ y_D &= y \cdot (1 + K_1 \cdot r^2 + K_2 \cdot r^4) \end{aligned} \tag{11}$$

Kde *K* sú koeficienty skreslenia a *r* radiálna vzdialenosť od stredu obrazu daný súradnicami c_x a c_y.

$$r = \sqrt{(x - c_x)^2 + \left(y - c_y\right)^2} \tag{12}$$

Praktická ukážka kalibrácie touto metódou a kalibračných objektov je znázornená na obr. 4, kde za a) je znázornené zadefinovanie trojrozmerných objektov pre kalibráciu, jedná sa o kocky s rozmermi 30×30, 50×50 a 80×80 mm. Okrem toho je znázornené aj umiestnenie kamery. Na obr. 4 b) je vstupný s miernym skreslením a na obr. 4 c) je výstupný obraz s korekciou skreslenia a so znázornením detekcie kalibračných objektov, kde červenou sú vyznačené detegované hrany referenčných objektov a modrou je vyznačená očakávaná poloha týchto hrán referenčných objektov.



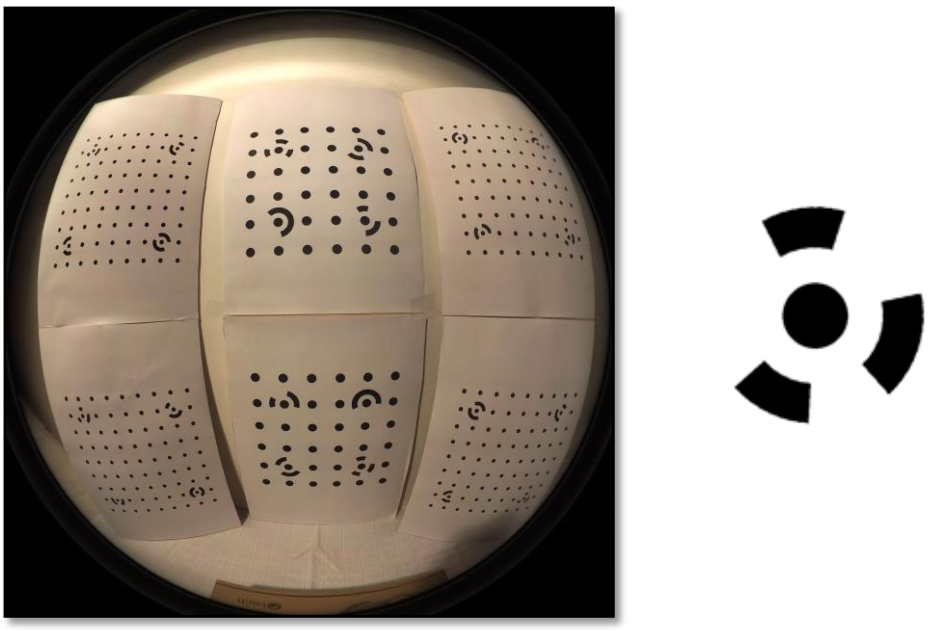
a) b) c)

Obr. 4 Praktická realizácia kalibrácie kamery použitím 3D objektov a Tsaiovej metódy a) Definícia polohy a veľkosti kalibračných objektov b) vstupný skreslený obraz c) výstupný obraz bez skreslenia

Nevýhodou tejto metódy je obmedzenie z hľadiska veľkosti skreslenia šošovky, kde pri vyššom skreslení a šošovkách typu rybie oko má táto metóda problémy s detekciou objektov a výpočtom vnútorných parametrov kamery [9].

IV. HYBRIDNÉ METÓDY KALIBRÁCIE KAMERY

Hybridné metódy kalibrácie kamery sú metódy, ktoré využívajú kombinácie algoritmov iných metód napríklad za účelom zvýšenia presnosti odhadu vnútorných parametrov kamery. Príkladom je metóda kalibrácie kamery pomocou testovacieho poľa, ktoré je znázornené na obr.5 a). Táto metóda kombinuje algoritmy Zhangovej a Tsaiovej metódy kalibrácie kamery [10].

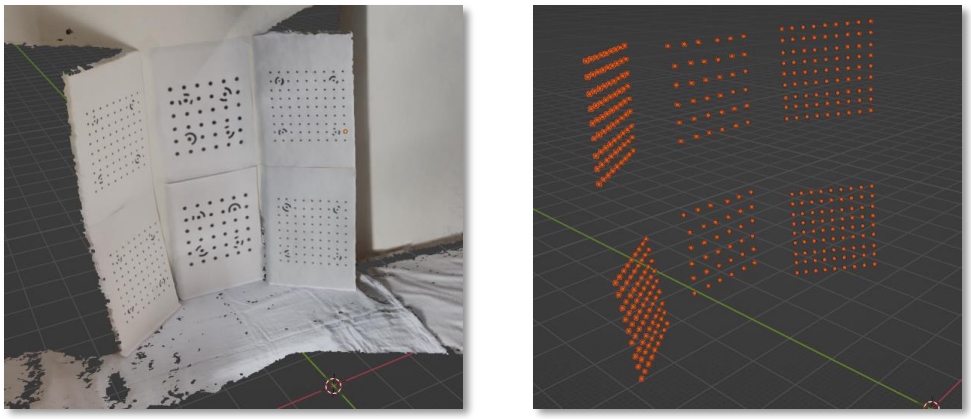


a) b)

Obr. 5 a) Testovacie pole pre výpočet vnútorných parametrov kamery b) Vlícovací bod

Pre výpočet odhadu vnútorných parametrov táto metóda používa podobné kalibračné vzory ako pri sú použité Zhangovej metóde, konkrétne sa jedná o rovnomerne rozmiestnené body na určitej ploche. Rozdiel ale spočíva vo výpočte priamej lineárnej transformácie, do ktorého sú zahrnuté súradnice každého bodu v trojrozmernom priestore, tento krok je charakteristický pre algoritmus Tsaiovej metódy. Navyše testovacie pole sa skladá z niekoľkých kalibračných vzoriek umiestnených na pevnom podklade - panely, pričom samotné panely sú rozmiestnené pod rôznymi uhlami tak, aby zachytili celý záber kamery a práve preto je potrebné poznať presnú polohu každého bodu v priestore [10].

Hlavnou výhodou je podobne, ako pri Tsaiovej metóde to, že na odhad parametrov kamery postačí len jedna snímka a to za predpokladu, že je známa presná poloha a rotácia kamery v priestore vzhľadom na kalibračné pole. Ako bolo uvedené vyššie, je potrebné zadefinovať presné súradnice každého bodu v testovacom poli. Ručné meranie každého bodu by bolo veľmi zdĺhavé a nepresné, práve preto je na získanie súradníc týchto bodov potrebné vytvoriť trojrozmerný model testovacieho poľa pomocou fotogrametrie, ideálne už s presne kalibrovaným fotoaparátom. Precízna fotogrametria zahŕňa vyhotovenie niekoľkých desiatok až stoviek snímok pre vygenerovanie presného modelu z rôznych uhlov a vzdialeností. Pri tomto procese nie je vždy zachytený celý objekt a práve z tohto dôvodu sú aspoň na okrajoch jednotlivých panelov umiestnené unikátne značky, tzv. vlícovacie body, pomocou ktorý dokáže algoritmus fotogrametrie spájať jednotlivé snímky. Z vygenerovaného trojrozmerného digitálneho modelu sú následne extrahované tieto body do formy *pointcloud*, v preklade *mračno bodov*. Mračno bodov predstavuje rozmiestnenie bezrozmerných bodov v 3D priestore, pričom každý bod nesie informáciu o svojej polohe a v niektorých prípadoch môže taktiež niesť informáciu o farbe. Štandardne je tento súbor ukladaný do súboru s príponou *.xyz*. Jedným z profesionálnych softvérov, ktorého výstup je súbor s príponou .xyz je napríklad *Scanreference*, ktorý v sebe zahŕňa aj algoritmus fotogrametrie. Alternatívou je použitie open-source programu *Blender*, avšak ktorého vstup už musí byť samotný 3D objekt – sken bodov [11]. Body kalibračného poľa sú nahrané do meracieho systému spolu aj s údajmi polohy kamery. Na obr. 6 a) je 3D model kalibračného poľa vygenerovaný fotogrametriou a na obr. 6 b) je samotné mračno extrahovaných bodov z 3D objektu [12].



a) b)

Obr. 6 a) 3D model kalibračného poľa získaný fotogrametriou b) Mračno bodov získané z 3D modelu kalibračného poľa

V. ZÁVER

Cieľom tejto práce bolo popísať problematiku vnútorných a vonkajších parametrov digitálnej kamery a predstaviť možné metódy kalibrácie kamery. Boli popísané tri metódy odhadu vnútorných parametrov kamery a to Zhangová, Tsaiová a hybridná metóda použitím kalibračného poľa a algoritmov oboch predchádzajúcich metód.

Najbežnejšie používaná Zhangova metóda sa vyznačuje najmä univerzálnosťou - je vhodná pre rôzne typy kamier a šošoviek, jednoduchosťou na realizáciu - je potrebné mať len dosku s kalibračným vzorom a pomocou tejto metódy je možné odmerať aj tangenciálne skreslenie. Nevýhodou je nízka presnosť a vyššia variácia odhadnutých parametrov pri opakovaných meraniach s tou istou kamerou.

Tsaiova metóda využíva na kalibráciu trojrozmerné objekty, výhodou je možnosť kalibrácie kamery použitím jednej fotografie, dôsledkom čoho je odhad vnútorných parametrov presnejší, rýchlejší a vykazuje stabilnejšie hodnoty pri opakovaných meraniach s tou istou kamerou. Nevýhodou je nemožnosť kalibrácie kamier so silným skreslením.

Metóda kalibrácie kamery pomocou pevne definovaného testovacieho poľa je náročnejšia na realizáciu a testovacie pole je väčšinou stavané len pre jeden typ kamier, resp. kamier s podobnými vlastnosťami šošoviek, ako je uhol záberu a pod. Využíva princípy odhadu vnútorných parametrov kamery z oboch predošlých metód. Výhodou je presný odhad vnútorných parametrov kamery.

Okrem vyššie spomenutých metód existuje ešte samostatná skupina metód kalibrácie kamery nazývaná ako autokalibrácia kamery. Metódy založené na autokalibrácii nepotrebujú na odhad vnútorných parametrov žiadne kalibračné polia a referencie ale postačí im napríklad zopár fotiek scenérie, na základe ktorej dokážu odhadnúť mieru skreslenia.

VI. CITOVANÉ DIELA

[1] Z. Zhang, A Flexible New Technique, IEEE Xplore, 2000.

[2] Intrinsic camera parameters calibration, London: University College London, 2019.

[3] MathWorks, What Is Camera Calibration?.

[4] D. Samper, J. Santolaria a e. al., Analysis of Tsai calibration method using twoand three-dimensional calibration objects, Springer, 2013.

[5] H. Bay, T. Tuytelaars a L. V. Gool, SURF: Speeded Up Robust Features, ETH Zürich.

[6] J. Xing, Z. Miao a J. Chen, An Algorithm for Seamless Image Stitching and Its Application, Springer, 2007.

[7] C. Stachniss, Camera Calibration: Zhang's Method, 2020.

[8] B. K. Horn, Tsai's camera calibration method revisited, 2000.

[9] R. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, IEEE Xplore, 1987.

[10] C. Zhang, Z. Y a Z. Huang, CS231A Project Final Report: Comparing Multiple Methods for Camera Calibration, 2022.

[11] H. Tobias, Atomic Blender (PDB/XYZ), 2019.

[12] D. Scaramuzza, Vision Algorithms for Mobile Robotics, ETH Zurich.

Štandardné videokodeky

Miroslav HERBERT, Ondrej KOVÁĆ

Katedra Technológií v Elektronike, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenská Republika

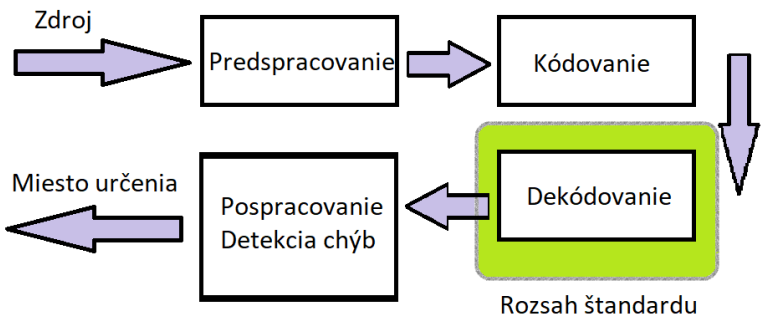
miroslav.herbert@student.tuke.sk; ondrej.kovac@tuke.sk

***Abstrakt* — Štandardné videokodeky predstavujú neoddeliteľnú súčasť digitálneho sveta, kde sú multimédiá čoraz viac v popredí našich každodenných zážitkov. Tieto kľúčové technologické komponenty sú zodpovedné za kompresiu a dekompresiu video obsahu, čo umožňuje efektívne streamovanie, ukladanie a prenos videí. Táto práca sa zameriava na prieskum týchto kódovacích štandardov, rozoberá ich funkcie, výhody a nevýhody, a sleduje ich vývoj v kontexte dynamického sveta digitálnej mediálnej produkcie. Dôkladnejšie sú opísané štandardy H.264 a H.266 a taktiež je venovaný čas organizáciám videokodekových štandardizácií.**

***Kľúčové slová* — H.264, H.266, štandardizácia, videokodeky**

I. Úvod

V súčasnosti existuje niekoľko štandardných videokodekov, ktoré sa používajú na celosvetovej úrovni a majú významný vplyv na kvalitu, veľkosť a dostupnosť digitálneho videa. Kompresné techniky digitálneho videa majú dôležitú úlohu v oblasti telekomunikácií a multimediálnych systémov, kde je dôležitým parametrom šírka pásma. Kódovacie techniky pre video sú kľúčové pri redukcii informačného obsahu potrebného pre sekvenciu obrázkov, pričom sa snažia udržať vysokú kvalitu z pohľadu diváka. Súčasné kompresné techniky zahrňujú zložité elektronické obvody, a ich náklady neumožňujú udržateľnú výrobu čipov LSI (Large-scale integration) iba veľkosériovou výrobou. Štandardizácia kompresných techník pre video je preto nevyhnutná. V 70. rokoch 20. storočia bolo zavedené priame PCM (pulzné kódové modulácie) kódovanie TV signálov rýchlosťou 140 Mbit/s, čo vyhovovalo digitálnej hierarchii hlavne pre viackanálovú telefóniu. Vysoká bitová rýchlosť však obmedzovala jeho použitie na distribúciu TV programov a štúdiové úpravy. S nástupom digitálnej TV pre satelity sa stala atraktívnou vďaka kompatibilite so systémami viacnásobného prístupu s časovým delením. V 80. rokoch boli experimentálne systémy s bitovými rýchlosťami približne 45 Mbit/s pre signály NTSC a 60 Mbit/s pre štandard PAL. Videotelefónne systémy v 60. rokoch neboli životaschopné. V 70. rokoch sa objavila myšlienka identifikácie rečníka, čo viedlo k vývoju kódovania videa s nízkou bitovou rýchlosťou. V 80. rokoch sa COST211 video kodek založený na diferenčnej pulznej kódovej modulácii stal štandardom podľa H.120. Dôležité je rozlišovať medzi kodekom a štandardom, pretože tieto pojmy sú často zamieňané. Kodek môže byť považovaný za výsledný produkt riešený hardwarovo alebo softwarovo, zahŕňajúci kodér a dekodér. Organizácie ako ITU-T VCEG a ISO/IEC MPEG sa podieľajú na štandardizácii videokodekov, pričom sa zameriavajú predovšetkým na syntax a dekodér. Cieľom tejto štandardizácie je umožniť jednotný rámec, znížiť implementačnú zložitosť a zabezpečiť minimálnu záruku kvality. Obr. 1 vizualizuje rozsah štandardizácie videokodekov [1] [2].



Obr. 1 Rozsah štandardizácie videokodekov

II. ORGANIZÁCIE VIDEOKODEKOVÝCH ŠTANDARDIZÁCII

Ako bolo uvedené, dve hlavné organizácie, ISO a ITU-T, dominujú videokodekovým štandardizačným prostredím. Napriek tomu sa do tejto oblasti v nedávnej dobe zapojili aj spoločnosti ako Google, Microsoft a Apple, ktoré sa snažia presadzovať vlastné normy v oblasti videokódovania. Aktuálnym trendom v tejto oblasti je vytváranie otvorených a voľne dostupných verzií videokodekov. Vzhľadom na licenčné problémy spojené s HEVC vznikla pre najnovší videokodek VVC skupina s názvom Media Coding Industry Forum (MC-IF). Táto skupina dohliada na licencovanie, ale nemá rozhodujúci vplyv na štandardizačný proces. Nasledujúca časť sa stručne zameria na dve organizácie spomenuté vyššie, konkrétne ISO/IEC a ITU-T.

A. ITU-T

Skupiny pre štúdium v sektore telekomunikačnej štandardizácie ITU-T zjednocujú expertov z celého sveta s cieľom vypracovať medzinárodné normy známe ako odporúčania ITU-T. Tieto odporúčania slúžia ako kľúčové prvky v globálnej infraštruktúre informačných a komunikačných technológií (ICT). Od svojho vzniku v roku 1865 ITU-T usmerňuje prístup založený na príspevkoch a dosiahnutí konsenzu v procese tvorby štandardov, v ktorom majú všetky krajiny a spoločnosti, bez ohľadu na svoju veľkosť, rovnaké právo ovplyvňovať vývoj odporúčaní ITU-T. Začínajúc ako orgán, ktorý štandardizoval medzinárodnú telegrafnú výmenu, ITU-T sa postupne stalo dôležitým hráčom v oblasti telekomunikácií a dnešného konvergovaného ekosystému ICT. Poskytuje globálnej komunite pre normalizáciu svetové normy a stále zohráva úlohu globálneho orgánu pre normy v oblasti IKT. Od roku 1984 sa ITU-T zaoberá štandardizáciou videokodekov [4].

B. ISO/IEC

ISO predstavuje celosvetovú sieť vedúcich štandardizačných organizácií. Prostredníctvom jej členov, ktorí pôsobia ako národné normalizačné orgány v 168 krajinách, spojuje odborníkov z celého sveta s cieľom vytvárať medzinárodné normy. Okrem monitorovania tisícok dokumentov prostredníctvom procesov ako návrhy, kontrola, hlasovanie a publikovanie, ISO poskytuje rôznorodé služby, ktoré sú zamerané na podporu ich strategických cieľov. Ich úsilie zahŕňa aj prácu na zvyšovaní povedomia verejnosti o normách a normalizácii. V spolupráci s inými organizáciami, ako sú IEC a ITU, sa snažia organizovať každoročný Svetový deň noriem. Tento deň slúži na oslavu normalizačných úspechov po celom svete a na sledovanie, ako normy adresujú súčasné výzvy spoločnosti. ISO taktiež angažuje verejnosť prostredníctvom spotrebiteľských výborov pre vývoj noriem [5]. História ISO sa začala v roku 1946, keď sa na Inštitúte stavebných inžinierov v Londýne stretli delegáti z 25 krajín [6]. IEC bola založená v roku 1906, a jej tajomníkom v čase druhej svetovej vojny bol britský inžinier Charles Le Maistre. V roku 1947 sa IEC rozhodla pripojiť k ISO a súčasne spolupracovať ako elektrická divízia ISO, pričom si udržala názov a technický postup [7]. Prvým spoločným krokom v oblasti videokodekov bol štandard MPEG-1 part 2 v roku 1993. Na Obr. 3 sú znázornené logá ITU-T a ISO.



Obr. 2 Logo spoločnosti ITU (a) *[4]* a ISO (b) *[7]*.

III. ŠTANDARDY VIDEOKODEKOV

Táto kapitola je venovaná stručnému súhrnu a rozboru jednotlivých, najvýznamnejších videokodekových štandardov. Ich chronologické zobrazenie je znázornené na Obr. 2. Je možné vidieť, že MPEG-2 a MPEG-4 vznikli spoluprácou ITU-T a ISO/IEC.

A. H.120

Prvým digitálnym videokodekovým štandardom bol H.120, schválený Medzinárodnou telekomunikačnou úniou (ITU-T) v roku 1984. V jeho počiatkoch v roku 1984 využíval systém DPCM (Differential Pulse Code Modulation), ktorý umožňoval kódovanie s premennou dĺžkou

slova. Výrazným vylepšením bola druhá verzia v roku 1988, ktorá pridala kompenzáciu pohybu a predikciu pozadia. Štandard pracoval pri bitových rýchlostiach 1544 kbit/s pre NTSC a 2048 kbit/s pre PAL. Táto technológia sa rýchlo vyvinula a dnes je H.120 prakticky nepoužívaný [2] [3].

B. H.261

Náhradou H.120 bol H.261, s ktorým bol predstavený zásadný posun v digitálnom kódovaní videa pomocou diskrétnej kosínusovej transformácie (DCT) v roku 1988, čo bola kľúčová zmena od predchádzajúcej DPCM. Dizajn z roku 1988 stanovil štruktúru, ktorá dominuje dodnes, vrátane kompenzácie pohybu makrobloku 16x16, 8x8 DCT, skalárneho kvantovania a iných techník. Druhá verzia v roku 1993 pridala spätne kompatibilný grafický režim s vysokým rozlíšením. S bitovými rýchlosťami od 64 do 2048 kbit/s sa H.261 používa dodnes, najmä ako funkcia spätnej kompatibility [2] [3].

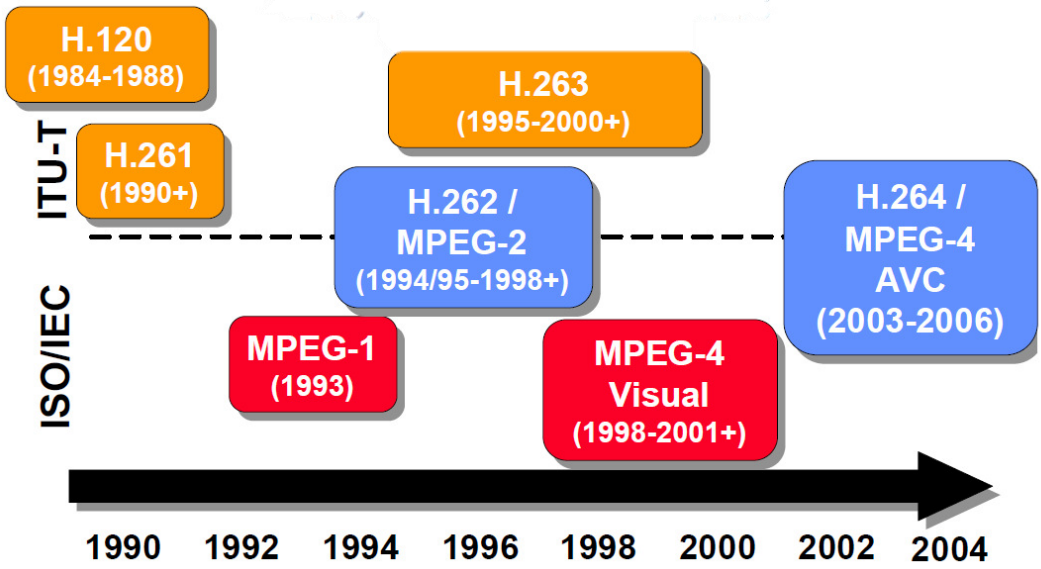
C. MPEG-1 a MPEG-2

Skupina expertov pre pohyblivý obraz (MPEG) bola založená v roku 1988 s cieľom štandardizovať videokodeky pre pamäťové média. Prvým výsledkom bola séria MPEG-1, schválená v roku 1992, s úvodným štandardom ISO 11172. Tento formát bol určený pre bitové rýchlosti do 1,5 Mbit/s a získal masovú akceptáciu vďaka svojej schopnosti kódovať CD-kvalitné audio a video na jednom kompaktnom disku [2] [3].

MPEG-2, štandardizovaný v roku 1996, rozšíril pôsobnosť na vyššie bitové rýchlosti a kódovanie HDTV. Jeho algoritmus bol zdokonalený a prispôsobený pre kódovanie televízneho obrazu s prekladaným riadkovaním [2] [3].

D. MPEG-4 Part 2

V roku 1999 bola predstavená verzia MPEG-4 Part 2, známa aj ako ISO/IEC 14496-2. Tento štandard základne vychádzal z H.263, ale priniesol zvýšenú efektivitu kódovania, najmä pri nižších bitových rýchlostiach. Zaviedol kreatívne doplnky, ako je vylepšená odolnosť voči chybám, segmentované kódovanie tvarov, kódovanie syntetického a polysyntetického obsahu a iné. Druhá a tretia verzia priniesli ďalšie vylepšenia v začiatkoch 21. storočia [2] [3].



Obr. 3 Chronológia medzinárodných štandardov kódovania videa

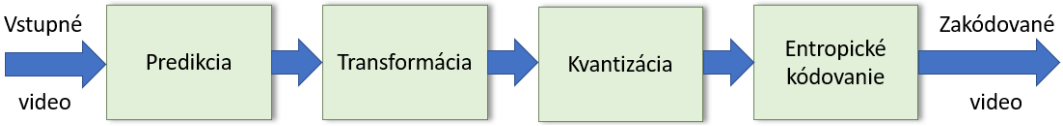
E. H.264/MPEG-4 AVC

H.264/MPEG-4 Advanced Video Coding (AVC) je spoločný produkt ITU-T a ISO/IEC, vyvinutý od 1999 do 2003 a predstavuje pravdepodobne najrozšírenejší štandard videokompresie na svete. Tento štandard je široko využívaný v aplikáciách ako HDTV vysielanie, mobilné siete a internet, vrátane videokomunikácií v reálnom čase. Svojím pôsobením od roku 2003 sa stal jedným z najpopulárnejších videokodekov na svete [2] [3]. Kompresná technológia H.264 sa využíva pri vysielaní, streamovaní a dokonca aj na bežných optických diskoch.

Digitálny video signál je tvorený sekvenciou obrazov, nazývaných snímky, ktoré obsahujú pixely rozložené v dvoch rozmeroch a sú reprezentované v troch hlavných farbách: červenej, zelenej a modrej. Typicky sa tieto pixely v RGB formáte konvertujú na Y, U, V. V štandardnom

systéme blokového kódovania sú pixely v celej snímke rozdelené na makrobloky, pričom komponenty Y majú rozmery 16 x 16 pixelov, a komponenty U a V majú rozmery 8 x 8 pixelov. H.264/AVC je navrhnutý tak, aby zvládal rozmanité aplikačné požiadavky, pretože ponúka bohatú súpravu nástrojov na kódovanie videa. Rôzne aplikácie si zvyčajne vyžadujú rôzne funkcie, ako napríklad odolnosť voči chybám, kompresnú efektivitu, nízke oneskorenie a zložitosť. Preto existujú tri profily: základný, rozšírený a hlavný profil, ktoré sú prispôsobené na riešenie špecifických požiadaviek [8].

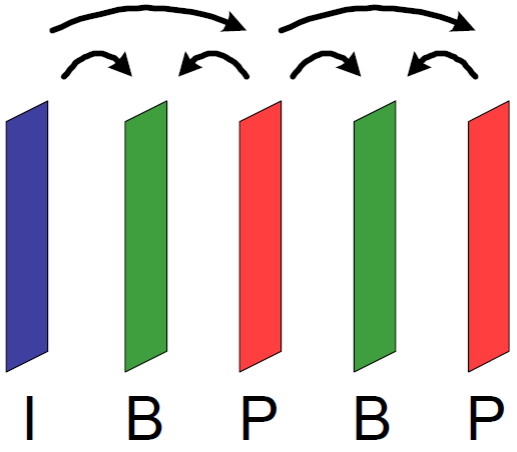
H.264/AVC dosahuje výrazné zlepšenie v oblasti kompresného výkonu, poskytujúc sieťovo prijateľné video reprezentácie určené pre rôzne aplikácie, vrátane ukladania, šírenia, streamovania a videotelefónie. Táto metóda kompresie, ktorá sa nachádza vo väčšine bežných štandardov kódovania videa, pozostáva zo štyroch kľúčových etáp: prediktívneho kódovania, transformácie, kvantovania a entropického kódovania, ako je znázornené na Obr. 4.



Obr. 4 Princíp kompresnej metódy pre H.264

Keďže sú po sebe nasledujúce snímky veľmi podobné, tak sa pre minimalizáciu dočasnej redundancie využíva technika prediktívneho kódovania. Snímky sú vzorkované a transformované do frekvenčnej domény prostredníctvom transformačného procesu, čo pomáha znížiť priestorovú redundanciu. Transformácia zanecháva len malý počet významných transformačných koeficientov a veľký počet nedôležitých koeficientov. Nepodstatné koeficienty sú nastavené na nulu kvantizačným procesom, kde každý koeficient je delený kladným celým číslom a výsledok je zaokrúhlený na najbližšiu celočíselnú hodnotu. Kódovanie nenulových a nulových koeficientov sa vykonáva pomocou kódov s nerovnomernou dĺžkou. Predikcia makroblokov sa uskutočňuje pomocou blokov s veľkosťou 4 × 4 alebo 16 × 16 v prípade predikcie v rámci snímky, pričom sa vychádza z predpovedí z okolitých blokov v tej istej snímke. Rôzne veľkosti blokov od 4 × 4 do 16 × 16 sa zohľadňujú aj v prípade predikcie medzi snímkami, kde sa využívajú predpovede z predtým kódovaných snímok. Vnútrobloková predikcia, ktorá je aplikovaná na predpovedanie susedných blokov v aktuálnom rámci, blokov v predchádzajúcom rámci a blokov až z dvoch predchádzajúcich rámcov, je následne diskretizovaná do veľkostí 16 × 16, 8 × 8 alebo 4 × 4 pixelov v závislosti od efektivity predikcie [8].

V dôsledku toho je potrebné zakódovať iba rozdiel medzi nimi, namiesto zakódovania celého súčasného makrobloku, a to v ideálnom prípade s malým rozdielom. Typ predikcie sa klasifikuje do dvoch kategórií: predikcia medzi snímkami a predikcia v rámci snímky, v závislosti od toho, odkiaľ pochádza referenčný makroblok. V prípade medzisnímkovej predikcie (P alebo B) sa referenčný makroblok nachádza niekde pred alebo za aktuálnou snímkou, kde sa nachádza aktuálny makroblok. Referenčný makroblok môže byť získaný ako vážená funkcia makroblokov z viacerých snímok. V režime vnútroblokovej predikcie (I) je referenčný makroblok obvykle vypočítaný pomocou matematických funkcií susedných pixelov aktuálneho makrobloku [9]. Proces predikcie snímok je znázornený na Obr. 5.



Obr. 5 Vnútroblokové (I) a medzisnímkové predikčné (B, P) snímky [2]

V prípade medzisnímkovej predikcie sa referenčná snímka, obsahujúca predikčný blok, vyberie z predtým dekódovaných snímok. Pohybový vektor pre aktuálny makroblok sa získa z posunu predikčného bloku alebo z predtým kódovaných pohybových vektorov. Nakoniec sa vypočíta predikčná snímka pomocou vektora pohybu a referenčnej snímky [9]. Zvyšok (rezíduum) medzi referenčnou a predikčnou snímkou sa uloží a tento proces sa nazýva pohybová kompenzácia (MC), ako je znázornené na Obr. 6.



Obr. 6 Pohybová kompenzácia (MC)

Tieto pohybové vektory môžu byť voliteľne vážené s cieľom zohľadniť časovú blízkosť medzi snímkami a následne sú zaslané do dátového toku. Na každú snímku sa obvykle aplikuje deblokovací filter, ktorý sa ukladá v dekódovanom formáte, aby mohol byť neskôr použitý pri predpovediach medzi snímkami a vyhladzovaní ostrých hrán, ktoré môžu vzniknúť v dôsledku blokového kódovania. Transformácia zahrňuje premenu blokov na reprezentácie vo frekvenčnej oblasti a kvantovanie koeficientov. Krok diskrétnej kosínusovej transformácie (DCT) je definovaný transformáciou bloku X maticou A na DCT koeficienty Y pre každý makroblok. V prípade 4 × 4 blokov sú tieto matice X a $Y \in$ R4 × 4 vyjadrené ako

$$Y = AXA^T = \begin{bmatrix} a & a & a & a \\ b & c & -c & -b \\ a & -a & -a & a \\ c & -b & b & -c \end{bmatrix} X \begin{bmatrix} a & b & a & c \\ a & c & -a & -b \\ a & -c & -a & b \\ a & -b & a & c \end{bmatrix} \quad (1)$$

kde a = 1/2, b = √(1/2) cos(π/8), c = √(1/2) cos(3π/8). Riadky A sú ortonormálne. Pre výpočet rovnice (1) na reálnom procesore je potrebná aproximácia pre b a c. Táto aproximácia je dosiahnutá použitím metódy pevného bodu, ktorá ekvivalentne zmenší každý riadok matice A o 2,5 a potom zaokrúhli na najbližšie celé číslo. Jadrová transformácia C_{f4} transformuje každý člen matice A o 2,5 a zaokrúhľuje na najbližšie celé číslo. Škálovacia matica S_{f4} obnovuje normy riadkov C_{f4} na hodnotu 1 pomocou škálovania takto

$$C_{f4} = \begin{bmatrix} c_1 & c_1 & c_1 & c_1 \\ c_2 & c_1 & -c_1 & -c_2 \\ c_1 & -c_1 & -c_1 & c_1 \\ c_1 & -c_1 & c_2 & -c_1 \end{bmatrix}, \quad S_{f4} \begin{bmatrix} s_1 & s_2 & s_1 & s_2 \\ s_2 & s_3 & s_2 & s_3 \\ s_1 & s_2 & s_1 & s_2 \\ s_2 & s_3 & s_2 & s_3 \end{bmatrix}, \quad (2)$$

kde c_1 = 1, c_2 = 2, s_1 = 1/4, s_2 = 1/(2√10) a s_3 = 1/10. Matica Y sa potom určí ako

$$Y = \left[C_{f4} X C_{f4}^T\right] S_{f4}. \quad (3)$$

Transformácia DCT prevedie zostávajúce dáta z časovej domény na frekvenčnú doménu. Keďže väčšina obrázkov a videí obsahuje nízkofrekvenčné informácie, DCT umožňuje sústrediť kódovacie informácie. Hlavným účelom kvantovania je zmenšiť transformované koeficienty a kódovacie informácie. Vzhľadom na to, že ľudské oko je menej citlivé na vysokofrekvenčné zložky obrazu, niektoré štandardy kompresie videa a obrazu môžu pre vysokofrekvenčné údaje používať vyššie hodnoty kvantizácie (kvantizačné parametre). Štandard H.264 využíva 4x4 celočíselnú DCT. Entropický kódovač konvertuje syntaxové prvky (kvantizované koeficienty a ďalšie informácie, ako sú pohybové vektory, predikčné režimy atď.) do bitového toku, ktorý je následne možné získať entropickým dekodérom. H.264 implementuje dve metódy entropického kódovania: Kontextové adaptívne kódovanie s premenlivou dĺžkou (CAVLC) a kontextové adaptívne binárne aritmetické kódovanie (CABAC) [9].

F. H.266/VVC

H.266/VVC (Versatile Video Coding) predstavuje najnovší štandard videokodeku, navrhnutý na pokračovanie úspešných predchádzajúcich štandardov ako H.264/MPEG-4 AVC a H.265/HEVC. Jeho cieľom je poskytnúť efektívnejšie kódovanie videa pri zachovaní vysokej vizuálnej kvality. H.266/VVC bol vyvinutý v reakcii na narastajúce nároky na šírku pásma a nárast objemu dát v súvislosti so sťahovaním a streamovaním videa v rôznych aplikáciách, vrátane vysokého rozlíšenia a nových technológií, ako je 360-stupňové video a rozšírená realita (AR). H.266/VVC prináša niekoľko vylepšení oproti svojim predchodcom [10]. Používa sa pre najviac používané streamovacie platformy ako Netflix, YouTube, Vivo a ďalšie. Jeho hlavné rysy zahŕňajú:

1) Vysoká kompresná efektivita:

H.266/VVC sa vyznačuje výrazným pokrokom v oblasti kompresnej efektivity. To znamená, že dokáže poskytnúť vysokú kvalitu videa pri nižších dátových tokoch. Pre koncových užívateľov to môže znamenať efektívnejšie streamovanie a sťahovanie videí, pričom sa minimalizuje zaťaženie sieťového pripojenia a ukladanie dát.

2) Podpora pre rôzne aplikácie:

Navrhnutý so zreteľom na rôzne aplikácie, H.266/VVC ponúka flexibilitu a prispôsobivosť pre moderné videoplatformy. Od mobilných zariadení až po vysoké rozlíšenie a nové formy multimédií, tento štandard sa snaží efektívne pokrývať širokú škálu potrieb spotrebiteľov.

3) Podpora pre nové technológie:

H.266/VVC aktívne zohľadňuje nové technologické trendy. Schopnosť efektívneho kódovania 360-stupňových videí, rozšírenej reality a ďalších inovácií otvára dvere pre nové formy obsahu, ktoré sú čoraz populárnejšie medzi užívateľmi.

4) Zachovanie spätnej kompatibility:

V rámci svojich vlastností H.266/VVC dbá na spätnú kompatibilitu s predchádzajúcimi štandardmi. To znamená, že môže dekódovať aj videá kódované pomocou starších štandardov, čím zabezpečuje plynulý prechod a integrovanie s existujúcimi multimediálnymi systémami [10].

IV. ZÁVER

V tejto práci bol v stručnosti opísaný historický vývoj videokodekov a boli opísané významné videokodekové štandardy ako H.264/MPEG-4 AVC a H.266/VCC a iné. H.264 zohral kľúčovú úlohu v oblasti videokompresie a nastavil štruktúru pre budúce videkodeky. Štandardizačné organizácie ITU a ISO hrali významnú rolu pri formovaní týchto noriem. S nástupom technologického pokroku vznikla potreba efektívnejších riešení, ktoré momentálne spĺňa H.266/VVC. Tento štandard prináša výrazné zlepšenia v kompresnej efektivite, súčasne s podporou nových technologických trendov od mobilných zariadení po rozšírenú realitu. S neustálym vývojom a novými štandardmi, ako je H.266/VVC, je možné očakávať ďalšie inovácie v oblasti kvality, efektivity a podpory pre nové technologické výzvy.

ZDROJE

[1] M. Ghanbari, "Standard Codecs: Image compression to advanced video coding," The Institution of Engineering and Technology, 2003. ISBN: 978-0-85296-710-2.
[2] G. J. Sullivan, "Overview of International Video Coding Standards (preceding H.264/AVC)," from ITU-T VICA Workshop, 22-23 July 2005, ITU Headquarter, Geneva.
[3] Ebrahimi,T.: Image and Video Compression Standards. Signal Processing Laboratory, Swiss Federal Institute of Technology, Lausanne. 2000.
[4] ITU, "About ITU-T," online: https://www.itu.int/en/ITU-T/about/Pages/default.aspx.
[5] ISO, "What we do," online: https://www.iso.org/what-we-do.html.
[6] ISO, "About us," online: https://www.iso.org/about-us.html.
[7] L. D. Eicher, W. Kuen, R. Marechal and coll. "Friendship Among Equals: Recollections from ISO's first fifty years," 1997. ISBN 92-67-10260-5.
[8] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003, doi: 10.1109/TCSVT.2003.815165.
[9] Jayaratne, M., Gunawardhana, L. K. P. D., & Samarathunga, U. (2022), Comparison of H.264 and H.265. In: Engineering and Technology Quarterly Reviews, Vol.5, No.1, 17-24.
[10] Schierl, T., Nguyen, T., & Suhring, K. (2020). H.266/VVC Standard for High-Efficiency Video Coding. IEEE Transactions on Circuits and Systems for Video Technology, 30(11), 4292–4306. https://doi.org/10.1109/TCSVT.2020.3018896

Minimalistická implementácia TLS klienta pre vstavané aplikácie

[1]*Jaroslav PAJKOŠ,* [2]*Miloš DRUTAROVSKÝ*

[1,2]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]jaroslav.pajkos@student.tuke.sk, [2]milos.drutarovsky@tuke.sk

Abstrakt – **Tento článok sa zaoberá základnou analýzou protokolu TLS 1.3 (Transport Layer Security verzie 1.3), s dôrazom na jeho kryptografické mechanizmy a integráciu s protokolom TCP/IP. Dôraz bol kladený na vytvorenie minimalistického TLS 1.3 klienta, ktorý podporuje len minimálnu sadu kryptografických algoritmov pre zabezpečenú komunikáciu medzi klientom a serverom, avšak dostatočnú na pripojenie k väčšine zabezpečených webových serverov. Implementácia kryptografických funkcií bola nahradená z iných kryptografických knižníc, primárne pre vstavané zariadenia a napísané v programovacích jazykoch C/C++. Minimalistický TLS 1.3 klient bol upravený pre jednoduchú portovateľnosť primárne pre vstavané platformy, akou je napríklad zariadenie ESP32-C6. V článku sú uvedené experimentálne zistené doby trvania jednolitvých kryptografických funkcií v rámci minimalistického TLS 1.3 klienta.**

Kľúčové slová – **CSPRNG, ESP32-C6, IoT, kryptografická knižnica, senzorový uzol, TCP/IP, TLS 1.3**

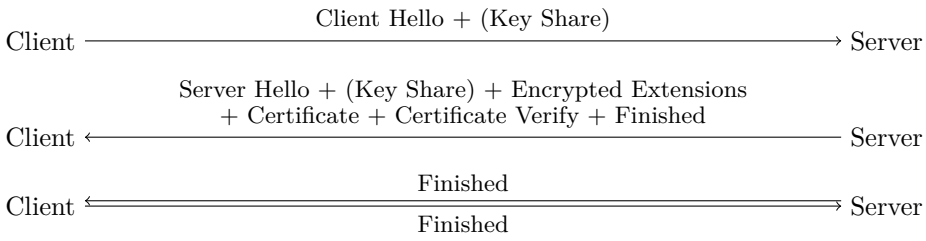
I. Úvod

V súčasnej dobe je bezpečnosť v sieťových technológiách považovaná za nevyhnutnú súčasť bezpečnej počítačovej komunikácie. Pre zabezpečenie komunikácie na internete alebo aj v interných sietiach je potrebné využívať vždy overené, známe a bezpečné spôsoby komunikácie, ktoré garantujú, že súčasný technologický vývoj v nasledujúcich rokoch neumožní kompromitáciu aktuálne používaných metód zabezpečenia. Pri sieťovej bezpečnosti je obzvlášť dôležité brať na vedomie kybernetické útoky, ktoré môžu nastať vo vnútri počítačovej siete. Kybernetický útok vo vnútri počítačovej siete, môže patriť do kategórie pasívnych útokov, kedy potencionálny útočník odpočúva komunikáciu pre získanie nejakých neopravených informácií, avšak existujú aj aktívne útoky, pri ktorých potencionálny útočník sa snaží priamo zasiahnuť do komunikácie, tým, že pozmení jej obsah, pre dosiahnutie kompromitácie, najčastejšie v podobe úniku dát alebo neoprávneného prístupu. V súčasnosti je nevyhnutné zabezpečiť aj IoT (Internet of Things) zariadenia, medzi ktoré patria napríklad aj senzorové uzly. Pri zabezpečení internej komunikácie sa v súčasnosti najčastejšie využíva protokol TLS 1.3 (Transport Layer Security, verzia 1.3)[1], ktorý zabezpečuje šifrovanú komunikáciu medzi dvoma stranami na internete. Tento protokol je navrhnutý tak, aby poskytoval dôvernosť a integritu údajov v komunikácii medzi klientom a serverom. Cieľom tohto článku bolo analyzovať protokol TLS 1.3 zo strany využitia kryptografických algoritmov a následného vytvorenia minimalistického TLS 1.3 klienta, ktorý obsahuje len nevyhnutné algoritmy potrebné na pripojenie k väčšine zabezpečeným webovým serverom (protokol HTTPS). Obmedzená sada kryptografických algoritmov využívaná v minimalistickom TLS 1.3 klientovi, bola použitá z iných kryptografických knižníc, primárne určených pre vstavané zariadenia a napísaných v jazyku C. Minimalistický TLS 1.3 klient umožňuje portovateľnosť na operačné systémy Windows, Linux a zariadenie ESP32-C6. V článku sú uvedené experimentálne zistené doby trvania kryptografických funkcií pri nadviazaní TLS spojenia boli vykonané na zariadení ESP32-C6. Bežné používateľské počítače využívajú robustné kryptografické knižnice. Minimalistický TLS klient je primárne určený pre zariadenia s obmedzenými zdrojmi, akým je napríklad ESP32-C6, kde je doba trvania meraných funkcií, uvedených v tomto článku je kritická.

[1]https://www.allaboutcircuits.com/news/what-you-need-to-know-about-the-tls-1.3-protocol-and-wolfssls-ssl-tls-libra/

II. Základné vlastnosti protokolu TLS 1.3

Bezpečná komunikácia v počítačovej sieti musí spĺňať kritéria utajenia, integrity a dôvernosti dát. Na zabezpečenú komunikáciu medzi klientom a serverom sa od augusta 2018 využíva protokol TLS 1.3, ktorý predstavuje bezpečnostné a rýchlostné zlepšenia v porovnaní s jeho predchádzajúcimi verziami. Tento protokol sa najčastejšie používa na zabezpečenie dát prenášaných prostredníctvom protokolov HTTP (HTTPS), VPN (Virtual Private Network) alebo VoIP (Voice over Internet Protocol). Pred komunikáciou je dôležité, aby klient a server uskutočnili nadviazanie komunikácie, tzv. handshake, ktorý je v protokole TLS 1.3 rýchlejší vďaka skrátenému procesu dohody. Proces nadviazania komunikácie medzi klientom a serverom zabezpečí dohodu na podporovaných kryptografických algoritmoch a overenie, že správy neboli pozmenené nepovolanou osobou.



Obr. 1 Fázy nadviazania komunikácie medzi klientom a serverom v protokole TLS 1.3.

V protokole TLS 1.3 sa bezpečná komunikácia medzi klientom a serverom začína procesom handshake, znázorneným na Obr. 1[2]. Tento proces zahŕňa výmenu správ medzi klientom a serverom, ktoré obsahujú informácie o podporovaných kryptografických algoritmoch, overenie identity a vytvorenie tajných kľúčov, ktoré slúžia na zabezpečenie prenášaných dát počas nadviazania komunikácie medzi klientom a serverom.

Client Hello inicializuje handshake, kde klient posiela informácie o podporovaných verziách TLS a kryptografických algoritmoch.

Server Hello predstavuje odpoveď servera, ktorá obsahuje vybrané verzie a algoritmy zo skupiny poskytnutej klientom.

Encrypted Extensions obsahujú dodatočné informácie potrebné pre pripojenie.

Certificate je certifikát servera podpísaný certifikačnou autoritou, ktorý overuje identitu servera.

Certificate Verify je správa, pomocou ktorej server preukazuje, že je skutočným vlastníkom certifikátu.

Server Handshake Finished a **Client Handshake Finished** sú finálne správy, kde server a klient vypočítajú a porovnajú haš zo všetkých vymenených správ, aby sa zabezpečilo, že počas procesu nedošlo k pozmeneniu dát.

Po úspešnom ukončení handshaku, klient a server vypočítajú aplikačné kľúče (application keys), ktoré budú používané pri ochrane dát počas komunikácie. V mnohých prípadoch server a klient akceptujú aj použitie tzv. resumption ticketu, ktorý umožňuje rýchle opätovné spojenie a udržiavanie stavu relácie. Z pohľadu fáz nadviazania komunikácie sa pri použití resumption ticketu správy Certificate a Certificate Verify vynechávajú. TLS 1.3 má niekoľko výhod oproti TLS 1.2, ako napríklad:

- všetky správy po Server Hello sú zašifrované,
- TLS 1.3 podporuje len 5 autentizačných algoritmov, TLS 1.2 ich podporuje 37 [1].

III. Kryptografická knižnica TIIGER TLS

Základ nášho riešenia je založený na kryptografickej knižnici TIIGER TLS[3], ktorú vytvoril Michael Scott a obsahuje zdrojové kódy pre TLS 1.3 klienta v programovaciom jazyku C++ a TLS 1.3 server v programovaciom jazyku Rust. Knižnica TIIGER TLS je závislá na externej knižnici Miracl Core[4] a voliteľnej externej knižnici Libosdium[5]. Knižnicu TIIGER TLS sme použili spolu s externou knižnicou Miracl Core pre otestovanie podpory vybraných kryptografických algoritmov, pre zmenšenie potrebnej sady pre vytvorenie minimalistického TLS 1.3 klienta. Knižnica TIIGER TLS obsahuje vrstvu SAL (Security Abstraction Level),

[2] https://www.wolfssl.com/tls-1-3-performance-part-2-full-handshake-2/
[3] https://github.com/Crypto-TII/TLS1.3
[4] https://github.com/miracl/core
[5] https://github.com/jedisct1/libsodium

ktorá obsahuje volania všetkých kryptografických funkcií a je navrhnutá a zdokumentovaná tak, aby umožňovala nielen zmenu podporovaných algoritmov, ale aj nahradenie existujúcich implementácií kryptografických algoritmov vlastnými alebo inými z rôznych kryptografických knižníc.

IV. Analýza najčastejšie využívaných kryptografických algoritmov

Pri experimentálnom testovaní najčastejšie používaných kryptografických algoritmov sme sa zameriavali na podporu autentizačného šifrovania, metódy výmeny kľúčov a overovania digitálnych certifikátov serverov zo strany webových serverov. V Tabuľke 1 sú uvedené konkrétne názvy webových serverov, ktoré boli otestované, a výsledky súvisiace s používanými kryptografickými algoritmami. Kryptografické algoritmy, ktoré sa využívajú najčastejšie vzhľadom na vykonané experimenty sú: x25519, AES_128_GCM (TLS_AES_128_GCM_SHA_256), RSA 2048/4096 ECDSA- SECP256r1- SHA256 (EC-ECP(R)-SHA-256/384).

Názov servera	**Algoritmus pre výmenu kľúčov**	**Algoritmus pre autentizačné šifrovanie**	**Algoritmus digitálneho certifikátu**
google.com	x25519	AES_128_GCM	EC-SECP(R)-SHA-256
facebook.com	x25519	AES_128_GCM	EC-SECP(R)-SHA-256
instagram.com	x25519	AES_128_GCM	EC-SECP(R)-SHA-256
cnn.com	x25519	AES_128_GCM	RSA 2048
youtube.com	x25519	AES_128_GCM	EC-SECP(R)-SHA-256
twitter.com	x25519	AES_128_GCM	RSA 2048
netflix.com	x25519	AES_128_GCM	EC-SECP(R)-SHA-256
microsoft.com	x25519	AES_128_GCM	RSA 2048
malwarebytes.com	x25519	AES_128_GCM	RSA 2048
espressif.com	x25519	AES_128_GCM	RSA 2048
hbo.com	x25519	AES_128_GCM	RSA 2048
bbc.com	x25519	AES_128_GCM	RSA 2048
proton.me	x25519	AES_128_GCM	RSA 4096
tuke.sk	x25519	AES_128_GCM	RSA 2048
upjs.sk	x25519	AES_128_GCM	RSA 2048
refresher.sk	x25519	AES_128_GCM	RSA 2048
csfd.cz	x25519	AES_128_GCM	RSA 2048
alza.sk	x25519	AES_128_GCM	EC-SECP(R)-SHA-256
eset.sk	x25519	AES_128_GCM	EC-SECP(R)-SHA-256
datart.sk	x25519	AES_128_GCM	EC-SECP(R)-SHA-384

Tabuľka 1
Kryptografické algoritmy podporované experimentálne otestovanými webovými servermi.

V. Náhrada kryptografických algoritmov v SAL vrstve

Protokol TLS 1.3 obsahuje viacero kryptografických algoritmov, avšak na dosiahnutie konektivity na typické webové servery je postačujúce využiť obmedzenú množinu týchto algoritmov. Všetky kryptografické algoritmy sú platformovo nezávislé. Vo všeobecnosti jediná platformová závislosť je pri sieťovej konektivite a generovaní entropie. Vrstva SAL obsahuje volania všetkých kryptografických funkcií, vrátane generovania entropie (TRNG) a kryptograficky bezpečných pseudonáhodných čísel (CSPRNG). Pri nahradení bolo potrebné analyzovať implementácie jednotlivých funkcií v rôznych kryptografických knižniciach a následne vytvoriť zdrojové kódy, ktoré obsahujú len funkcie, ktoré sú potrebné na vytvorenie jednotlivých funkcionalít (napr. pri eliptických krivkách a algoritme RSA je potrebné len overovať digitálny podpis, nie podpisovať). Týmto spôsobom bol vytvorený minimalistický TLS 1.3 klient, ktorý obsahuje minimum potrebných algoritmov na dosiahnutie konektivity na väčšinu webových serverov. Medzi kryptografické knižnice určené pre zariadenia s obmedzenými zdrojmi patria knižnice PicoTLS[6], TinyTLS[7] a knižnice TiniCrypt[8]. Z týchto knižníc boli využité implementácie potrebných kryptografických funkcií pre vytvorenie minimalistického TLS 1.3 klienta. Podrobný opis procesu vytvorenia minimalistického TLS 1.3 klienta sa nachádza v [2].

[6]https://github.com/h2o/picotls
[7]https://github.com/Anthony-Mai/TinyTls
[8]https://github.com/intel/tinycrypt

Nahradenie hašovacích funkcií – V minimalistickej verzii TLS 1.3 sa využívajú funkcie SHA256 a SHA384 z rodiny SHA2 [3]. Použité implementácie sú z knižnice PicoTLS. Interný stav algoritmu je reprezentovaný štruktúrou `cf_sha256_context`, ktorá spracováva 64-bajtové bloky dát a sleduje počet spracovaných blokov a bajtov čakajúcich na spracovanie. Pre SHA384 sa používa štruktúra `cf_sha512_context`, ktorá spracováva 128-bajtové bloky dát.

Nahradenie autentizačného šifrovania – Autentizačné šifrovanie v protokole TLS 1.3 poskytuje ochranu dát a integrity pomocou autentizačného tagu a šifrovania. Implementácia využíva mód GCM [4] s blokovou šifrou AES, ktorá spracováva otvorený text po 128 bitoch. Táto implementácia, nazývaná TLS_AES_128_GCM_SHA256, je prevzatá z knižnice PicoTLS. Kľúčové informácie pre proces šifrovania sa nachádzajú v štruktúre `cf_aes_context`. Táto štruktúra obsahuje informácie o počte kôl šifrovania, súbor kľúčov generovaných z pôvodného kľúča, inicializačný vektor, typ autentizačného šifrovania, kryptografický kľúč a autentizačný tag.

Nahradenie algoritmu HMAC – Algoritmus HMAC (Hash-based Message Authentication Code) kombinuje MAC a hašovacie funkcie pre zabezpečenie integrity a autenticity správ. Prevzatá implementácia je z knižnice PicoTLS a využíva algoritmus HMAC, ktorý je tvorený funkciou `SAL_HMAC`. Táto funkcia začína s prázdnym autentizačným tagom nastaveným na 32 bajtov a obsahuje informácie o kľúči rovnakej dĺžky. Funkcia spracováva otvorený text.

Nahradenie algoritmu HKDF pre odvodenie zdieľaného tajomstva – Algoritmus HKDF (HMAC-based Extract-and-Expand Key Derivation Function) [5] slúži na generovanie a expanziu kľúčov. Skladá sa z dvoch krokov – extrakcie a expanzie. Funkcia pre extrakciu `SAL_hkdfExtract` transformuje vstupný kľúčový materiál na pseudonáhodný kľúč s použitím soli (salt). Funkcia `SAL_hkdfExpand` generuje výstupný kľúč požadovanej dĺžky využitím pseudonáhodného kľúča a dodatočných informácií.

Nahradenie algoritmu pre výmenu kľúčov – Algoritmus x25519 [6] sa nazýva Diffie-Hellmanov algoritmus pre výmenu kľúčov na báze eliptickej krivky Curve25519. Táto eliptická krivka je definovaná nad konečným poľom $\mathbb{GF}(2^{255} - 19)$. Prevzatá implementácia je z knižnice PicoTLS. Táto implementácia využíva dve funkcie `SAL_generateKeyPair` a `x25519_create_keypair` na generovanie kľúčov a odvodenie zdieľaného tajomstva. Tieto funkcie overujú, či je spôsob výmeny kľúčov nastavený na x25519, generujú pseudonáhodný kľúč, vykonávajú potrebné operácie na zabezpečenie súkromného kľúča a kontrolujú výsledný zdieľaný kľúč.

Overenie digitálneho certifikátu na báze algoritmu RSA – RSA [7] je najvyužívanejší algoritmus v asymetrickej kryptografii v protokole TLS 1.3, používaný na overenie digitálnych certifikátov. Prevzatá implementácia je z knižnice TinyTLS a pracuje s 2048-bitovými a 4096-bitovými kľúčmi, pričom využíva výplňové schémy PKCS 1.5 [8] a RSA-PSS [9]. Funkcie `RSA_2048_PKCS15_VERIFY` a `RSA_2048_PSS_RSAE_VERIFY` zabezpečujú vytvorenie štruktúr pre overenie digitálneho podpisu, zabezpečenie súkromného kľúča a výpočet verejného kľúča. Funkcia `PSS_VERIFY` overuje podpis a vráti hodnotu 1, ak je overenie úspešné.

Overenie digitálneho certifikátu na báze eliptickej krivky SECP256R1 – Eliptická krivka SECP256R1[9] patrí do skupiny Weierstrassových kriviek. Zvolená impelentácia tejto eliptickej krivky je z knižnice TinyCrypt. Funkcia `SECP256R1_ECDSA_VERIFY` má argumenty typu hašovacej funkcie, štruktúry reprezentujúcej certifikát, štruktúry reprezentujúcej digitálny podpis a verejný kľúč. V rámci tejto funkcie sa vytvára haš (SHA256) z certifikátu.

A. Sieťová konektivita na vstavanej platforme ESP32-C6

V informačných technológiách sa termín sieťový socket používa na označenie koncového bodu, ktorý je pripojený prostredníctvom počítačovej siete. Pri volaní sieťového rozhrania na zabezpečenie komunikácie medzi dvoma alebo viacerými bodmi v sieti, dochádza k vytvoreniu socketu, ktorý je jedinečne identifikovaný kombináciou IP adresy a čísla portu. Vývojová doska ESP32-C6 využíva sieťovú konektivitu rovnako, ako operačný systém Linux. Pre zabezpečenie funkcionality Wi-Fi konektivity v zariadení ESP32-C6 je potrebné začleniť hlavičkový súbor `wifi.h`. Tento súbor poskytuje potrebné rozhrania a definície, ktoré umožňujú zariadeniu komunikovať prostredníctvom Wi-Fi siete. Tiež poskytuje možnosť konfigurovať sieťové parametre, ako napríklad priradenie statickej súkromnej IP adresy. Kľúčovým prvkom v konfigurácii je štruktúra `wifi_config_t`, ktorá obsahuje makrá `WIFI`

[9] https://neuromancer.sk/std/secg/secp256r1

a `PASSWORD`. Tieto makrá definujú názov Wi-Fi siete a príslušné heslo. Dôležité je aj zdôraznenie autentikačného módu WIFI_AUTH_WPA_PSK, ktorý je zásadný pre zabezpečenú komunikáciu medzi zariadením a Wi-Fi sieťou.

B. Generovanie náhodných čísel

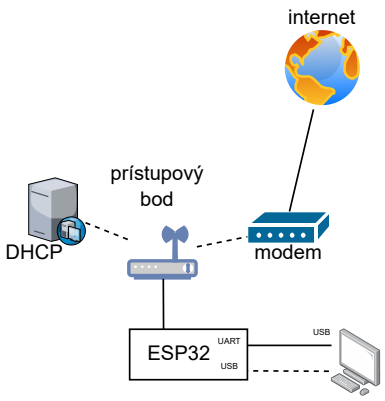
Generátory náhodných čísel sú kľúčové v zabezpečenej komunikácii na tvorbu soli a generovanie súkromných kľúčov. Existujú dva hlavné typy: pseudonáhodné (PRNG) a skutočne náhodné (TRNG). PRNG používa algoritmy na tvorbu deterministických sekvencií, ktoré vytvárajú pseudonáhodné dáta, zatiaľ čo TRNG využíva fyzikálne procesy pre vytvorenie entropie. V kryptografii sa často používa CSPRNG (Cryptographically Secure Pseudorandom Number Generator), ktorý je navrhnutý na zabezpečenie nepredvídateľnosti. Tento typ generátora zvyčajne využíva entropiu z kvalitného zdroja, ako je API (Aplication Programming Interface) operačného systému.

V rámci optimalizácie pseudonáhodného generátora náhodných čísel z knižnice PicoTLS bola vykonaná zmena, ktorá umožňuje flexibilnejšie a efektívnejšie generovanie pseudonáhodných hodnôt. Pôvodný limit maximálneho počtu vygenerovaných hodnôt sa upravil na 10 000 v súlade s knižnicou MbedTLS. V prípade prekročenia tohto limitu dôjde ku opätovnej inicializácií entropie, čím môže generovanie pokračovať namiesto ukončenia programu. Táto zmena je výhodná pre vstavané zariadenia, kde by ukončenie aplikácie mohlo mať závažné dôsledky. Generovanie pseudonáhodných hodnôt sa vykonáva prostredníctvom funkcie `SAL_randomOctad`, ktorá bola navrhnutá tak, aby generovala hodnoty a ukladala ich do štruktúry `octad`. Ak je požadovaná dĺžka menšia alebo rovná maximálnej dĺžke, ktorú môže generovať funkcia `cf_hash_drbg_sha256_gen`, generujú sa hodnoty priamo. Ak je požadovaná dĺžka väčšia, generovanie prebieha v cykle, pričom sa po každom cykle kontroluje, či aktuálny index dosiahol požadovanú dĺžku. Ak áno, vykoná sa opätovná inicializácia generátora. Týmto spôsobom funkcia `SAL_randomOctad` umožňuje flexibilné generovanie pseudonáhodných hodnôt s požadovanou dĺžkou.

Na platforme ESP32-C6 sa pre generovanie entropie využíva funkcia `esp_fill_random` z hlavičkového súboru `esp_system.h`. Táto funkcia naplní buffer náhodnými bajtmi z hardvérového generátora. Generovanie entropie je závislé od povolenia periférií Wi-Fi, Bluetooth a interného zdroja entropie funkciou `bootloader_random_enable`. Generovanie entropie je založené na fyzickom šume a spomenuté podmienky sú kľúčové pre jeho kvalitu. ESP32-C6 obsahuje´aj druhé štádium bootloadera, ktoré získava aj informácie pre entropiu.

VI. Analýza rýchlosti kryptografických operácií v TLS 1.3 na ESP32-C6

Na zariadení ESP32-C6 je dôležitá doba trvania kryptografických operácií, ako je výpočet HKDF, overenie digitálneho podpisu a autentizačné šifrovanie. Pre experimentálne overenie aplikácie minimalistického TLS 1.3 klienta je nevyhnutné pripojiť vývojovú dosku ESP32-C6 cez UART rozhranie k počítaču prostredníctvom USB portu. Sieťová konektivita, vrátane Wi-Fi, je opísaná v sekcii IV, podsekcii A. Schéma zapojenia zariadenia ESP32-C6 sa nachádza na Obr. 2. Zariadenie ESP32-C6 je pripojené na štandardný smerovač (router), ktorý má v sebe zabudovaný DHCP server a modem. DHCP server pridelí privátnu IP adresu zariadeniu ESP32-C6 a modem zabezpečí pripojenie z domácej siete do internetu.



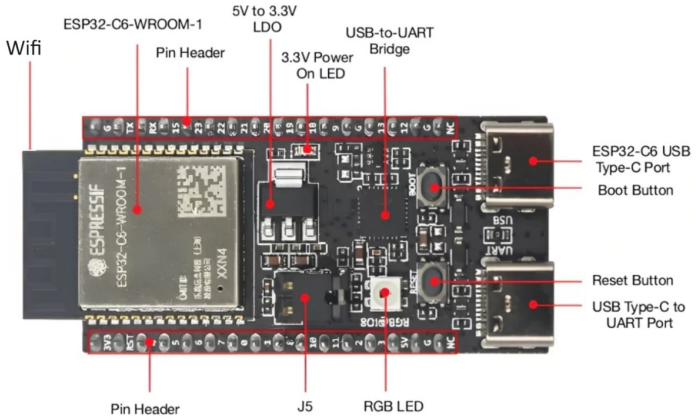
Obr. 2 Schéma zapojenia vývojového kitu na báze ESP32-C6 k štandardnému smerovaču.

A. Vývojový kit na báze ESP32-C6

ESP32-C6 je nízkopríkonový systém na čipe (SoC), pričom konkrétny model ESP32-C6-DevKitC-1-N8 je vývojová doska založená na tomto SoC. Dané zariadenie je znázornené na Obr. 3[10]. Tento systém je vybavený jedno-jadrovým mikroprocesorom typu RISC-V. Niektoré z jeho kľúčových parametrov sú:

- **procesor**: ESP32-C6-DevKitC-1-N8 je nízkonákladová vývojová doska, ktorá využíva modul ESP32-C6-WROOM-1, ktorý obsahuje 32-bitový jedno-jadrový RISC-V procesor. Procesor ESP32-C6 môže pracovať s frekvenciou až 160 MHz,
- **RAM**: modul obsahuje 512 kB HP (High-Power) SRAM (Static Random Access Memory) a 16 kB LP (Low-Power) SRAM,
- **Wi-Fi**: podporuje Wi-Fi 6 v pásme 2,4 GHz,
- **bluetooth**: podporuje Bluetooth 5,
- **vybrané periférie**: podporuje IEEE 802.15.4 (Zigbee 3.0 a Thread 1.3), má USB (Universal Serial Bus) Type-C port.

Procesor s jadrom RISC-V využíva open-souce inštrukčnú sadu (ISA (Instruction set architecture)). ISA je základnou súčasťou návrhu procesora, ktorá definuje, ako komunikuje s hardvérom a softvérom. RISC-V je perspektívnym jadrom v oblasti mikroprocesorov.



Obr. 3 Vývojový kit na báze ESP32-C6.

B. Experimentálne meranie doby trvania vybraných kryptografických funkcií

Na meranie času bola využitá funkcia `esp_timer_get_time()`, ktorá využíva interný časovač procesora s vysokým rozlíšením. Táto funkcia číta 64-bitový hardvérový časovač sledujúci hodnoty času od štartu systému a konvertuje ich na mikrosekundy.

```
int64_t start_time = esp_timer_get_time();
meranaFunkcia();
int64_t end_time = esp_timer_get_time();
printf("Namerany cas v mikrosekundach %lld\n",
        end_time-start_time);
```

Ako príklad merania je uvedená funkcia `meranaFunkcia()`, ktorej doba trvania sa vypočíta odčítaním dvoch 64-bitových premenných `end_time` a `start_time`, ktoré sú vyčítané funkciou pre zaznamenanie aktuálneho času v mikrosekundách.

[10]https://www.luisllamas.es/wp-content/uploads/2023/01/esp32-c6-details.jpg

kryptografický algoritmus	nameraný čas (ms)
generovanie entropie (64 bajtov)	1,05
generovanie pseudonáhodného bajtu	0,16
x25519 (generovanie kľúčového páru)	216,69
x25519 (výpočet zdieľaného tajomstva)	214,81
HKDF-extrakcia (96 bajtov)	[0,19 - 0,33]
HKDF-expanzia ([56 - 118] bajtov)	[0,19 - 0,36]
overenie digitálneho podpisu RSA (2048)	253,14
overenie digitálneho podpisu RSA (4096)	316,33
overenie digitálneho podpisu SECP-256-R1	179,95
AEAD - šifrovanie ([3 - 38] bajtov)	[17,47 - 28,51]
AEAD - dešifrovanie ([31 - 1025] bajtov)	[22,98 - 363,63]

Tabuľka 2
Nameraný čas vykonania kryptografických algoritmov na zariadení ESP32-C6 s nastavenou frekvenciou procesora 160 MHz.

Tabuľka 2 uvádza namerané časy trvania kryptografických algoritmov. Najrýchlejšia je generácia pseudonáhodného bajtu, najpomalšie sú autentizačné dešifrovanie a overenie RSA podpisu s 4096-bitovým kľúčom. Overenie digitálneho podpisu RSA s 2048-bitovým kľúčom je 1,1-krát pomalšie v porovnaní s meraním, ktoré je uvedené na stránke WolfSSL[11].

VII. Záver

Tento článok poskytol základnú analýzu protokolu TLS 1.3 so zameraním na jeho kryptografické algoritmy. Hlavným cieľom bolo vytvoriť minimalistickú implementáciu TLS 1.3 klienta pomocou programovacích jazykov C/C++ a kryptografickej knižnice TIIGER TLS. Minimalizácia TLS 1.3 klienta zahŕňala obmedzenie podporovanej sady kryptografických algoritmov a následné použitie kryptografických funkcií z iných kryptografických knižníc prevažne pre vstavané zariadenia. Týmto spôsobom bol vytvorený minimalistický TLS 1.3 klient, ktorý nie je závislý na žiadnych externých knižniciach. Tento minimalistický TLS klient je navrhnutý primárne pre vstavané zariadenia s obmedzenými zdrojmi, ako je napríklad zariadenie ESP32-C6. Implementácia bola otestovaná na tomto zariadení a tiež boli vykonané časové meranie kryptografických funkcií v minimalistickom TLS 1.3 klientovi. Minimalizácia TLS 1.3 klienta poskytuje možnosť ďalšieho rozvoja, respektíve pridania alebo zmenenia kryptografických algoritmov.

VIII. Poďakovanie

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-22-0400.

Literatúra

[1] S. B. Alemu, "The transformation of tls from version 1.2 to 1.3: Efficiency vs security vs interoperability," in *Bachelor's Thesis*. ETH Zürich, Institute of Information Security, Department of Computer Science, Mar. 2020.
[2] J. Pajkoš, "Bezpečná komunikácia klient-server s využitím protokolu tls 1.3," TUKE, FEI Košice, máj 2024, Bakalárska práca.
[3] R. Chaves, G. Kuzmanov, L. Sousa, and S. Vassiliadis, "Improving sha-2 hardware implementations," in *Cryptographic Hardware and Embedded Systems - CHES 2006, 8th International Workshop*, vol. 4249. Springer, 2006, pp. 298–310. [Online]. Available: https://iacr.org/archive/ches2006/24/24.pdf
[4] TCC, "Theory of Cryptography Conference," in *Advances in Cryptology - TCC 2016: 14th International Conference on the Theory and Application of Cryptology and Information Security, Taipei, Taiwan, January 24-28, 2016, Proceedings, Part I*, ser. Lecture Notes in Computer Science, vol. 9562. Springer, 2016, pp. 1–508. [Online]. Available: https://www.iacr.org/archive/crypto2016/98140241/98140241.pdf
[5] H. Krawczyk, "Cryptographic extraction and key derivation: The hkdf scheme," in *Advances in Cryptology – CRYPTO 2010*, T. Rabin, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 631–648. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-14623-7_34
[6] M. Kleppmann, "Implementing curve25519/x25519: A tutorial on elliptic curve cryptography," 2021, available at https://www.cl.cam.ac.uk/teaching/2122/Crypto/curve25519.pdf.
[7] P. M. de Zeeuw and E. J. van Asselt, "The convergence rate of multi-level algorithms applied to the convection-diffusion equation," *SIAM Journal on Scientific and Statistical Computing*, vol. 6, no. 2, pp. 492–503, 1985. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/0906034
[8] B. Kaliski, "PKCS 1: RSA Encryption Version 1.5," RSA Laboratories East, Request for Comments 2313, March 1998. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc2313
[9] S. A. Kakvi, "On the security of rsa-pss in the wild," Cryptology ePrint Archive, Paper 2019/1268, 2019, https://eprint.iacr.org/2019/1268. [Online]. Available: https://eprint.iacr.org/2019/1268

[11] https://www.wolfssl.com/espressif-risc-v-hardware-accelerated-cryptographic-functions-up-to-1000-faster-than-software/

Možnosti použitia polymérnych kompozitných materiálov v konštrukcii FSR senzorov

Ľubomír LIVOVSKÝ

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

lubomir.livovsky@tuke.sk

***Abstrakt* — Nedávne pokroky vo vývoji polymérnych kompozitov otvorili priestor pre cenovo efektívne riešenia v oblasti flexibilných hmatových, tlakových senzorov. Tento článok skúma vhodnosť polyetylén-uhlíkového kompozitného materiálu pre návrh flexibilných FSR (Force Sensitive Resistor) senzorov. Štúdia sa zameriava na analýzu správania sa materiálov s ohľadom na zmeny rezistivity vyvolané vonkajšími silami, najmä skúmanie úloh perkolačných javov a kvantového tunelovania. Poznatky získané z tejto štúdie prispievajú k pochopeniu potenciálnych aplikácií kompozitných materiálov v senzorovej technológii a podčiarkujú jeho účinnosť pri výrobe flexibilných senzorov.**

***Kľúčové slová* — polymérny kompozitný materiál, Linqstat, Velostat®, rezistivita, FSR senzor.**

I. Úvod

Materiály a kompozity na báze polymérov sú perspektívne pre vývoj rôznych taktilných senzorov. Takéto senzory si vyžadujú stabilné a spoľahlivé polymérne materiály, aby bolo možné vytvárať účinné systémy s transformáciou sily na elektrický signál [1]. Existujú rôzne fyzikálne javy, ktoré využívajú senzory citlivé na pôsobenie sily: a) tenzometrický jav – využíva zmenu odporu deformovaného vodiča, b) kvantové tunelovanie – využíva tunelový efekt, ktorý spôsobuje vodivosť v kompozite s vodivými plnivami rozptýlenými vo vnútri polymérnej matrice, c) perkolácia – ktorá pôsobením sily zabezpečuje prechod z izolovaného do vodivého stavu. Materiály, ktoré transformujú zmenu sily na zmenu elektrickej vodivosti označujeme ako piezorezistívne materiály [2].

Materiály Linqstat a Velostat® sú primárne navrhnuté ako antistatický materiál, pričom ich hlavnou funkciou je eliminovať nežiadúce prejavy statického elektrického náboja. Tento článok poukazuje na to, že aj u týchto polyetylén-uhlíkových kompozitných materiálov je možné určiť prah perkolácie a kvantovania vodivých častíc. Kompozitné materiály je tak možné použiť napr. na konštrukciu jednoduchých a cenovo dostupných interdigitálnych FSR senzorov [11], ako sú senzory tlaku, hmotnosti, deformácie.

II. Princíp elektrickej vodivosti kompozitov

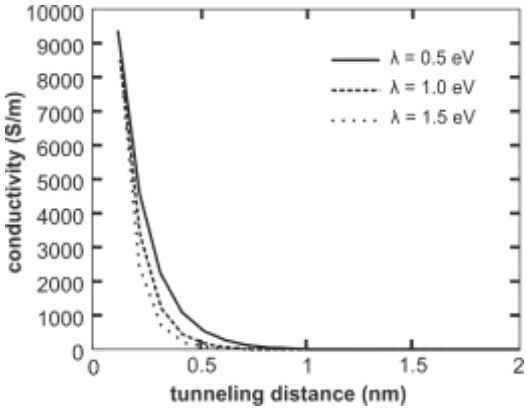
Elektrické spojenie medzi vodivými časticami v kompozitnom materiály vzniká dvoma rôznymi mechanizmami: kvantovým tunelovaním elektrónov a perkoláciou [3]. Polymérne kompozity na báze polyetylén-uhlíkových materiálov sú výhodné z hľadiska tunelovania a perkolácie najmä preto, že polyetylén je polymér, ktorý je známy svojou flexibilitou a formovateľnosťou. Táto vlastnosť umožňuje ľahkú integráciu uhlíkových častíc do matrice polyetylénu a tvorbu kompozitov s rôznymi tvarmi a veľkosťami. Uhlíkové častice v polyetylénovej matrici vytvárajú v kompozite sieť vodivých častíc, ktorá môže uľahčiť tunelovanie elektrónov medzi jednotlivými časticami.

A. Elektrická vodivosť - kvantové tunelovanie

Teória kvantového tunelovania predpokladá výskyt rozptýlených vodivých častíc v matrici izolátora, ktoré môžu vytvoriť elektrické prepojenie tak, že elektróny napriek izolačnej bariére môžu prechádzať z jednej častice do druhej. To znamená, že aj keď častice nemajú fyzickú interakciu, interagujú „elektricky“ prostredníctvom efektu tunelovej vodivosti medzi časticami.

Kvantové tunelovanie spôsobuje nelineárnu napäťovo-prúdovú odozvu vodivých častíc

nanokompozitov. Zo závislosti na Obr. 1 vyplýva, že ak šírka bariéry izolátora $\lambda >> e\Delta V$, tunelová vodivosť klesá s rastúcou dĺžkou tunela [4].

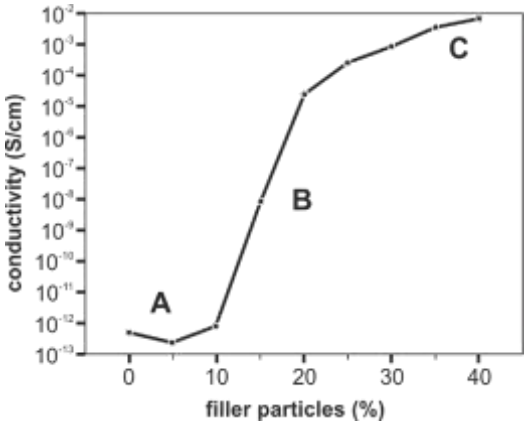


Obr. 1: Závislosť vodivosti nanokompozitu na tunelovej vzdialenosti v izolačnej matrici [4]

B. Elektrická vodivosť – perkolácia

Teória perkolácie popisuje vlastnosti kompozitných polymérnych materiálov s vodivými časticami, pri ktorých dochádza k zmenám z izolovaného do vodivého stavu a naopak. Stav, pri ktorom sa prvýkrát vytvorí perkolačná (vodivá) sieť, sa nazýva perkolačný prah.

Krivku závislosti elektrickej vodivosti na koncentrácii plniva (Obr. 2) je možné rozdeliť na tri oblasti [6]. Oblasť (A) predstavuje nízku koncentráciu plniva, elektrická vodivosť je o 1 až 2 rády väčšia, ako vodivosti polymérnej matrice bez plniva. V oblasti (B) dochádza ku vzniku súvislých vodivých dráh v celom priereze kompozitu a v pomerne úzkom intervale koncentrácií plniva elektrická vodivosť rastie o niekoľko rádov (4 až 10). Oblasť (C) už len v malej miere prispieva k elektrickej vodivosti (2 až 3 rády), nakoľko transport elektrónov je už zabezpečený predtým vytvorenou vodivou sieťou.



Obr.2: Závislosť elektrickej vodivosti polymérnych kompozitov na obsahu častíc [6]

III. TESTOVANÉ KOMPOZITNÉ MATERIÁLY

Na trhu polymérnych kompozitných materiálov vhodných pre konštrukciu FSR senzorov sú materiály od firmy 3M s označením Velostat® a od firmy Caplinq s označením Linqstat. Oba materiály sú vyrobené rovnakou technológiou, pričom Velostat® je patentovo chránený názov materiálu a Linqstat je obchodný názov pre generický materiál vyrábaný a predávaný firmou Caplinq. Oba materiály by sa mali vyznačovať podobnými vlastnosťami, pretože používajú rovnaký technologický postup výroby. Hlavnými výhodami týchto materiálov sú flexibilita, mechanická a chemická stabilita a relatívne nízka cena. Oba materiály sú kompozitné polymérne materiály tvorené uhlíkom impregnovaným do polyetylénu. Prítomnosť uhlíkového prášku transformuje pôvodne dielektrický polyetylén na elektricky vodivý kompozitný materiál, ktorý je možné zaradiť medzi piezorezistívne materiály [2]. Velostat® a Linqstat sú navrhnuté s ohľadom na antistatické a elektrostatické vlastnosti. Okrem už spomenutých vlastností kompozitných materiálov sa vyznačujú podobnými charakteristikami, vrátane flexibilnej povahy, mechanického základu a najmä relatívne nízkej ceny.

Flexibilné piezorezistívne materiály majú tendenciu deformovať sa časom a nevracajú sa do

pôvodného stavu. Okrem toho sa piezorezistívne senzory zvyčajne vyznačujú nelinearitou a vykazujú významnú hysteréziu [10]. Nejednotná, resp. nehomogénna štruktúra materiálu negatívne ovplyvňuje opakovateľnosť nameraných údajov, čo vedie k nízkej presnosti a opakovateľnosti meraní tlaku alebo sily [2].

Hoci implementácie týchto materiálov ako citlivej vrstvy senzorov boli predmetom mnohých štúdií, hodnotenie elektrických a mechanických charakteristík tohto kompozitného materiálu je stále nedostatočné. Citlivosť polymérneho kompozitu na aplikovaný tlak je definovaná hlavne kvantovým tunelovaním a perkoláciou. Oba tieto javy sa v skutočnosti vyskytujú súčasne, ale v závislosti od koncentrácie častíc dominuje jeden mechanizmus vedenia prúdu nad druhým. Podľa uskutočnených štúdií by mali tieto materiály meniť svoj elektrický odpor v dôsledku týchto efektov pri mechanickom namáhaní [10].

Cieľom tohto článku je otestovať vlastnosť 5 vzoriek polymérnych materiálov a posúdiť ich vhodnosť pre aplikáciu v flexibilných tlakových senzoroch. Materiály sa budú posudzovať podľa nameraného povrchového odporu pred pôsobením tlaku a po tlakovom zaťažení. Praktické meranie tlaku testovaných materiálov bude na realizované na konštrukcii interdigitálneho tlakového senzora.

IV. Meranie rezistivity

Experimentálne môžeme rezistivitu ρ určiť pomocou štvorbodovej metódy (4P), ktorá je založená na použití štyroch sond, pričom pokles napätia V sa meria medzi dvoma vnútornými sondami, zatiaľ čo prúd I je do vzorky dodávaný cez dve vonkajšie elektródy (in-line). V ideálnom prípade 3D objemového polonekonečného materiálu je na jeho povrchu rezistivita ρ určená vzťahom:

$$\rho_{3D}^{line} = 2\pi s \frac{V}{I}\ [\Omega m], \tag{1}$$

kde s je vzdialenosť medzi dvojicou sond [9]. Ak je hrúbka vzorky t oveľa menšia ako vzdialenosť sond s (t << s), potom môžeme materiál považovať 2D plochu a prúd v sonde sa nebude šíriť sféricky ale pozdĺžne. Rezistivitu, potom môžeme určiť:

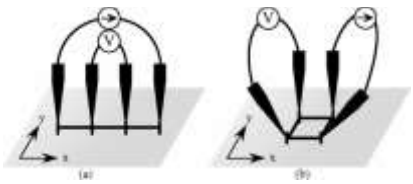
$$\rho_{2D}^{line} = \frac{\pi t}{ln2}\frac{V}{I}\ [\Omega \mathrm{m}]. \tag{2}$$

Okrem lineárneho usporiadania sond je možné umiestniť sondy do štvorca, pričom sa na meranie využíva menšia plocha (Obr. 3). Rezistivitu materiálu meranú štvorbodovou metódou s usporiadaním sond do štvorca môžeme vypočítať podľa:

$$\rho_{2D}^{squre} = \frac{2\pi t}{ln2}\frac{V}{I}\ [\Omega \mathrm{m}]. \tag{3}$$

V prípade homogénnej a nekonečne hrubej vzorky je možné povrchový merný odpor nahradiť plošným odporom (odporom na jednotku plochy) definovaným ako:

$$R_{SH} = \frac{\rho}{t}\,[\Omega]. \tag{4}$$



Obr.3: Usporiadanie sond 4P metódy merania rezistivity a) in-line, b) do štvorca [ ]

V. Experimenty

Cieľom experimentov bolo otestovať rôzne kompozitné materiály s označením Linqstat a Velostat®, ktoré sú bežne dostupné a používané aj pri výrobe senzorov, ktoré sú publikované v mnohých článkoch [12][13][14]. Napriek tomu, že autori článkov sa zameriavajú na vyhodnocovanie vlastností senzorov ako celku, často sa neuvádzajú štúdie deformačných vlastností kompozitného materiálu použitého pri konštrukcii flexibilného senzora a jeho vplyv týchto deformácií na hodnoty meranej veličiny.

Tento článok je zameraný na výskyt javov prerkolácie resp. kvantového tunelovania u týchto materiálov. Tieto javy boli analyzované prostredníctvom merania zmeny rezistivity kompozitných materiálov pred a po aplikovaní izostatického tlaku p = 20 MPa počas 20 minút pri teplote 35 °C. Hodnota tlaku mimo intervalu odporúčaných hodnôt 0 N až 3 N [10] bola zvolená za účelom vytvorenia nevratnej deformácie polymérnej matrice vodivých častíc

s následným vytvorením trvalých vodivých prepojení, čo by sa malo prejaviť poklesom rezistivity v porovnaní s pôvodnou hodnotou.

Na experimenty bolo použitých 5 vzoriek (4 vzorky Linqstat a 1 vzorka Velostat®) s rozmermi 140 mm x 140 mm. Na každej vzorke bolo vykonaných 36 meraní rezistivity (na rôznych miestach) bez aplikácie tlaku a potom po aplikovaní vonkajšieho izostatického tlaku. Merania rezistivity boli vykonané pomocou 4P metódy s umiestnením sond podľa Obr. 3b so vzdialenosťou strany štvorca s = 5 mm. Rezistivita bola meraná meracím prístrojom Agilent 34420A. Na zabezpečenie rovnakých podmienok merania, bolo použité pohyblivé zariadenie CHECKSUM, ktoré zaručovalo fixné umiestnenie meracích sond a rovnakú prítlačnú silu meracích sond ku kompozitnému materiálu.

Odmerané hodnoty rezisitivity ρ boli prepočítané na hodnotu R_{SH} podľa vzťahov (3) a (4). Výsledky sú prezentované v Tabuľke 1. Odchýlky v hodnotách R_{SH} môžu byť spôsobené nerovnomerným rozptylom vodivých častíc v polymérnej matrici, a/alebo nerovnomernou hrúbkou materiálu v meraných bodoch.

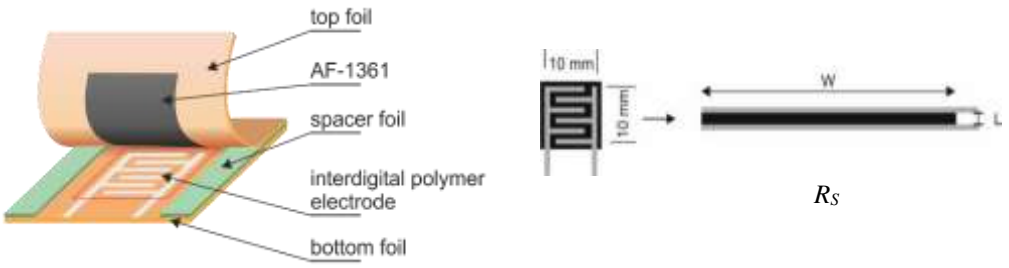
Tabuľka 1
Zmeny hodnoty plošného odporu R_{SH} testovaných kompozitných materiálov

Materiál	Hrúbka [cm]	R_{SH} [Ω]	ρ[Ωcm]	Tlak R_{SH}[Ω]	Tlak ρ[Ωcm]	ΔR_{SH} [Ω]	ΔR_{SH} [%]
Linqstat							
MVCF-8S10K	0,02	5785 ± 168	116	2863 ± 194	57	-2922	-50,51
MVCF-4S50K	0,01	19499 ± 1048	195	10626 ± 915	106	-8873	-45,50
MVCF-8S50K	0,02	10202 ± 338	204	9509 ± 430	190	-693	-6,79
VCF-2S200K	0,02	21935 ± 5771	439	23735 ± 5808	475	1800	8,21
Velostat®							
AF-1361	0,01	4863 ± 342	46	4831 ± 352	48	-32	-0,66

Následne boli na vzorkách vykonané testy vplyvu pôsobenia izostatického tlaku p = 20 MPa po dobu 20 minút pri teplote 35 °C. Tieto testy boli vykonané s cieľom zistiť, či nastane dodatočná trvalá deformácia polymérnej matrice s uhlíkovými časticami po aplikovaní tlaku, ktorý by spôsobil priblíženie vodivých častíc a tým vytvorenie dodatočných vodivých ciest. Predpoklad bol, že dôjde k výraznému poklesu hodnoty R_{SH} v porovnaní s pôvodnou hodnotou.

Z vykonaných experimentov vyplýva, že k významnému poklesu hodnoty R_{SH} (Tabuľka 1.) došlo pri vzorkách MVCF-8S10K a MVCF-8S10K, materiál MVCF-8S50K zaznamenal pokles. Naopak pri materiáloch VCF-2S200K a AF-1361 nebol zaznamenaný pokles.

Napriek tomu že niektoré kompozitné materiály v predchádzajúcom teste nevykazovali zmenu rezistivity v závislosti na tlaku, boli aj tieto materiály podrobené testom pôsobenia tlaku na FSR senzor. Ako citlivý prvok FSR senzora boli použité testované materiály. Konštrukcia FSR senzora s kompozitným materiálom AF-1361 je znázornená na Obr. 4 [11].



a) b)

Obr.4: Konštrukcia interdigitálneho FSR senzora na báze kompozitného materiálu AF-1361. a) konštrukcia senzora, b) náhradný obdĺžnikový snímací rezistor R_S [11]

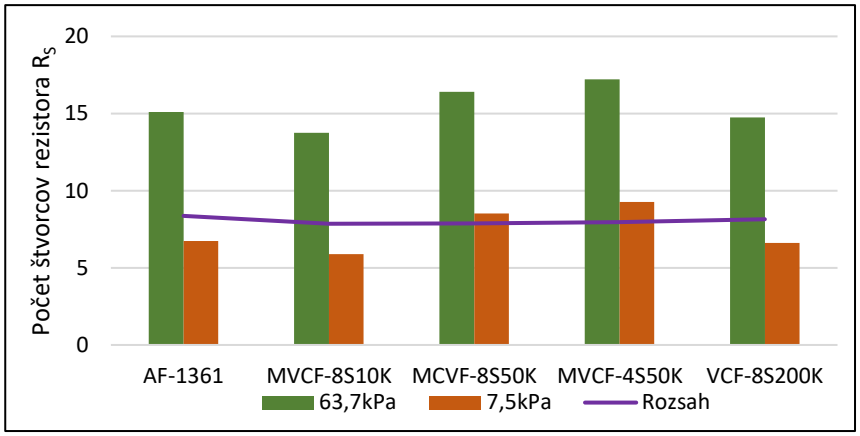
Ak poznáme hodnotu plošného odporu kompozitného materiálu R_{SH} a počet štvorcov, ktoré tvoria snímací rezistor R_S, je možné určiť jeho hodnotu podľa:

$$R_S = R_{SH} * n\ [\Omega], \tag{5}$$

kde n je počet štvorcových rezistorov, pričom platí, že n = W/L, kde L je dĺžka a W šírka snímacieho rezistora (Obr. 4b). Ak L = 1,25 mm a W = 50,0 mm, potom počet štvorcových rezistorov, ktoré tvoria snímací rezistor je n = 40.

Zväčšovaním tlaku na interdigitálny senzor dochádza k postupnému „vytváraniu“ snímacieho rezistora R_S „pridávaním“ väčšieho počtu štvorcových odporov. Ak poznáme hodnotu

snímacieho rezistora R_S , potom je možné určiť počet štvorcových rezistorov, ktoré tento rezistor tvoria. Parameter počtu štvorcových rezistor je možné použiť na vyhodnocovanie vplyvu rôznych kompozitných materiálov na FSR senzor. Tento parameter je možné použiť na porovnanie vplyvu kompozitných materiálov s rôznymi hodnotami rezistivity ρ na výslednú odozvu FSR senzora na pôsobiaci tlak. Vykonané experimenty testovaných materiálov sú zhrnuté na Obr. 5.



Obr.5: Zmena počtu štvorcov snímacieho rezistora interdigitálneho FSR senzora v závislosti od tlaku pre rôzne kompozitné materiály.

Výsledky z Obr. 5 je možné interpretovať ako pôsobenie kompozitného materiálu len vo forme vodivej fólie, ktorá sa podieľa na vytvorení snímacieho rezistora a vplyv perkolácie alebo elektrónového tunelovania sa na zmene vodivosti kompozitného materiálu neprejavuje. Hodnota „Rozsah“ je rozdiel počtu štvorcov rezistora pre aplikované hodnoty tlaku, pričom priemerná hodnota 8 bola rovnaká pre všetky testované kompozitné materiály.

Podľa doterajších skúseností s interdigitálnymi FSR senzormi, zmena odporu pochádza hlavne od zmeny veľkosti plochy medzi vodivými elektródami a vodivou polymérnou kompozitnou fóliou. Senzor realizovaný podľa Obr. 4 vykazuje závislosť R = f(p) zmeny rezistivity na pôsobiacom tlaku. Túto závislosť je možné vysvetliť mechanickou deformáciou vrchnej fólie senzora s polymérnym kompozitným materiálom, nerovnomernosťami elektród senzora ako aj hrúbky separátora v konštrukcii senzora.

VI. Záver

Vlastnosti kompozitných materiálov pri trvalom zaťažení závisia od deformácie jeho štruktúry na báze polyméru. Čas relaxácie kompozitných materiálov po odstránení záťaže je kľúčovým aspektom pre ich implementáciu ako snímacieho prvku v akomkoľvek type FSR senzorov. Vytvorenie perkolačnej vodivej siete v kompozitnom materiály je predpokladom, aby sa z izolovaného materiálu stal vodivý.

Vykonané experimenty s rôznymi polymérnymi kompozitnými materiálmi ukazujú, že zmena rezistivity týchto materiálov vplyvom perkolácie v závislosti od pôsobiaceho tlaku v konštrukcii interdigitálneho FSR senzora je diskutabilná. Kompozitné materiály, ktoré zmenili svoju hodnotu rezistivity pôsobením izostatického tlaku 20 MPa, ako aj materiály, u ktorých k takejto zmene nedošlo, zaznamenali rovnakú zmenu pri teste tlakom na FSR senzor.

K zmene rezistivity interdigitálnych FSR senzorov dochádza hlavne zmenou aktívnej plochy vodivých elektród a kompozitného materiálu. Zmenšovanie rezistivity snímacieho prvku je spôsobované zväčšovaním plochy kontaktu a z toho vyplývajúcej zmeny prechodového odporu medzi vodivými elektródami a kompozitným materiálom. Testy ukázali, že kompozitné materiály použité v interdigitálnom type FSR senzora plnia len úlohu flexibilnej vodivej fólie.

Poďakovanie

Tento príspevok vznikol vďaka projektu Vega 1/0049/24 „Výskum a vývoj nových metód charakterizácie a spoľahlivosti miniaturizovaných modulov elektroniky s využitím nových typov povrchových úprav DPS“, projektu Kega 011TUKE-4/2023 „Synergia edukačného procesu, výskumu, inovácií a priemyselného sveta v kontexte prestavby študijného programu automobilová elektronika“, projekt ITMS 313011T557 „Podpora výskumno-vývojového potenciálu v oblasti dopravných prostriedkov“ a projektu APVV 22-0289 „Synergia nových foriem a riešení spájkovania a ich vplyv na kvalitu v kontexte nízkoenergetického a environmentálneho prístupu pri výrobe výkonových modulov elektroniky“

LITERATÚRA

[1] B. Ates, S. Koytepe, A. Ulu, C. Gurses, V.K. Thakur, "Chemistry, structures, and advanced applications of nanocomposites from biorenewable resources". Chem. Rev. 2020, 120, 9304–9362.

[2] D. Giovanelli, E. Farella, "Force sensing resistor and evaluation of technology for wearable body pressure sensing," J. Sens. 2016, 1–14.

[3] A.B. Oskouyi, U. Sundararaj, P. Mertiny, "Tunneling Conductivity and Piezoresistivity of Composites Containing Randomly Dispersed Conductive Nano-Platelets," Materials (Basel), 2014.

[4] J.G. Simmons, "Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film," J. Appl. Phys. 1963;34:1793–1803.

[5] C. Jie, Z. Xinxing, "Modulating the percolation network of polymer nanocomposites for flexible sensors," Journal of Applied Physic 128, 220901, 2020

[6] *I. J. Youngs, "Exploring the universal nature of electrical percolation exponents by genetic algorithm fitting with general effective medium theory," Journal of Physics D, 2002, 35, 3127-3137*

[7] M. Kalantari, J. Dargahi, J. Kövecses, M. G. Mardasi, S. Nouri, "A New Approach for Modeling Piezoresistive Force Sensors Based on Semiconductive Polymer Composites," IEEE/ASME Transactions on Mechatronics, 17(3), 572–581, 2012.

[8] R. Holm, "Electric Contacts," New York: Springer, 1967.

[9] I. Miccoli, F. Edler, H. Pfnür, C. Tegenkamp, "The 100th anniversary of the four-point probe technique: the role of probe geometries in isotropic and anisotropic systems," J Phys Condens Matter, 2015.

[10] A. Dzedzickis, E. Sutinys, V. Bucinskas, U. Samukaite-Bubniene, B. Jakstys, A. Ramanavicius, I. Morkvenaite-Vilkonciene, "Polyethylene-Carbon Composite (Velostat®) Based Tactile Sensor," *Polymers* 2020, *12*, 2905.

[11] I. Vehec, L. Livovsky, "Flexible Resistive Sensor Based on Velostat," *2020 43rd International Spring Seminar on Electronics Technology (ISSE)*, Demanovska Valley, Slovakia, 2020, pp. 1-6.

[12] K. Kim, K.R. Lee, W.H. Kim, K.B. Park, T.H. Kim, J.S. Kim, J.J. Pak, "Polymer-based flexible tactile sensor up to 32 × 32 arrays integrated with interconnection terminals," Sens. Actuators A Phys. 2009, 156, 284-291

[13] H. Eun-Soo, S. Jeong-Hoon, K. A. Yong-Jun,"Polymer-Based Flexible Tactile Sensor for Normal and Shear Load Detection," In Proceedings of the 19th IEEE International Conference on Micro Electro Mechanical Systems, Istanbul, Turkey, 22–26 January 2006; pp. 714–717.

[14] T. Ules, A. Hausberger, M. Grießer, S. Schlögl, D.P. Gruber, "Introduction of a new in-situ measurement system for the study of touch-feel relevant surface properties," Polym. 2020, 12, 1380.

[15] T. Kanagawa, R. Hobara, I. Matsuda, T. Tanikawa, A. Natori, S. Hasegawa, "Anisotropy in conductance of a quasi-one-dimensional metallic surface state measured by a square micro-four-point probe method", Phys Rev Lett. 2003, 91(3):036805.

Aplikácia grafénu vo flexibilnej elektronike

[1]*Martin MATI,* [2]*Peter LUKÁCS*

[1,2]Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

[1]martin.mati@student.tuke.sk, [2]peter.lukacs@tuke.sk

***Abstrakt* — Táto práca sa zaoberá možnosťami využitia grafénu v oblasti flexibilnej elektroniky. Možnosti využitia biodegradovateľných vodivých materiálov je čoraz viac skúmaný vedeckou komunitou po celom svete. Práve náhrada konvenčných medených flexibilných dosiek plošných spojov podporuje znižovania elektronického odpadu, ktorého likvidácia je finančne, časovo a ekologicky náročná. V práci sú popísané vybrané aplikácie grafénu ako náhrady za anorganický vodič. Taktiež sú opísané výrobné metódy grafénu. V práci sú uvedené konkrétne príklady z oblasti senzorov, zberačov energie, solárnych článkov a OLED displejov, kde grafén predstavuje sľubnú náhradu za bežné anorganické materiály.**

***Kľúčové slová* — Grafén, grafit, flexibilná elektronika, CVD**

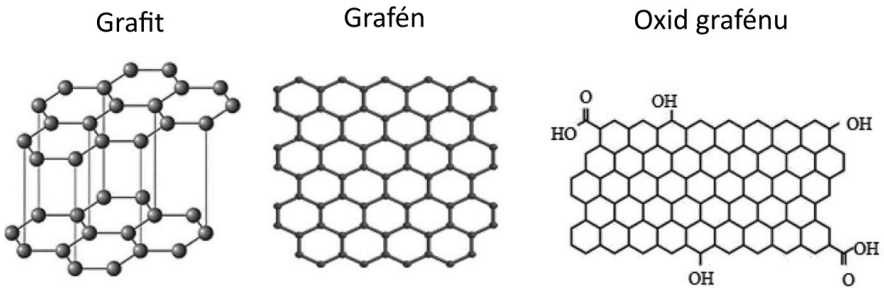
I. Úvod

Výskum v oblasti materiálov pre flexibilnú elektroniku zaznamenal v posledných rokoch množstvo úspechov. Jedným z takýchto materiálov je napríklad grafén, ktorý je označovaný ako materiál budúcnosti, kvôli je všestrannému využitiu v rôznych oblastiach. Najväčší potenciál vykazuje v oblasti smartfónov, solárnych článkov, či nositeľnej elektroniky. Predpokladá sa, že grafén by mohol nahradiť komerčne používaný ITO (Indium tin oxide – Indiom obohatený oxid cínu), keďže indium je vzácne a jeho cena je vysoká. Grafén ďalej nahrádza anorganické materiály vo flexibilnej, ktoré nemajú veľmi dobré mechanické vlastnosti ako napríklad ohybnosť, či mechanická odolnosť v ťahu. Okrem výborných mechanický vlastností je grafén elektricky vodivý a priehľadný, a preto je vhodné ho použiť v oblasti flexibilnej elektroniky [1].

II. Grafén a jeho vlastnosti

A. Grafén

Grafén je jednoatómová planárna 2D vrstva grafitu, kde sú atómy uhlíka v hybridnom stave sp^2 . Hrúbka tohto materiálu je teda úmerná veľkosti jedného atómu uhlíka. Jednotlivé atómy uhlíka sú usporiadané do šesťuholníkového tvaru. Dĺžka väzieb medzi jednotlivými atómami je 0,142 nm. S myšlienkou vytvorenia jednoatómovej štruktúry prišiel v 40-tych rokoch 20. storočia Phillip Russel Wallace, no prvýkrát bola táto štruktúra vytvorená až v roku 2004. Práve v tomto roku ruskí výskumníci Geim a Novoselov na univerzite v Manchestri po prvýkrát izolovali vrstvu grafénu na vrstve SiO_2 (oxid kremičitý) z grafitu metódou lepiacej pásky. Za tento objav dostali obidvaja vedci v roku 2010 Nobelovu cenu. Táto metóda sa však v súčasnosti používa iba na vytvorenie malého množstva grafénu pre účely výskumu [2]. Na Obr. 1 je grafické znázornenie atómov uhlíka v štruktúre grafitu, grafénu a oxidu grafénu.

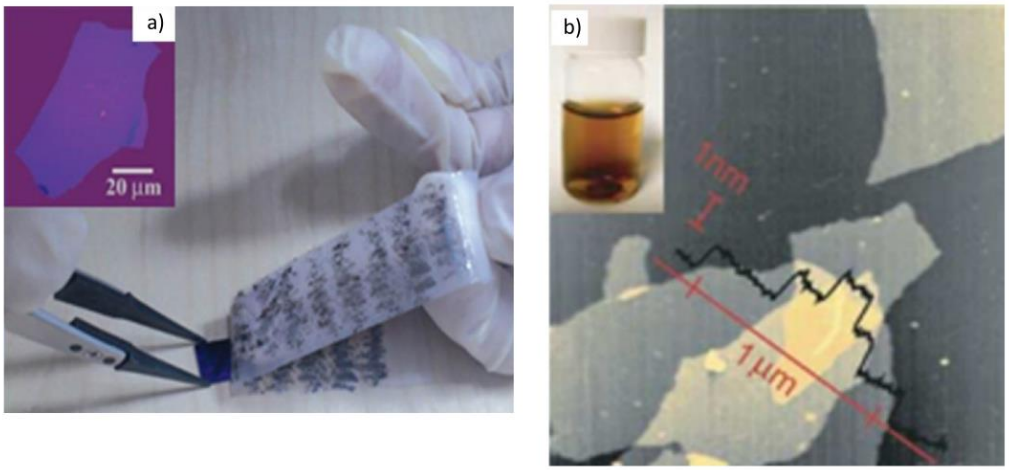


Obr. 1 Grafické znázornenie atómov uhlíka v štruktúre grafitu, grafénu a oxidu grafénu

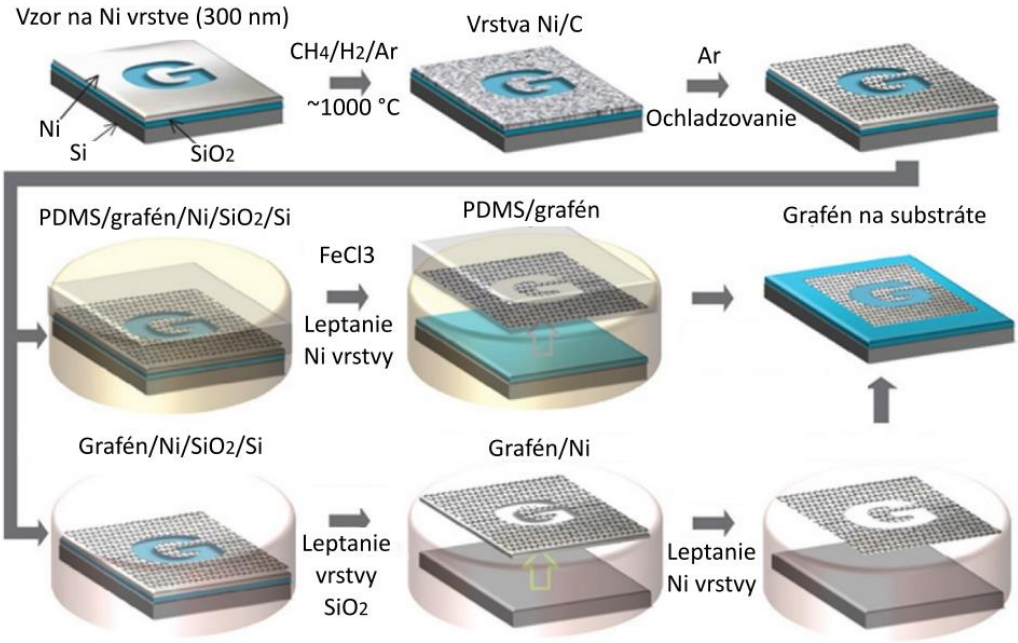
Ako už bolo spomenuté, grafén má dobré elektricky vodivé vlastnosti, no zároveň má aj dobré tepelne vodivé vlastnosti. Grafén je mechanicky pevný a pružný, dokáže sa roztiahnuť až na 20% svojej pôvodnej veľkosti [2].

B. Metódy výroby grafénu

Najjednoduchšou metódou pre vytváranie grafénu je tzv. metóda lepiacej pásky (Obr. 2a). Podstatou tejto metódy je odštiepenie jedinej vrstvy grafitu. To je možné dosiahnuť pomocou obyčajnej lepiacej pásky. Postupným nalepovaním lepiacej pásky na povrch grafitu je možné odlúpať postupne tenšie kúsky grafitu až kým sa nedosiahne požadovaná hrúbka jedného atómu, čo predstavuje jednu vrstvu grafénu. Keďže grafit je tvorený vrstvami grafénu v hybridnom stave sp^2, tak Van der Waalsove sily medzi jednotlivými vrstvami grafénu sú pomerne slabé, čo umožňuje odštiepiť vrstvu grafénu pomocou lepiacej pásky. Aj keď je táto metóda jednoduchá, nie je možné ju využiť na veľkoplošnú výrobu, pretože produkuje iba malé množstvo grafénu. Napriek tomu, je grafén vyrobený touto metódou kvalitný a jeho štruktúra neobsahuje veľké množstvo nedostatkov (defektov) [1],[2],[3].



Obr. 2 a) Mechanické odlupovanie (exfoliácia) vrstvy grafénu pomocou lepiacej pásky, b) chemicky vyrobený oxid grafénu



Obr. 3 Znázornenie procesu výroby grafénu metódou CVD

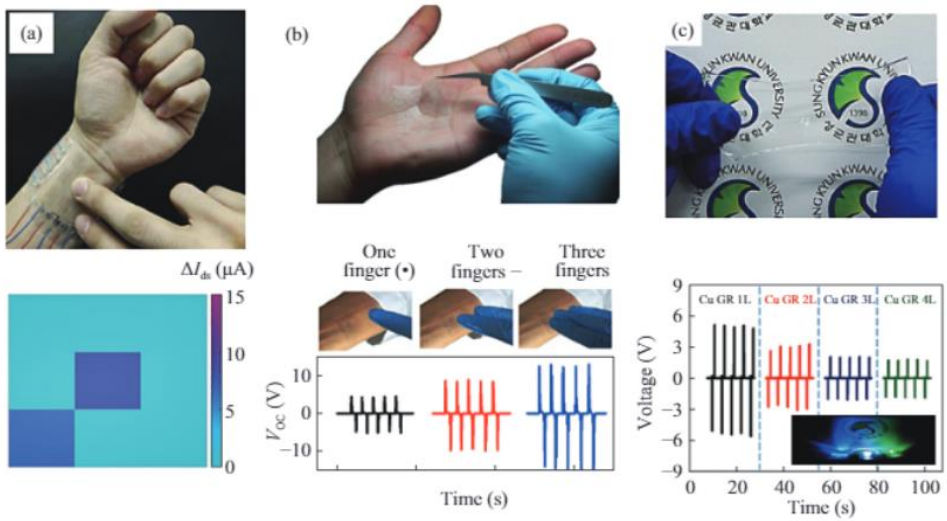
Ďalšou metódou je tzv. chemická syntéza (Obr. 2b). Chemická syntéza najbežnejšie využíva tzv. Hummerovu a Brodieho metódu. Pomocou týchto metód je možné produkovať grafén veľkoplošne. Podstatou oboch metód je oxidácia grafitu a následná redukcia tzv. oxidu grafénu. Grafit je oxidovaný pomocou kyseliny a oxidačného činidla. Ako kyselina sa najčastejšie využíva kyselina dusičná (HNO_3) alebo kyselina sírová (H_2SO_4) a ako oxidačné činidlá napríklad manganistan draselný ($KMnO_4$) alebo dusičnan sodný ($NaNO_3$). Na redukcia oxidu grafénu je

možné požiť hydrazín, kyselinu jodovodíkovú (HI) alebo Hexametyléntetraamín. Keďže oxid grafénu je hydrofilný, tak vzniknuté molekuly vody sa dostávajú medzi jednotlivé vrstvy grafénu. Tým sa zvyšuje vzdialenosť medzi paralelnými vrstvami grafénu, čo umožňuje ich oddelenie. V poslednom kroku sa na definitívne odlúpnutie grafénu používa ultrazvuk [1],[3],[4],[5].

Najpoužívanejšou metódou je tzv. chemická depozícia pár (CVD – Chemical vapor deposition). Podstatou tejto metódy je rast grafénovej vrstvy na kovovom substráte/vrstve (Ni, Cu, Pt) za pomoci uhľovodíkov (CH_4, H_2 alebo alternatívne Ar). Katalytický kov, ako je napríklad nikel, absorbuje uhlík, ktorý vzniká rozkladom uhľovodíkov pri vysokej teplote (približne 1000 °C). Atómy uhlíka sa usadzujú na povrchu kovovej vrstvy. V procese chladnutia postupne vzniká grafén. Na kovový substrát môže obsahovať motív, ktorý môže byť prenesený na grafénovú vrstvu. Po odleptaní kovovej vrstvy dostávame grafénovú vrstvu s motívom, ako je napríklad na Obr.3. Hlavnou nevýhodou CVD metódy je potenciálna prítomnosť defektov v katalytickom materiáli, nedostatočná kontrola nad formovanou hrúbkou grafénovej vrstvy a cena katalytických kovov. Komerčne je takto možné vyrábať grafén pomocou technológie roll-to-roll [1],[3].

III. FLEXIBILNÉ ZBERAČE ENERGIE

Ľudské telo generuje prebytočnú mechanickú energiu, ktorú je možné premeniť na elektrickú energiu slúžiacu napríklad na napájanie rôznych senzorov na ľudskom tele respektíve nositeľnej elektroniky. Prebytočná mechanická energia vzniká napríklad pohybom končatín, svalov, dýchaním ale aj napríklad tlkotom srdca. Technológia zberačov energie nahrádza klasické batérie, ktoré majú limitovanú životnosť. Takéto zberače energie môžu napájať aj senzory, ktoré sa nachádzajú v ľudskom, kde nie je možné umiestniť klasické batérie. V súčasnosti sa ako zberače energie používajú anorganické materiály, ktoré nemajú veľmi dobré vlastnosti, ako sú napríklad priehľadnosť či pružnosť, pre aplikácie v oblasti flexibilných senzorov na ľudskom tele. Keďže grafén je flexibilný a priehľadný, tak môže slúžiť ako elektróda pre flexibilné zberače energie. Flexibilné zberače energie väčšinou pracujú na piezoelektrickom alebo triboelektrickom jave [1].



Obr. 4 a) Taktilný senzor s grafénovými elektródami založený na triboelektrickom jave, b) flexibilný zberač energie založený na triboelektrickom jave, spodná časť ukazuje napäťové úrovne v závislosti od počtu priložených prstov, c) flexibilita grafénu a napäťové úrovne v závislosti od počtu grafénových vrstiev

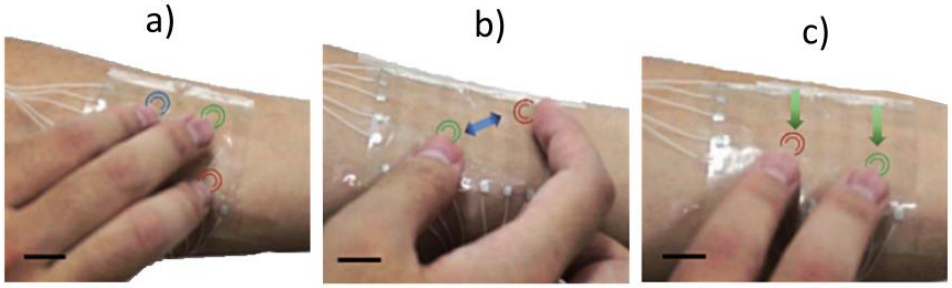
Flexibilné zberače energie založené na piezoelektrickom jave zvyčajne využívajú piezoelektrické materiály ako sú napríklad oxid zinočnatý (ZnO), sulfid kademnatý (CdS) alebo nitrid gália (GaN) v kombinácii s grafénovou elektródou. Príkladom je aj zberač zložený s nanodrôtov ZnO a grafénu. Takýto zberač energie má štruktúru superkondenzátora, v ktorom uchováva energiu získanú z mechanického pohybu alebo slnečnej energie.

Zberače založené na triboelektrickom jave sú ovplyvňované niekoľkými faktormi ako sú napríklad chemická štruktúra použitých materiálov, drsnosť povrchu alebo vlhkosť prostredia. Ak prídu do fyzického kontaktu dva dostatočne odlišné materiály s opačne nabitým nábojom, tak vzniká dostatočne veľký rozdiel potenciálov, čo umožňuje pohyb voľných elektrónov nachádzajúcich sa v oboch materiáloch [1],[4].

IV. FLEXIBILNÉ SENZORY

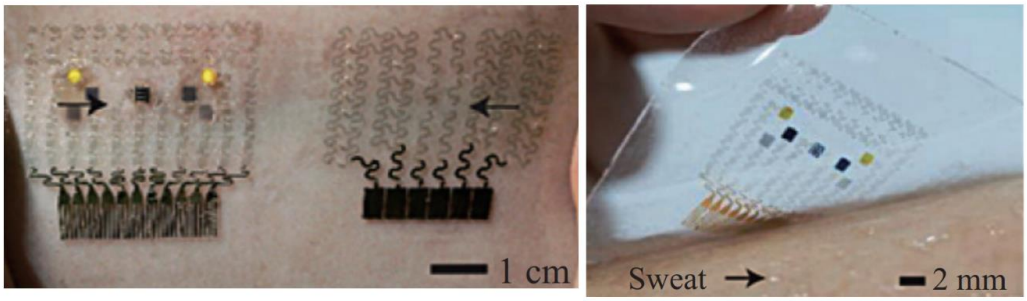
Keďže grafén nemá len výborné mechanické a elektrické vlastnosti, ale vykazuje aj isté piezoelektrické, či elektrochemické vlastnosti. Vďaka tomu spolu so s tým, že grafén je priehľadný, flexibilný a má malú hrúbku, je možné ho využiť na vytváranie nositeľných senzorov, ktoré len do malej miery znižujú komfort nositeľa [1].

Na Obr. 5 je možné vidieť priehľadný taktilný flexibilný senzor umiestnený na ruke. Tento senzor dokáže rozlíšiť, v ktorom mieste naň pôsobí sila. Senzor môže byť umiestnený napríklad aj na rukavici, kedy dokáže rozlíšiť pohyb jednotlivých prstov, čo ukazuje potenciálne využitie napríklad v hernom priemysle. Senzor funguje tak, že sa v danom mieste zmení elektrický odpor (piezorezistívny senzor) alebo kapacita (kapacitný senzor), Následne je možné tieto veličiny odmerať, z čoho je následne možné zistiť intenzitu a typ pohybu napríklad prsta. Podobný typ senzora, umiestnený na zápästí, dokáže zachytiť tep srdca a následne ho monitorovať, čo preukazuje jeho využitie v medicíne. Existujú aj tzv. grafénové elektronické tetovania, ktoré umožňujú meranie teploty či vlhkosti pokožky [1].



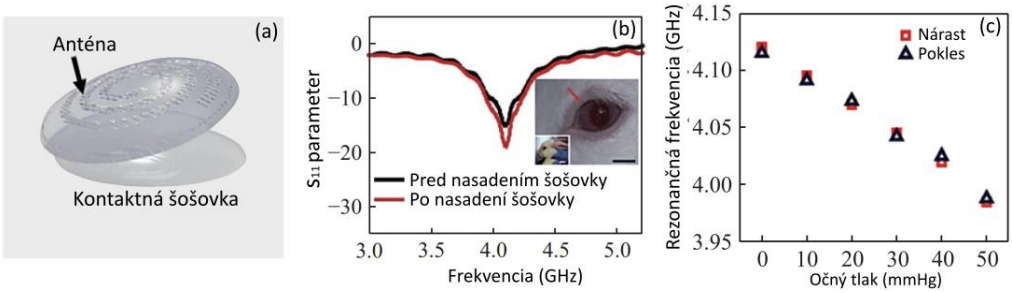
Obr. 5 Taktilný flexibilný senzor, a) dokáže rozlíšiť niekoľkých dotykov naraz, b) naťahovanie senzora, c) potiahnutie prstov po senzore jedným smerom

Na Obr. 6 je znázornený flexibilný senzor, ktorý dokáže z potu zistiť obsah cukru v krvi a následne dokáže aj hladinu cukru regulovať. Senzor využíva práve elektrochemické vlastnosti grafénu. Senzor taktiež dokáže merať teplotu, glukózu, vlhkosť, pH, ale taktiež obsahuje drobné ihly, ktoré dodávajú potrebné látky do krvného obehu [1].



Obr. 6 Flexibilný senzor na monitorovanie cukrovky

Na Obr. 7 a) je možné vidieť senzor na meranie glukózy zo slzy v oku a na meranie očného tlaku umiestnený na kontaktnej šošovke. Senzor obsahuje anténu, ktorá mení svoje vlastnosti v závislosti od meranej veličiny. Obr.7 b) ukazuje zmenu dopredného odrazového koeficientu pred a po nasadení kontaktnej šošovky v závislosti od obsahu glukózy v slze. Na Obr.7 c) je zase znázornená zmena rezonančnej frekvencie antény v závislosti od veľkosti očného tlaku [1].



Obr. 7 Grafénová kontaktná šošovka na monitorovanie glukózy zo slzy v oku

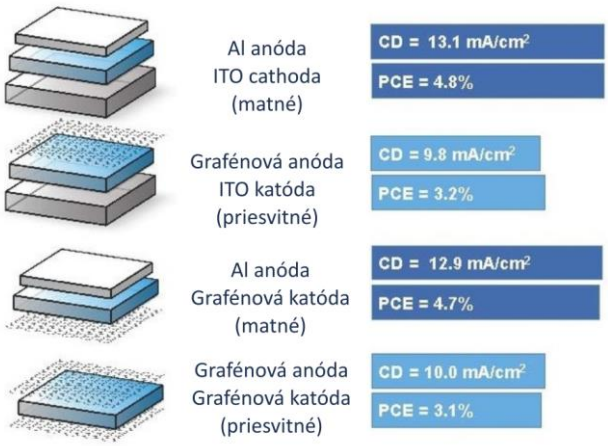
V. FLEXIBILNÉ SOLÁRNE ČLÁNKY

Solárny článok je zariadenie, ktoré vytvára dvojice z voľných elektrónov a dier, nachádzajúcich sa v materiáli článku, pri dopade slnečného žiarenia na jeho povrch. Tieto nosiče náboja sú potom zbierané na vrchnej a spodnej elektróde článku. Ak vznikne dostatočné množstvo náboja, tak medzi elektródami vzniká napätie. Celou podstatou solárneho článku je teda ako efektívne generovať nosiče náboja a následne ich efektívne zbierať pred tým, než dôjde k ich rekombinácii. Preto sú v týchto článkoch veľmi dôležité elektródy, ktoré musia byť

priesvitné, musia mať dobrú elektrickú vodivosť a v tomto prípade musia byť aj ohybné. Všetky tieto vlastnosti spĺňa práve grafén [1].

Prvý grafénový solárny článok bol vyrobený v roku 2008 a používal grafén vyrobený metódou lepiacej pásky. Následne bol článok skonštruovaný aj z grafénu vyrobeného chemickou syntézou, no neboli dosiahnuté priaznivé výsledky, pretože takto vyrobený grafén má vysoký plošný odpor a tým pádom je nevhodný ako elektróda. Najlepšie výsledky, z hľadiska výroby a efektivity, dosahujú články, kde je grafén vyrobený metódou CVD. No oproti klasickým, komerčne dostupným rigidným článkom, kde sa efektivita článku pohybuje okolo 25%., majú grafénové solárne články podstatne nižšiu efektivitu [1].

Na výrobu flexibilných solárnych článkov sa v dnešnej dobe využíva ITO. Tento materiál je, podobne ako grafén, vodivý a priesvitný, ale je krehký a nie je veľmi flexibilný. Pri ohybe článku teda môže dôjsť k jeho zničeniu. Ďalšou nevýhodou ITO , ako už bolo spomenuté, je vzácnosť india. Tým pádom je aj cena takého článku dosť vysoká. Preto sa ako vhodná náhrada za ITO ukazuje práve grafén [6].

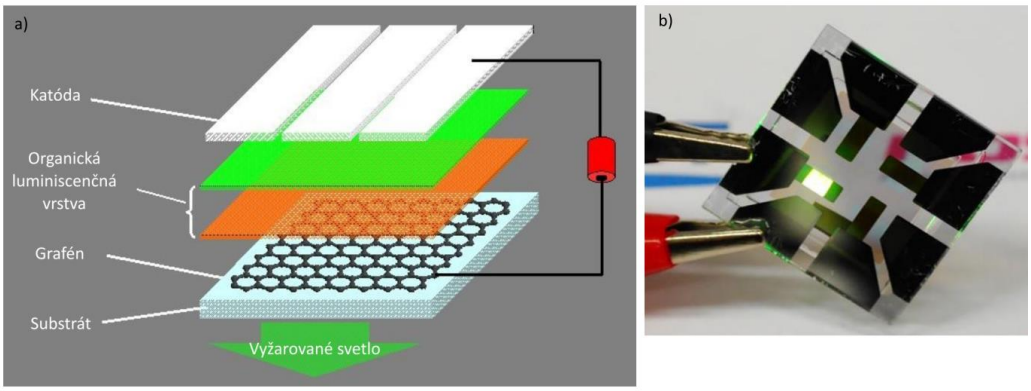


Obr. 8 Porovnanie efektivity (PCE) solárnych článkov s kombinovanými elektródami

Z Obr.8 je možné vidieť, že kombináciou rôznych typov materiálov elektródy je možné dosiahnuť rôzny výsledky efektivity. Najlepšie výsledky dosahuje hliníková elektróda v kombinácii s ITO a grafénom. Použitím grafénu na výrobu oboch elektród boli dosiahnuté najhoršie výsledky. Na zvýšenie efektivity grafénových solárnych článkov sa využíva dopovanie grafénu napríklad chloridom zlatitým ($AuCl_3$). Efektivitu je možné zvýšiť aj použitím viacerých vrstiev grafénu. Dopovaním a vrstvením grafénu je možné dosiahnuť efektivitu až 7% [1],[6].

VI. FLEXIBILNÉ OLED DISPLEJE

Ako už napovedá samotný názov technológie, displeje využívajú na vytváranie obrazu organické elektroluminiscenčné diódy. Tie oproti tekutým kryštálom v LCD nepotrebujú podsvietenie, pretože svetlo sami generujú . V porovnaní s LCD tak majú OLED displeje výrazne nižšiu spotrebu pri jasnejšom obraze. Na druhú stranu sa elektroluminiscenčné diódy pomerne rýchlo opotrebujú. Životnosť týchto displejov je preto až 2-krát kratšia ako v prípade LCD. Ich výhodou je však to, že sú ľahké, flexibilné a ich hrúbka je menšia ako u ostatných typov displejov [1], [7].



Obr. 9 a) Jednotlivé vrstvy OLED displeja s grafénovou elektródou, b) grafénová LED

OLED displej sa skladá z organickej polymérnej luminiscenčnej vrstvy, ktorá je samotným zdrojom svetla. Táto vrstva sa nazýva tiež aktívnou vrstvou a nachádza sa medzi dvoma

elektródami, z ktorých jedna musí byť priesvitná, aby sa vzniknuté svetlo dostalo vonku. V súčasnosti sa ako priesvitná elektróda používa ITO. No podobne ako u solárnych článkoch vzniká problém s vysokou cenou india. ITO taktiež nie je vhodný materiálom na flexibilné OLED, kvôli svojim vlastnostiam. ITO je taktiež príčinou kratšej životnosti, pretože dochádza k difúzii medzi luminiscenčnou a ITO vrstvou, čím sa znižuje kvalita obrazu [1], [8].

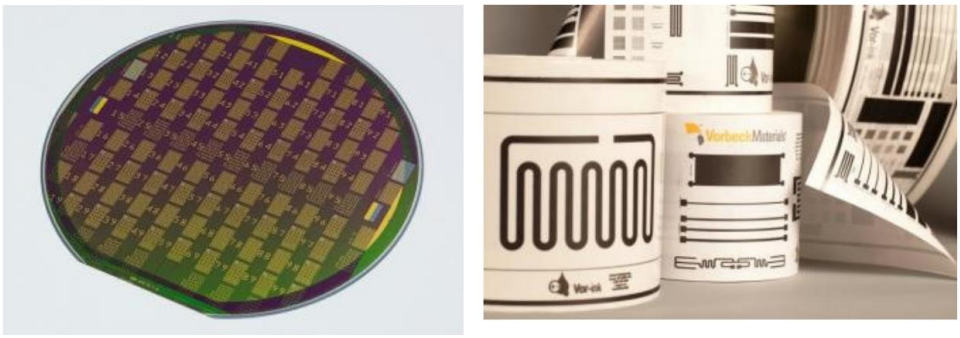
ITO elektródu je teda možné nahradiť grafénovou elektródou. Vrstva grafénu sa vytvára na priehľadnom substráte pomocou MOCVD (Metaloraganic CVD). MOCVD je výrobný proces podobný CVD, bez použitia katalytických kovov. Pomocou fotolitografie je potom možné upraviť tvar grafénovej vrstvy. Vodivé vlastnosti grafénovej elektródy je možné zlepšiť dopovaním kyselinou dusičnou (HNO_3) [7].

VII. Záver

Grafén poskytuje mechanickú a chemickú stabilitu, ktoré sú dôležitými faktormi pre aplikácie pre flexibilnú elektroniku. Okrem vybraných aplikácií sa grafén vo flexibilnej elektronike využíva v širokom rozsahu. Rozvoj flexibilnej elektroniky neustále zvyšuje požiadavky na používané materiály a trendom sú v dnešnej dobe práve 2-rozmerné štruktúry ako je napríklad grafén. Zaujímavou aplikáciou sú tzv. grafénové atramenty, ktoré umožňujú vytváranie tlačiteľnej elektroniky. V tejto oblasti sa tieto atramenty využívajú na vytváranie flexibilných antén pre smartfóny alebo na vytváranie tenko-vrstvových tranzistorov (TFT).

Grafén taktiež spĺňa požiadavky na využitie v biomedicíne, ktorými sú napríklad malá veľkosť, flexibilita, chemická stabilita vo fyziologických podmienkach alebo schopnosť fungovať v nepriaznivých podmienkach po dlhý čas. Grafén je preto možné využiť na vytváranie zariadení, ktoré budú monitorovať životné funkcie alebo priamo na liečbu rôznych ochorení.

Ďalšou aplikáciou grafénu je výroba integrovaných obvodov. Výskum sa stále častejšie zaoberá touto myšlienkou a existuje predpoklad, že grafén nahradí v tejto oblasti rozsiahlo využívaný kremík V posledných rokoch investujú veľké spoločnosti práve do vývoja grafénových zariadení, no za komerčnými zariadeniami je potrebné ešte pár rokov počkať.



Obr. 10 Grafénový čip (vľavo), grafénové antény pre mobilné aplikácie (vpravo)

Materiály na výrobu elektród konvenčných batérií môžu byť obohatené grafénom. Takéto batérie sú potom ľahké a odolné. Potenciálne sa môže aj skrátiť čas nabíjania a predĺžiť životnosť batérie. Výhodou je aj flexibilita grafénu. Je predpoklad, že sa v blízkej budúcnosti upustí vývoj od klasických elektrochemických článkov a prejde sa na tzv. superkondenzátory. Keďže grafén je veľmi tenký, tak je možné vytvoriť superkondenzátor s vysokou kapacitou, s oveľa menšími rozmermi.

Napriek vysokému potenciálu grafénu vo flexibilnej elektronike, je nutné zlepšiť proces výroby grafénu. Problémom je vyrobenie veľkého množstva grafénu, ktorý by mal rovnaké vlastnosti a stabilitu po celej ploche. V tejto dobe je vývoj výrobných procesov grafénu ešte stále len v počiatočných štádiách.

Poďakovanie

Tento článok bol podporený projektmi KEGA no. 011TUKE-4/2023, APVV 22-0289 a projektmi VEGA no. 1/0049/24 a no. 1/0278/24, ITMS kód 313011T557 („Podpora výskumno-vývojového potenciálu v oblasti dopravných prostriedkov“).

Použitá literatúra

[1] Das, Tanmoy & Sharma, Bhupendra & Katiyar, Ajit & Ahn, Jong-Hyun. (2018). Graphene-based flexible and wearable electronics. Journal of Semiconductors. 39. 011007. 10.1088/1674-4926/39/1/011007.

[2] PPC group. Grafen – co to je a k čemu se používá? [Online] https://www.products.pcc.eu/cs/blog/grafen-co-to-je-a-k-cemu-se-pouziva/

[3] Tae-Hee Han, Hobeom Kim, Sung-Joo Kwon, Tae-Woo Lee,Graphene-based flexible electronic devices, 2017, Pages 1-43,ISSN 0927-796X, https://doi.org/10.1016/j.mser.2017.05.001.

[4] Díez-Pascual, A.M.; Rahdar, A. Graphene-Based Polymer Composites for Flexible Electronic Applications. Micromachines 2022, 13, 1123. https://doi.org/10.3390/ mi13071123
[5] Tae-Hee Han, Hobeom Kim, Sung-Joo Kwon, Tae-Woo Lee, Graphene-based flexible electronic devices, Materials Science and Engineering: 2017, Pages 1-43, ISSN 0927-796X, https://doi.org/10.1016/j.mser.2017.05.001.
[6] Weng, Z., Dixon, S. C., Lee, L. Y., Humphreys, C. J., Guiney, I., Fenwick, O., Gillin, W. P., Wafer-Scale Graphene Anodes Replace Indium Tin Oxide in Organic Light-Emitting Diodes. Adv. Optical Mater. 2022, 10, 2101675. https://doi.org/10.1002/adom.202101675
[7] Roni Peleg, Graphene could replace ITO in OLED displays, 2022 [Online]. https://www.graphene-info.com/graphene-could-replace-ito-oled-displays
[8] Nancy W. Stauffer, Transparent, flexible solar cells, 2017, [Online]. https://news.mit.edu/2017/mit-researchers-develop-graphene-based-transparent-flexible-solar-cells-0728
[9] Graphene -info, Graphene for the mobile industry,[Online]. https://www.graphene-info.com/graphene-mobile-industry

Vplyv opakovaného reflow procesu na zmáčavosť povrchov pred osadením DPS

[1]*Alena PIETRIKOVÁ*

[1] Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovensko

[1]alena.pietrikova@tuke.sk

***Abstrakt* — Tento článok analyzuje odolnosť novo-vyvinutých povrchových úprav dosiek plošných spojov (DPS) voči opakovanému reflow procesu z pohľadu ich oxidácie a zmáčavosti ešte pred osadením súčiastkami. Tento v priemysle často sa vyskytujúci prípad nastáva u obojstranných DPS, kedy je v prvom kroku DPS osadená súčiastkami len z jednej strany a druhá neosadená strana DPS je zaťažená reflow procesom ešte pred následným nanesením spájkovacej pasty a osadením súčiastkami. Každá povrchová úprava odoláva tomuto vplyvu teploty počas reflow procesu inak. Odolnosť povrchových úprav voči opakovanému reflow procesu bola analyzovaná u štandardných povrchových úprav DPS ako je ENIG, Cu, ImSn a tiež u novo-vyvinutých povrchových úprav na báze SnAg (SnAg7, SnAg7DOC, SnAg5, SnAg5ENi). Bol analyzovaný vplyv opakovaného reflow procesu na zmáčavosť povrchov. Výsledky meraní ukazujú, že povrchové úpravy na báze nových povrchových úprav SnAg5 and SnAg7 majú vynikajúcu spájkovateľnosť a tiež vhodnosť pre použitie tam, kde sa vyžaduje opakovaný reflow ešte pred osadením DPS súčiastkami.**

***Kľúčové slová* — opakovaný reflow, odolnosť povrchu voči oxidácii, zmáčavosť povrchu, SnAg, povrchová úprava**

I. ÚVOD

Povrchové úpravy dosiek plošných spojov (DPS) poskytujú ochranu voči oxidácii a zaisťujú tiež dlhodobú životnosť DPS pri ich skladovaní. Povrchové úpravy DPS tiež zabezpečujú ich spájkovateľnosť a majú tak principiálny vplyv vytvorenie kvalitného spoja alebo potlačenie tvorby porúch v závislosti od aplikovanej technológie. Zmáčanie povrchov DPS, ktoré je základným predpokladom pre vytvorenie kvalitného spoja, závisí do značnej miery od typu povrchovej úpravy. Typ povrchovej úpravy DPS má vplyv na ich odolnosť voči opakovaným reflow procesom, tiež na kvalitu a dlhodobú životnosť spojov z pohľadu ich odolnosti voči opakovanému ohrevu a ochladzovaniu [1], [2].

Výber a aplikácia vhodnej povrchovej úpravy sú dôležité na zaistenie optimálneho výkonu dosky či už pre prípad aplikácie vysokých frekvencií alebo ich činnosti v drsných prevádzkových podmienkach. Výber povrchovej úpravy závisí na špecifických požiadavkách každej aplikácie. Správna voľba povrchovej úpravy určuje, či bude DPS fungovať optimálne počas celej plánovanej životnosti. Vzhľadom na množstvo aplikácií a prostredie, kde sa DPS nachádzajú, nie je jedna univerzálna voľba povrchovej úpravy dostačujúca.

Existuje mnoho možností povrchových úprav dostupných pre DPS. Každý typ povrchovej úpravy prináša svoj jedinečný súbor vlastností, výhod a potenciálnych výziev. Štandardné kritériá zohľadňujú ochranu povrchu DPS počas skladovania, plochosť (vhodnosť pre súčiastky s jemným rozstupom), výrobné náklady, zložitosť výrobného procesu alebo kompatibilitu s montážnymi technikami. Vzhľadom na množstvo aplikácií a prostredie, kde sa dosky plošných spojov nachádzajú, nie je jedna univerzálna voľba povrchovej úpravy dostačujúca. Voľba povrchovej úpravy závisí na špecifických požiadavkách každej aplikácie. Táto voľba často určuje, či bude DPS fungovať optimálne počas celej plánovanej životnosti. Správny výber povrchovej úpravy môže výrazne ovplyvniť optimálny výkon dosky, jej dlhodobú životnosť a spoľahlivosť.

V našom článku porovnávame štandardné povrchové úpravy ako sú ENIG, SnHASL, Cu, ImSn (imerzný cín) s novo vyvinutými povrchovými úpravami na báze SnAg (SnAg7, SnAg7DOC, SnAg5, SnAg5ENi) z hľadiska odolnosti povrchov voči opakovaným reflow procesom, pričom

sa zameriavame najmä na zmenu zmáčavosti ako kľúčového ukazovateľa zmien v povrchovej štruktúre. Opierame sa pri tom o výsledky odolnosti povrchov v základnom stave, ktoré sme analyzovali v literatúre [3], [4].

II. MATERIÁL A METÓDA EXPERIMENTOV

Pre experimenty sme použili 7 typov povrchových úprav: 3 štandardné povrchové úpravy: čistá meď, ENIG a ImSn a 4 typy novo-vyvinutých povrchových úprav DPS na báze SnAg, ktoré boli vyrobené elektrolytickým vylúčením z vodného alebo nevodného roztoku (Tab. 1). Zloženie zliatiny na povrchu DPS, vytvorené vo forme povrchovej úpravy, je takmer eutektické. To znamená, že zliatina SnAg7 obsahuje 7 % striebra, zliatina SnAg5 obsahuje 5 % striebra. Tieto zliatiny sa významne líšia bodom tavenia aj od eutektickej zliatiny (obsahuje 3,5 %Ag), čo má zásadný vplyv na tvorbu adhézneho spoja na medzifázovom rozhraní.

Tabuľka 1
Neštandardné povrchové úpravy DPS na báze SnAg

Typ povrchu	Difúzna bariéra	Metóda vylúčenia 1,5 µm vrstvy SnAg	T tavenia / T tuhnutia (°C)
SnAg7	chemická Ni (1,5 µm)	Z nevodného roztoku	270 / 221
SnAg7DOC	bez difúznej bariéry	Z nevodného roztoku	270 / 221
SnAg5	chemická Ni (1,5 µm)	Z vodného roztoku	245 / 221
SnAg5ENi	elektrolytická Ni (1,5 µm)	Z vodného roztoku	245 / 221

Pre stanovenie odolnosti povrchov DPS voči opakovaným reflow procesom bola aplikovaná metóda testov zmáčania povrchov roztečenou spájkou, ktorú sme realizovali prostredníctvom svojpomocne vyvinutého laboratórneho zariadenia. Zariadenie umožňuje in –situ kamerové snímanie zmeny tvaru a roztečenia SAC guľôčky spájky (priemer 760 µm) počas jej tavenia a to pri aplikácii pretavovacieho profilu RSS (Ramp Soak Spike) s dĺžkou trvania cca 300 sekúnd a max. teplotou 250 °C. Zotrvanie na najvyššej teplote v rozsahu nad teplotou tavenia (oblasť TAL), teda interval, kedy bola guľôčka spájky v roztavenom stave, bolo približne 50 sekúnd. V tejto oblasti bol vykonaný kamerový záznam z bočného pohľadu. Výhodou merania kontaktného uhla v reálnom čase realizovaného malým laboratórnym zariadením je schopnosť merať rýchlosť roztečenia optickým skúmaním priemeru a výšky roztavenej guľôčky spájky a možnosť merania kontaktného uhla po ukončení reflow procesu ale aj počas tohto procesu.

Testy zmáčania povrchov umožňovali tak stanoviť okrem kontaktného uhla po ukončení procesu reflow tiež koeficient zmáčania po ukončení procesu reflow a čas pre aktívne zmáčanie povrchu.

Pre meranie a vyhodnotenie kontaktného uhla a geometrických parametrov potrebných pre výpočet koeficientu zmáčania po roztečení spájky bol použitý software ImageJ. Tento softvér umožňuje metódou sférickej aproximácie vyhodnocovať každý záber (mikroskopické obrázky z bočného pohľadu) každej vzorky samostatne bez automatizácie v samostatných časových krokoch. Touto metódou je možné spoľahlivo a presne v reálnom čase zmerať kontaktný uhol [5], [6]. Obrázok 1 zobrazuje efekt roztečenia guľôčky spájky po reflow procese na Cu a SnAg5 povrchu DPS.



Cu SnAg5

Obr. 1 Efekt roztečenia guľôčky spájky po reflow na Cu a SnAg5 povrchu DPS.

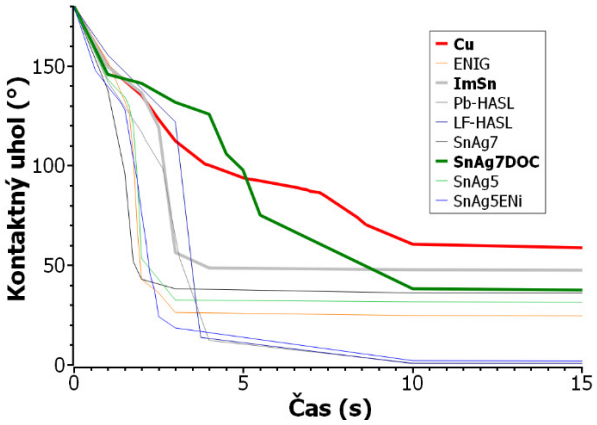
Pomer geometrických parametrov (priemer "d" a výška "h" guľôčky) po roztečení spájky na povrchu a kontaktný uhol sú veľmi dôležité parametre popisujúce zmáčavosť substrátu roztavenou spájkou.

Koeficient zmáčania S_R môže byť vypočítaný ako funkcia výšky a priemeru guľôčky spájky *d* [2]:

$$S_R = \frac{d-h}{d} 100\ \%. \qquad (1)$$

Vzorky s označením 0 boli vystavené prvému, základnému procesu reflow. Vzorky pre stanovenie odolnosti povrchovej úpravy DPS po prvom opakovanom reflow procese boli označené číslom 1, po druhom opakovanom reflow procese číslom 2. Pri každom type testu bolo použitých minimálne 5 vzoriek.

Merania kontaktného uhla v reálnom čase poukázali na výrazné rozdiely v rýchlosti roztečenia guľôčky spájky v oblasti TAL a to v závislosti od typu povrchovej úpravy DPS (Obr. 2.). Zmáčanie všetkých typov povrchov pri tavení guľôčky spájky v oblasti TAL prebieha na počiatku veľmi rýchlo, až kým jej roztečenie nedosiahne 80 % z maximálnej hodnoty.

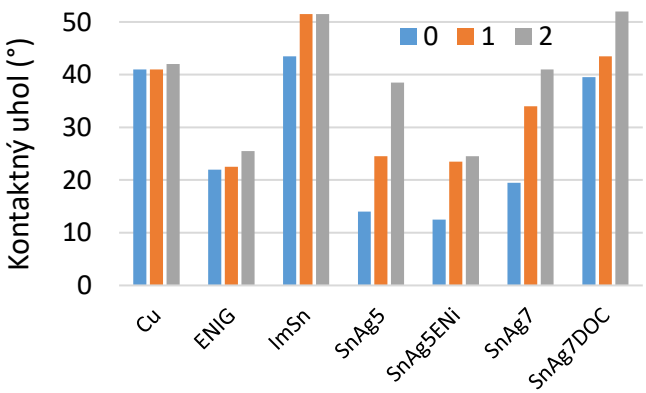


Obr. 2 Závislosť kontaktného uhla od času a typu povrchovej úpravy po prvom procese reflow [3], [4].

Nad touto hodnotou už nedochádza k významným zmenám zdanlivého kontaktného uhla. Čas, kedy roztečenie spájky dosiahne 80 %, sa však u rôznych typov povrchových úprav výrazne líši. To nás priviedlo k zadefinovaniu nového parametra pre analýzu zmáčania povrchov. Zadefinovali sme čas pre aktívne zmáčanie, teda časový interval, počas ktorého sa guľôčka spájky začne taviť a okamžite dochádza v rapídnym zmenám kontaktného uhla, až kým jej roztečenie nedosiahne 80 %.

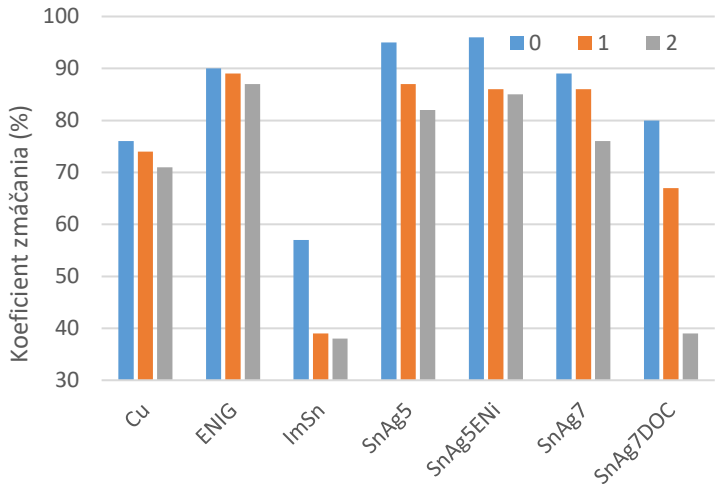
III. VÝSLEDKY A DISKUSIA

Cieľom tohto príspevku je zvýrazniť praktický problém so zmáčaním povrchu pri pretavovaní druhej strany obojstrannej DPS, ktorý je často analyzovaný v priemysle montážnych technológií. Obrázok 3 zobrazuje hodnotenie zmeny kontaktného uhla roztavených guľôčok spájky na rôznych povrchoch DPS po 1, 2 a 3 cykloch reflow. Po prvom procese reflow sa zdá, že nová povrchová úprava SnAg5 je najvýhodnejšia, pretože vykazuje dokonca lepšie zmáčanie ako ENIG. Naopak, najhoršie výsledky boli dosiahnuté pri povrchovej úprave ImSn. Celkovo možno skonštatovať, že povrchové úpravy obsahujúce Ni difúznu bariéru (ENIG, SnAg5, SnAg7, SnAg7ENi) majú najmenší kontaktný uhol. Po 2. a 3. reflow procese sa okrem medi pozoroval nárast kontaktného uhla u všetkých vzoriek, pričom najvýraznejší nárast bol zaznamenaný u povrchových úprav na báze SnAg. Napriek tomuto nárastu kontaktného uhla u vzoriek na báze SnAg bol výsledný kontaktný uhol po 3. pretavení nižší ako u Cu po prvom pretavení. Z hľadiska vyhodnocovania kontaktného uhla je najstabilnejšia a pre opakovaný reflow proces najviac vhodná úprava ENIG. Vrstva SnAg5ENi vykazuje porovnateľné vlastnosti.



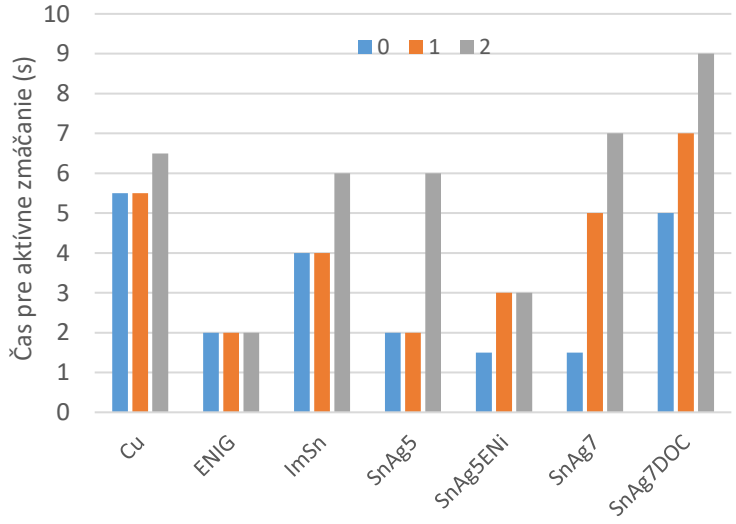
Obr. 3 Kontaktný uhol roztavených guľôčok spájky na povrchoch DPS po 1, 2 a 3 cykloch reflow

Pôvodne sa mohlo zdať, že vyhodnocovanie rýchlosti zmáčania je podobné ako hodnotenie pomocou kontaktného uhla. Avšak dôsledná analýza odhalila (Obr.4), že táto metóda je v skutočnosti citlivejšia. Najvýraznejší pokles zmáčavosti bol pozorovaný pri povrchu ImSn, hoci tento povrch mal už pôvodne najnižšie hodnoty rýchlosti zmáčania. Podobne, povrch SnAg7DOC (ide o povrchovú úpravu v zložení SnAg7 nanesenú priamo na povrch Cu na DPS bez aplikácie difúznej vrstvy Ni) mal malú odolnosť, čo ho robí nevhodným pre opakované reflow procesy. Všetky vzorky vykazovali pokles hodnôt rýchlosti zmáčania, avšak u povrchov ENIG, SnAg5 a SnAg5ENi bol tento pokles minimálny. Tieto povrchy sa tak javia ako najodolnejšie po opakovaných reflow procesoch.



Obr. 4 Koeficient zmáčania spájky na rôznych povrchoch DPS po 1, 2 a 3 cykloch reflow.

Vzorky s niklovou bariérou (ENIG, SnAg7, SnAg5 a SnAg5ENi) dosiahli najkratší čas pre aktívne zmáčanie, pričom ENIG, SnAg5ENi a SnAg5 po 2. zaťažení dosiahli najnižšie hodnoty (Obr. 5). Naopak, vzorky, kde sa tvoril spoj priamo s meďou, teda bez prítomnosti difúznej bariéry (Cu, ImSn, SnAg7DOC), mali dlhší čas pre aktívne zmáčanie. Okrem vzoriek ENIG došlo k nárastu tohto času, najmä po 3 reflow procesoch. Dlhší čas pre aktívne zmáčanie zvyčajne sprevádza znížená odolnosť voči opakovanému teplotnému zaťaženiu. Vzorky s najdlhším časom pre aktívne zmáčanie po viacnásobnom teplotnom zaťažení nie sú vhodné pre použitie v technológii, kde sa pred osadením elektronickými komponentmi očakáva opakovaný reflow.



Obr. 5 Čas pre aktívne zmáčanie po viacnásobnom teplotnom zaťažení (po 1, 2 a 3 cykloch reflow)

IV. ZÁVER

V rámci nášho výskumu sme porovnávali štandardné povrchové úpravy DPS (ENIG, Cu, ImSn) s novo vyvinutými na báze SnAg: SnAg7, SnAg7DOC, SnAg5, SnAg5ENi z hľadiska odolnosti povrchov voči opakovaným reflow procesom. Zistili sme, že povrchové úpravy obsahujúce Ni difúznu bariéru (ENIG, SnAg5, SnAg7, SnAg7ENi) majú najmenší kontaktný uhol. Najstabilnejšia a najodolnejšia povrchová úprava pre opakované reflow procesy je ENIG, porovnateľné vlastnosti má aj vrstva SnAg5ENi.

Celkovo možno konštatovať, že nová povrchová úprava SnAg5 vykazuje najlepšie vlastnosti v porovnaní s ostatnými povrchmi, dokonca vykazuje lepšie zmáčanie ako ENIG. Najhoršie vlastnosti sme pozorovali pri povrchovej úprave ImSn.

Opakovaný reflow čistých DPS s povrchovými úpravami spôsobil zhoršenie zmáčavosti u každej povrchovej úpravy v rôznej miere. Najmenšie zmeny boli zaznamenané u povrchovej úpravy ENIG. Podobne, takmer žiadna zmena nebola pozorovaná v prípade dostatočnej odolnosti povrchu SnAg5 a SnAg5ENi po prvom opakovanom cykle. Dokonca aj po druhom opakovanom cykle reflow mal povrch SnAg5ENi lepšiu zmáčavosť ako povrch ImSn. Povrchové úpravy s niklovou difúznou bariérou dosiahli výrazne kratší čas pre aktívne zmáčanie než povrchové úpravy bez takéhoto bariéry. Predpokladá sa, že skrátenie času pre aktívne zmáčanie bude mať za následok menšie hrúbky IMC, čo bolo potvrdené aj u vzoriek, ktoré neboli podrobené procesu reflow [8]. V konečnom dôsledku zhoršenie odolnosti povrchu po opakovanom reflow cykle negatívne ovplyvní kvalitu spájkovaného spoja.

Nové povrchové úpravy na báze SnAg s takmer eutektickým zložením a s Ni bariérovou vrstvou začínajú byť tiež populárne a tvoria konkurenciu k povrchu typu ENIG. Vyznačujú sa vynikajúcou spájkovateľnosťou a dokonale rovným povrchom, čo ich robí vhodnými pre celý rad aplikácií, hlavne pri pokračujúcej miniaturizácii v elektronike. Pri výbere typu povrchovej úpravy DPS je nevyhnutné zvážiť náklady na materiál, zložitosť procesu a kompatibilitu s montážnymi technikami. Je potrebné zvážiť aj faktory ako je prevádzkové prostredie, teplotný rozsah, elektrická vodivosť a integrita signálu. Niektoré úpravy môžu byť vhodné pre drsné prevádzkové podmienky, zatiaľ čo iné môžu byť vhodnejšie pre lacnú spotrebnú elektroniku.

Vzhľadom na široké spektrum aplikácií, v ktorých sú dosky plošných spojov používané, je otázkou, či stačí jedna typická povrchová úprava. Realita ukazuje, že žiadna univerzálna možnosť neexistuje. Voľba najvhodnejšej povrchovej úpravy je zásadná v závislosti na konkrétnych požiadavkách a výzvach každej aplikácie. Táto voľba často určuje, či bude doska plošných spojov fungovať optimálne počas celej svojej plánovanej životnosti. ENIG sa v posledných rokoch presadil, najmä pre dosky s vysokou hustotou. Zlato chráni nikel počas skladovania, zatiaľ čo nikel poskytuje robustný a dobre spájkovateľný povrch. Planarita povrchovej úpravy ENIG ho robí vhodným pre komponenty s jemným rozostupom, avšak jeho nevýhodou je vyššia cena.

Poďakovanie

Tento článok bol napísaný s podporou projektov KEGA 011TUKE-4/2023, VEGA 1/0049/24, VEGA 1/0278/24, APVV 22-0289 a ITMS s kódom 313011T557.

References

[1] V. Wirth, K. Rendl, and F. Steiner, “Effect of multiple reflow cycles on intermetallic compound creation,” in *2015 38th International Spring Seminar on Electronics Technology (ISSE)*, Eger, Hungary: IEEE, May 2015, pp. 226–230. doi: 10.1109/ISSE.2015.7247995.

[2] S. Tahat, A. Alakayleh, M. E. A. Belhadi, A. Alahmar, A. Yaeger, and S. Hamasha, “Effects of Multiple Reflows on IMC and Shear Strength of Individual Solder Joints,” in *2023 22nd IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, Orlando, FL, USA: IEEE, May 2023, pp. 1–9. doi: 10.1109/ITherm55368.2023.10177561.

[3] A. Pietrikova, D. Dzivy, O. Vrublenskaya, and M. A. Shikun, “Wetting of New SnAg Surface Finish,” in *2022 45th International Spring Seminar on Electronics Technology (ISSE)*, Vienna, Austria: IEEE, May 2022, pp. 1–6. doi: 10.1109/ISSE54558.2022.9812833.

[4] D. Dzivy and A. Pietrikova, “Real-time contact angle’s measurement of molten solder balls in laboratory conditions,” *Microelectron. Int.*, vol. 39, no. 3, pp. 132–138, Jun. 2022, doi: 10.1108/MI-12-2021-0122.

[5] E. Bormashenko, “Apparent contact angles for reactive wetting of smooth, rough, and heterogeneous surfaces calculated from the variational principles,” *J. Colloid Interface Sci.*, vol. 537, pp. 597–603, Mar. 2019, doi: 10.1016/j.jcis.2018.11.068.

[6] M.-W. Yang and S.-Y. Lin, “A method for correcting the contact angle from the θ/2 method,” *Colloids Surf. Physicochem. Eng. Asp.*, vol. 220, no. 1–3, pp. 199–210, Jun. 2003, doi: 10.1016/S0927-7757(03)00064-5.

[7] C. F. Coombs and H. T. Holden, Eds., *Printed circuits handbook*, Seventh edition. New York: McGraw-Hill Education, 2016.

[8] A. Pietrikova *et al.*, “Effect of non-standard SnAg surface finishes on properties of solder joints,” *Appl. Surf. Sci. Adv.*, vol. 18, p. 100483, Dec. 2023, doi: 10.1016/j.apsadv.2023.100483.

[9] D. Dzivy, A. Pietrikova, O. Vrublevskaya, and M. A. Shikun, “Influence of Multiple Reflows and Surface Finishes on Solder Joint Resistivity,” *Acta Electrotech. Inform.*, vol. 21, no. 4, pp. 3–8, Dec. 2021, doi: 10.2478/aei-2021-0001.

Nenávistné prejavy na sociálnych sieťach a ich detekcia pomocou vysvetliteľných modelov

[1]*Miroslava MATEJOVÁ,* [2]*Ján PARALIČ,* [3]*Adam KOVÁČ*

[1,2,3] Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]miroslava.matejova@tuke.sk, [2]jan.paralic@tuke.sk, [3]adam.kovac.2@student.tuke.sk

***Abstrakt* — Cieľom tejto štúdie je analyzovať vysvetliteľné modely umelej inteligencie pre detekciu nenávistných prejavov na sociálnych sieťach. Je zameraná na vytvorenie modelov, ktoré nielen presne klasifikujú texty, ale poskytujú aj zrozumiteľné vysvetlenia svojich rozhodnutí, čím prispievajú k transparentnosti a dôveryhodnosti modelov. V praktickej časti je dôraz kladený na implementáciu a testovanie rôznych modelov ako BERT, CNN a XGBoost, pričom sa využívajú vysvetliteľné metódy LIME a SHAP. Na hodnotenie vysvetliteľnosti jednotlivých riešení sa používa nástroj DoX. Modely boli trénované na datasete HateXplain, ktorý poskytuje relevantné dáta pre analýzu nenávistných prejavov. Výsledky práce ukazujú, že kombinácie pokročilých modelov a metód vysvetliteľnej umelej inteligencie poskytujú sľubné riešenie pre presnejšie a viac transparentné výsledky.**

***Kľúčové slová* — nenávistné prejavy, sociálne siete, vysvetliteľná umelá inteligencia, vysvetliteľnosť**

I. Úvod

V súčasnej dobe je analýza textových dát neoddeliteľnou súčasťou mnohých oblastí výskumu a praxe, pričom jednou z nich je aj detekcia nenávistných prejavov na sociálnych sieťach. Tieto prejavy môžu mať závažné dôsledky pre jednotlivcov aj komunity, čo vyvoláva potrebu efektívnych nástrojov na ich rozpoznávanie a analýzu. Vysvetliteľná umelá inteligencia (XAI) zohráva kľúčovú úlohu v tomto procese, pretože umožňuje nielen identifikovať problematický obsah, ale aj pochopiť dôvody, prečo bol určitý text klasifikovaný ako nenávistný alebo ofenzívny. Použitie vysvetliteľných modelov je zásadné pre zvýšenie dôveryhodnosti a transparentnosti výsledkov, čo je nevyhnutné v aplikáciách s významnými sociálnymi a etickými dopadmi.

Problémom je zložitosť modelov umelej inteligencie, ktoré často pôsobia pre používateľa veľmi nezrozumiteľne, čo môže markantne ovplyvniť ich dôveryhodnosť. Takéto modely nazývame black box (čierna skrinka). V kontexte umelej inteligencie (AI) sa termínom čierna skrinka označujú modely, ktorých vnútorné mechanizmy a rozhodovacie procesy sú pre ľudí nezrozumiteľné alebo neprístupné. V týchto modeloch nie je možné jednoducho získať alebo analyzovať informácie o tom, ako sú výstupy generované na základe vstupných údajov. Čierne skrinky sú často vysoko komplexné ako napr. hlboké neurónové siete, ktoré môžu mať milióny parametrov, ktoré bežný človek nemá šancu pochopiť. Napriek tomu, často dosahujú vysokú úroveň presnosti a preto sú využívané v mnohých doménach.

Dôležitosť XAI je zakotvená v dôveryhodnosti, zodpovednosti, regulačných požiadavkách, etických otázkach a v možnosti ladenia a zlepšovania modelu. V akomkoľvek technologickom systéme je dôvera základným pilierom. Ak používatelia rozumejú, ako systém funguje, s väčšou pravdepodobnosťou mu budú dôverovať. V tomto prípade nejde len o technickú záležitosť, je to aj o tom ako ľudia vnímajú technológie. Ak je systém transparentný a je možné pochopiť jeho rozhodovanie, rýchlejšie sa adaptuje a používatelia budú viac naklonení k jeho začleňovaniu do procesov a praxe.

V oblastiach, kde rozhodnutia priamo ovplyvňujú ľudské životy, je morálna a etická zodpovednosť za tieto rozhodnutia kľúčová. Ak sa vyskytne chyba, transparentný a vysvetliteľný systém nám umožňuje rýchlo identifikovať príčinu a opraviť ju. V mnohých prípadoch to môže prispieť k ďalšiemu rozvoju systému a vyvarovaniu sa podobných chýb, ktoré by spôsobili väčšiu škodu ako osoh.

II. DETEKCIA TOXICKÝCH PREJAVOV

Toxické prejavy sú negatívne alebo škodlivé výroky, ktoré môžu byť urážlivé, hanlivé, rasistické alebo inak nevhodné. V digitálnej dobe, kde sa väčšia časť komunikácie často odohráva online, sa stáva detekcia a moderácia toxických komentárov dôležitým nástrojom na ochranu komunít pred škodlivým obsahom. Takýto obsah má mnohokrát ničivý dopad na jednotlivcov a často aj na komunity, na ktoré poukazuje. Tento problém je veľmi aktuálny a celkový tón diskusií alebo komentárov na internete je znehodnocovaný práve toxickým obsahom. Takéto komentáre môžu spôsobiť dopad na psychiku nielen mladých ľudí, ale všetkých vekových kategórií používajúcich internet. Mnohokrát odradzuje jednotlivcov zapojiť sa do diskusie alebo vyjadriť svoj názor.

Hlavným cieľom štúdie [1] bola interpretácia a vysvetlenie rozhodnutí urobených komplexnými modelmi umelej inteligencie s cieľom pochopiť ich rozhodovací proces pri detekcii nenávistných prejavov. Na tento účel sa na dataset Google Jigsaw použili už existujúce modely ako neurónová sieť, viacvrstvový perceptrón, rozhodovacie stromy, KNN, náhodný les, naivný Bayes, logistická regresia a LSTM (long short-term memory) a varianty modelu BERT (Bidirectional Encoder Representations from Transformers), ako napr. BERT + ANN (umelá neurónová sieť) a BERT + MLP (viacvrstvový perceptrón), aby sa dosiahla najlepšia presnosť. Ďalej bol použitý dataset HateXplain. Najlepšiu správnosť (accuracy) dosiahli autori pre dataset Google Jigsaw pre model LSTM 97.6 %. Pre dataset HateXplain to bol model BERT + MLP so správnosťou 93.67 %. Vysvetliteľnosť modelov hodnotili autori pomocou metódy LIME (Local interpretable model-agnostic explanations) a benchmarku ERASER. BERT + MLP bol modelom s najlepším výkonom z hľadiska hodnovernosti. Najlepšie výsledky s najvyššou komplexnosťou vykázal model BERT + ANN. Tieto poznatky sú zásadné, pretože naznačujú, že pokročilé modely založené na transformátoroch, ako je BERT, môžu poskytovať nielen vyššiu správnosť, ale tiež lepšie nástroje na interpretáciu a pochopenie výsledkov, čo môže byť zvlášť užitočné v oblastiach vyžadujúcich transparentnosť a vysvetliteľnosť.

Výskum robený tvorcami datasetu HateXplain [2] je tiež prínosom pre problematiku toxických komentárov. Použili modely CNN-GRU, BiRNN, BiRNN-Attention a BERT. Každý model má dve verzie, jednu, v ktorej sú modely trénované iba s použitím označenia triedy a druhú, kde sa modely trénujú pomocou označení triedy a pozornosti. V tomto výskume dosiahol najvyššiu správnosť model BERT-HateXplain (bola využitá pozornosť získaná zo základnej pravdy naučenej z datasetu HateXplain), s hodnotou 69.8 %.

Na základe analýzy súčasného stavu sme prišli k záveru, že problematika toxických komentárov sa dostáva do popredia globálneho záujmu. Uvedomujeme si závažnosť tohto problému a smerujeme k efektívnym metódam na minimalizáciu ich negatívneho dopadu na online komunity. Rozhodli sme sa pre dataset HateXplain v spojení s často používanými modelmi ako sú CNN, XGBoost a BERT. Inovatívnym je vyhodnotenie vysvetliteľnosti pomocou metódy DoX, ktorá podľa našich informácii ešte nebola implementovaná v súvislosti s vysvetľovaním detekcie toxického prejavu.

III. MODELY DETEKCIE

V tejto kapitole bližšie priblížime modely, ktoré boli využité pre detekciu nenávistných komentárov v rámci tohto výskumu.

A. CNN (Konvolučná neurónová sieť)

Konvolučná neurónová sieť je typ siete, ktorá svoje uplatnenie našla hlavne v oblasti počítačového videnia, ale úspešne sa využíva aj v spracovaní reči a prirodzeného jazyka. Konvolučné siete sú založené na konvolúciách, čo sú matematické operácie, ktoré umožňujú aplikovať filter na vstupné dáta. Systém je koncepčne podobný biologickému systému neurónov a synapsií mozgu, odkiaľ pochádza aj jeho názov. Medzi hlavné vrstvy konvolučných neurónových sietí patria plne-prepojené (fully-connected), konvolučné (convolutional) a spájacie (pooling) vrstvy.

B. XGBoost (eXtreme Gradient Boosting)

XGBoost je algoritmus strojového učenia, ktorý patrí do kategórie súborového učenia, konkrétne skupiny využívajúcej zvyšovanie gradientu. Využíva rozhodovacie stromy ako základný model a regularizačné techniky na zlepšenie zovšeobecňovania modelov. XGBoost, známy svojou výpočtovou efektívnosťou, sa široko používa na úlohy regresie aj klasifikácie. Algoritmus funguje tak, že do súboru postupne pridáva slabé modely, pričom každý nový model sa zameriava na opravu chýb, ktoré urobili predchádzajúce modely. Využíva techniku optimalizácie gradientu na minimalizáciu preddefinovanej stratovej funkcie počas trénovania.

XGBoost je široko používaný v rôznych doménach vďaka svojmu vysokému prediktívnemu výkonu a všestrannosti naprieč rôznymi súbormi údajov.

C. *BERT (Bidirectional encoder representations from transformers)*

Model BERT je jazykový model, ktorý bol predstavený v článku spoločnosti Google [3] a dosiahol špičkové výsledky v oblasti spracovania prirodzeného jazyka. Predstavuje prelom najmä vďaka svojej bidirekcionálnej štruktúre, ktorá rozumie kontextu daného slova tým, že ho vyhodnocuje v rámci obsahu celého textu. Tento prístup prináša výrazný posun od iných modelov, ktoré zvyčajne spracúvajú text lineárne. Základná znalostná vrstva BERTu je odvodená z prvotného tréningu, ktorý sa realizuje na obrovských textoch a poskytuje modelu rozsiahle všeobecné znalosti jazyka. Na prvotný tréning sa používajú napr. maskovaný jazykový model (MLM) alebo predpovedanie nasledujúcej vety (NSP). MLM je konfrontovaný s vetami, z ktorých boli určité slová nahradené maskovacím tokenom, čo núti model predpovedať pôvodné slovo na základe porozumenia okolitého kontextu, čím sa rozvíja hlbšie chápanie syntaktickej a sémantickej štruktúry jazyka. Pri NSP sa model snaží predpovedať, či je druhá veta prirodzeným pokračovaním prvej, čo je úloha, ktorá má široké uplatnenie, najmä v prípadoch, kedy systém musí rozpoznať logickú alebo tematickú spojitosť medzi vetami.

IV. METÓDY VYSVETLITEĽNOSTI A ICH VYHODNOTENIE

A. *Metódy vysvetliteľnosti*

1) *LIME (Local interpretable model-agnostic explanations)*

LIME [4] je metóda navrhnutá na poskytovanie lokálnych vysvetlení predikcií modelov strojového učenia. Hlavným cieľom je umožniť používateľom lepšie porozumieť, prečo model urobil určitú predikciu, a to bez ohľadu na to, či ide o text, obrázok alebo video. LIME funguje tak, že vytvára perturbácie vo vstupných dátach (malé zmeny v okolí analyzovanej inštancie) a sleduje, ako na tieto zmeny model reaguje. Na základe týchto rozrušených dátových bodov, ktorým je pridelená váha podľa ich blízkosti k pôvodnému príkladu, je tento model schopný sa učiť a aproximovať predikcie pôvodného modelu. Metóda je modelovo-agnostická, teda poskytuje vysvetlenia pre mnoho modelov, čím umožňuje širokú aplikáciu v praxi.

2) *SHAP (Shapley additive explanations)*

SHAP [5] je metóda, ktorá pozostáva hlavne z výpočtu tzv. Shapleyho hodnoty pre každý atribút v dátach. Shapleyho hodnota ukazuje, aký veľký vplyv má atribút na predpoveď, ktorú model poskytuje. Táto metóda vychádza z teórie kooperatívnych hier a jej cieľom je zmerať príspevok každého "hráča" (v tomto prípade atribútu) k celkovému výsledku. Tree SHAP [6] je špeciálna verzia metódy SHAP, ktorá je navrhnutá na interpretáciu rozhodovacích stromov a ich odvodených modelov. Pri použití Tree SHAP sa berie do úvahy špecifická štruktúra rozhodovacích stromov. Rozhodovacie stromy pracujú tak, že rozdeľujú dáta na základe určitých pravidiel. Tree SHAP efektívne analyzuje tieto rozdelenia a určuje, aký veľký príspevok má každá vlastnosť (alebo otázka) ku konečnej predpovedi modelu.

B. *Vyhodnotenie vysvetliteľnosti*

Rastúca popularita XAI vyvolala otázku, ako hodnotiť vysvetlenia modelov strojového učenia. Novým pohľadom na vysvetliteľnosť sa zaoberá objektívna metrika DoX [7]. Stratégia hodnotenia DoX je založená na špecifickom teoretickom modeli vysvetlenia známom ako Achinsteinova teória vysvetlení ako aj Carnapových kritériách adekvátnosti vysvetlenia. Táto teória zdôrazňuje, že pragmatická odpoveď na otázku v širokom, ale aplikovateľnom zmysle vedie k ilokučnému aktu (t.j. vysvetlenie). V dôsledku toho sú vysvetlenia odpoveďami na množstvo základných otázok (archetypov), pričom každá z nich poskytuje jedinečný pohľad na diskutované koncepty. DoX pre vypočítanie vysvetliteľnosti používa metriku priemerný DoX. Táto metrika kombinuje výpočty kumulatívnej pertinencie a ilokučného vysvetlenia. Výsledný priemerný DoX preto môže slúžiť ako metrika na posúdenie, či je vysvetliteľnosť systému vyššia, rovnaká alebo nižšia ako iná. Pre lepšie vysvetlenie použijeme príklad. Predstavte si, že máte vetu "Tento komentár je úplne nezmyselný a dezinformačný" a súbor relevantných aspektov {urážka, agresivita, diskriminácia, vulgarizmus, neúcta, toxicita}. Niektoré z týchto aspektov môžu byť považované za mierne negatívne alebo kritické, ale žiaden z nich explicitne neobsahuje prvky ako urážka, agresivita, diskriminácia, vulgarizmus, alebo neúcta. V dôsledku toho by priemerná kumulatívna pertinencia pre každý archetyp/otázku bola nízka alebo nulová, čo by viedlo k nízkemu alebo nulovému skóre DoX pre toxicitu. Na druhej strane, keby sme použili vetu "Tento komentár je hlúpy a autor je idiot" situácia by bola iná. Táto veta obsahuje prvky urážky a neúcty, ktoré sú relevantné pre súbor aspektov, ako napríklad urážka a neúcta. Tieto aspekty by mali vyššiu priemernú kumulatívnu pertinenciu. Hoci DoX nie je možné priamo použiť na model

čiernej skrinky, môže sa však použiť na výstup XAI algoritmov alebo akýchkoľvek iných vysvetliteľných informácií (napríklad dokumentácia, vedecké články, knihy), aby sa pochopilo, ako môže byť táto informácia použitá na vysvetlenie.

V. NÁVRH RIEŠENIA

Postup riešenia sleduje metodológiu CRISP-DM doplnenú o ďalšie fázy venované vysvetliteľnosti modelov a jej vyhodnoteniu.

A. Pochopenie problému

Výskum sa zameriava na vytvorenie modelov, ktoré by dokázali nie len presne klasifikovať texty, ale aj poskytnúť vysvetlenie svojich rozhodnutí, čím by prispeli k transparentnosti a pochopeniu umelej inteligencie v aplikáciách spracovania nenávistných prejavov. To umožní nielen lepšie pochopenie toho, ako modely pracujú, ale tiež poskytne základ pre ďalšie vylepšenia v oblasti automatického rozpoznávania.

B. Pochopenie dát

HateXplain [2], prvý referenčný dataset pre nenávistné prejavy, ktorý obsahuje anotácie na úrovni slov a fráz, zachytávajúce aj odôvodnenia pre dané označenia. Príspevky pochádzajú z Twitteru (9 055 príspevkov) a Gabu (11 093 príspevkov). Anotácia prebiehala pomocou pracovníkov z Amazon Mechanical Turk (MTurk). Títo anotátori klasifikovali každý príspevok ako nenávistný (hatespeech), urážlivý (offensive) alebo normálny a zároveň identifikovali cieľové komunity, na ktoré sa príspevky vzťahujú. Dataset zahŕňa desať komunít, s tým, že najčastejšie sú terčom Afričania, moslimovia a Židia. Pri urážlivých prejavoch sú hlavnými cieľmi ženy, Afričania a homosexuáli. Anotátori boli požiadaní o vyznačenie častí textu, ktoré odôvodňujú ich rozhodnutie o klasifikácii. Tieto dodatočné anotácie umožňujú hlbšie skúmať, ako sa prejavuje nenávistný alebo urážlivý prejav. Každý príspevok bol anotovaný tromi anotátormi a konečná klasifikácia bola určená väčšinovým hlasovaním.

C. Príprava dát

Dáta sme rozdelili na trénovaciu a testovaciu časť v pomere 80:20. Toto umožňuje modelu trénovať na veľkej časti datasetu a potom overiť jeho schopnosť na nevidených dátach.

Dataset neobsahuje chýbajúce hodnoty a dáta sú očistené od interpunkčných znamienok. Ďalším krokom je tokenizácia, kde sa rozdelí text na jednotlivé prvky (tokeny), ktoré môžu byť slová, čísla alebo symboly. Tento proces nám umožňuje rozložiť veľké množstvo textu na základné stavebné bloky, na ktoré sa môžeme zamerať pri ďalšej analýze. Text v našom datasete je už vopred rozdelený na tokeny. Po tokenizácii nasleduje proces vektorizácie. Tu prevádzame tokeny na numerickú reprezentáciu využívajúcu metódu TF-IDF, ktorú sme zvolili kvôli modelom ako XGBoost. TF-IDF vektorizácia sa používa na premenu textov na vektory čísel, ktoré môžu byť vstupom pre algoritmy strojového učenia. TF (Term Frequency) označuje frekvenciu výskytu slova v danom dokumente. Jednoducho povedané, čím častejšie sa slovo v dokumente vyskytuje, tým je považované za dôležitejšie. Existuje viacero spôsobov, ako TF vypočítať, ale väčšinou sa používa počet výskytov daného slova v dokumente delený celkovým počtom slov v dokumente. IDF (Inverse Document Frequency) znižuje váhu slov, ktoré sa vyskytujú príliš často naprieč rôznymi dokumentmi a tým pádom sú menej informatívne. Slovo, ktoré sa vyskytuje v mnohých dokumentoch, nemá veľkú hodnotu a dostáva nižšie IDF skóre. TF-IDF skóre slova v dokumente je potom vypočítané ako súčin jeho TF a IDF skóre. Týmto spôsobom sú odmenené slová s vysokou frekvenciou v danej vete/danom dokumente, ale penalizované slová, ktoré sú bežné naprieč mnohými dokumentmi. Výsledkom je, že každý dokument je reprezentovaný ako vektor, kde každý element tohto vektora predstavuje jedinečnú váhu priradenú slovu v kontexte celého datasetu.

D. Modelovanie

V tejto fáze sa venujeme samotnému trénovaniu modelov, ktoré predstavuje kľúčovú časť procesu. Výber konkrétnych modelov je motivovaný a podporený už existujúcimi výskumami. Trénovanie modelu zahŕňa niekoľko krokov vrátane výberu typu modelu, nastavenia parametrov a samotného procesu učenia.

1) CNN

Ako začiatočný model sme vybrali CNN hlavne pre jeho rýchlosť. Model je trénovaný na trénovacej sade s použitím aj validačnej sady, aby sa overila jeho presnosť a predišlo sa pretrénovaniu. Neurónová sieť bola zložená z niekoľkých vrstiev, kde výstupná vrstva používa

softmax aktivačnú funkciu a umožňuje klasifikáciu do troch tried.

2) XGBoost

Ako druhý bol vybraný model XGBoost. Použité boli parametre use_label_encoder=False, čo je vhodné, keďže predikcia je už kódovaná a eval_metric='mlogloss', ktorá pomáha monitorovať stratovú funkciu pre viactriednu klasifikáciu.

3) BERT

Pre hlbšiu analýzu a porozumenie kontextu sme sa rozhodli implementovať model BERT. Pri modelovaní v tomto prípade využívame iné vektorové reprezentácie ako napr. Token embeddings (reprezentácia vstupných textov, ktoré zachytávajú jazykový kontext a sémantiku), Positional embeddings (informácie o pozícii tokenov, umožňujúce modelu zachytiť poradie slov v texte), Segment embeddings (rozlišovanie rôznych segmentov vstupu, napr. v prípade porovnania dvoch textov).

E. Vyhodnotenie modelovania

Pre každý z predstavených modelov bola vypočítané metriky presnosť, návratnosť a F1-skóre pre každú z troch klasifikačných tried. Najlepší výsledok danej metriky je v tabuľkách vyznačený hrubým písmom. V presnosti vyšiel najlepšie model BERT, ktorý dosiahol pri normálnych príspevkoch hodnotu 75 %, pri klasifikácii urážlivých príspevkoch presnosť 54 % a pri nenávistných príspevkoch 73 %. Výsledky presnosti pre všetky modely sú zobrazené v Tabuľka 1. Nižšie hodnoty môžu byť spôsobené nedostatočným ladením hyperparametrov modelov.

Tabuľka 1

Presnosť modelov detekcie toxických prejavov

Model	Presnosť - normálny príspevok	Presnosť - urážlivý príspevok	Presnosť - nenávistný príspevok
CNN	58 %	40 %	57 %
XGBoost	60 %	51 %	71 %
BERT	**75 %**	**54 %**	**73 %**

Normálne príspevky na základe metriky návratnosť najlepšie klasifikoval model XGBoost s hodnotou 79 %. Urážlivé a nenávistné príspevky podľa návratnosti najlepšie vyhodnotil model BERT (52 % a 80 %). V nasledujúcej Tabuľka 2 sú zhrnuté všetky výsledky pre metriku návratnosť.

Tabuľka 2

Návratnosť modelov detekcie toxických prejavov

Model	Návratnosť - normálny príspevok	Návratnosť - urážlivý príspevok	Návratnosť - nenávistný príspevok
CNN	59 %	36 %	62 %
XGBoost	**79 %**	38 %	60 %
BERT	71 %	**52 %**	**80 %**

F1-skóre predstavuje harmonický priemer presnosti a návratnosti. V tejto metrike najlepšie výsledky dosiahol model BERT, ktorý dosiahol hodnoty postupne 73 %, 53 % a 77 %. Zvyšné výsledky je možné nájsť v Tabuľka 3.

Tabuľka 3

F1-skóre modelov detekcie toxických prejavov

Model	F1-skóre - normálny príspevok	F1-skóre - urážlivý príspevok	F1-skóre - nenávistný príspevok
CNN	59 %	38 %	60 %
XGBoost	68 %	43 %	65 %
BERT	**73 %**	**53 %**	**77 %**

XGBoost má najlepšie výsledky pri kategórii normálnych príspevkov, ale problémy s klasifikáciou nenávistných a urážlivých príspevkov. Najlepšie vyvážená presnosť medzi všetkými kategóriami patrí BERTu, s najnižšími chybami pri nenávistných príspevkoch. CNN dosahuje dobrú presnosť pri odhaľovaní nenávisti, ale vyššie chyby pri normálnych a urážlivých príspevkoch v porovnaní s BERT. Na základe týchto výsledkov sa zdá, že BERT poskytuje najlepšiu rovnováhu medzi presnosťou a chybovosťou naprieč všetkými kategóriami, čo ho robí najvhodnejším modelom pre klasifikáciu nenávistného prejavu v tomto kontexte.

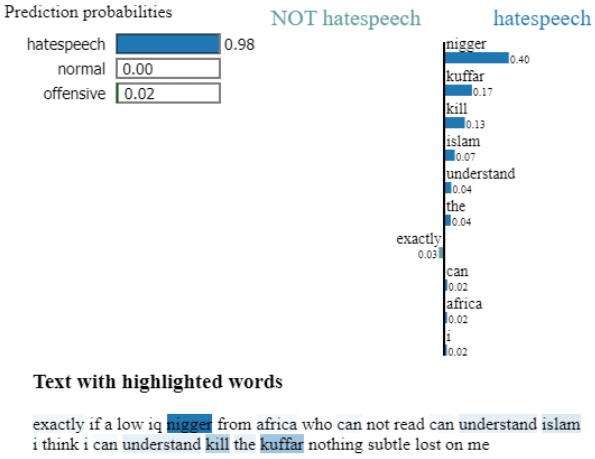
F. Vysvetliteľnosť

V tejto časti sa zameriame na analýzu výsledkov SHAP a LIME. Vysvetliteľnosť je kľúčovým aspektom, pretože umožňuje lepšie porozumieť tomu, ako modely prijímajú rozhodnutia, a tým

zvyšuje dôveru v ich použitie. Na tento účel sme využili kombináciu rôznych modelov a techník vysvetliteľnosti, konkrétne XGBoost a SHAP, CNN a SHAP, BERT a LIME. Najlepšie výsledky dosiahol model BERT, preto v nasledujúcej časti zobrazíme konkrétne výsledky vysvetliteľnosti v spojení s týmto modelom. Pre naše analýzy sme použili tri konkrétne vety vybrané z datasetu, ktoré predstavujú rôzne typy textových vstupov.

Veta 1: *"Exactly if a low iq nigger from africa who can not read can understand islam i think i can understand kill the kuffar nothing subtle lost on me".*

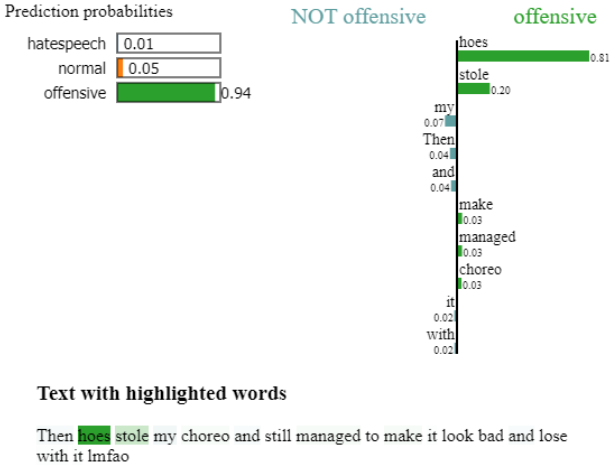
Táto veta obsahuje rasistické a nenávistné prejavy, čo nám umožňuje skúmať, ako modely identifikujú a klasifikujú takéto správanie. Podľa hodnôt LIME mali najväčší vplyv na klasifikáciu tejto vety ako nenávistnej slová: *nigger* (0.398), *kuffar* (0.172), *kill* (0.127), *islam* (0.066), *understand* (0.044). Bližšie sú LIME hodnoty zobrazené na Obr. 1.



Obr. 1 LIME vysvetlenie pre prvú ukážkovú vetu

Veta 2: *"Then hoes stole my choreo and still managed to make it look bad and lose with it lmfao".*

Táto veta obsahuje vulgárne a urážlivé slová, čo nám poskytuje príklad, ako modely reagujú na ofenzívny obsah. Podľa hodnôt LIME mali najväčší vplyv na klasifikáciu tejto vety ako urážlivej slová: hoes (0.813), stole (0.203). Ďalšie hodnoty sú zobrazené na Obr. 2, kde sú tieto slová aj zvýraznené priamo vo vete.



Obr. 2 LIME vysvetlenie pre druhú ukážkovú vetu

Veta 3: *"I am still sad i have been trying to make my phone work so it will take pictures the rooftops are all white in my neighborhood and it so lovely i hate my phone right now"*

Táto veta je normálna a neobsahuje nenávistné prejavy, čo nám pomáha pri analýze, ako modely odlišujú normálny obsah od nenávistného alebo urážlivého. Podľa hodnôt LIME mali najväčší vplyv na klasifikáciu tejto vety ako normálnej slová: phone (0.004), now (0,003), pictures (0.003). Tieto hodnoty sú pomerne nízke, čo nasvedčuje neutralite tejto vety.

Pri analýze rôznych typov textových vstupov sme identifikovali vplyv jednotlivých slov na klasifikáciu textov. Ukázalo sa, že kontext je veľmi dôležitý. Modely vedia označiť množstvo nenávistných prejavov aj bez kontextu vety. Potom sú tu aj slová ako "africa" a "islam", ktoré by sa mohli zdať neutrálnymi no môžu mať v závislosti od kontextu negatívny vplyv na klasifikáciu, čo naznačuje, že modely dokážu identifikovať aj skryté významy. Model BERT v kombinácii s LIME dokázal v porovnaní s ostatnými testovanými modelmi lepší prehľad v kontexte.

G. Hodnotenie vysvetliteľnosti

V tejto časti popíšeme výsledky hodnotenia vysvetliteľnosti modelu pomocou metriky DoX. Toto vyhodnotenie pracuje s hodnotami SHAP, takže je zamerané len na modely CNN a XGBoost, lepšie výsledky dosiahol XGBoost a tie sú bližšie popísané. Vysvetliteľné aspekty predstavujú rôzne komponenty alebo charakteristiky predikcií, ktoré chceme podrobne analyzovať. V kontexte metriky DoX tieto aspekty poskytujú základ pre meranie vysvetliteľnosti modelu. Ako aspekty {X} sme zvolili atribúty, SHAP hodnoty, hodnoty predikcie a referenčný bod (priemerná hodnota predikcií). Výsledkom je ohodnotenie otázok týkajúcich sa vysvetliteľných aspektov. Hodnoty DoX pre jednotlivé otázky ukazujú mieru vysvetliteľnosti pre každý aspekt modelu. Vyššie hodnoty DoX naznačujú lepšiu vysvetliteľnosť, zatiaľ čo nižšie hodnoty naznačujú slabšiu vysvetliteľnosť. Hodnoty sú nadobudnuté z prvých 100 inštancií vybraných z kolekcie textových vysvetlení kvôli zníženiu výpočtovej náročnosti. Pre model XGBoost dosiahli najlepšie výsledky otázky:

- {X}, unless what? : 0.3997
- How is {X} expressed? : 0.3978
- What is similar to {X}? : 0.3862

Výsledky naznačujú, že model poskytuje relatívne dobré vysvetlenia týkajúce sa podmienok, za ktorých je koncept platný, spôsobu vyjadrenia a podobnosti s inými konceptmi. Pre model XGBoost dosiahli najnižšie hodnoty otázky

- Except in what context, is {X} considered acceptable? : 0.2388
- While discussing what topics does {X} arise? : 0.2435
- Which groups are affected by {X}?: 0.2551

Tieto otázky dosiahli najnižšie hodnoty DoX. Môže byť užitočné zamerať sa na zlepšenie vysvetliteľnosti modelu v týchto aspektoch. Priemerná hodnota DoX pre všetky otázky a aspekty je 0.322, ktorá naznačuje strednú úroveň vysvetliteľnosti, čo znamená, že model je schopný poskytovať relatívne zrozumiteľné vysvetlenia, ale existuje priestor na zlepšenie.

VI. ZÁVER

Táto štúdia bola zameraná na detekciu nenávistných prejavov pomocou XAI. Implementovali sme a hodnotili rôzne modely, konkrétne CNN, XGBoost a BERT v kombinácii s metódami LIME a SHAP na vysvetlenie rozhodnutí modelov. Pri analýze rôznych textových vstupov sme identifikovali vplyv jednotlivých slov na klasifikáciu textov. Model BERT v kombinácii s LIME preukázal vynikajúci prehľad v kontexte a dokázal identifikovať nenávistné prejavy s najlepšou presnosťou spomedzi využitých modelov, ale existuje ešte priestor na zlepšenie. Na záver možno povedať, že XAI má veľký potenciál v oblasti detekcie nenávistných prejavov. Naša práca prispieva k rozvoju tohto prístupu a otvára nové možnosti pre aplikáciu AI v boji proti nenávistným prejavom a urážlivému obsahu na internete. Pre budúce výskumy odporúčame zamerať sa na zlepšenie vysvetliteľnosti komplexnejších modelov, ako je BERT a rozšíriť rámec hodnotenia vysvetliteľnosti na ďalšie modely a metriky.

POĎAKOVANIE

Táto práca vznikla vďaka podpore Agentúry na podporu výskumu a vývoja na základe Zmluvy č. APVV-22-0414 a Vedeckej grantovej agentúry MŠVVaŠ SR projekt č. 1/0685/21.

POUŽITÁ LITERATÚRA

[1] H. Mehta and K. Passi, “Social Media Hate Speech Detection Using Explainable Artificial Intelligence (XAI),” Algorithms 2022, Vol. 15, Page 291, vol. 15, no. 8, p. 291, Aug. 2022, doi: 10.3390/A15080291.

[2] B. Mathew, P. Saha, S. M. Yimam, C. Biemann, P. Goyal, and A. Mukherjee, “HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection,” 35th AAAI Conference on Artificial Intelligence, AAAI 2021, vol. 17A, pp. 14867–14875, 2021, doi: 10.1609/AAAI.V35I17.17745.

[3] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, vol. 1, pp. 4171–4186, 2019.

[4] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why should I trust you?’ Explaining the predictions of any classifier,” 2016, Accessed: Apr. 07, 2021. [Online]. Available: https://arxiv.org/abs/1602.04938?source=post_page----------

[5] S. M. Lundberg and S. I. Lee, “A Unified Approach to Interpreting Model Predictions,” Adv Neural Inf Process Syst, vol. 2017-December, pp. 4766–4775, May 2017, Accessed: Jan. 22, 2024. [Online]. Available: https://arxiv.org/abs/1705.07874v2

[6] S. M. Lundberg, G. G. Erion, and S.-I. Lee, “Consistent Individualized Feature Attribution for Tree Ensembles,” Feb. 2018, Accessed: May 29, 2024. [Online]. Available: https://arxiv.org/abs/1802.03888v3

[7] F. Sovrano and F. Vitali, “An objective metric for Explainable AI: How and why to estimate the degree of explainability,” Knowl Based Syst, vol. 278, p. 110866, Oct. 2023, doi: 10.1016/J.KNOSYS.2023.110866.

The process of digitalization of the territory and DSML for the application domain of spatial planning and environment

[1]*Štefan KOREČKO,* [2]*Michal PUSTÝ,* [3]*Branislav SOBOTA*

[1,2,3] Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical university of Košice, Letná 9, 04200 Košice, Slovak republic

[1]stefan.korecko@tuke.sk

***Abstract* — This article focuses on the design, implementation, and use of a domain-specific modeling language for the application domain of spatial planning and environmental management. It was necessary to analyze existing solutions in this area. Domain-specific modeling languages and the possibilities of using the WebGME tool for metamodeling were analyzed. The current state in the field of spatial planning and environmental management was also analyzed, focusing on processes, information, and existing systems. Subsequently, a domain-specific modeling language was designed and implemented for the given domain. Its services were described, requirements were characterized and defined. Finally, a domain-specific modeling language was created and tested on selected projects to determine whether it can effectively facilitate the work of planners and identify possible shortcomings. The goal of this work was to create a domain-specific modeling language for spatial planning and environmental management to effectively facilitate the work of planners in urban and non-urban areas at the same time, the possibilities of use as a basis for human-robot communication are discussed here.**

***Keywords* — Domain-specific modeling language, environmental management, spatial planning, WebGME, digital territory**

I. Introduction

Virtual reality (VR) became a vital aspect of our lives today. In addition to VR, related technologies such as mixed reality (MR) or augmented reality (AR) to associated in the XR

The digitization of the territory is a very interesting but also a complex and process demanding area [1]. The result of this process has a subsequent impact both in the economic, social and ecological spheres, but also for possible future use as a basis for e.g. virtualization (digital twin of the territory) and the subsequent implementation of testing, navigation, prognostic or training applications. Spatial planning is also a part of this process [2].

Currently, spatial planning is a complex process involving the management, development, and utilization of land to achieve desired outcomes such as urban development, natural resource conservation, or disaster risk reduction. Its main task is to coordinate planned changes in the area with the interests of the public and private sectors, while also safeguarding the environment and cultural heritage or it prepare a knowledge base for future possible use of artificial intelligence, machine learning [3] [4] or robotics technologies (e.g. autonomous cars movement, automatic delivery services or trusted assistance services [5] [6] [7]). The environment is crucial in spatial planning as it ensures sustainable use of natural resources and protects ecosystems. Effective planning and management of these spaces require spatial planners to analyze and interpret large amounts of data, which can be a challenging task without the right tools and techniques. However, this opens up opportunities for the use of DSML (Domain-Specific Modeling Language) [8], which, if properly tailored to the domain of spatial planning, can improve the resulting product according to the requirements and goals it serves.

Compared to general-purpose programming languages for spatial planning, DSML offers several advantages. Firstly, it adapts to the needs of professional planners. Customized systems contain specialized constructions and abstractions that facilitate the modeling of complex spatial systems [9] (e.g. also ecologic, traffic or agricultural). This can help reduce the time and effort required to create accurate and detailed models of spatial planning. By using appropriate CASE (Computer-Aided Software Engineering) tools [10], we can enable collaborative work on projects

for spatial planners. Through their use, planners and engineers can more easily share and discuss their ideas and collaborate in finding optimal solutions to complex spatial problems [11]. Using a model-based approach, experts can simulate various scenarios and assess the potential impact of their decisions on the environment, urbanization, and the residents themselves.

Simulation using models created in a properly designed DSML can show how these systems will function, how they will interact with each other, what properties they will have, and how they will meet user and other requirements and goals. Models in simulation should demonstrate the impact on the environment by simulating interactions and dynamics of environmental variables, such as pollution levels, habitat degradation, or resource depletion, to assess the potential consequences of various scenarios or interventions [12]. This can help identify potential risks and compromises while ensuring that planning decisions are made transparently and responsibly.

Overall, the use of DSML in spatial planning can help improve the accuracy and efficiency of the planning process. DSML can assist spatial planning and the environment by providing a specialized and accurate framework for representing and analyzing spatial planning data, enabling more effective decision-making and understanding of concepts [13]. By leveraging the strengths of modeling and simulation, spatial planning experts can create a better future, not only for the residents of the community but also contribute to better environmental protection in the community and its surroundings.

II. Design of application domain

The main focus of the application is to streamline spatial planning and environmental protection in order to achieve increased accuracy, simplify reuse, and improve project comprehensibility in the resulting product. It is necessary to define potential risks and weaknesses of the application domain in advance, while also planning strategies to mitigate them. Additionally, it is important to analyze the services and functions that our created DSML should support. We characterize the model itself and describe the individual metamodels that our overall system should contain.

A. Description of services

We have determined the services that our DSML should provide. These services and functions should address the comprehensive needs of this area and provide solutions for current projects. Among the basic services and functions that our DSML should provide are:

- *Spatial planning modeling:* The language should allow precise and comprehensive modeling of spatial plans and regulations, land use planning, infrastructure planning in both uninhabited and inhabited areas. This should enable users to intuitively create and modify plans according to specific requirements and regulations.
- *Analytical tools for the environment:* The language should allow integration of analytical tools for assessing and monitoring the environmental impacts of spatial planning. These tools should include assessments of biodiversity impact, air and water quality, and the impact of environmental changes on the surrounding ecosystem.
- *Recommendations for stakeholder management:* The language should allow the implementation of tools for managing and engaging stakeholders, enabling effective collaboration and communication among various stakeholders, including local residents, government institutions, and environmental groups.
- *Integration of regulatory requirements:* The language should support the integration of local and international regulatory requirements and standards for environmental and spatial planning. This ensures that the created plans are aligned with legal and environmental regulations applicable to the country or specific area.
- *Simulation of infrastructure impact:* Providing means for simulating the impact of new infrastructure projects on the environment and the community. These simulations should enable the prediction of impacts on transportation, pollution, and quality of life in the given area.
- *Visualization of geographic data:* Providing tools for visualizing geographic data and creating interactive maps and models. These tools should allow users to better understand the geographic aspects of their plans and display their outputs for easier comparison and analysis.
- *Base for artificial intelligence, machine learning and robotic technologies:* Providing knowledge base for possible use of these technologies in the future e.g. virtual reality (digital twin of the territory), testing, navigation, prognostic or training applications

B. Metamodel

In the field of systems engineering, a well-defined metamodel serves as the backbone for a comprehensive understanding, design, and implementation of integrated systems. The metamodel is structured into five characteristic sections, each encompassing a critical aspect of the system:

- *Architecture metamodel*
- *Security metamodel*
- *Deployment metamodel*
- *Roles and permissions metamodel*
- *UI metamodel*

Together, these sections form a cohesive representation of interconnected components and relationships. The aim of this chapter is to provide an overview of the created metamodel and then delve into the complexity of each individual metamodel. Through this exploration, we seek to clarify the reasons for the chosen design, which is taken into consideration in spatial planning and environmental protection. We will also clarify the role that each metamodel plays in the functioning of the system.

Understanding the metamodel enables us to gain valuable insights into the underlying structure, which facilitates effective decision-making, collaboration, and sustainable development in the context of spatial planning and environmental protection.

Components in the metamodel include attributes, constraints, and aspects. Attributes represent properties, characteristics, or elements of a component. Constraints define rules or conditions that a component must meet. And finally, aspects represent various perspectives of the component. They capture the behavior, structure, functionality, or qualitative properties of the component.

Metamodels also include connections between components, which include:

- *Base Relationship/Inheritance:* Represented by a red line. It defines the structure and behavior of models. It establishes a relationship between the base node and its instances or derived nodes.
- *Pointer:* Represented by a blue line. It signifies a type of relationship between two nodes, where one node refers to another node without implying inheritance or containment. Pointers can be used to represent various types of relationships, such as associations, references, dependencies, or connections between different parts of the model.
- *Containment:* Represented by a black line. It defines where a node can be created in the hierarchy. The one-to-many relationship allows creating a node in multiple places, with cardinality possibly limited (e.g., 0..* means any number of children).

The section of the metamodel diagram that focuses on the security metamodel is shown in the figure Fig. 1.

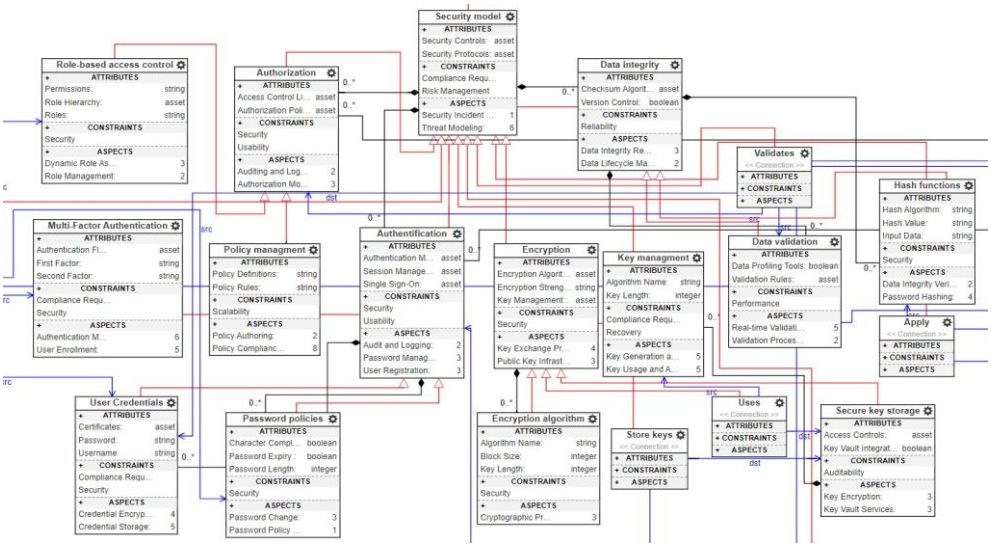


Fig. 1 Security metamodel

III. Implementation and Verification of the Proposed DSML

The main goal of this implementation is to provide efficient and intuitive tools for modeling in a specific domain context. To achieve this goal, we have developed a detailed use case scenario, which will help us in the implementation and validation of the DSML's functionality.

The use case scenario thoroughly describes typical situations and tasks that our modeling platform addresses, while also defining interactions between the user and the DSML. In this way, we can explore and demonstrate relevant aspects of the modeling process within a specific domain.

Furthermore, we will specify requirements for the selected target system and describe it's specific functional and non-functional requirements. The aim is to clearly specify the strategy and expectations behind the development of the DSML, focusing on functionalities that enable effective modeling in specific domain contexts, as well as aspects that ensure it is reliable and sustainable operation. Analyzing these goals and requirements will provide a solid foundation for evaluating the success of the DSML implementation within a specific domain.

In addition to specifying requirements for the target system, we will also focus on verifying the practical use of the DSML. We will analyze specific examples and situations where our language can provide significant added value compared to existing modeling tools. We will also evaluate the performance of the DSML in different contexts and its ability to successfully address the specific needs of users in the given domain environment.

Alongside describing the use case scenario and verifying the functionality of the DSML, we will also present diagrams illustrating the architecture and interactions within our modeling language. These diagrams will provide insight into the internal structure of the DSML and its components, aiding in better understanding the implementation and connections between various parts of our modeling tool.

Overall, we will provide a comprehensive view of the implementation, validation, and structure of our domain-specific modeling language, allowing us to assess its value and contribution in a specific domain environment.

A. Functional requirements

Functional requirements of the DSML include the ability to define entities and their properties, create and manage relationships between them, and provide mechanisms for defining and executing specified rules. Additionally, readability and comprehensibility of the created models are considered key objectives of our DSML system.

1. Land Planning
    a. Users should have access to an interactive map with detailed land zoning and the ability to search for information about specific plans and restrictions in the area.
    b. Users should be informed about the process of creating and implementing plans and strategies and have the opportunity to actively participate in consultations and public meetings to express their opinions and requirements.
    c. Users should have access to information about planned urban projects and construction in their vicinity, including details about new constructions, infrastructure renovations, and potential environmental and community impacts.
2. Monitoring and Data Recording of the Territory
    a. Users should have access to public data regarding land use and the environment in their area, including information about emissions, water usage, and biodiversity status.
    b. Users should have access to online mapping tools with layers displaying land use, river and watercourse routes, protected areas, and other important environmental characteristics.
3. Ensuring Compliance with Legislation
    a. Users should be informed about planned and executed inspections, and in case of discrepancies or violations of legislation, they should be able to report them through a user-friendly platform.
    b. Users should have access to professional assistance and advice regarding environmental protection legislation. They should be able to consult with specialists and obtain information about how specific regulations and rules affect their area and environment. This advice should be available through a user-friendly platform or personal communication.
4. Continuous Development and Innovation
    a. Users should have access to information about new technologies and initiatives related to environmental protection and could participate in educational events promoting environmentally responsible behavior.
    b. Users should have the opportunity to actively participate in initiatives and projects related to environmental protection in their area and be informed about opportunities to engage in volunteer activities and local campaigns.

B. Non-functional requirements

Non-functional requirements for the DSML emphasize performance and scalability to handle manipulation with extensive models. Additionally, performance and scalability are priorities to ensure that the DSML can flexibly grow and adapt to changing user needs. Security and data protection are inseparable parts of our non-functional requirements, ensuring that handling sensitive information in the models is controlled and secure.

1. Security
    a. Data security and sensitive information regarding land planning and the environment must be safeguarded. The system should employ high-level data encryption to minimize unauthorized manipulation or access to this data.
    b. The system must effectively identify and neutralize potential security threats. Implementing robust attack detection systems and strict authentication measures should be a priority in building the system.
    c. A comprehensive permission management system is necessary, allowing access only to authorized users based on their roles and permissions. This ensures that sensitive data is accessible only to those individuals who have authorization and a legitimate need to work with it.
2. Availability and Reliability
    a. The system must be available to users 24 hours a day, 7 days a week, with minimal planned downtime for maintenance and updates. This ensures that users have uninterrupted access to the system when needed.
    b. Monitoring tools must be implemented to track system performance and availability. These tools should be capable of automatically alerting administrators in case of downtime or reduced availability, allowing them to react promptly and minimize downtime.
    c. A backup and recovery system must be in place to regularly back up all data and information and be capable of restoring them in case of any failure. This minimizes the risk of data loss and ensures data continuity.
3. Performance and Scalability
    a. The system should efficiently process large volumes of data and information without significant performance degradation. This includes optimizing database structures, utilizing high-performance servers, and employing efficient data processing algorithms.
    b. Implementing an automatic scaling mechanism is important, allowing the system to adapt to increased loads and data volumes without disrupting ongoing operations. This minimizes the risk of system overload and ensures smooth operation even during periods of increased activity.
    c. Performance monitoring mechanisms should be implemented to track various aspects of the system, such as memory usage, computational power, and network bandwidth. These tools allow for identifying and addressing potential bottlenecks and optimizing system performance as a whole.
    d. The system must ensure high response speed to user requests regardless of the volume of processed data. It should be optimized to minimize response time and maximize user satisfaction.

IV. Verification of The Proposed DSML on Selected Projects

To verify the proposed methodology and DSML, we will use three different projects. Specifically, these are projects for planning urban green spaces, transportation infrastructure, and monitoring air quality.

In the WebGME tool, we utilized the Composition mode, which provided us with a means to easily create the necessary diagrams. This mode allows for the creation of various types of diagrams by creating new instances of a specific part of the metamodel that the diagram focuses on. Upon switching to Composition mode within this instance, we were able to start creating the diagram and insert individual nodes into it. Thanks to WebGME, it's possible to precisely define which nodes are relevant for a particular part of the metamodel and integrate them appropriately into the created diagram. The WebGME tool only allows connecting diagram nodes that have defined relationships within the metamodel. This means that the creation of diagrams is closely tied to the structure and constraints of the metamodel from which the diagrams originate. This approach is applicable to all diagrams intended for verifying the methodology within this project, ensuring consistency and reliability in the created models.

In the architecture model diagram, the fundamental component is the Cloud Server. The server processes all user requests and accesses. The diagram adheres to the standards defined in the deployment metamodel. In the image below, you can see how we managed to create a deployment model based on the standards defined in the created metamodel in Composition mode. The diagram in the image is specifically based on the project for planning suburban transportation infrastructure in Enfield, Adelaide, Australia. Here, we see the implementation of various elements such as the Cloud Server, a user with the role of developer, a continuous deployment tool and a continuous integration tool.

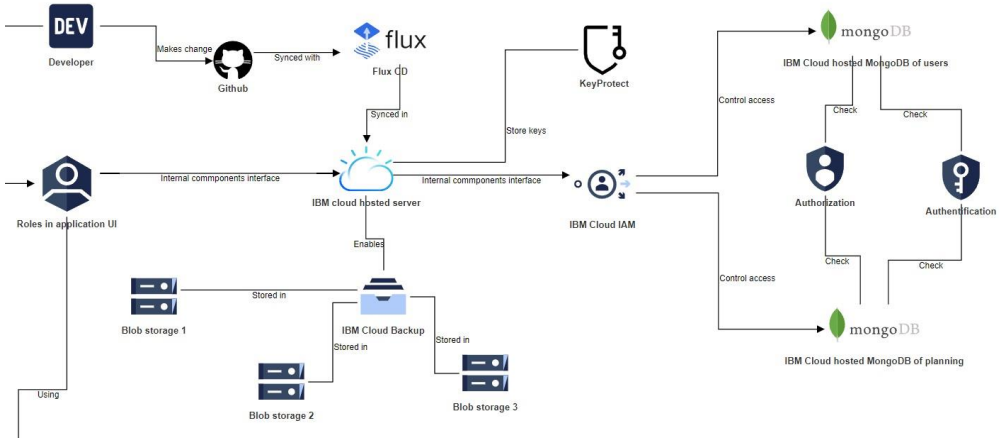


Fig. 2 Diagram of the deployment model

V. SUMMARY AND CONCLUSION

The aim of this work was to design and implement a DSML for the domain of urban planning and environmental protection using the WebGME tool and to verify it's implementation on various projects within this domain. Before the actual design of the DSML, it was necessary to analyze the current state and perspectives in the domain-specific modeling in software engineering. Similarly, it was necessary to analyze the possibilities of using WebGME for metamodeling and implementing DSML for the domain of urban planning and the environment. Several projects and systems currently used in the selected domain area were explored.

After analyzing the processes, information, systems, and surroundings in the domain, the design of the DSML itself was approached. The application domain was thoroughly characterized, and its services were described. As part of our work, a security analysis of the application domain was also conducted, allowing us to describe in more detail the areas in the system where potential security threats and attacks could occur. This systematically revealed potential security threats that might otherwise have gone unnoticed.

An important step was the creation of the metamodel in the WebGME tool. The metamodel was divided into five parts, each covering a critical aspect of the system. These parts together form a coherent representation of interrelated components that define relationships and rules, enabling the efficient creation of models for individual projects in the domain of urban planning and environmental protection. After creating the metamodel, the implementation phase followed, where the defined projects were further dissected. Various usage scenarios were described for each project, outlining typical situations and tasks addressed in each project. Similarly, requirements for the selected target system were specified, describing specific functional and non-functional requirements of each project.

After implementing the created DSML, the verification phase followed, where specific examples and situations were analyzed to demonstrate how our language could provide significant added value compared to existing modeling tools. Three different projects were used for verification, each focusing on a different sector within the application domain of urban planning and environmental protection. The created metamodel was validated using the WebGME tool. This validation revealed deficiencies and errors in the created metamodel, which were subsequently corrected, and the metamodel was revalidated.

Finally, within the model diagrams, tools and resources used in each project were described, creating model diagrams that serve to better understand and visualize each project. These detailed descriptions allow not only urban planning professionals but also all stakeholders to better understand the complexity of the project environment and adapt their decisions and actions in future change planning or system development. Overall, the DSML we developed can assist urban planning and environmental protection professionals with efficient and precise modeling of complex relationships between various environmental factors, enabling them to better anticipate and manage the impacts of unexpected changes and also help them develop projects in this domain.

In conclusion, it is necessary to say, in the home laboratory of the authors (the LIRKIS laboratory at the DCI FEEI TU Košice) work is carried out related to the possible use of eXtended Reality and associated technologies in the field of their use in simulation or training of a processes for trustworthy human-robot interaction (HRI, [14]). Digitization of the real territory is one of the ways to prepare the basis for possible reliable, predictable, safe and independent movement of the robot on such territory, especially in the case of its use in reliable assistance services. Pilot data e.g. to create a simulation of the risk environment in such a territory were also mentioned,

for example, in [15]. In addition, high-quality virtualization of the territory with the use of our collaborative virtual reality technology [16] can also bring further progress in building especially training systems on this basis in the future. And that could be one of the ways of our future research.

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] Pomeroy, J.: Cities of Opportunities; Routledge Press, Taylor & Francis Group, London, 2020, 1st edition, ps.246, ISBN 9780367900656

[2] Pomeroy, J.: Hardware, Software, heartware: Digital twinning for more sustainable built environments; Routledge Press, Taylor & Francis Group, London, 2024, 1st edition, ps.247, ISBN 9781032569383

[3] Portugal, Ivens;Alencar, Paulo;Cowan, Donald: “A Preliminary Survey on Domain-Specific Languages for Machine Learning in Big Data”. In: 2016 IEEE International Conference on Software Science, Technology and Engineering (SWSTE). 2016, s. 108–110. available doi: 10.1109/SWSTE.2016.23.

[4] Mohasseb, Alaa; Bader-El-Den, Mohamed; Liu, Han; Cocea, Mihaela: Domain specific syntax based approach for text classification in machine learning context. In: 2017 International Conference on Machine Learning and Cybernetics (ICMLC). 2017, zv. 2, s. 658–663. available doi: 10.1109/ICMLC.2017.8108983.

[5] Taeihagh, Araz; Lim, Hazel Si Min: "Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks". Transport Reviews. 39 (1): pp. 103–128. ISSN 0144-1647, available doi:10.1080/01441647.2018.1494640.

[6] Hang and Gao Zhao, et al. : TNT: Target-driven Trajectory Prediction. In Proceedings of the 2020 Conference on Robot Learning (CoRL), pMLR Press, 2021. Vol 155, pp. 895—904, ISSN 1532-4435

[7] Kong, Xianjuan; Dang, Anrong; Li, Gongli: Research on evaluation of location planning for urban public service facilities based on GIS spatial analysis. In: 2010 IEEE International Geoscience and Remote Sensing Symposium. 2010, pp. 4220–4223. available doi: 10.1109/IGARSS.2010.5648863.

[8] Kelly, Steven; Tolvanen, Juha-Pekka: Domain-Specific Modeling: Enabling Full Code Generation. 2008. isbn 978-0-470-03666-2. available doi: 10.1002/9780470249260.

[9] Ciski, Mateusz; Rzasa, Krzysztof: Cultural Parks in the Spatial Planning System in Poland. In: 2018 Baltic Geodetic Congress (BGC Geomatics). 2018, pp. 145–149. available doi: 10.1109/BGC-Geomatics.2018.00033.

[10] Arnowitz, Jonathan; Arent, Michael; Berger, Nevin: Effective Prototyping for Software Makers. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2006. ISBN 0120885689.

[11] Hashim, M Amal; Sreedevi, A G: Model-driven engineering: A survey of the state of the art. International Journal of Computer Applications. 2018, vol. 179, No. 44, pp. 31–39.

[12] Asha, P.; Natrayan, L.; Geetha, B.T.; Beulah, J. Rene; Sumathy, R.; Varalakshmi, G.; Neelakandan, S. IoT enabled environmental toxicology for air pollution monitoring using AI techniques. Environmental Research. 2022, roč. 205, s. 112574. ISSN 0013-9351. available doi: https: //doi.org/10.1016/j.envres.2021.112574.

[13] Wegeler, Timo; Gutzeit, Friederike; Destailleur,Aurcle; Dock, Bernhard: Evaluating the Benefits of Using Domain-Specific Modeling Languages: An Experience Report. In: Proceedings of the 2013 ACM Workshop on Domain-Specific Modeling. Indianapolis, Indiana, USA: Association for Computing Machinery, 2013, s. 7–12. DSM ’13. ISBN 9781450326001. available doi: 10.1145/2541928.2541930.

[14] Lasota P.A., Song T., Shah J.A.: A Survey of Methods for Safe Human-Robot Interaction. Now Foundations and Trends, 2017, available doi: https://doi.org/10.1561/2300000052

[15] Sobota B., Guzan M., Kírešová S., Korečko Š.: Towards Real-World Data Supported XR Training of Trustworthy Human-Robot Interaction in a Risky Environment; In: IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics: Proceedings. - Danvers (USA): Institute of Electrical and Electronics Engineers, 2024, pp. 365-369. ISBN 979-8-3503-1720-6

[16] Hudák M., Korečko Š., Sobota, B.: Advanced user interaction for web-based collaborative virtual reality. In 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 2020, 343–348, doi:10.1109/CogInfoCom50765.2020.9237899.

Aktívne balansovanie so spínanými kondenzátormi

[1]*Daniel MARCIN,* [2]*Milan LACKO*

[1] [2]Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]daniel.marcin@tuke.sk, [2]milan.lacko@tuke.sk

***Abstrakt* — Táto práca skúma metódy aktívneho balansovania batériových článkov pomocou spínaných kondenzátorov. Simulácie ukazujú, že dvojúrovňové zapojenie spínaných kondenzátorov zvyšuje rýchlosť balansovania a efektívnosť prenosu energie medzi článkami oproti jednoúrovňovému zapojeniu. Pridanie dodatočného kondenzátora môže celý proces spomaliť, no na konci simulácie sa výsledky naopak zlepšili.**

***Kľúčové slová* — aktívne balansovanie, batéria, Li-ion, spínané kondenzátory.**

I. Úvod

Aktívne balansovanie batériových článkov je inovatívna metóda na minimalizáciu energetických strát, ktoré sa vyskytujú pri pasívnom balansovaní. Aktívne balansovanie umožňuje efektívnejšie využitie energie v batériových systémoch, avšak vyžaduje pre svoju činnosť zložitejšie obvody. V rámci tejto práce sa zameriavame na topológiu aktívneho balansovania so spínanými kondenzátormi. Tá pre svoju činnosť využíva PWM signály na riadenie spínačov, čo umožňuje opakované pripojenie kondenzátorov k dvojici článkov a prenos energie medzi nimi. Tento prístup je jednoduchý na riadenie, ale môže byť časovo náročný. Výsledky simulácií ukazujú potenciál a efektívnosť tejto metódy.

II. Aktívne balansovanie

Kvôli minimalizácii energetických strát pri pasívnom balansovaní boli vyvinuté metódy aktívneho balansovania. Tieto metódy zahŕňajú použitie komponentov na uskladnenie energie, ako sú kondenzátory, cievky alebo transformátory. Používajú sa na uskladnenie energie z článkov s vysokým SOC a ich prenos do článkov s nízkym SOC. Aktívne balansovanie umožňuje efektívnejšie využitie energie batérií v určitých aplikáciách, ale v porovnaní s pasívnym balansovaním si vyžaduje zložitejšie obvody a riadiace algoritmy [1], [2]. Okrem metód s uskladnením energie existujú aj metódy s DC/DC meničmi, pre priamy tok energie či paralelne zapojené tranzistory pre premostenie (z angl. shunting) článkov v obvode [3]. V tomto článku sa venujeme topológii založenej na spínaných kondenzátoroch.

A. Zapojenie spínaných kondenzátorov

Balansovanie spínaných kondenzátorov funguje tak, že na všetky spínače sa aplikuje dvojica komplementárnych PWM signálov, ktoré spôsobujú ich striedanie medzi zapnutým a vypnutým stavom [4], [5]. Každý kondenzátor je opakovane pripojený k dvojici článkov, čo umožňuje prenos energie z článkov s vyšším napätím do článkov s nižším napätím. Toto zapojenie tvoria kondenzátory C1, C2 a C3 na Obr. 1. Metóda riadenia balansovania spínaných kondenzátorov je jednoduchá a nevyžaduje informácie o SOC článku. Proces balansovania však môže byť časovo náročný z dvoch hlavných dôvodov. Po prvé, keď sa balansovanie blíži ku koncu, rozdiel napätí medzi susednými článkami sa stáva minimálnym, čo spôsobuje, náročnosť nabíjať kondenzátory a prenášať energiu. Po druhé, energiu možno prenášať len medzi dvoma susednými článkami v rámci jedného balansovacieho cyklu.

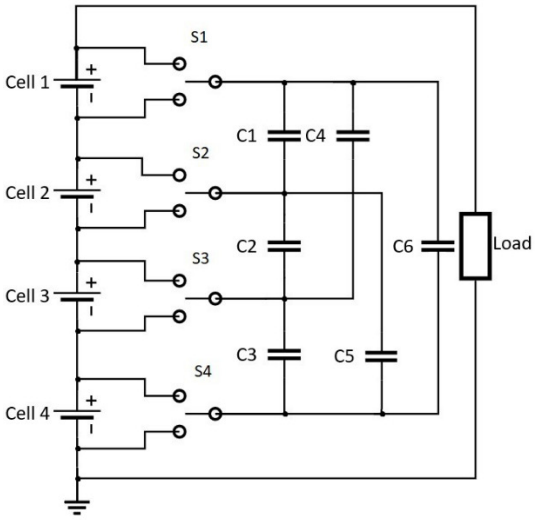
B. Zapojenie dvojúrovňových spínaných kondenzátorov

Dvojúrovňové balansovanie spínaných kondenzátorov zahŕňa použitie dvoch úrovní kondenzátorov, ako je znázornené na Obr. 1 [6]. Prvú úroveň tvoria kondenzátory C1, C2 a C3, zatiaľ čo druhú úroveň tvoria kondenzátory C4 a C5. Podobne ako pri metóde balansovania

spínaných kondenzátorov sa na všetky spínače aplikuje dvojica komplementárnych PWM signálov. V prípade rozdielu napätí medzi článkom 1 a článkom 3 sa energia môže prenášať priamo cez kondenzátor C4 namiesto toho, aby prechádzala cez C1 a C2. Využitím týchto „skratiek" prenosu energie možno zvýšiť rýchlosť až na jednu štvrtinu [7] a zlepšiť účinnosť balansovania znížením počtu kondenzátorov pre prenos energie zo vzdialených článkov.

C. Zapojenie dvojúrovňových spínaných kondenzátorov s pridaným kondenzátorom

Pridaním jedného kondenzátora navyše k predchádzajúcej schéme vytvoríme ďalšiu topológiu balansovania. Tento kondenzátor tvorí priame spojenie medzi prvým a posledným článkom v packu. V tomto prípade to je kondenzátor C6 na Obr. 1. Svoje opodstatnenie a využitie má hlavne vtedy, pokiaľ je rozdiel v napätí medzi prvým a posledným článkom.



Obr. 1 Schéma zapojenia balansovania so spínanými kondenzátormi

III. SIMULAČNÉ VÝSLEDKY

A. Simulácie pre porovnanie rýchlosti balansovania so spínanými kondenzátormi

S cieľom vyvinúť modulárny batériový systém sme vykonali simulácie rôznych obvodov aktívneho balansovania článkov na báze kondenzátorov. Spomedzi rôznych sme si vybrali obvod so spínanými kondenzátormi. Ako druhé sme simulovali obvod s dvojúrovňovými spínanými kondenzátormi. Napokon sme druhú simuláciu rozšírili pridaním ďalšieho kondenzátora ktorý pridáva prepojenie s prvým a posledným článkom. Tento obvod je rovnako zobrazený aj v simulačnom modeli na Obr. 2. Simulácie boli vykonané v programe MATLAB, Simulink.

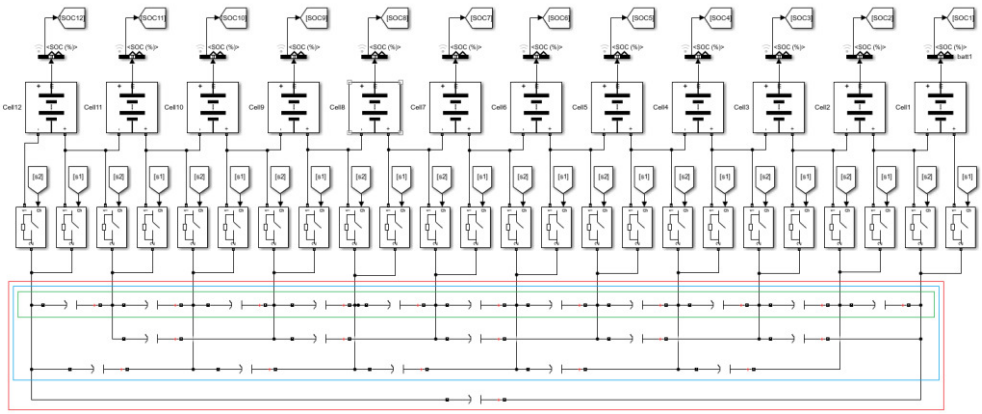
B. Modelovanie, parametre a simulačné dáta

Na simuláciu parametrov batériového článku boli použité údaje výrobcu MOLICEL s označením modelu INR18650-P26A. Pre náš účel sme simulovali na 12 článkoch zapojených do série. Pre účely porovnania týchto balansovacích obvodov boli použité modely batériového článku, ideálneho spínača a kondenzátorov z prostredia Simscape. Simulačné parametre sú uvedené v Tab. 1.

Tab. 1 Simulačné parametre

Parameter	Hodnota	Jednotka
Nominálne napätie článku	3,6	V
Kapacita článku	2,6	Ah
Čas odozvy článku	0,5	s
Impedancia článku (1kHz)	13,85	mΩ
Kapacita kondenzátora	50	mF
Spínacia frekvencia spínačov	1	kHz
Čas simulácie	2000	s

Spínače v obvode sú pripojené k dvom komplementárnym signálom PWM s názvom s1 a s2, ktoré pracujú s frekvenciou 1 kHz. Signál s1 ovláda spínače s nepárnymi číslami, zatiaľ čo signál s2 ovláda spínače so párnymi číslami. Simulačná schéma je zobrazené na Obr. 2.

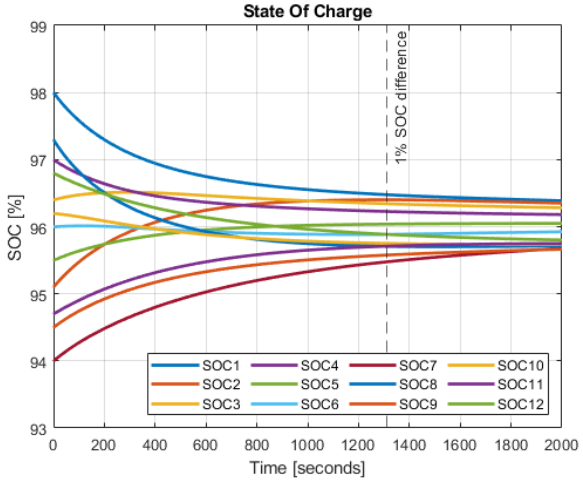


Obr. 2 Simulačná schéma pre 12 článkov v sérii v prostredí MATLAB Simulink s dvojúrovňovým zapojením spínaných kondenzátorov s pridaným kondenzátorom

Simulácia zahŕňala články usporiadané v náhodnom poradí. Počiatočný maximálny rozdiel v SOC predstavoval 4 percentuálne body. Z dôvodu vytvorenia náročnejších podmienok pre balansovanie, články s najvyšším a najnižším SOC neboli umiestnené vedľa seba. Počiatočné hodnoty SOC boli v rozmedzí 95 % až 98 % a sú rovnaké pre všetky simulácie. Simulácie trvali 2000 s. Na záver sme výsledky simulácií porovnávali, keď bol rozdiel medzi najnižšou a najvyššou hodnotou SOC 1 %.

C. Výsledky balansovania so zapojením spínaných kondenzátorov

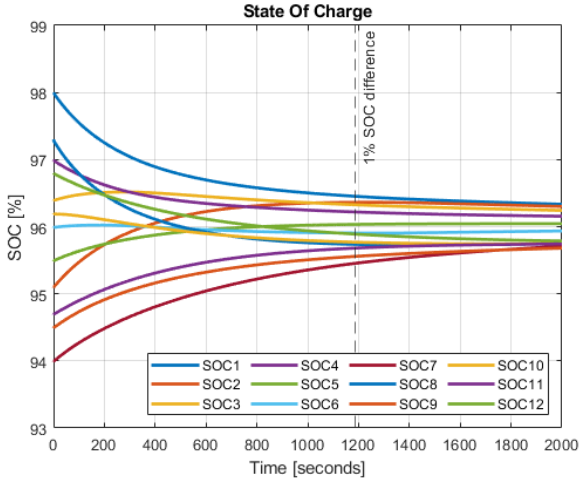
V prvej simulácii bol pomocou programu MATLAB Simulink implementovaný obvod so spínanými kondenzátormi. Model z Obr. 2 bol modifikovaný aby zodpovedal tomuto obvodu (zelená farba). Výsledky simulácie ukázali, že 1 % odchýlka v SOC článkov sa dosiahne za 1312 sekúnd. Článkom s najvyšším SOC bol článok 1 a článkom s najnižším SOC bol článok 7. Tieto výsledky simulácie sú znázornené na Obr. 3.



Obr. 3 Priebeh SOC 12S BP so zapojením spínaných kondenzátorov

D. Výsledky balansovania s dvojúrovňovým zapojením spínaných kondenzátorov

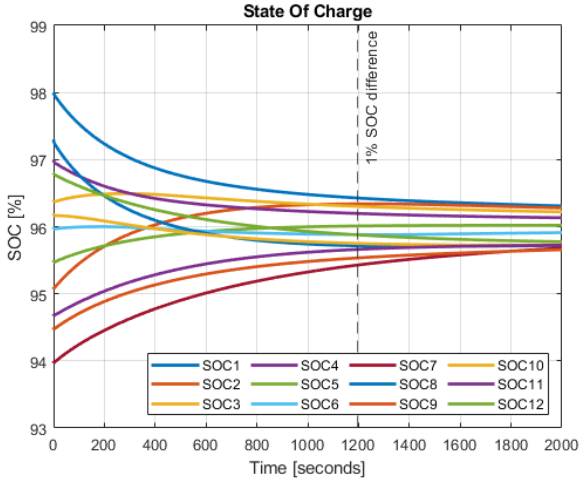
V druhej simulácii sme podobne implementovali dvojúrovňové zapojenie spínaných kondenzátorov do prostredia MATLAB Simulink (modrá farba Obr. 3). Simulácie ukázali, že 1 % odchýlka sa dosiahne za 1186 sekúnd, ako je znázornené na Obr. 4. Podľa očakávania sa tento obvod s ďalšími úrovňami kondenzátorov ukázal byť rýchlejší ako obvod so spínanými kondenzátormi. Rozdiel v čase balansovania bol 126 sekúnd a správanie SOC článkov bolo podobné.



Obr. 4 Priebeh SOC 12S BP s dvojúrovňovým zapojením spínaných kondenzátorov

E. Výsledky balansovania s dvojúrovňovým zapojením spínaných kondenzátorov s pridaným kondenzátorom

Nakoniec sme do modelu MATLAB Simulink implementovali ďalšiu úroveň kondenzátorov, ako je znázornené v simulačnej schéme na Obr. 2 (červená farba). Očakávalo sa, že tento spôsob balansovania bude najrýchlejší vzhľadom na prítomnosť dodatočného kondenzátora pripojeného k BP. Výsledky simulácie (Obr. 5) však ukázali, že pri rovnakej hodnote kondenzátora ako pri ostatných sa v skutočnosti balansovanie mierne spomalilo. Čas balansovania na dosiahnutie 1 % odchýlky v SOC bol 1195 sekúnd, čo bolo o 9 sekúnd pomalšie ako bez pridaného kondenzátora.



Obr. 5 Priebeh SOC 12S BP s dvojúrovňovým zapojením spínaných kondenzátorov s pridaným kondenzátorom

Nakoniec sme porovnali nielen výsledky simulácií pri 1 % odchýlke SOC, ale aj na konci času simulácie. Pri prvej simulácii s použitím obvodu so spínanými kondenzátormi sa dosiahol rozdiel SOC 0,72 %. Simulácia s dvojúrovňovými spínanými kondenzátormi viedla k rozdielu SOC 0,68 %, zatiaľ čo tretí obvod s pridaným kondenzátorom dosiahol rozdiel SOC 0,65 %. Tieto výsledky naznačujú, že zapojenie pridaného kondenzátora má potenciál urýchliť proces balansovania.

IV. Záver

Táto práca predstavuje metódy aktívneho balansovania článkov s využitím kondenzátorov. Pri výstavbe modulárneho batériového systému, ktorý sa skladá z viacerých batériových modulov, je nevyhnutné prehodnotiť metódu aktívneho balansovania. Diskutovali sme niekoľko metód balansovania a pre zhodnotenie ich výkonu sme ich implementovali do simulácií pomocou programu MATLAB Simulink. Metóda so zapojenými kondenzátormi je najjednoduchšia, ale aj najpomalšia. Pri batériovom module so 12 článkami v sérii trvalo balansovanie 1312 sekúnd na dosiahnutie rozdielu SOC o 1 % medzi najnižším a najvyšším článkom. Dvojvrstvová metóda so zapojenými kondenzátormi dosiahla rovnaký 1 % rozdiel za 1186 sekúnd. Zaujímavosťou je, že pridanie dodatočného kondenzátora do obvodu s dvojitým vrstvením spôsobilo mierne dlhšie

balansovanie, a to 1195 sekúnd. Na konci simulácie (2000 sekúnd) bol rozdiel SOC medzi najvyšším a najnižším článkom 0,75 % pri použití metódy so spínanými kondenzátormi. Pri dvojvrstvovej metóde bol rozdiel 0,68 % a s dodatočným kondenzátorom 0,65 %. Prvý test s pridaným kondenzátorom síce nepriniesol najlepšie výsledky, ale na konci simulácie bol rýchlejší ako bez kondenzátora. Výsledky simulácií sú dôležitým základným výskumom pre budúce experimentálne štúdie.

Poďakovanie

Táto práca bola podporená z projektov APVV-18-0436, VEGA 1/0363/23 and KEGA 059TUKE-4/2024.

Referencie

[1] J. Qi and D. Dah-Chuan Lu, 'Review of battery cell balancing techniques', in 2014 Australasian Universities Power Engineering Conference (AUPEC), IEEE, Sep. 2014, pp. 1–6. doi: 10.1109/AUPEC.2014.6966514.

[2] J. Carter, Z. Fan, and J. Cao, 'Cell equalisation circuits: A review', J Power Sources, vol. 448, p. 227489, Feb. 2020, doi: 10.1016/j.jpowsour.2019.227489.

[3] Z. B. Omariba, L. Zhang, and D. Sun, 'Review of Battery Cell Balancing Methodologies for Optimizing Battery Pack Performance in Electric Vehicles', IEEE Access, vol. 7, pp. 129335–129352, 2019, doi: 10.1109/ACCESS.2019.2940090.

[4] C. Pascual and P. T. Krein, 'Switched capacitor system for automatic series battery equalization', in Proceedings of APEC 97 - Applied Power Electronics Conference, IEEE, 1997, pp. 848–854. doi: 10.1109/APEC.1997.575744.

[5] U. K. Das et al., 'Advancement of lithium-ion battery cells voltage equalization techniques: A review', Renewable and Sustainable Energy Reviews, vol. 134, p. 110227, Dec. 2020, doi: 10.1016/j.rser.2020.110227.

[6] Y. Takeda and H. Koizumi, 'Modularized double-tiered switched capacitor voltage equalizer', in IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, IEEE, Oct. 2017, pp. 2782–2787. doi: 10.1109/IECON.2017.8216469.

[7] A. C. Baughman and M. Ferdowsi, 'Double-Tiered Switched-Capacitor Battery Charge Equalization Technique', IEEE Transactions on Industrial Electronics, vol. 55, no. 6, pp. 2277–2285, Jun. 2008, doi: 10.1109/TIE.2008.918401.

Vplyv povrchových úprav dosiek plošných spojov na fraktografickú morfológiu a mechanické vlastnosti spájkovaných spojov

Daniel DZIVÝ

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

daniel.dzivy@tuke.sk

***Abstrakt* — Cieľom tohto článku bolo študovať mechanizmy lomu spájkovaných spojov po mechanickej záťaži v strihu a preskúmať vzťahy medzi mechanickou pevnosťou a príslušnými fraktografickými morfológiami. V našom článku sme porovnali 4 štandardné typy povrchov DPS (ENIG, ImSn, SnHASL, Cu). Najmenšiu náchylnosť k krehkému zlyhaniu a starnutiu v porovnaní s ostatnými úpravami preukázala fraktografická analýza lomových povrchov na báze ENIG.**

***Kľúčové slová* — mechanická pevnosť v strihu, povrchové úpravy, fraktografická morfológia**

I. Úvod

Výskumné práce zaoberajúce sa kvalitou spájkovaných spojov sa zvyčajne zameriavajú na vplyv typu spájkovej zliatiny a procesných podmienok na kvalitu a spoľahlivosť spojov. Kvalita spojov sa len zriedka posudzuje z hľadiska povrchových úprav. Je známe, že povrchová úprava DPS tvorí kritické rozhranie medzi DPS a komponentom. Povrchová úprava DPS má dve hlavné funkcie, t. j. chrániť meď pred oxidáciou na DPS a zabezpečiť dobrú zmáčavosť. Zmáčavosť povrchov DPS, dĺžka času zmáčania, charakterizácia rozhraní a mechanické vlastnosti rôznych povrchových úprav sú dôležité na určenie optimálneho času a teploty počas tvorby spojov (počas procesu pretavenia spájky) [1], [2]. Tieto faktory ovplyvňujú rýchlosť tvorby a rastu IMC, odolnosť spoja voči mechanickému namáhaniu a starnutiu, ako aj elektrické vlastnosti spájkovaných spojov. Povrchové úpravy majú tiež vplyv na kvalitu spájkovaných spojov z hľadiska fraktografickej morfológie, adhézie, tvorby a hrúbky IMC na rozhraní, strižnej pevnosti, húževnatosti. Najnáchylnejšie na starnutie sú spájkované spoje, ktoré sú krehké alebo kvázi-krehké, pretože IMC sa počas starnutia zväčšujú a prispievajú k zníženiu celkovej pevnosti spájkovaného spoja [3], [4].

Počas strižného testu sa spájkovaný spoj pretrhne. Porucha môže nastať v objeme spájky (húževnaté porušenie) alebo na rozhraní medzi spájkou a povrchovou úpravou (krehké porušenie) [5], [6]. Porušenie však môže byť kvázi-tvárne alebo kvázi-krehké, keď lomová plocha obsahuje kombináciu húževnatej a krehkej oblasti porušenia. Ak sa spájkovaný spoj poruší v oblasti IMC, porušenie je klasifikované ako krehké, čo zodpovedá norme JESD22-B117B.

Existuje nedostatok štúdií skúmajúcich vplyv rôznych povrchových úprav na fraktografickú morfológiu. Dosiahnutie spoľahlivých spájkovaných spojov spočíva v prítomnosti drobných precipitátov IMC rovnomerne rozptýlených do objemu spájky. Nadmerný rast IMC, charakterizovaný jeho krehkosťou, vedie k degradácii tepelnej únavovej životnosti, pevnosti v ťahu a húževnatosti pri lome. Neúplne vyplnené spoje vedú k tvorbe spájkových dutín, čo sú prázdne miesta spôsobené nedostatočným množstvom spájky alebo nedostatočným zmáčaním. Povrchová úprava má významný vplyv na tvorbu dutín a zlepšuje zmáčanie počas procesu spájkovania. Použitie povrchových úprav na báze niklu môže zlepšiť spoľahlivosť spájkovaných spojov a znížiť tvorbu dutín. Vrstva Ni pôsobí ako bariéra, ktorá zabraňuje difúzii Cu z medeného podložia.

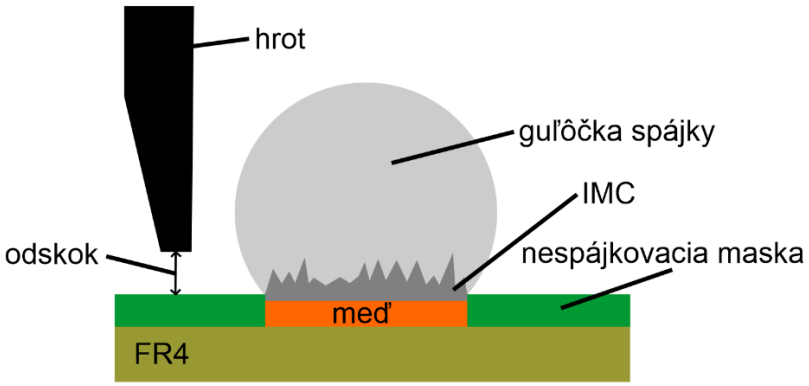
Súčasné trendy v oblasti výskumu spájkovaných spojov sú zamerané najmä na zmenu materiálového zloženia spájkovacej zliatiny. V našej štúdii sme sa zamerali na povrchovú úpravu a ako povrchová úprava ovplyvňuje mechanické vlastnosti spájkovaného spoja. V návrhu tohto experimentu sme vychádzali z našich predchádzajúcich výsledkov analýzy vplyvu povrchových úprav na kvalitu spájkovaných spojov. Zistili sme, že rozhranie medzi povrchovou úpravou a spájkou sa výrazne líši v závislosti od prítomnosti niklovej difúznej bariérovej vrstvy. Je známe, že typ IMC závisí od typu materiálov (spájka, povrchová úprava). Tvorba IMC na báze $Ni_{(X)}Sn_{(X)}$ je výsledkom reakcie medzi vrstvou niklu a bezolovnatou spájkou. Spájkované spoje na báze $Ni_{(X)}Sn_{(X)}$ sú tvárnejšie než $Cu_{(X)}Sn_{(X)}$ po opakovanom pretavení alebo po starnutí. Dôvodom tohto javu je pomalší rast IMC na báze $Ni_{(X)}Sn_{(X)}$ než $Cu_{(X)}Sn_{(X)}$ [4], [7], [8].

V našej štúdii sme sa zamerali na analýzu morfológie lomových povrchov a porovnanie konvenčných povrchových úprav DPS (čistá Cu, SnHASL, ENIG). Mechanizmy lomu boli skúmané pozorovaním materiálových charakteristík lomových plôch pomocou optického mikroskopu a skenovacieho elektrónového mikroskopu [9], [10].

II. Experimentálna časť

Spájkované spoje boli testované použitím spájkovacích guľôčok SAC305 s priemerom 760 µm. Každá testovaná vzorka obsahovala 100 plôšok s priemerom 660 µm. Pred spájkovaním bola každá vzorka chemicky očistená. Po čistiacom procese bolo na povrch DPS nanesené tavivo ROL0 CVP-390 pomocou šablónovej tlače. Guľôčky spájky boli ručne umiestnené na DPS. Spájkovanie bolo vykonané pomocou IR pece s preddefinovaným profilom optimalizovaným pre spájkové guličky SAC305 a použité tavivo. Následne boli vzorky testované mechanicky pomocou strihovej skúšky pomocou testera DAGE4000. Výsledky získané zo strihovej skúšky boli použité na výpočet energie potrebnej pre porušenie spoja, ktorá je určená veľkosťou plochy pod krivkou sily a posunutia. Tieto údaje poskytujú komplexné hodnotenie pevnosti spoja na základe maximálnej sily a dĺžky posunu počas strihovej skúšky.

Výška strihu bola 80 µm, ako je uvedené na Obr. 1, a rýchlosť zaťaženia bola 0,1 mm/s. Podľa normy JESD22-B117B je táto rýchlosť zaťaženia považovaná za nízku, čo je optimálne pre húževnaté typy porušenia. Počas testovania môžu nastať tri režimy porúch. V prvom režime môže dôjsť k poruche v objeme spájky nad vrstvou IMC. V druhom režime dochádza k poruche v objeme spájky aj v IMC. V treťom režime je zrejmá porucha priamo vo vrstve IMC.

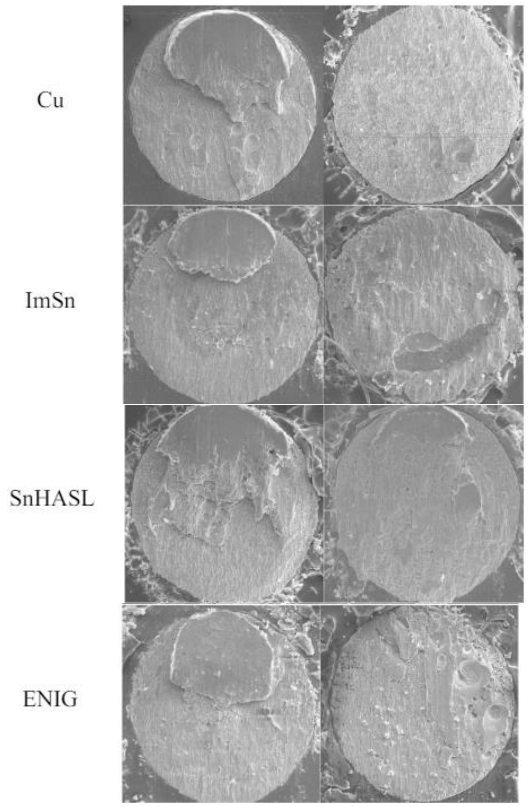


Obr. 1 Princíp mechanického namáhania v strihu

Ďalej boli vzorky skúmané pomocou skenovacieho elektrónového mikroskopu JEOL JSM-35 CF pri urýchľovacom napätí 25 kV. Keďže vzorky boli nevodivé, boli naprášené zlatom pomocou JEOL ION SPUTTER JFC-1100 pri napätí 1,2 kV po dobu 5 minút. Hlavným problémom bolo nastaviť mikroskop SEM tak, aby zachytil podrobné snímky s dôrazom na oblasti porúch.

III. Výsledky a diskusia

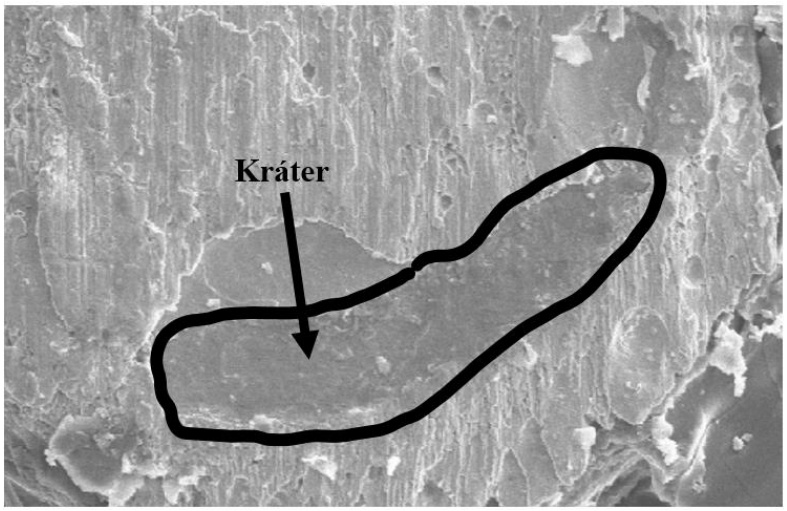
Po vykonaní strihovej skúšky boli vypočítané priemerné hodnoty maximálnej pevnosti v šmyku, posunutia a energie potrebnej pre porušenie spoja. Vzorky, ktoré najviac zodpovedali priemerným hodnotám vzhľadom na priebeh sily a posunutia a hodnoty týchto parametrov, boli vybrané pre skenovaciu elektronovú mikroskopiou (SEM) na obrázku 2.



Obr. 2 Snímky lomových plôch po jednom a piatich pretaveniach pre skúmané povrchové úpravy

Táto štúdia sa primárne zameriavala na fraktografickú morfológiu spájok po strihovej skúške po prvom a piatom cykle reflow. Numerické výsledky získané zo strihovej skúšky nie sú tak presné pri stanovení náchylnosti spájkovaných spojov ku krehkosti ako analýza fraktografickej morfológie.

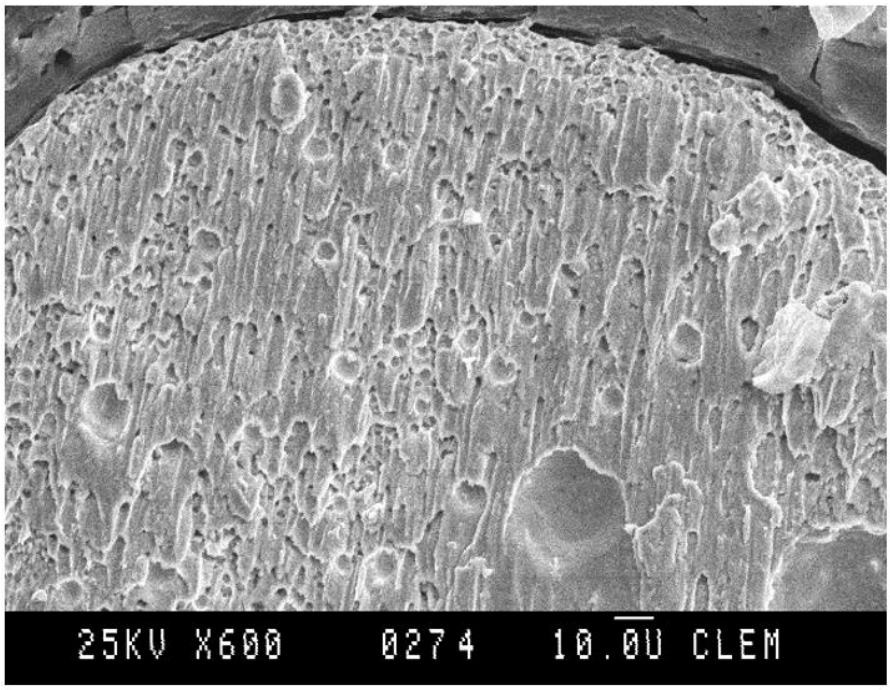
Z fraktografických obrázkov na obrázku 2 je viditeľné 3D rozloženie spájky na mieste lomu. Na lomovej ploche je v niektorých oblastiach výrazný zrnitý vzhľad, čo potvrdzuje prítomnosť IMC (intermetalických zlúčenín). Húževnaté krky okolo mikroskopických jamiek sú typické pre húževnato-krehký lom. Okrem toho je viditeľná prítomnosť húževnatej lomovej oblasti, charakterizovanej vláknitými, zvýšenými oblasťami na lomovej ploche. Vzhľad najvyšších bodov na lomovej ploche obsahuje len stopy po pohybe hrotu, bez príznakov zrnitej (krehkej) časti, typickej pre čisto krehký lom. Takýto vzhľad lomu bol potvrdený u vzoriek označených Cu, ImSn, SnHASL, ENIG. Pomer zrnitej k húževnatej lomovej časti sa líši v závislosti na type povrchovej úpravy a počtu pretavení, s výraznejšími rozdielmi pozorovanými po 5 pretaveniach. Pri bližšom skúmaní je viditeľný významný výskyt malých jamiek s krehkým typom porušenia. Tieto vzorky sú viac náchylné k krehkému lomu, aj keď ich lom sa dá charakterizovať ako kvázihúževnatý.



Obr. 3 Detailný snímok lomovej plochy spájkovaného spoja s povrchovou úpravou ImSn po piatich pretaveniach

Po piatom pretavení spájky vznikli veľké krátery, ktoré sú zodpovedné za krehké zlyhanie v týchto oblastiach. Pri spájkovaných spojoch s povrchovými úpravami Cu a ImSn došlo k lomom v IMC na väčších plochách (ako je znázornené na obrázku 3), avšak nie v takom rozsahu, aby tieto porušenia bolo možné identifikovať ako krehké alebo kvázi-krehké porušenie.

K ďalšiemu podobnému porušeniu došlo aj pre Cu. Podrobná snímka fraktografickej morfológie je na obrázku 4. V rámci lomovej oblasti sú viditeľné malé jamky medzi dlhými krkmi spájky. Tieto jamky vznikli lomom v oblasti IMC na dne týchto malých jamiek a pretože rýchlosť strihu bola príliš nízka na to, aby sa zvýraznili krehké porušenia spájkovaných spojov. Priemerné hodnoty energií potrebných pre porušenie spoja a ich rozptyly zo skúšok šmykového namáhania pre spájkované spoje s rôznymi povrchovými úpravami sú uvedené v tabuľke 1.



Obr. 4 Detailný snímok lomovej plochy spájkovaného spoja s povrchovou úpravou Cu po piatich pretaveniach

Najnižšia energia po prvom procese pretavenia je možné pozorovať práce pri vzorce Cu. Ako je znázornené na obrázku 4, lomová oblasť zahŕňa množstvo väčších jamiek vzniknutých lomením v rámci vrstvy IMC. Po piatom pretavení sa energia znížila avšak nie tak rapídne ako pri povrchovej úprave ImSn. Ako aj vyplýva z obrázku 3, pri povrchovej úprave ImSn došlo k zlyhaniu na omnoho väčšej oblasti vo vrstve IMC, než pri ostatných povrchových úpravách. Najstabilnejšími povrchovými úpravami z pohľadu energie potrebnej na porušenie spoja sa zdá bať povrchová úprava SnHASL a ENIG, nakoľko ani po piatich pretaveniach pri týchto spájkovaných spojoch nedošlo k výraznému poklesu energie a taktiež nedošlo k výraznejším krehkým porušeniam.

Tabuľka 1
Priemerná energia spájkovaných spojov po skúške v strihu

Povrchová úprava	Energia v strihu po prvom pretavení [mJ]	Energia v strihu po piatom pretavení [mJ]
Cu	2,65 ± 0,22	2,19 ± 0,23
ImSn	3,47 ± 0,12	1,97 ± 0,20
SnHASL	3,76 ± 0,02	3,38 ± 0,26
ENIG	3,68 ± 0,18	3,15 ± 0,10

IV. ZÁVER

Podrobne štúdium fraktografických morfológií a mechanických vlastností spájkovaných spojov po strihovej skúške odhalilo spoločný záver: húževnatosť spájkovaných spojov klesá naprieč rôznymi povrchovými úpravami.

Je potrebné podotknúť, že povrchová úprava ENIG obsahuje niklovú difúznu bariéru, ktorá chráni podkladovú meď od zrýchlenej difúzie do objemu spájky, čím je spoj menej náchylný na krehké porušenie v oblasti IMC. Povrchová úprava ENIG si udržiava stabilitu aj po viacnásobných reflow procesoch, ako sa očakávalo. Povrchová úprava SnHASL je taktiež dostatočne spoľahlivá aj po piatich procesoch pretavenia, avšak spôsob, akým je vyrobená, neumožňuje túto povrchovú úpravu použiť pri komponentoch s malými rozostupmi.

Analýza fraktografickej morfológie a skúšok smykového napätia ukazuje, že len povrchové úpravy SnHASL a ENIG preukazujú smykovú energiu presahujúcu 3 mJ po piatom procese pretavenia. Tento výsledok naznačuje húževnatosť spájkovaných spojov a naznačuje stabilitu týchto povrchových úprav aj po viacnásobných procesoch pretavenia. Povrchová úprava ImSn síce spĺňa túto podmienku po prvom pretavení, avšak po piatok pretavení došlo pri tejto povrchovej úprave k poklesu húževnatosti, čím sa zvýšilo riziko krehkého porušenia v oblasti IMC.

Analýza fraktografickej morfológie spájkovaných spojov poukazuje na dôležitosť analýzy húževnatosti v závislosti od typu povrchovej úpravy. Kombinácia výsledkov skúšok mechanickej pevnosti s výsledkami fraktografickej morfológie poskytuje komplexné porozumenie spoľahlivosti spájkovaných spojov.

Poďakovanie

Tento článok bol podporovaný projektom KEGA 011TUKE4/2023, projektom VEGA 1/0049/24, projektom VEGA 1/0278/24, projektom APVV 22-0289 a projektom ITMS 313011T557.

Literatúra

[1] T. Xu, X. Hu, J. Li, a Q. Li, "Fracture behavior and mechanical strength of sandwich structure solder joints with Cu–Ni(P) coating during thermal aging", *J. Mater. Sci. Mater. Electron.*, roč. 31, č. 5, s. 3876–3889, mar. 2020, doi: 10.1007/s10854-020-02933-7.

[2] J. Wang, H. Su, D. Mu, X. Kong, Y. Jin, a X. Shi, "Growth behavior of IMCs in Sn–1.0Ag–0.5Cu–xBi/Ni joints during isothermal aging", *J. Mater. Sci. Mater. Electron.*, roč. 32, č. 15, s. 20777–20792, aug. 2021, doi: 10.1007/s10854-021-06591-1.

[3] Y.-J. Seo a J.-W. Yoon, "Effects of solder volume and size on microstructures and mechanical properties of Sn-3.0Ag-0.5Cu solder joints", *J. Mater. Sci. Mater. Electron.*, roč. 33, č. 20, s. 16700–16709, júl. 2022, doi: 10.1007/s10854-022-08567-1.

[4] H. Qiu, X. Hu, S. Li, Y. Wan, a Q. Li, "Shear strength and fracture surface analysis of lead-free solder joints with high fraction of IMCs", *Vacuum*, roč. 180, s. 109611, okt. 2020, doi: 10.1016/j.vacuum.2020.109611.

[5] D.-H. Lee, M.-S. Jeong, a J.-W. Yoon, "Comparative study of interfacial reaction and bonding property of laser- and reflow-soldered Sn–Ag–Cu/Cu joints", *J. Mater. Sci. Mater. Electron.*, roč. 33, č. 10, s. 7983–7994, apr. 2022, doi: 10.1007/s10854-022-07948-w.

[6] T.-Y. Lee, J. Kim, S.-Y. Jun, B.-J. Ahn, D.-G. Han, a S. Yoo, "Interfacial microstructure and brittle fracture behavior of Sn-based solder on novel nickel-less DEG and EPIG surface finishes", *J. Mater. Sci. Mater. Electron.*, roč. 34, č. 2, s. 129, jan. 2023, doi: 10.1007/s10854-022-09510-0.

[7] S. Tahat, A. Alakayleh, M. E. A. Belhadi, A. Alahmar, A. Yaeger, a S. Hamasha, "Effects of Multiple Reflows on IMC and Shear Strength of Individual Solder Joints", v *2023 22nd IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, Orlando, FL, USA: IEEE, máj. 2023, s. 1–9. doi: 10.1109/ITherm55368.2023.10177561.

[8] X. Lu, L. Zhang, C. Chen, a X. Wang, "Microstructure and orientation evolution of β-Sn and interfacial Cu6Sn5 IMC grains in SAC105 solder joints modified by Si3N4 nanowires", *J. Mater. Res. Technol.*, roč. 26, s. 4723–4738, sep. 2023, doi: 10.1016/j.jmrt.2023.08.186.

[9] A. Pietrikova, D. Dzivy, O. Vrublenskaya, a M. A. Shikun, "Wetting of New SnAg Surface Finish", v *2022 45th International Spring Seminar on Electronics Technology (ISSE)*, Vienna, Austria: IEEE, máj. 2022, s. 1–6. doi: 10.1109/ISSE54558.2022.9812833.

[10] A. Pietrikova *et al.*, "Effect of non-standard SnAg surface finishes on properties of solder joints", *Appl. Surf. Sci. Adv.*, roč. 18, s. 100483, dec. 2023, doi: 10.1016/j.apsadv.2023.100483.

Určenie spätných vplyvov obnoviteľných zdrojov energie na distribučnú sústavu

[1]*Kristián ELIÁŠ,* [2]*Ľubomír BEŇA*

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

[1] kristian.elias@student.tuke.sk, [2] lubomir.bena@tuke.sk

***Abstrakt* — Boj proti zmene klímy, diverzifikácia energetických zdrojov či udržateľné hospodárenie so zdrojmi sú len jedny z mnohých faktorov, prečo v dnešnej dobe dochádza k prudkému rozvoju využívania obnoviteľných zdrojov energie (OZE). Aj preto dochádza v súčasnosti k neustálemu zvyšovaniu počtu zdrojov elektrickej energie využívajúcich OZE pripojených do distribučných sústav (DS) nízkeho napätia. Najrozšírenejšími zdrojmi elektrickej energie inštalovanými v DS nízkeho napätia využívajúcimi OZE sú fotovoltické elektrárne (FVE), ktorých výkony sa môžu pohybovať v jednotkách až desiatkach kW. Takáto decentralizovaná výroba však pre prevádzkovateľov distribučných sústav môže spôsobovať problémy týkajúce sa prevádzky DS.**

Tento príspevok sa zaoberá analýzou spätného vplyvu jednofázových a trojfázových FVE na DS nízkeho napätia a tiež návrhom možných opatrení na zníženie negatívnych účinkov FVE na DS, do ktorej sú pripojené.

***Kľúčové slová* — EA-PSM, distribučná sústava, obnoviteľné zdroje energie, kvalita elektrickej energie**

I. Úvod

Elektrická energia ako produkt musí spĺňať prísne kritériá kvality. Povinnosťou prevádzkovateľov distribučných sústav je preto zaistiť, aby elektrická energia v mieste pripojenia odberateľa mala požadovanú kvalitu. Pod pojmom kvalita elektrickej energie sa rozumie najmä kvalita napätia. V Európe sa kvalitou elektrickej energie zaoberá norma EN 50160 (slovenská verzia tejto normy je STN EN 50160). Touto normou sú definované, opísané a špecifikované hlavné charakteristiky napätia, ktoré sa nachádza na napájacích svorkách používateľov vo verejných nn, vn a vvn striedavých sieťach pri normálnych prevádzkových podmienkach (norma sa nevzťahuje na mimoriadne prevádzkové situácie ako napríklad likvidácia poruchy, výnimočné situácie a pod.). [1][2][3]

Pripojené zariadenia užívateľov DS sa však vyznačujú spätnými vplyvmi, ktorými ovplyvňujú kvalitu elektrickej energie v DS. Prevádzkovatelia distribučných sústav preto stanovili najvyššie prípustné spätné vplyvy zo strany užívateľa na DS.

Posudzovanie spätného vplyvu jednofázových a trojfázových FVE na DS nízkeho napätia sa teda vykonávalo na základe požiadaviek normy STN EN 50160: 2011, ale aj na základe požiadaviek prevádzkovateľa distribučnej sústavy.

II. Sledované parametre pre posudzovanie spätného vplyvu FVE na DS

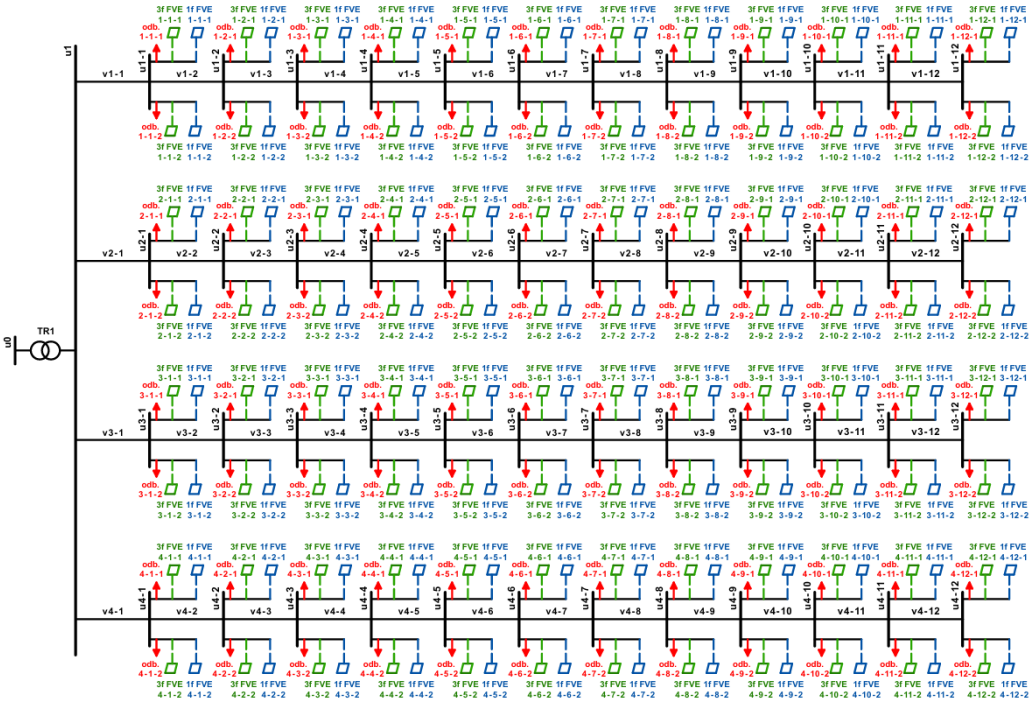
Limitné kritériá pre posudzovanie spätného vplyvu FVE na DS nízkeho napätia:

- veľkosť fázového napätia má byť v rozsahu $U_n \pm 10$ %, kde $U_n = 230$ V, tzn. v rozmedzí od 207 V do 253 V,
- hodnota faktora nesymetrie napájacieho napätia k', ktorá vyjadruje percentuálny pomer spätnej a súslednej zložky fázového napätia má byť v rozsahu 0 % až 2 %,
- zmena napätia spôsobená pripojením FVE nesmie v mieste pripojenia prekročiť ± 3 % v porovnaní s hodnotou napätia, ak FVE nie sú pripojené, t. j. $|\Delta u| \leq 3$ %,
- najvyššia úroveň napäťovej nesymetrie k_u, ktorá vznikne pripojením jedného zariadenia užívateľa do DS môže byť 0,7 %,
- limitné prúdové zaťaženie jednotlivých vedení odpovedá 80 % menovitého prúdu jednotlivých vedení,

- limitné prúdové zaťaženie primárneho a sekundárneho vinutia transformátora odpovedá 100 % menovitého primárneho a sekundárneho prúdu transformátora.

III. Model nízkonapäťovej distribučnej siete

Model nízkonapäťovej DS pozostáva z transformátora vn/nn a štyroch vedení vychádzajúcich z jedného uzla. Jednotlivé vedenia sú vyhotovené káblami rôznych typov (vedenia v1 a v2 káblom pre nadzemné vedenia a vedenia v3 a v4 káblom pre podzemné vedenia) a prierezov (viď Tabuľka 1). Každé vedenie obsahuje 12 uzlov, pričom v každom uzle sa nachádzajú dve odberné miesta. V každom odbernom mieste sa nachádza trojfázový odber o veľkosti 3,464 kW a účinníku cos φ = 1. Všeobecná topológia modelovanej siete s označením všetkých jej prvkov sa nachádza na Obr. 1, kde odbery sú označené červenou farbou, trojfázové FVE zelenou farbou a jednofázové FVE modrou farbou.



Obr. 1 Všeobecná topológia modelovanej siete

Tabuľka 1
Parametre vedení

vedenie	typ	R [Ω/km]	X [Ω/km]	I_n [A]	dĺžka [m]
v1	NFA2X 4x95	0,32	0,083	240	600
v2	NFA2X 4x120	0,25	0,082	280	600
v3	NAYY-J 4x150	0,206	0,08	254	600
v4	NAYY-J 4x240	0,129	0,08	332	600

IV. Modelované stavy

Celkovo bolo namodelovaných sedem stavov, z ktorých bol jeden referenčný (A0), kedy do DS neboli pripojené žiadne FVE.

Tabuľka 2
Prehľad modelovaných stavov

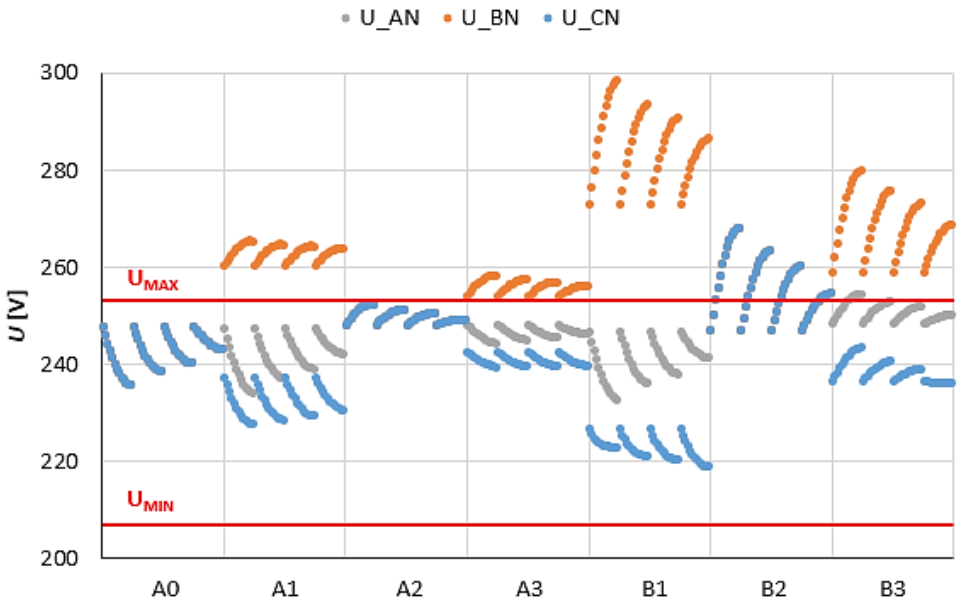
	Pripojenie iba 1-fázových FVE (do fázy A) s dodávkou 3,68 kW, cos φ_L = 0,97	Pripojenie iba 3-fázových FVE s dodávkou 11 kW, cos φ_L = 0,97	Pripojenie mixu 1-fázových FVE (3,68 kW do fázy A) a 3-fázových FVE (11 kW), cos φ_L = 0,97, pričom v jednom uzle je jedna 1-fázová FVE a jedna 3-fázová FVE
FVE v 1/2 odberných miest	A1	A2	A3
FVE vo všetkých odberných miestach	B1	B2	B3

Ostatných šesť stavov sa odlišovalo hustotou inštalácie FVE (buď v ½ uzlov alebo vo všetkých uzloch) alebo pripojenými FVE, tzn. buď boli do DS pripojené len jednofázové FVE, len trojfázové FVE alebo mix jednofázových a trojfázových FVE. Prehľad modelovaných stavov a ich označenia sa nachádzajú v Tabuľka 2.

V. SÚHRN DOSIAHNUTÝCH VÝSLEDKOV

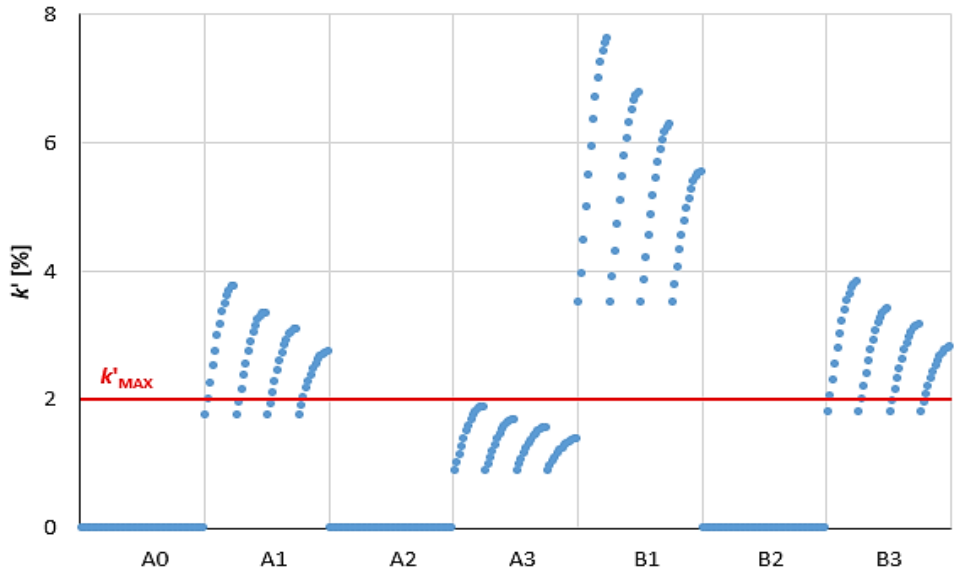
Na Obr. 2, Obr. 3 a Obr. 4 sú znázornené posudzované kvalitatívne parametre napätia vo všetkých uzloch siete pre jednotlivé modelované stavy. V grafoch sú vyznačené aj limitné hodnoty týchto kvalitatívnych parametrov, ktoré by sa pri normálnych prevádzkových podmienkach nemali prekročiť. Limitnými hodnotami sú:

- U_{MAX} = 253 V a U_{MIN} = 207 V,
- k'_{MAX} = 2 %,
- $+\Delta u_{MAX}$ = + 3 % a $-\Delta u_{MAX}$ = - 3 %.



Obr. 2 Súhrnné grafické znázornenie fázových napätí vo všetkých uzloch siete pre jednotlivé modelované stavy

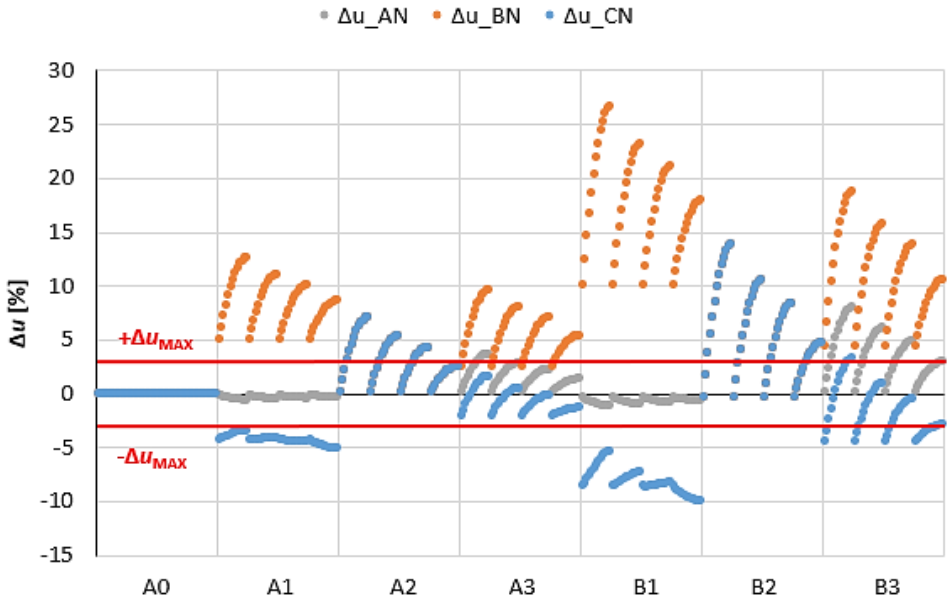
Z Obr. 2 je vidieť, že veľkosť fázového napätia je vo všetkých troch fázach vo všetkých uzloch na všetkých vedeniach v rámci dovoleného rozsahu len v prípade modelovaného stavu A2.



Obr. 3 Súhrnné grafické znázornenie faktora napäťovej nesymetrie vo všetkých uzloch siete pre jednotlivé modelované stavy

Z Obr. 3 je vidieť, že v stavoch, kedy sú v DS inštalované iba trojfázové FVE (stavy A2 a B2)

je faktor napäťovej nesymetrie vo všetkých uzloch DS nula. V ostatných stavoch, kedy sú do DS pripojené jednofázové FVE dochádza k vzniku nesymetrie, ktorá je v rámci dovolených hraníc vo všetkých uzloch na jednotlivých vedeniach iba v prípade modelovaného stavu A3.



Obr. 4 Súhrnné grafické znázornenie relatívnych zmien napätia vo všetkých uzloch siete pre jednotlivé modelované stavy

Z Obr. 4 je vidieť, že po pripojení FVE dochádza v niektorých resp. vo všetkých uzloch na všetkých štyroch vedeniach vo všetkých stavoch (okrem vedenia v4 v stave A2) k relatívnym zmenám napätia, či už v jednej, dvoch alebo aj všetkých troch fázach, ktoré prekračujú dovolené hranice.

Tabuľka 3
Posúdenie spätného vplyvu FVE na DS pre jednotlivé modelované stavy

stav	vedenie	veľkosť fázového napätia U_f = 230 V ± 10 %	faktor nesymetrie napätia k' ≤ 2 %	zmeny napätia po pripojení/odpojení všetkých FVE $\|\Delta u\|$ ≤ 3 %	prúdové zaťaženie vedenia do 80 % menovitej zaťažiteľnosti	prúdové zaťaženie transformátora do 100 % menovitej zaťažiteľnosti
A1	v1	✗	✗	✗	✗	✗
	v2	✗	✗	✗	✓	
	v3	✗	✗	✗	✓	
	v4	✗	✗	✗	✓	
A2	v1	✓	✓	✗	✓	✓
	v2	✓	✓	✗	✓	
	v3	✓	✓	✗	✓	
	v4	✓	✓	✓	✓	
A3	v1	✗	✓	✗	✓	✓
	v2	✗	✓	✗	✓	
	v3	✗	✓	✗	✓	
	v4	✗	✓	✗	✓	
B1	v1	✗	✗	✗	✗	✗
	v2	✗	✗	✗	✗	
	v3	✗	✗	✗	✗	
	v4	✗	✗	✗	✗	
B2	v1	✗	✓	✗	✗	✗
	v2	✗	✓	✗	✗	
	v3	✗	✓	✗	✗	
	v4	✗	✓	✗	✓	
B3	v1	✗	✗	✗	✗	✗
	v2	✗	✗	✗	✗	
	v3	✗	✗	✗	✗	
	v4	✗	✗	✗	✓	

Poznámka 1: ✗ – úroveň spätného vplyvu je mimo dovolených hraníc,
✓ – úroveň spätného vplyvu je v rámci dovolených hraníc.

Poznámka 2: Ak je posudzovaný parameter mimo dovoleného rozsahu aspoň v jednej fáze v ktoromkoľvek uzle alebo úseku na vedení, potom je v tabuľke ✗.

Z Tabuľka 3 vyplýva, že spätné vplyvy pripojených FVE na DS sú v dovolených medziach len v prípade vedenia v4 v modelovanom stave A2, kedy sú v ½ odberných miest pripojené dve trojfázové FVE, každá s dodávkou 11 kW a účinníkom cos φ_L = 0,97. Vo všetkých ostatných prípadoch dochádza k negatívnemu pôsobeniu pripojených FVE na DS v takej miere, že kvalitatívne parametre napätia v uzloch na jednotlivých vedeniach prekračujú limitné hodnoty alebo dochádza k prúdovému preťažovaniu vedení či transformátora vn/nn.

Tabuľka 4
Posúdenie dodržania limitnej úrovne napäťovej nesymetrie spôsobenej jedným zariadením užívateľa DS

vedenie	faktor nesymetrie napätia po pripojení jednej 1-fázovej FVE (3,68 kW, cos φ_L = 0,97) k_u ≤ 0,7 %
v1	✓
v2	✓
v3	✓
v4	✓

Z Tabuľka 4 vyplýva, že pripojením jedného zariadenia užívateľa DS (v tomto prípade jednofázovej FVE 3,68 kW, cos φ_L = 0,97) do ktoréhokoľvek uzla DS nedôjde k prekročeniu najvyššej dovolenej úrovne napäťovej nesymetrie spôsobenej užívateľom DS v danom uzle, ani v žiadnom inom uzle v rámci DS.

VI. Opatrenia na zníženie negatívneho vplyvu FVE na DS nízkeho napätia

Pre stavy B1 a B2, kedy je počet pripojených FVE do DS nízkeho napätia najvyšší, boli navrhnuté opatrenia, ktorých aplikácia má pozitívny vplyv na zníženie negatívneho spätného účinku FVE na DS.

Stav B1, kedy boli v DS inštalované iba jednofázové FVE (všetky boli pripojené do fázy A), sa vyznačoval vznikom nesymetrie, ktorá spôsobila, že veľkosť napätia v uzloch na vedeniach bola v jednotlivých fázach rozdielna. Okrem toho nesymetria spôsobovala aj tok prúdu v neutrálnom vodiči, ktorého veľkosť v niektorých častiach DS dokonca prekračovala najvyššiu dovolenú hodnotu. Opatrením pre obmedzenie vzniknutej nesymetrie v DS je, aby jednotlivé jednofázové FVE boli pripájané rovnomerne do všetkých troch fáz. Prijatím tohto opatrenia dôjde k zlepšeniu všetkých posudzovaných parametrov v uzloch na jednotlivých vedeniach DS. Ďalšími opatreniami, ktoré spolu v kombinácií so symetrizáciou pripájania jednofázových FVE priaznivo pôsobia na znižovanie negatívneho pôsobenia jednofázových FVE na DS sú zníženie účinníka FVE (do určitej hodnoty) či skrátenie dĺžky jednotlivých vedení. Bez symetrizácie pripájania jednofázových FVE s výkonom 3,68 kW a cos $\varphi_L = 0,97$ však nie je možné dosiahnuť splnenie požadovaných kritérií prijatím žiadneho iného opatrenia. Symetrické pripájanie jednofázových FVE je ale v praxi ťažko realizovateľné, pretože prevádzkovateľ distribučnej sústavy nemá možnosť ovplyvniť, do ktorej fázy užívateľ DS pripojí svoj jednofázový zdroj. Jediným opatrením, ktoré zaistí, aby spätný vplyv jednofázových FVE na DS inštalovaných vo všetkých uzloch modelovanej siete bol v dovolených medziach, pričom by bol zachovaný počet pripojených FVE a pripojenie jednotlivých FVE by nebolo symetrizované, je zníženie jednotkového výkonu jednofázových FVE z 3,68 kW na 0,4 kW (pri cos $\varphi_L = 0,97$).

Pri rovnakej hustote pripojenia, t. j. vo všetkých uzloch modelovanej siete, no inštalácií trojfázových FVE namiesto jednofázových FVE (stav B2) je maximálny možný výkon trojfázových FVE, pri ktorom nedochádza k spätnému vplyvu v nedovolenom rozsahu 2 kW, čo je päťkrát viac. V stave B2 je možné znížiť negatívny spätný vplyv FVE na DS na požadovanú úroveň aj skrátením dĺžok jednotlivých vedení.

VII. Záver

Z dosiahnutých výsledkov vyplýva, že z pohľadu prevádzkovateľa distribučnej sústavy je výhodnejšie, ak sú FVE pripájané do DS trojfázovo, pretože:

- pri podobných spätných vplyvoch je možné trojfázovými FVE dodávať x-násobne vyšší výkon v porovnaní s jednofázovými FVE,
- trojfázové FVE neprispievajú k nesymetrii v sieti,
- pri trojfázových FVE je možné nežiadúce spätné vplyvy dostať na prijateľnú úroveň jednoduchším spôsobom (pri jednofázových je potrebná symetrizácia pripojenia, ktorá je v praxi ťažko realizovateľná).

Zoznam použitej literatúry

[1] Kovács, Zoltán – Szathmáry, Peter: Spoľahlivosť a kvalita elektrickej energie. 302 s. ISBN 978 80-8073-864-8.
[2] Dolník, Bystrík – Pitoňák, Ján – Beňa, Ľubomír: Kvalita a spoľahlivosť dodávky elektrickej energie. 1. vyd. Košice: Technická univerzita v Košiciach, 2022. 216 s. ISBN 978-80-553-4107 1.
[3] STN EN 50160: 2011, Charakteristiky napätia elektrickej energie dodávanej z verejnej elektrickej siete.

Vplyv modifikácie počtu a dĺžky ukazovateľov na dcérske uzly na veľkosť binárnej reprezentácie SVDAG

[1]*Anastasiia TOVSTOLIS,* [2]*Branislav MADOŠ,* [3]*Heidar KHORSHIDIYEH*

[1,2,3] Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]anastasiia.tovstolis@student.tuke.sk, [2]branislav.mados@tuke.sk, [3]heidar.khorshidiyeh@tuke.sk

Abstrakt **– Článok je venovaný problematike reprezentácie troj-dimenzionálnych scén, konkrétnejšie ich geometrie, s využitím doménovo-špecifických hierarchických dátových štruktúr. Tie sú založené na hierarchickej dekompozícii scény na oktanty a v závislosti od toho, či je možná úsporná reprezentácia v prípade existencie identických podstromov sa klasifikujú na dátové štruktúry založené na sparse voxel octrees a sparse voxel directed acyclic graphs. V článku je predstavená modifikácia dátovej štruktúry sparse voxel directed acyclic graphs, ktorá oproti originálnej verzii disponuje viacerými dĺžkami ukazovateľov na dcérske uzly. To na jednej strane predlžuje binárnu reprezentáciu masky dcérskych uzlov na dvojnásobok, no na druhej strane umožňuje zaviesť optimalizáciu dĺžky ukazovateľov na dcérske uzly spolu s ich preusporiadaním podľa frekvencie s akou sú referencované. Článok teda predstavuje modifikovanú sparse voxel directed acyclic graphs dátovú štruktúru a porovnáva ju s originálnou verziou tejto dátovej štruktúry a s dátovou štruktúrou sparse voxel octrees, z hľadiska veľkosti ich binárnej reprezentácie pri ukladaní informácie o geometrii voxelizovanej 3D scény.**

Keywords **— troj-dimenzionálna voxelizovaná scéna, geometria voxelizovanej troj-dimenzionálnej scény, hierarchická dátová štruktúra, sparse voxel octrees, sparse voxel directed acyclic graphs, bezstratová kompresia dát**

I. Úvod

Nástup informačných a komunikačných technológií (IKT) zasiahol do vývoja mnohých odvetví ľudskej činnosti ako aj do súkromného života ľudí. V mnohých prípadoch uľahčil a zrýchlil ľuďom prácu, umožnil im efektívnejšiu komunikáciu a v niektorých prípadoch dokonca umožnil činnosti, ktoré by inak nebolo možné realizovať. Na druhej strane IKT produkuje enormné množstvá dát, ktorých objemy sa navyše z roka na rok zväčšujú. Tieto dáta sú produkované, prenášané, konzumované a uchovávané v takých objemoch, že ich obhospodarovávanie sa stáva samostatným problémom.

S cieľom umožniť ukladanie a prenos čoraz väčších objemov dát, vzniká požiadavka na ich úspornú reprezentáciu. Jedným z riešení je ich úsporné ukladanie s využitím ich kompresie. Tá môže byť delená vo všeobecnosti do viacerých kategórií, ako je napríklad delenie na využitie bezstratových kompresných algoritmov, ako je to v prípade textových informácií, ktoré umožňujú plne zrekonštruovať bez strát pôvodnú informáciu a na využitie stratových kompresných algoritmov, ktoré dosahujú vyššie kompresné pomery, avšak za cenu straty určitého objemu informácií, ktoré už nie je možné spätne zrekonštruovať, ako je to napríklad v prípade multimédií ako je obraz, zvuk alebo video.

Ďalšie delenie možno realizovať pri porovnaní časovej zložitosti prípadne absolútneho času kompresie a dekompresie dát, kde existujú algoritmy, ktorých časová zložitosť a/alebo absolútny čas realizácie výpočtov je zhodný pre kompresiu ako aj pre dekompresiu dát a naopak, algoritmy, kde sú tieto časy rozdielne, pričom spravidla je časová zložitosť vyššia a absolútny čas dlhší v prípade kompresie.

Napokon možno spomenúť klasifikáciu kompresných algoritmov na tie, ktoré sú pre všeobecné použitie a dokážu pomerne dobre komprimovať prakticky akékoľvek dáta a na tie, ktoré sú určené primárne pre určitú doménu, ako je napríklad zvuk, kde sa uplatňuje kompresia do formátu MP3 alebo celá plejáda kompresných algoritmov pre videosekvencie. Takéto algoritmy a s nimi súvisiace dátové štruktúry, ktoré ich využitím vznikajú, nazývame doménovo-špecifické.

Tento článok sa zaoberá problematikou reprezentácie geometrie troj-dimenzionálnych scén s využitím doménovo-špecifických hierarchických dátových štruktúr, ktoré vznikajú aplikovaním doménovo-špecifických algoritmov, a ktorých traverzovanie si vyžaduje špecifické postupy. Keďže výsledkom aplikovania týchto algoritmov je reprezentácia geometrie 3D scény prostredníctvom hierarchických dátových štruktúr úspornejším spôsobom, možno algoritmy, ktoré umožňujú ich vznik, považovať za kompresné a algoritmy pre ich traverzovanie zas za dekompresné. Súčasne, väčšina takýchto hierarchických dátových štruktúr zachytáva geometriu troj-dimenzionálnej scény bez straty akýchkoľvek detailov a teda ide o využitie bezstratovej kompresie dát.

Tento článok sa zaoberá preskúmaním vplyvu zavedenia viacerých dĺžok ukazovateľov na dcérske uzly v rámci Sparse Voxel Directed Acyclic Graphs (SVDAGs) na dĺžku jej binárnej reprezentácie.

II. SÚVISIACE PRÁCE

Dátová štruktúra Sparse Voxel Octrees (SVO), s implementáciou ukazovateľov na dcérske uzly, ako aj jej verzia Pointerless Sparse Voxel Octrees (PSVO) bez implementácie ukazovateľov na dcérske uzly, patria k základným doménovo-špecifickým HDS, ktoré slúžia na reprezentáciu geometrie 3D scén. V roku 2022 navrhol Madoš a kol. v [1] hierarchickú dátová štruktúru Clustered Sparse Voxel Octrees (CSVOs), ktorá je zložená až z troch rôznych typov uzlov, keď okrem bežných interných uzlov – Internal Node (INODE) a listových uzlov – Leaf Node (LNODE) implementuje aj tzv. prelistové uzly – Preleaf Node (PLNODE). CSVO disponuje aj zložitejším spôsobom používania ukazovateľov na dcérske uzly, a pre každý typ uzla má iný počet ukazovateľov rôznych dĺžok ich binárnej reprezentácie.

Z SVO, ktorá je stromom, bola odvodená dátová štruktúra Sparse Voxel Directed Acyclic Graphs (SVDAGs), ktorá je orientovaným acyklickým grafom - Directed Acyclic Graph (DAG). Je to tak preto, že umožňuje spájanie spoločných podstromov – Common Subtree Merge (CSM), ktoré umožňuje dva a viac identických podstromov danej dátovej štruktúry reprezentovať úsporne, takým spôsobom, že je v nej úplná reprezentácia len jednej inštancie takéhoto podstromu, ktorá je viacnásobne referencovaná. SVDAG dátová štruktúra môže mať zhodnú konštrukciu s SVO tak konceptuálne ako aj pri kódovaní na binárnej úrovni, jediným rozdielom je, že zatiaľ čo SVO neumožňuje, aby mal uzol viacero materských uzlov, SVDAG to umožňuje. Obidve dátové štruktúry pri ich zhodnej konštrukcii na binárnej úrovni môžu byť traverzované tým istým traverzovacím algoritmom a jeho programovou implementáciou. SVDAG navrhol v roku 2010 v [2] Kämpe a kol.

Dátovú štruktúru, ktorá je založená na cyklickom orientovanom grafe a teda umožňuje CSM ale súčasne neobsahuje vo svojej implementácii ukazovatele na dcérske uzly, predstavili v roku 2020 Vokorokos a kol. ako Pointerless Sparse Voxel Directed Acyclic Graphs (PSVDAGs) [3]. Táto dátová štruktúra je kompaktná, avšak nie je možné ju rýchlo traverzovať, preto, tak ako všetky tieto dátové štruktúry bez ukazovateľov na dcérske uzly, je určená skôr na archivačné účely. V roku 2021 Madoš a Ádám preto predstavili v [4] rýchly algoritmus pre konverziu PSVDAG dátovej štruktúry do SVDAG dátovej štruktúry.

V roku 2016 predstavil Villanueva a kol. v [5,6] svoju hierarchickú dátovú štruktúru s označením Symmetry-aware Sparse Voxel Directed Acyclic Graphs (SSVDAGs). Zatiaľ čo SVDAG je zarovnávaná na 32b, SSVDAG je zarovnávaná na 16b. Jej maska dcérskych uzlov má, na rozdiel od SVO alebo SVDAG až dvojnásobný počet bitov, teda 16b, pretože každý z ôsmich Header Tagov (HT) je 2b. Je to tak preto, že SSVDAG zavádza vo svojich interných uzloch až tri rôzne ukazovatele na dcérske uzly. Popri HT indikujúcom prázdny suboktant hodnotou „00“, existuje aj tzv. krátky ukazovateľ na dcérsky uzol, ktorý má iba 16b a je kódovaný pomocou HT , ktorý je nastavený na hodnotu „01“ a tzv. dlhý ukazovateľ, ktorý je 32b a ak je kódovaný ako HT nastavený na hodnotu „10“ je mu predradený 33b adresy nastavený na hodnotu 0 a ak je mu predradený 33b adresy nastavený na hodnotu 1, je kódovaný ako HT s hodnotou „11“. SSVDAG umožňuje CSM aj vtedy, ak je identickosť dosiahnutá len vtedy, ak sa použije transformácia zrkadlením a to v ktorejkoľvek z troch osí obrazu nezávisle (môže byť zrkadlený aj vo všetkých troch osiach naraz). Aby bolo možné indikovať, či bola použitá táto transformácia zrkadlením, sú 3b ukazovateľa na dcérske uzly obetované pre reprezentáciu zrkadlenia, každý bit pre jednu z troch osí obrazu. Aby sa zvýšila miera kompaktácie tejto dátovej štruktúry, listové uzly sú realizované ako mriežky voxelov s rozmerom 4^3 vox. Ukazovatele na dcérske uzly v prípade, ak ukazujú na interné dcérske uzly, adresujú 16b slová, ak ukazujú na listové uzly, adresujú až 64b slová. Podobne ako PSVDAG aj SSVDAG využíva kompaktáciu založenú na frekvencii referencovania uzlov – Frequency Based Compaction (FBC).

III. MODIFIKOVANÁ SVDAG

Medzi hierarchické dátové štruktúry, ktorým je v tomto článku venovaná pozornosť, patrí Sparse Voxel Octrees (SVO). Je to v súlade s názvom tejto dátovej štruktúry strom, to znamená, že dcérsky uzol môže mať vždy iba jeden materský uzol. Nie je teda možné realizovať spájanie spoločných podstromov (CSM). Naopak, jeden uzol môže mať od 1 do 8 dcérskych uzlov. Ak je reprezentovaná geometria s rozmerom N^3, musí pre N platiť, že sa dá zapísať ako 2^m, kde m je prirodzené číslo. Potom dátová štruktúra má uzly rozdelené do m vrstiev, pričom tie sú označené číslami z intervalu <0;m-1>. Vo vrstve 0 je uložený len jeden uzol, ktorý je označený ako koreňový. Uzly vo vrstvách 0 až m-2 sú označené ako interné – Internal Nodes (INODEs). Uzly vo vrstve m-1 sú označené ako listové uzly – leaf Nodes (LNODEs).

Interné uzly sú zložené z masky dcérskych uzlov (CHNM), ktorá je zložená z 8 Header Tags (HTs), pričom každý má veľkosť 1b. Ak je tento bit nastavený na hodnotu 0, neexistuje k nemu prislúchajúci dcérsky uzol. Ak je tento bit nastavený na hodnotu 1, dcérsky uzol exituje. Po CHNM je konkatenované pole ukazovateľov na dcérske uzly. Materský uzol môže mať od 1 do 8 aktívnych dcérskych uzlov, preto má aj pole ukazovateľov variabilnú dĺžku. Dĺžku ukazovateľov možno nastaviť variabilne, v tomto článku sú použité 32b ukazovatele.

Hierarchická dátová štruktúra SVDAG má zhodnú reprezentáciu pomocou Backus-Naurovej formy, a jej konštrukcia je aj na binárnej úrovni zhodná s binárnou reprezentáciou SVO HDS. Jediným, no veľmi podstatným rozdielom je, že konkrétny uzol, či už interný alebo listový, môže mať viac ako jeden materský uzol. Inými slovami, dcérske uzly môžu byť referencované viacnásobne z rôznych materských uzlov. Preto nie je možné terminologicky nazývať túto dátovú štruktúru stromom a je nazývaná ako orientovaný acyklický graf – Directed Acyclic Graph (DAG). Vďaka tejto úprave možno výskyt viacerých, plne rozvinutých podstromov, ktoré sú navzájom zhodné, nahradiť výskytom iba jedného takéhoto podstromu, ktorý je viacnásobne referencovaný.

Na rozdiel od SVO, kde je možné veľmi jednoducho pripraviť verziu tejto dátovej štruktúry bez ukazovateľov na dcérske uzly – Pointerless Sparse Voxel Octrees (PSVOs), nie je možné jednoduchým spôsobom vynechať z konštrukcie SVDAG ukazovatele na dcérske uzly a tak zostaviť Pointerless Sparse Voxel Directed Acyclic Graphs (PSVDAGs). Táto dátová štruktúra exituje, ale musela vyriešiť mechanizmus spájania zhodných podstromov sofistikovanejším spôsobom. Na druhej strane, spájanie spoločných podstromov v SVDAG je tak efektívne, že nielen, že dokáže uložiť dáta kompaktnejšie v porovnaní s SVO, ale dokonca aj v porovnaní s PSVO.

Zatiaľ čo SVDAG má iba jediný druh ukazovateľov na dcérske uzly, ktorých dĺžka je v našom článku 32b, prináša v tomto článku navrhnutá modifikovaná SVDAG, pre účely tohto článku označená ako MSVDAG, až tri rôzne dĺžky binárnej reprezentácie ukazovateľov na dcérske uzly. Super krátke ukazovatele – Super Short Pointer (SSP) majú dĺžku binárnej reprezentácie 8b, krátke ukazovatele – Short Pointer (SP) majú dĺžku binárnej reprezentácie 16b a dlhé ukazovatele - Long Pointer (LP) majú dĺžku binárnej reprezentácie až 32b. V súlade s tým bola rozšírená dĺžka CHNM na 16b, pretože každý HT je uložený na 2b. Ak je HT nastavený na hodnotu „00“, indikuje neaktívny suboktant, ku ktorému teda neexistuje aktívny dcérsky uzol. Ak je HT nastavený na hodnoty „01“, „10“ a „11“, existujú dcérske uzly daného materského uzla a dĺžky ukazovateľov na ne sú potom 8b pre HT = 01, 16b pre HT=10 a 32b pre HT=11.

```
SVO ::= (n)<NODE>
NODE ::= <INODE>|<LNODE>
INODE ::= <CHNM>(p)<BIT><PTS>
LNODE ::= <CHNM>(q)<BIT>
CHNM ::= (8)<HT>
PTS ::= (1)*(8)<PT>
PT ::= (r)<BIT>
HT ::= "0" | "1"
BIT ::= “0” | “1”
```

a)

```
SVDAG ::= (n)<NODE>
NODE ::= <INODE>|<LNODE>
INODE ::= <CHNM>(p)<BIT><PTS>
LNODE ::= <CHNM>(q)<BIT>
CHNM ::= (8)<HT>
PTS ::= (1)*(8)<PT>
PT ::= (r)<BIT>
HT ::= "0" | "1"
BIT ::= “0” | “1”
```

b)

```
MSVDAG ::= (n)<NODE>
NODE ::= <INODE> | <LNODE>
INODE ::= <ICHNM><PTS>
LNODE ::= <LCHNM>
ICHNM ::= (8)<LHT>
LCHNM ::= (8)<HT>
PTS ::= (1)*(8)<PT>
PT ::= (8)<BIT> | (16)<BIT> | (32)<BIT>
LHT ::= "00" | "01" | "10" | "11"
HT ::= "0" | "1"
BIT ::= “0” | “1”
```

c)

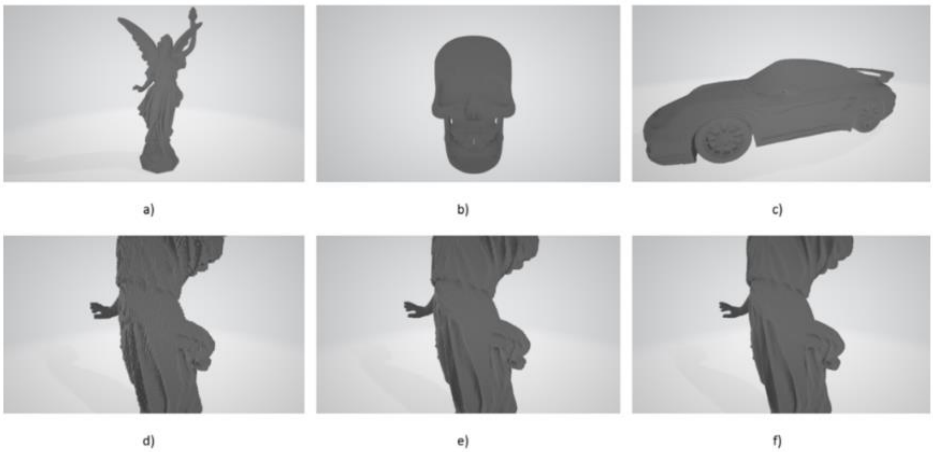
kde platí nasledovné:
(SYM) - povinný neterminálny symbol SYM ,
”sym” - terminálny symbol SYM,
(n)(SYM) - symbol SYM uvedený n-krát,
(n) * (m)(SYM) - symbol SYM, uvedený n až m-krát,
| - alternatíva,
p určuje počet rezervovaných bitov pripojených k CHNM v INODE
q určuje počet rezervovaných bitov pripojených k CHNM v LNODE
r určuje počet bitov ukazovateľa na CHNM

Obr. 1 Backus-Naurova forma pre a) SVO HDS, b) SVDAG HDS a c) MSVDAG HDS

IV. VÝSLEDKY TESTOV

Realizované testy boli zamerané na porovnanie veľkosti binárnej reprezentácie troch vyššie spomínaných HDS, teda SVO, SVDAG a MSVDAG, pri ukladaní geometrie tej istej voxelizovanej trojdimenzionálnej scény, s predpokladom, že najlepšie výsledky budú dosiahnuté pri použití MSVDAG HDS. Testy boli realizované na počítači s konfiguráciou Intel(R) Core(TM) i5-3470 CPU @ 3.20 GHz a 8 GB RAM, s operačným systémom Debian Linux verzie 4.19.0-6 a gcc verzie 8.3.0.

Na testovacie účely boli použité modely Angel Lucy, Porsche a Skull, ktoré boli pôvodne reprezentované ako polygonálne povrchové modely, zložené z trojuholníkov a uložené v dátovom formáte WaveFront OBJ (ich vizualizácie možno vidieť na Obr. 2). Následne boli zasadené do scén a tie boli voxelizované do celkovo štyroch rozlíšení, 128^3, 256^3, 512^3 a 1024^3. Vzniklo teda 12 testovacích scén.



Obr. 2 Vizualizácia voxelizovaných scén použitých na účely testovania: a) „Angel Lucy" pri 512^3; b) „Skull" pri 512^3; c) „Porsche" pri 512^3; d) „Angel Lucy" pri 256^3; e) „Angel Lucy" pri 512^3; f) „Angel Lucy" pri 1024^3.

Následne bola ich geometria uložená vo formáte SVO, SVDAG a MSVDAG a boli empiricky zisťované veľkosti týchto dátových štruktúr v počte bajtov, ktoré zaberajú na sekundárnom úložnom zariadení ako aj v počte uzlov, ktoré tieto dátové štruktúry tvoria a to pre všetky modely a pre všetky ich rozlíšenia. Získané výsledky možno vidieť v Tabuľke 1, 2 a 3.

Tabuľka 1 Výsledky testov pre model Angel Lucy voxelizovaný do rozlíšenia 128^3 a SVO, SVDAG a MSVDAG HDS

	Angel Lucy 128^3					
	SVO		SVDAG		MSVDAG	
Úroveň	Uzly	Veľkosť [B]	Uzly	Veľkosť [B]	Uzly	Veľkosť [B]
13	5 291	21 164	163	652	163	163
12	1 193	25 936	1 054	24 432	1 054	7 162
11	267	5 840	267	5 840	267	2 717
10	62	1 316	62	1 316	62	643
9	16	312	16	312	16	140
8	3	76	3	76	3	22
7	1	16	1	16	1	5
6	1	8	1	8	1	3
5	1	8	1	8	1	3
4	1	8	1	8	1	3
3	1	8	1	8	1	3
2	1	8	1	8	1	3
1	1	8	1	8	1	3
0	1	8	1	8	1	3

Tabuľka 2 Výsledky testov pre model Porsche voxelizovaný do rozlíšenia 128^3 a SVO, SVDAG a MSVDAG HDS

Porsche 128^3						
	SVO		SVDAG		MSVDAG	
Úroveň	Uzly	Veľkosť [B]	Uzly	Veľkosť [B]	Uzly	Veľkosť [B]
13	12 062	48 248	170	680	170	170
12	2 379	57 764	1 713	44 088	1 713	12 735
11	421	11 200	411	10 992	411	5 035
10	73	1 976	73	1 976	73	959
9	12	340	12	340	12	155
8	2	56	2	56	2	16
7	1	12	1	12	1	4
6	1	8	1	8	1	3
5	1	8	1	8	1	3
4	1	8	1	8	1	3
3	1	8	1	8	1	3
2	1	8	1	8	1	3
1	1	8	1	8	1	3
0	1	8	1	8	1	3

Tabuľka 3 Výsledky testov pre model Skull voxelizovaný do rozlíšenia 128^3 a SVO, SVDAG a MSVDAG HDS

Skull 128^3						
	SVO		SVDAG		MSVDAG	
Úroveň	Uzly	Veľkosť [B]	Uzly	Veľkosť [B]	Uzly	Veľkosť [B]
13	18 170	72 680	132	528	132	132
12	4 343	90 052	3 278	74 876	3 278	21 997
11	1 004	21 388	974	21 056	974	9 841
10	235	4 956	235	4 956	235	2 418
9	47	1 128	47	1 128	47	549
8	8	220	8	220	8	92
7	1	36	1	36	1	10
6	1	8	1	8	1	3
5	1	8	1	8	1	3
4	1	8	1	8	1	3
3	1	8	1	8	1	3
2	1	8	1	8	1	3
1	1	8	1	8	1	3
0	1	8	1	8	1	3

Ako je možné vidieť z tabuľky výsledkov, v prípade počtu uzlov dochádza k výraznej redukcii ich počtu vo vrstvách smerom k listovej vrstve uzlov, pričom k tejto redukcii dochádza pri prechode z SVO na SVDAG reprezentáciu. Medzi SVDAG a MSVDAG v tomto ohľade nie je žiadny rozdiel, pretože počet uzlov sa medzi týmito dvomi dátovými štruktúrami nemení. Ak sa do úvahy berie veľkosť binárnej reprezentácie jednotlivých dátových štruktúr, dochádza k postupnej redukcii veľkosti binárnej reprezentácie jednotlivých vrstiev uzlov v smere od SVO k SVDAG a k MSVDAG, pričom toto sa deje vo výraznejšej miere najmä pri listovej vrstve a k nej priľahlým vrstvám hierarchickej dátovej štruktúry.

Tabuľka 4 Veľkosti binárnych reprezentácií jednotlivých vrstiev uzlov pre model Angel Lucy 1024^3 v SVO, SVDAG a MSVDAG HDS a kompresný pomer medzi SVDAG a MSVDAG

Angel Lucy 1024^3				
	SVO	SVDAG	MSVDAG	
Úroveň	Veľkosť [B]	Veľkosť [B]	Veľkosť [B]	Kompresia
13	1 466 324	812	203	4,00
12	1 832 392	994 608	288 018	3,45
11	455 992	423 760	257 276	1,65
10	111 088	107 840	84 211	1,28
9	25 936	25 712	16 156	1,59
8	5 840	5 840	2 891	2,02
7	1 316	1 316	643	2,05
6	312	312	140	2,23
5	76	76	22	3,45
4	16	16	5	3,20
3	8	8	3	2,67
2	8	8	3	2,67
1	8	8	3	2,67
0	8	8	3	2,67

Tabuľka 5 Veľkosti binárnych reprezentácií jednotlivých vrstiev uzlov pre model Porsche 1024^3 v SVO, SVDAG a MSVDAG HDS a kompresný pomer medzi SVDAG a MSVDAG

Porsche 1024^3				
	SVO	SVDAG	MSVDAG	
Úroveň	Veľkosť [B]	Veľkosť [B]	Veľkosť [B]	Kompresia
13	3 876 452	976	244	4,00
12	4 808 612	1 512 708	437 365	3,46
11	1 148 972	869 544	503 138	1,73
10	265 060	238 856	183 954	1,30
9	57 764	56 640	43 171	1,31
8	11 200	11 200	5 536	2,02
7	1 976	1 976	973	2,03
6	340	340	155	2,19
5	56	56	16	3,50
4	12	12	4	3,00
3	8	8	3	2,67
2	8	8	3	2,67
1	8	8	3	2,67
0	8	8	3	2,67

Tabuľka 6 Veľkosti binárnych reprezentácií jednotlivých vrstiev uzlov pre model Skull 1024^3 v SVO, SVDAG a MSVDAG HDS a kompresný pomer medzi SVDAG a MSVDAG

Skull 1024^3				
	SVO	SVDAG	MSVDAG	
Úroveň	Veľkosť [B]	Veľkosť [B]	Veľkosť [B]	Kompresia
13	4 768 176	916	229	4,00
12	5 963 556	1 960 888	565 204	3,47
11	1 491 792	1 375 084	830 495	1,66
10	369 092	363 836	294 519	1,24
9	90 052	89 804	73 809	1,22
8	21 388	21 388	12 341	1,73
7	4 956	4 956	2 465	2,01
6	1 128	1 128	549	2,05
5	220	220	92	2,39
4	36	36	10	3,60
3	8	8	3	2,67
2	8	8	3	2,67
1	8	8	3	2,67
0	8	8	3	2,67

Tabuľky 4, 5 a 6 sú pohľadom na dosiahnutú efektivitu reprezentácie geometrie voxelizovaných 3D scén pri využití MSVDAG HDS, keď v stĺpci kompresia možno vidieť koľko násobne menšia binárna reprezentácia týchto vrstiev vznikla pri aplikovaní MSVDAG HDS oproti SVDAG HDS. Možno vidieť, že bolo dosiahnuté zmenšenie binárnej reprezentácie vrstvy od 1,22 pri vrstve 4 modelu Skull 1024^3 až po 4,00 pri vrstve uzlov s poradovým číslom 13, teda listovej vrstve všetkých troch modelov voxelizovaných do rozlíšenia 1024^3 .

V. Záver

Článok sa zaoberal vplyvom zvýšenia počtu rôznych dĺžok ukazovateľov na dcérske uzly v SVDAG HDS na kompresný pomer pri ukladaní geometrie voxelizovaných troj-dimenzionálnych scén. Empiricky bolo zistené, že v rade SVO, SVDAG a MSVDAG stúpa efektivita ukladania tejto geometrie do HDS a teda najvýraznejšia kompresia je dosahovaná pri používaní MSVDAG HDS.

Poďakovanie

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

Referencie

[1] Madoš, B.; Chovancová, E.; Chovanec, M.; Ádám, N. CSVO: Clustered Sparse Voxel Octrees—A Hierarchical Data Structure for Geometry Representation of Voxelized 3D Scenes. Symmetry 2022, 14, 2114. https://doi.org/10.3390/sym14102114

[2] Kämpe, V.; Sintorn, E.; Assarson, U. High Resolution Sparse Voxel DAGs. ACM Transactions on Graphics 2013, 32, pp. 1–13. ISSN 0730-0301. https://doi.org/10.1145/2461912.2462024

[3] Vokorokos, L.; Madoš, B.; Bilanová, Z. PSVDAG: Compact Voxelized Representation of 3D Scenes Using Pointerless Sparse Voxel Directed Acyclic Graphs. Computing and Informatics, 2020, 39, pp. 587-616. ISSN 1335-9150 (print); 2585-8807 https://doi.org/10.31577/cai_2020_3_587

[4] Madoš, B.; Ádám, N. Transforming Hierarchical Data Structures—A PSVDAG—SVDAG Conversion Algorithm. Acta Polytechnica Hungarica, 2021, 18, pp.47-66. ISSN 1785-8860. https://doi.org/10.12700/APH.18.8.2021.8.3

[5] Villanueva, A.J.; Marton, F.; Gobetti, E. Symmetry-aware Sparse Voxel DAGs. In Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D '16), Redmond, WA, USA, 27–28 February, 2016; pp. 7–14. ISBN 978-1-4503-4043-4. https://doi.org/10.1145/2856400.2856420

[6] Villanueva, A.J.; Marton, F.; Gobetti, E. Symmetry-aware Sparse Voxel DAGs (SSVDAGs) for compression-domain tracing of high-resolution geometric scene. J. Comput. Graph. Tech. (JCGT), 2017, 6, pp. 30. http://jcgt.org/ published/0006/02/01

Quantitative and qualitative properties of LLM-based data analysis models in terms of their usability

[1]*Oliver LOHAJ,* [2]*Ján PARALIČ,* [3]*Daniel ONOFREJ*
[1,2,3] Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]oliver.lohaj@tuke.sk, [2]jan.paralic@tuke.sk, [3]daniel.onofrej@student.tuke.sk

***Abstract* — The increasing integration of Large Language Models (LLMs) into various applications necessitates a thorough understanding of their usability. This article researches quantitative and qualitative properties of LLMs, specifically focusing on the widely used Gemini and GPT models, to assess their effectiveness in generating human-quality text and their overall usability for real-world tasks. Through a series of experiments involving user surveys, chatbot interactions, and quantitative metric analysis, the research aims to identify the strengths and weaknesses of each model and provide valuable insights for improving LLM usability.**

***Keywords* — Large Language Models, LLM, quantitative properties, qualitative properties, usability**

I. Introduction

The rapid advancement of LLMs has opened doors to a plethora of possibilities in human-computer interaction and natural language processing. Chatbots, virtual assistants, and other LLM-powered applications are becoming increasingly prevalent, highlighting the critical need to ensure their usability for a wide range of users. This research delves into the core aspects of LLM usability, exploring both qualitative and quantitative properties to assess how effectively these models can generate text that is not only accurate but also relevant, informative, and engaging for users.

II. Large Language Models

LLMs are sophisticated AI models trained on vast amounts of text data to learn the intricacies of human language. They excel at tasks like text generation, translation, summarization, and question answering, making them valuable tools in various domains [1, 3, 6].

Transformer Architecture: The Backbone of LLMs: The Transformer architecture, with its innovative attention mechanisms, lies at the heart of LLMs [2]. This architecture allows models to effectively process and generate text by capturing the context and relationships between words, leading to more coherent and human-like outputs.

Tokenization: Breaking Down Language: Tokenization is a fundamental step in LLM processing, where text is divided into smaller units called tokens [3]. Different methods like BPE [4], WordPiece [5], and SentencePiece offer varying approaches to tokenization, influencing the model's understanding and generation of text.

Evolution of LLMs: From Early Models to Powerhouses: The field of LLMs has witnessed remarkable progress, with models like GPT-3 [7] and PaLM [8] showcasing groundbreaking capabilities in language understanding and generation. The evolution of LLMs highlights the increasing complexity and sophistication of these models, paving the way for more diverse and powerful applications [6].

A. Properties of LLMs

1) Qualitative Properties:

These properties delve into the subjective aspects of LLM outputs, focusing on how users perceive and interact with the generated text [9, 10, 11]. Key qualitative properties include:

- Thematic Consistency: The ability of the model to maintain relevance and coherence to a specific topic or conversation throughout interactions.

- Context Understanding and Generation: The effectiveness of the model in comprehending the context of a query and generating responses that are relevant and consistent with that context.
- Flexibility in Language Use: The model's adaptability to different styles, registers, and genres of language to cater to specific user needs and preferences.
- Tone and Style of Communication: The ability of the model to adjust its tone (e.g., formal, informal, friendly, authoritative) and style to match the input or user requirements.

2) Quantitative Properties:

These properties involve measurable aspects of LLM performance that can be objectively evaluated using specific metrics [10, 15]:

- Accuracy: The degree to which the model's responses are correct and relevant to the given query or reference text.
- Perplexity: A measure of how well the model predicts a sample, with lower perplexity indicating better predictive ability.
- BLEU Score: A metric for evaluating the quality of machine translation and text coherence by comparing the generated text with reference translations.
- ROUGE Score: A set of metrics used to assess the quality of summaries or translations by comparing them with reference summaries.
- F1 Score: The harmonic mean of precision and recall, providing a balanced evaluation of the model's performance.

We decided to include the top_p and temperature in part dedicated to quantitative properties. While adjusting top_p and temperature can influence the quality and style of the generated text, they are not metrics that measure the performance of the model. Instead, they are tools used to tune the model's output based on desired characteristics, which in turn can affect the quantitative properties.

- Top_p (Nucleus Sampling Threshold): This hyperparameter controls the diversity of the generated text by determining the subset of most likely words considered for the next token [19, 20].
- Temperature: This parameter influences the randomness of word selection within the chosen subset, with higher values leading to more creative but potentially less factual outputs [19, 20].

B. Usability

Defining Usability in the Context of LLMs: Usability, as defined by the ISO 9241-11 standard, focuses on how effectively users can interact with a system to achieve their goals [12, [13], 22]. For LLMs, this translates to the ease with which users can understand and utilize the generated text outputs, the efficiency of the interaction, and the overall satisfaction with the results.

Exploring LLM usability research articles were reviewed to gain insights into the current state of LLM usability evaluation [11, 14, 16]. These studies highlighted the importance of developing comprehensive frameworks and methods for assessing LLM usability, considering both qualitative and quantitative aspects.

C. Metrics for evaluation of usability

1) Qualitative Metrics:

- User Satisfaction: Overall satisfaction with the chatbot experience and the quality of generated text.
- Grammatical Correctness: How the model's outputs adhere to grammatical rules and avoid errors.
- Understandability and Informativeness: The clarity and informativeness of the generated text, ensuring users can easily comprehend the information.
- Relevance of Responses: How the model's outputs directly address the user's questions or requests.
- Simplicity of Responses: The level of simplicity and conciseness in the language used, avoiding unnecessary complexity.
- Trustworthiness of Responses: How the users perceive the information provided by the model as reliable and credible.

2) Quantitative Metrics:

- Robustness to Typos: Assessing the model's ability to handle typos and errors in the input text without compromising the accuracy and relevance of responses.
- Handling Question Variations: Evaluating the model's capability to understand and respond to different phrasings or variations of the same question.

- Impact of "top_p" and "temperature": Measuring how these hyperparameters affect the quality of generated text using similarity scores and BERT score (If we apply top_p and temperature used to tune the model's output based on desired characteristics, in return it affects the quantitative properties). BERT score is a metric for evaluating the quality of text generated by language models by comparing it to reference text using BERT embeddings. It measures the similarity of tokens between the generated and reference text based on contextual embeddings, providing a more nuanced assessment than traditional metrics like BLEU [23].

D. Testing methods

1) Qualitative Testing:

- User Surveys: Online surveys with Likert scale questions, ranging from 1 (strongly disagree) to 5 (strongly agree), were used to gather feedback on user satisfaction and preferences regarding various aspects of the chatbot interaction and the generated text.
- Chatbot Interaction: A chatbot web app was developed and deployed to allow users to interact directly with both GPT and Gemini models, providing a platform for observing user behavior and gathering qualitative data on their experience.

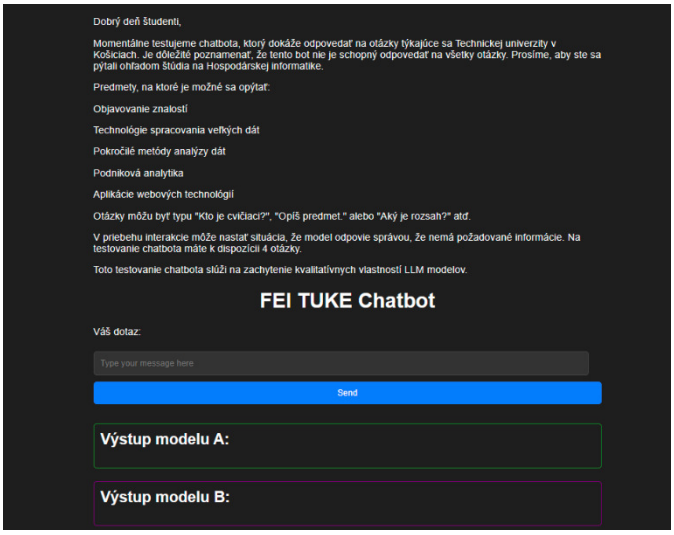


Fig. 1 Chatbot preview

2) Quantitative Testing:

- Python Scripts and Metrics: Python scripts were developed to automate the testing process and calculate quantitative metrics like similarity scores and BERT score [18]. This allowed for objective comparisons between GPT and Gemini models under different conditions.
- Controlled Experiments: Experiments were designed to evaluate the impact of "top_p", "temperature", and typos on the model outputs, comparing them to reference texts to assess accuracy, relevance, and overall quality.

E. Dataset for models

Dataset for GPT: A structured question-answer format dataset was created specifically for the GPT model using GPT-4. This involved generating potential questions that students might ask about the "Business Informatics" study program and extracting corresponding answers from the scraped web pages of the Faculty of Electrical Engineering and Informatics, particularly focusing on the Department of Cybernetics and Artificial Intelligence and its study program "Business Informatics" [16, [17]]. This approach aimed to provide the GPT model with a well-organized dataset to enhance its ability to generate accurate and relevant responses. The structured format ensured that the model could easily learn the patterns of questions and their appropriate answers, thereby improving its performance in providing precise information. We have been inspired by article [21].

Dataset for Gemini: In contrast, the dataset for the Gemini model utilized the raw scraped text directly without restructuring it into a question-answer format. This method allowed the Gemini model to leverage its advanced capabilities for understanding and generating text from a broader and more diverse knowledge base. By processing the unstructured data, the Gemini model could access a wider range of information, which enhanced its ability to understand context and generate more comprehensive and nuanced responses. This approach took advantage of the Gemini model's flexibility and adaptability, enabling it to provide richer and more contextually relevant answers based on the vast amount of raw data it was trained on.

F. Practical implementation

To facilitate interactions and usability testing of the GPT and Gemini models, a user-friendly chatbot interface was developed using the Flask framework. This interface was designed to allow students to ask questions related to the "Business Informatics" program, enabling them to interact seamlessly with both models. The chatbot served multiple purposes:

- Usability Testing: It provided a practical platform for evaluating the usability of each model in a realistic academic setting. By simulating real-world interactions, the chatbot allowed us to gather valuable feedback on how effectively each model could handle student queries.
- User Interaction: The intuitive design ensured that students could easily engage with the chatbot, ask their questions, and receive responses without any technical difficulties. This ease of use was crucial for obtaining genuine insights into the user experience.
- Performance Evaluation: The chatbot interface enabled the comparison of responses generated by the GPT and Gemini models, helping to identify their respective strengths and weaknesses in providing accurate, informative, and contextually appropriate answers.
- Data Collection: By tracking the interactions, we could collect data on various aspects of usability, such as response accuracy, user satisfaction, and the ability of each model to handle diverse and complex queries.

III. Evaluation process and gathering insights from users and metrics

Qualitative Evaluation: 22 students from the "Business Informatics" program participated in the testing phase, interacting with the live chatbot and providing feedback through Likert scale surveys. Their responses offered valuable insights into their satisfaction with each model, the perceived quality of the responses, and their overall experience using the chatbot.

Quantitative Evaluation: Python scripts were implemented to automatically evaluate the model outputs based on the chosen quantitative metrics. This involved analyzing the presence of typos, on the similarity scores and BERT score of the generated text compared to reference answers. Scripts also analyzed the impact of different "top_p" and "temperature" values on the final answer of the model, but the answers were collected during qualitative evaluation.

A. Results and analysis

1) Qualitative Findings: User Perceptions and Preferences

Overall Satisfaction: Both GPT and Gemini models received positive feedback from users regarding their ability to generate relevant and informative responses. GPT was slightly favored for its fluency and grammatical accuracy, while Gemini was commended for its informativeness and trustworthiness.

Preference for Factual and Concise Responses: User responses indicated a tendency to prefer factual and concise answers, even if they seemed less "human-like" in their language style. This suggests that users prioritize accuracy and efficiency over stylistic elements when seeking information or assistance.

2) Quantitative Findings: Measuring Model Performance

Robustness to Typos: Both models showed sensitivity to typos in the input, with GPT exhibiting a more pronounced decline in performance with increasing noise. This highlights the need for mechanisms to address typos and ensure accurate understanding of user queries.

Handling Question Variations: Gemini demonstrated superior performance in understanding and responding to variations of the same question, maintaining higher similarity scores across different phrasings. This suggests that Gemini's approach to dataset creation and its ability to leverage a broader knowledge base contribute to its effectiveness in handling diverse user queries.

We also looked into the impact of "top_p" and "temperature": The analysis revealed that the optimal range for these hyperparameters differed between GPT and Gemini. GPT performed better with lower values, generating more factual and focused responses, while Gemini benefited from slightly higher values, allowing for more creative and diverse outputs while maintaining relevance.

B. GPT vs. Gemini (resulting differences and testing outcomes)

1) Flexibility and Adaptability:

Testing Results: Gemini demonstrated superior performance in handling variations of the same question and adapting to diverse user queries. Its ability to process raw text allows it to identify relevant information even when the phrasing deviates from the training data. GPT, while capable of understanding variations to some extent, showed limitations due to its reliance on pre-defined question-answer pairs.

Impact on Usability: Gemini's flexibility makes it more user-friendly and adaptable to real-world scenarios where user queries may not always be perfectly phrased or predictable.

2) Comprehensiveness and Informativeness:

Testing Results: Gemini's responses were generally found to be more comprehensive and informative, providing additional context and insights beyond the direct answer to the query. GPT's responses, while accurate, tended to be more concise and focused, sometimes lacking the depth of information users desired.

Impact on Usability: Gemini's ability to provide more comprehensive information enhances its usability for users seeking a deeper understanding of the topic or exploring related concepts.

3) Accuracy and Factuality:

Testing Results: Both models demonstrated high levels of accuracy within their domain of knowledge. However, GPT showed a slight advantage in factual consistency due to its reliance on pre-defined answers. Gemini, while generally reliable, occasionally exhibited instances of unusual answers, generating information that was not factually correct.

Impact on Usability: While both models provide reliable information, Gemini's potential for factual errors may require additional verification steps or human oversight in certain applications.

4) Controllability and Style:

Testing Results: GPT's responses were generally perceived as more grammatically correct and stylistically consistent due to the controlled nature of its training data. Gemini's responses, while still grammatically sound, exhibited more variability in style and tone.

Impact on Usability: GPT's consistent style may be preferable in applications where formality and uniformity are desired. Gemini's stylistic variability offers a more natural and human-like conversational experience.

C. Overall usability considerations

The decision between GPT and Gemini depends on the specific requirements of your application and the desired user interaction level. GPT is ideal for applications that prioritize high accuracy, factual consistency, and a controlled response style. Conversely, Gemini is better suited for scenarios that require flexibility, adaptability, and the ability to deliver comprehensive and informative responses to a wide range of user queries.

IV. Conclusion

This research has provided valuable insights into the usability of LLMs, particularly focusing on the strengths and weaknesses of Gemini and GPT models. The findings emphasize the importance of a holistic approach to LLM evaluation, considering both qualitative and quantitative aspects to ensure these models effectively meet user needs and expectations.

Key takeaways from the research include:

- User-centric Design is Paramount: LLM development and implementation should prioritize user needs and preferences, incorporating feedback from user studies and surveys to ensure a positive and satisfying experience. This applies to the practical implementation of chatbot and interaction with it.
- Model-Specific Tuning is Essential: Each LLM model has unique characteristics and optimal performance is achieved through careful tuning of hyperparameters like "top_p" and "temperature" to align with the desired application and user expectations. In our research, we worked with different settings and GPT performed more satisfactory with lower values, while the Gemini performed better with slightly higher values.
- Dataset Quality Matters: The structure, content, and diversity of the dataset significantly impact the quality and relevance of the generated responses. Building comprehensive and well-structured datasets is crucial for effective LLM performance [18]. We created our own dataset in Slovak language by scraping the webpages of our Faculty.
- Robustness to Noise is Critical: LLMs are susceptible to typos and errors in user input. Implementing mechanisms to address these issues and improve robustness will ensure accurate understanding of user queries and prevent misinterpretations. These are important in case like ours, where we work with language where there are many acute accents and hiccups present, that can create potential noise.

Acknowledgment

This research was funded by the Slovak Research and Development Agency under the contracts No. APVV-22-0414, APVV-20-0232, APVV-17-0550 and by the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic and the Slovak Academy of Sciences under grants No. 1/0685/21 and 1/0259/24.

REFERENCES

[1] Douglas, Michael R.: *Large Language Models* [online]. July 2023. [cit. 2024-02-14]. Available at: https://arxiv.org/abs/2307.05782
[2] Vaswani, Ashish et al.: *Attention Is All You Need* [online]. June 2017. [cit. 2024-02-14]. Available at: https://arxiv.org/abs/1706.03762.
[3] Minaee, Shervin et al.*: Large Language Models: A Survey* [online]. February 2024. [cit. 2024-02-14]. Available at: https://arxiv.org/abs/2402.06196
[4] Chetna Khanna: *Byte-Pair Encoding: Subword-based tokenization algorithm* [online]. August 2021 [cit.2024-02-14]. Available at: https://towardsdatascience.com/byte-pair-encoding-subword-based-tokenization-algorithm-77828a70bee0
[5] Song, Xinying et al.: *Fast WordPiece Tokenization* [online]. December 2020. [cit. 2024-03-08]. Available at: https://arxiv.org/abs/2012.15524.
[6] Naveed, Humza et al.: *A Comprehensive Overview of Large Language Models* [online]. 2023. [cit. 2024-03-08]. Available at: https://arxiv.org/abs/2307.06435.
[7] Brown, Tom B. et al.: *Language Models are Few-Shot Learners* [online]. June 2020. [cit. 2024-04-08]. Available at: https://arxiv.org/abs/2005.14165.
[8] Chowdhery, Aakanksha et al.: *PaLM: Scaling Language Modeling with Pathways* [online]. April 2022. [cit. 2024-04-08]. Available at: https://arxiv.org/abs/2204.02311.
[9] Amirova, Aliya et al.: *Framework-Based Qualitative Analysis of Free Responses of Large Language Models: Algorithmic Fidelity* [online]. September 2023. [cit. 2024-02-08]. Available at: https://arxiv.org/abs/2309.06364.
[10] Bano, Muneera et al.: *Exploring Qualitative Research Using LLMs* [online]. June 2023. [cit. 2024-04-08]. Available at: https://arxiv.org/abs/2306.13298
[11] Wang, Jiayin et al.: *Understanding User Experience in Large Language Model Interactions* [online]. February 2024. [cit. 2024-02-18]. Available at: https://arxiv.org/abs/2401.08329.
[12] Bevan, Nigel et al.: *New ISO Standards for Usability, Usability Reports and Usability Measures* [online]. July 2016. [cit. 2024-02-18]. Available at: https://www.researchgate.net/publication/304107978_New_ISO_Standards_for_Usability_Usability_Reports_and_Usability_Measures.
[13] International Organization for Standardization & International Electrotechnical Commission. (2022). ISO/IEC 23053:2022, *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).* ISO/IEC. June 2022. [cit. 2024-04-08]. Available at: https://www.iso.org/standard/74438.html
[14] Torii, Maya Grace, Takahito Murakami, and Yoichi Ochiai.: *Expanding Horizons in HCI Research Through LLM-Driven Qualitative Analysis.* arXiv preprint arXiv:2401.04138 (2024). [online]. January 2024. [cit. 2024-02-08]. Available at: https://arxiv.org/abs/2309.06364.
[15] Murahari, Vishvak et al.: *QualEval: Qualitative Evaluation for Model Improvement* [online]. November 2023. [cit. 2024-04-08]. Available at: https://arxiv.org/abs/2311.02807.
[16] Urban, Eric: *How to generate question and answer pairs from your source dataset* [online]. February 2024. [cit. 2024-04-08]. Available at: https://learn.microsoft.com/en-us/azure/ai-studio/how-to/generate-data-qa
[17] Lewis, Patrick et al.: *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks* [online]. May 2020. [cit. 2024-04-08]. Available at: https://arxiv.org/abs/2005.11401.
[18] Han, J., Kamber, M., & Pei, J.: *Getting to know your data* [online]. December 2012. [cit. 2024-04-08]. Available at: https://www.researchgate.net/publication/288177518_Getting_to_Know_Your_Data
[19] Jakindah, Dixn: *Top P, Temperature and Other Parameters* [online]. May 2023. [cit. 2024-03-20]. Available at: https://medium.com/@dixnjakindah/top-p-temperature-and-other-parameters-1a53d2f8d7d7
[20] Shivhare, Gaurav: *Decoding GPT's Temperature Setting* [online]. November 2023. [cit. 2024-03-20]. Available at: https://www.linkedin.com/pulse/decoding-gpts-temperature-setting-gaurav-shivhare-ldqoe/
[21] Lotfi, Chaimaa et al.: *Web Scraping Techniques and Applications: A Literature Review* [online]. January 2021. [cit. 2024-04-05]. Available at: https://www.researchgate.net/publication/367719780_Web_Scraping_Techniques_and_Applications_A_Literature_Review.
[22] Lewis, James R. - Sauro, Jeff: *Usability and user experience: design and evaluation* [online]. August 2021. [cit. 2024-04-09]. Available at: https://www.researchgate.net/publication/373487143_USABILITY_AND_USER_EXPERIENCE_DESIGN_AND_EVALUATION.
[23] Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating Text Generation with BERT. *International Conference on Learning Representations (ICLR)*.

Možnosti vnorenia polymérnych hrubovrstvových rezistorov do 3D tlačených štruktúr

[1]*Igor VEHEC,* [2]*Lukáš GARNEK*

[1,2] Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]igor.vehec.2@tuke.sk, [2]lukas.garnek@student.tuke.sk

***Abstrakt* — Článok sa zaoberá možnosťou integrácie a vnorenia sieťotlačou deponovaných PTF (Polymer Thick Film) rezistorov do 3D tlačených štruktúr vytvorených štandardným FFF (Fused Filament Fabrication) procesom. Ako PTF vrstvy sú použité komerčne dostupné hrubovrstvové pasty: vodivá pasta ED3000 na báze striebra a odporová pasta ED7500. Pre 3D tlač sú použité štandardné termoplastické polymérne materiály ABS (akrylonitrilbutadiénstyrén) a PETG (polyetyléntereftalát-glykol). Analyzované sú možnosti zlepšenia kvality tlače PTF na 3D tlačený substrát pomocou dodatočnej tlače PTF dielektrickej vrstvy, ako aj možnosti následného vnorenia PTF vrstiev do 3D tlačenej štruktúry. Vyhodnocované sú mechanická pevnosť rozhrania vzniknutého pri spojení do kompaktnej 3D štruktúry, ako aj vplyv použitých technologických postupov na zmenu elektrického odporu PTF rezistorov.**

***Kľúčové slová* — 3D tlač, FFF, ABS, PETG, polymérne hrubé vrstvy, sieťotlačová technológia**

I. Úvod

Integrácia elektroniky do 3D tlačených štruktúr patrí medzi aktuálne témy. V súčasnosti existuje niekoľko prístupov k 3D tlačenej elektronike, ktoré zahŕňajú techniky ako sú vytláčanie materiálu, stereolitografia či jetting, pričom samotné prepojenie a integrácia elektronických komponentov je realizované pomocou rôznych techník [1], [2], [3], [4], [5]. Príkladom vnorenia elektronickej štruktúry do 3D tlačeného komponentu je implementácia vodivej vrstvy na báze Cu do 3D tlačenej štruktúry tvoriacich kompaktný RF systém, pri ktorej sa spojenie 3D tlačených častí realizovalo pomocou systému podobného systému LEGO [6]. Ďalším smerom, ktorým sa vývoj uberá je zameraný na materiály používané pre 3D tlač, ktoré sa skúmajú na viacerých úrovniach, vrátane snahy o zlepšenie ich elektrických vlastností či využitia recyklovaných (resp. recyklovateľných) materiálov [7],[8].

Polymérne hrubovrstvové vrstvy (PTF) predstavujú ďalšiu významnú oblasť elektroniky, kde sa hrubé vrstvy nanášajú rôznymi technológiami ako je napr. sieťotlač. Výhodou PTF technológie je relatívne nízka teplota spracovania funkčných PTF vrstiev (alebo spracovanie vrstiev bez tepla pomocou UV žiarenia), čo umožňuje aj ich aplikáciu na organické substráty, vrátane flexibilných materiálov. Predpokladom teda môže byť aj ich využitie na 3D tlačené substráty. PTF technológia nachádza v súčasnosti uplatnenie ako v klasickej elektronike (výroba elektrických prepojení, hrubovrstvových rezistorov), tak aj v moderných senzoroch [9], [10], [11], [12].

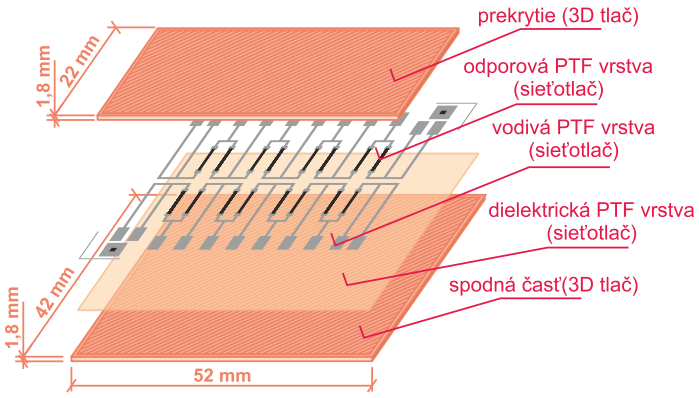
Obe technológie, 3D tlač aj sieťotlač, patria do aditívnej výroby (AM – Additive Manufacturing). Predkladaný článok sa zameriava na kombináciu týchto dvoch technológií s cieľom navrhnúť a overiť metódy deponovania vodivých a odporových PTF vrstiev na povrch 3D tlačených štruktúr s ich následným vnorením do 3D tlačených štruktúr vytvorených pomocou štandardnej FFF (Fused Filament Fabrication) metódy 3D tlače. Cieľom je vytvoriť cenovo dostupnú a efektívnu integráciu PTF elektroniky do 3D tlačených štruktúr.

II. Integrácia PTF rezistorov do 3D tlačenej štruktúry

Overenie možnosti vnorenia PTF rezistorov do 3D tlačenej štruktúry sa uskutočnilo na vzorkách, ktorých celkové usporiadanie je znázornené na Obr. 1. Výrobný postup predpokladá 3D tlač jednotlivých častí (spodnej časti a prekrytia), sieťotlač PTF vrstiev na spodnú 3D tlačenú časť, vytvrdenie PTF vrstiev a konečné zostavenie 3D tlačenej spodnej časti a prekrytia do jednej

kompaktnej štruktúry. V rámci procesu realizácie bolo potrebné navrhnúť a optimalizovať:

- *proces sieťotlače PTF vrstiev* (z hľadiska vplyvu drsnosti povrchu spôsobenej štruktúrou vlákien po 3D tlači na kvalitu tlače vodivých a odporových PTF pást),
- *proces spájania* spodnej časti a prekrytia do jednej kompaktnej štruktúry (s ohľadom na materiál použitý pre 3D tlač a možného vplyvu spájania na PTF rezistory).



Obr. 1 Usporiadanie rôznych typov vrstiev v experimentálnych vzorkách.

Pri experimentoch sa sledovali a vyhodnocovali zmeny elektrického odporu PTF rezistorov (s rozmermi 5,6 × 0,5 mm) vrátane odchýlok ich hodnôt. Topológia vodivých a odporových PTF vrstiev bola navrhnutá tak, aby elektrický odpor PTF rezistorov bolo možné merať štvorvodičovou (Kelvinovou) metódou. Výsledky boli porovnávané s PTF rezistormi vnorenými do štandardnej DPS na báze sklo-epoxidového materiálu FR4. Materiálmi použitými pre 3D tlač boli ABS (akrylonitrilbutadiénstyrén) a PETG (polyetyléntereftalát-glykol), oba s priemerom vlákna 1,75 mm. 3D tlačené časti boli vytvorené pomocou tlačiarne Prusa i3 MK2 so štandardnými parametrami tlače. Depozícia PTF vrstiev bola realizovaná sieťotlačovým zariadením Aurel 900PA. Všetky merania elektrického odporu sa uskutočnili pomocou multimetra Agilent 34411A.

A. Sieťotlač PTF vrstiev na 3D tlačený substrát

Ako vodivá polymérna pasta bola zvolená Ag termoplastická vodivá PTF pasta ED3000 s plošným odporom 30 mΩ/sq pri hrúbke 12 µm a s teplotou vytvrdzovania 90 až 120 °C počas 30 až 40 min. Pre PTF rezistory bola zvolená pasta ED7500, ktorá poskytuje široký rozsah hodnôt plošného odporu (od 1 Ω/sq do 1 MΩ/sq), s teplotou vytvrdzovania 150 až 200 ºC po dobu 30 min. Za účelom zlepšenia kvality tlače bola zvolená dielektrická pasta ED7500.

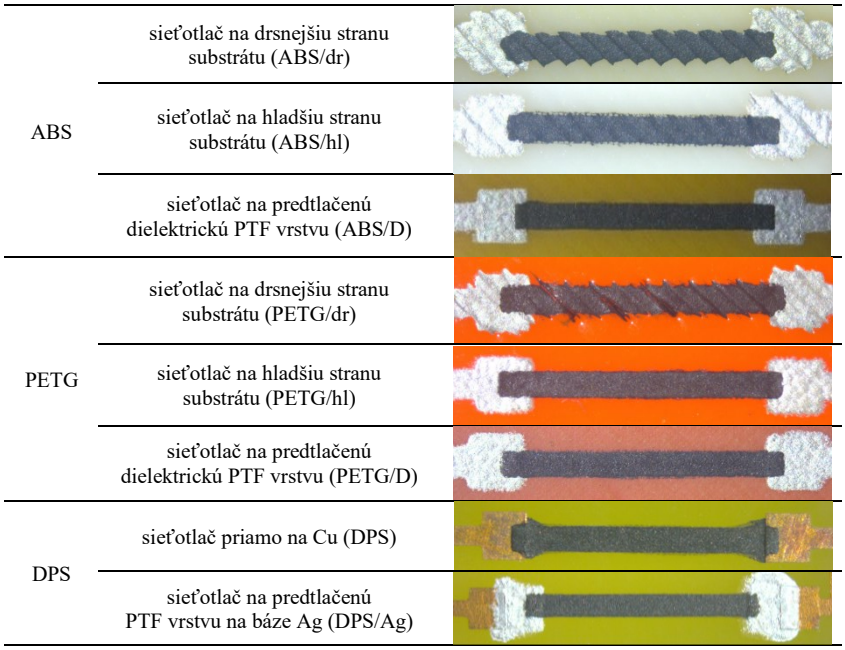
Hlavným problémom ovplyvňujúcim kvalitu tlače (a následne aj PTF rezistorov) je nerovný povrch 3D tlačeného substrátu, na ktorom sa opakuje štruktúra vlákien filamentu. Táto štruktúra je zvyčajne menej výrazná zo strany, ktorá je v kontakte s vyhrievanou doskou, na ktorej sa 3D tlač vykonáva. Na porovnanie kvality tlače boli preto vybrané 3 prístupy:

- sieťotlač vodivej PTF vrstvy (vytvrdzovanie 30 min pri 120 °C) a odporovej PTF vrstvy (vytvrdzovanie 30 min pri 150 °C) na hladšiu stranu 3D tlačenej štruktúry,
- sieťotlač vodivej a odporovej PTF vrstvy na drsnejšiu stranu 3D tlačenej štruktúry,
- dodatočná sieťotlač dielektrickej vrstvy (vytvrdzovanie 30 min pri 150 °C) na drsnejšiu strane 3D tlačenej štruktúry, ktorej úlohou bolo „vyrovnania“ nerovného povrchu kopírujúceho štruktúru vlákien filamentu, s následnou sieťotlačou vodivej a odporovej PTF vrstvy.

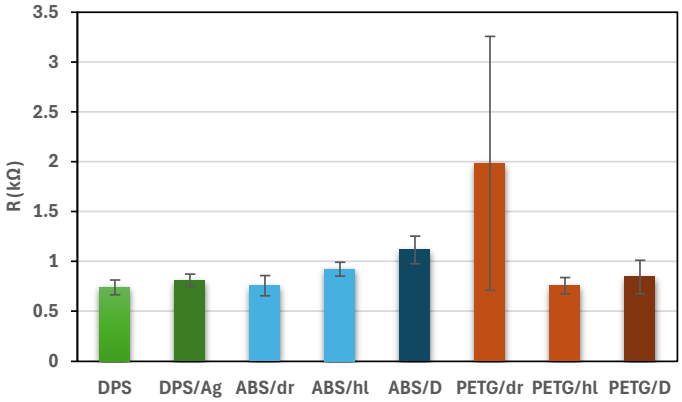
Výsledky kvality a ostrosti tlače realizovanej za rovnakých podmienok sú zobrazené na Obr. 2. Najlepšie výsledky z hľadiska ostrosti tlače sa dosiahli pre oba materiály (ABS a PETG) v prípade sieťotlače na hladšiu stranu substrátov (ABS/hl a PETG/hl), ako aj pri sieťotlači na povrchu s nanesenou dodatočnou dielektrickou PTF vrstvou (ABS/D a PETG/D). Kvalita tlače je pri nich porovnateľná s tlačou na DPS. Vzorky vytlačené na drsnejšej strane (ABS/dr a PETG/dr) vykazovali charakteristický vzor kopírujúci štruktúru vlákna po 3D tlači na povrchu substrátu. Viditeľný rozdiel medzi tlačou na ABS a PTEG bol spôsobený najmä rozdielnym charakterom (hĺbkou a šírkou) priehlbín medzi jednotlivými vláknami.

Charakteristickým znakom počas tepelného vytvrdzovania PTF pást bolo nerovnomerné zmršťovanie substrátov ABS aj PETG. Hlavné zmrštenie bolo v rovine x-y a bolo výraznejšie pri

prvom teplotnom spracovaní. Zmrštenie substrátov PETG (približne 6 %) bolo oproti materiálu ABS (približne 3,9 %) väčšie. Ďalšie tepelné spracovanie počas vytvrdzovania zvyšných PTF vrstiev malo len mierny vplyv na zmrštenie (približne 1 %). Dodatočná dielektrická vrstva síce viedla k zníženiu zmrštenia v rovine x-y (o cca 1,5 % pre PETG a o cca 1,7 % pre ABS), ale viedla tiež k dodatočným deformáciám (konkávne vydutie pri ABS, resp. zvlnenie povrchu pri PETG), čo prispelo k nerovnomernej tlači.



Obr. 2 Vplyv povrchu na kvalitu tlače PTF rezistorov pre rôzne typy substrátov.



Obr. 3 Vplyv povrchu na elektrický odpor PTF rezistorov pre rôzne typy substrátov.

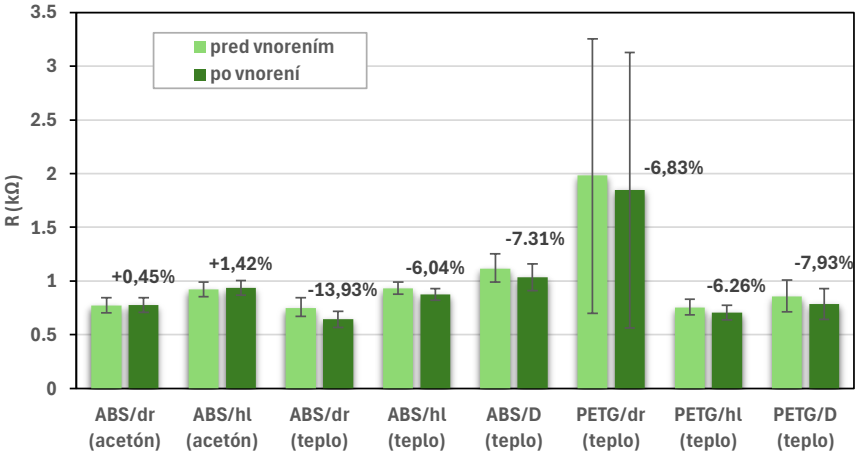
Výsledky merania elektrického odporu PTF rezistorov (Obr. 3) pre vzorky na báze materiálov PTEG aj ABS v porovnaní s PTF rezistormi na DPS ukazujú, že zmena elektrického odporu je výrazne ovplyvnená kvalitou tlače, a teda „drsnosťou“ povrchu. Okrem drsnosti povrchu môže výsledný elektrický odpor ovplyvniť aj zmrštenie substrátu počas vytvrdzovania PTF pást. PTF rezistory vytlačené na drsnejšej strane substrátu PETG vykazovali najvyšší elektrický odpor, ako aj rozptyl nameraných hodnôt. PTF rezistory vytlačené na dielektrickej PTF vrstve viedli k síce ostrejšej tlači, ale nerovnomernosť v podobe zvlnenia alebo vydutia spôsobila rozdielnu hrúbku nanesenej PTF vrstvy, čo prispieva k miernemu zvýšeniu rozptylu hodnôt elektrického odporu rezistorov. Rozptyly hodnôt PTF rezistorov vytlačených na hladšej strane 3D tlačeného substrátu sú porovnateľné s PTF rezistormi tlačenými na hladký povrch substrátu FR4 pre PCB.

B. Proces spájania častí do finálnej 3D štruktúry (*vnorenie PTF rezistorov*)

Vnorenie PTF rezistorov do 3D tlačenej štruktúry bolo odskúšané spojením vrchného 3D tlačeného prekrytia a spodnej 3D tlačenej časti s natlačenými PTF vrstvami pomocou dvoch metód v závislosti od použitého materiálu substrátu:

- pre ABS: spojenie teplom alebo acetónom (t. j. bez použitia tepla),
- pre PTEG: spojenie teplom.

Spojenie dvoch materiálov ABS sa testovalo pri rôznych teplotách, pričom za dostatočnú teplotu (spojenie bez nadmerných deformácii) sa určila teplota 150 ºC. V prípade materiálu PTEG sa ako najvhodnejšia teplota (spájanie materiálov bez nadmernej deformácie) stanovila tiež teplota 150 ºC. Pevnosť získaného spoja bola overená skúškou strihom a porovnaná s meraniami na plne tlačenej 3D štruktúre, kde bola mechanická pevnosť v strihu 13,8 MPa pre ABS, resp. 17,1 MPa pre PETG. Pevnosť v strihu dodatočne spájaných 3D tlačených častí najvýraznejšie poklesla pre prípady, v ktorých sa na povrchu nachádzala natlačená dodatočná dielektrická vrstva (pevnosť v strihu 0,5 MPa pre ABS a 0,9 MPa pre PETG), pričom spájanie acetónom pre materiál ABS v tomto prípade nebolo možné. Oproti tomu sa spájanie ABS materiálov bez dodatočnej dielektrickej vrstvy vyznačovalo pevnosťou v strihu 17,7 MPa, čo je porovnateľné s pevnosťou pre plne 3D tlačenú štruktúru. Celý problém s dielektrickou pastou je možné čiastočne obísť tlačou dielektrickej vrstvy len na miesta, kde sa predpokladá následná tlač vodivých a odporových PTF vrstiev.



Obr. 4 Vplyv spôsobu vnorenia na elektrický odpor PTF rezistorov pre rôzne typy substrátov.

Elektrický odpor PTF rezistorov po vnorení do 3D tlačenej štruktúre bol porovnávaný s ich elektrickým odporom pred vnorením (referenčná hodnota elektrického odporu). Výsledky sú znázornené v grafe na Obr. 4. PTF rezistory v prípade materiálu ABS vykazovali mierny nárast elektrického odporu po spojení vrchnej a spodnej vrstvy pomocou acetónu. V prípade spájania teplom došlo k zníženiu elektrického odporu približne o 6 až 14 % v prípade PTF rezistorov vytlačených sieťotlačou na drsnejšej strane substrátu. Z grafu teda možno usúdiť, že zmena odporu v dôsledku prekrývania (vnorenia) je spôsobená skôr dodatočným tepelným spracovaním, ktoré je pre PTF rezistory typické, a nie samotným prekrývaním. Pokles elektrického odporu po vnorení rezistorov PTF do 3D tlačených štruktúr bol podobný aj v prípade materiálu PETG a bol porovnateľný so zmenou elektrického odporu na kontrolných vzorkách DPS na báze materiálu FR4 (približne o 6,5 %).

III. Zhodnotenie výsledkov

Navrhovaný postup vnorenia PTF rezistorov do 3D tlačenej štruktúry zahŕňa niekoľko krokov: (1) 3D tlač jednotlivých častí, (2) sieťotlač a vytvrdzovanie PTF vrstiev a (3) spojenie jednotlivých častí 3D tlačenej štruktúry. Pri návrhu 3D tlačených komponentov ako aj topológie PTF vrstiev musí byť zohľadnené dodatočné zmrštenie substrátu spôsobené vytvrdzovaním PTF pást (cca 6 % pre PTEG a cca 3,9 % pre ABS, resp. 1,5 % pre PETG a 1,7 % pre ABS ak sa použije dodatočná dielektrická vrstva). Sieťotlač môže pozostávať z dvoch alebo troch krokov, pričom jednoduchší variant (2 tlače: vodivá + odporová PTF vrstva) je vhodnejší pre sieťotlač na hladší povrchu 3D tlačeného substrátu. Zvýšenie kvality tlače na drsnejšom povrchu 3D tlačeného substrátu je možné dosiahnuť dodatočnou (treťou) tlačou dielektrickej vrstvy, ktorá ale môže viesť k vydutiu alebo zvlneniu vzoriek. Alternatívou by bolo použitie PTF pást s nižšou teplotou vytvrdzovania alebo vytvrdzovaných UV svetlom, čo by tieto negatívne vplyvy mohlo zmierniť.

Samotné spájanie jednotlivých 3D tlačených častí do kompaktnej štruktúry závisí od zvoleného materiálu a od prítomnosti/neprítomnosti PTF dielektrickej vrstvy. V prípade materiálu ABS je možné spájanie teplom aj acetónom, materiál PETG je možné spojiť len pomocou tepla, pričom

dodatočná dielektrická vrstva výrazne znižuje mechanickú pevnosť vzniknutého spoja.

Celkovo možno zhodnotiť, že vnorenie vodivých aj odporových PTF vrstiev sieťotlačou do bežne dostupných 3D tlačených štruktúr na báze materiálov ABS a PETG je realizovateľné pre nenáročné aplikácie. Ide o relatívne lacnú alternatívu a jednoduchou úpravou procesu by sa mohli samostatné vodivé prepojenia na báze vodivých PTF pást použiť aj na prepojenie SMD aktívnych súčiastok, ktoré by sa mechanicky a elektricky prepojili napr. pomocou vodivých lepidiel.

IV. Záver

Vnorenie PTF vrstiev (na báze komerčných PTF pást ED3000 a ED7500) pomocou technológie sieťotlače do 3D tlačených štruktúr na báze materiálov ABS a PETG realizovaných metódou FFF 3D tlače je realizovateľné. Dodatočná teplotná úprava PTF rezistorov počas ich vytvrdzovania negatívne ovplyvňuje rozmerovú stabilitu substrátov, pri ktorej dochádza k ich zmršťovaniu. V prípade nerovného povrchu s pravidelnou štruktúrou vlákien po 3D tlači je možné nerovnosť povrchu zmierniť dodatočnou tlačou dielektrickej PTF vrstvy. V závislosti od typu materiálu je možné 3D vytlačené štruktúry spájať teplom (ABS a PTEG) alebo aj acetónom (v prípade ABS), pričom nedochádza k negatívnemu vplyvu na odporové PTF vrstvy. Dodatočná dielektrická PTF vrstva nepriaznivo ovplyvňuje následné spájanie 3D vytlačených vrstiev pomocou acetónu (nerealizovateľné) ale aj tepla (výrazné zníženie pevnosti spoja). Celkovo možno zhrnúť, že navrhovaný proces vnorenia PTF rezistorov pomocou kombinácie dvoch aditívnych výrobných technológií (3D tlače metódou FFF a sieťotlače PTF vrstiev) je realizovateľný pre nenáročné aplikácie elektroniky.

Poďakovanie

Tento príspevok vznikol vďaka projektu Vega 1/0049/24 „Výskum a vývoj nových metód charakterizácie a spoľahlivosti miniaturizovaných modulov elektroniky s využitím nových typov povrchových úprav DPS“, projektu Kega 011TUKE-4/2023 „Synergia edukačného procesu, výskumu, inovácií a priemyselného sveta v kontexte prestavby študijného programu automobilová elektronika“, projektu APVV 22-0289 „Synergia nových foriem a riešení spájkovania a ich vplyv na kvalitu v kontexte nízkoenergetického a environmentálneho prístupu pri výrobe výkonových modulov elektroniky“ a projektu ITMS 313011T557 „Podpora výskumno-vývojového potenciálu v oblasti dopravných prostriedkov“.

References

[1] S. Park, W. Shou, L. Makatura, W. Matusik, and K. Fu, "3D printing of polymer composites: Materials, processes, and applications," Matter, Vol. 5, Issue 1, 2022, pp. 43-76, ISSN 2590-2385, doi: 10.1016/j.matt.2021.10.018

[2] A. Joe Lopes, E. MacDonald, and R. B. Wicker, "Integrating stereolithography and direct print technologies for 3D structural electronics fabrication," Rapid Prototyping Journal, Vol. 18 No. 2, pp. 129-143, 2012. doi: 10.1108/13552541211212113.

[3] H.W. Tan, Y.Y. C. Choong, C.N. Kuo, H.Y. Low, C.K. Chua, "3D printed electronics: Processes, materials and future trends," Progress in Materials Science, vol. 127, 2022, doi: 10.1016/j.pmatsci.2022.100945

[4] B.H. Lu, H.B. Lan, H.Z. Liu, "Additive manufacturing frontier: 3D printing electronics," Opto-Electron Adv 1, 170004, 2018, doi: 10.29026/oea.2018.170004

[5] E. Macdonald et al., "3D Printing for the Rapid Prototyping of Structural Electronics," in IEEE Access, vol. 2, pp. 234-242, Dec. 2014, doi: 10.1109/ACCESS.2014.2311810

[6] M. I. Mohd Ghazali, S. Mondal, S. Karuppuswami and P. Chahal, "3D Printed Substrates for the Design of Compact RF Systems," 2019 IEEE 69th Electronic Components and Technology Conference (ECTC), Las Vegas, NV, USA, 2019, pp. 113-118, doi: 10.1109/ECTC.2019.00025.

[7] P. Veselý, J. Zdráhal, I. Králová, M. Klimtová and D. Froš, "Solderless Component Assembly: Novel Ecological Approach to Electronics Production," 2023 46th International Spring Seminar on Electronics Technology (ISSE), Timisoara, Romania, 2023, pp. 1-6, doi: 10.1109/ISSE57496.2023.10168409.

[8] V. Francis, and P. K. Jain, "3D printed polymer dielectric substrates with enhanced permittivity by nanoclay inclusion," Virtual and Physical Prototyping, Vol. 12, Issue 2, pp. 107-115, 2012 doi: 10.1080/17452759.2017.1312466.

[9] I. Vehec, T. Lenger, S. Kardos, P. Lukacs and K. Gontkovic, "Polymer Thick-film Piezoresistive Sensor Integrated into the PCB," 2023 46th International Spring Seminar on Electronics Technology (ISSE), Timisoara, Romania, 2023, pp. 1-8, doi: 10.1109/ISSE57496.2023.10168379.

[10] G. Harsányi, "Polymer films in sensor applications: a review of present uses and future possibilities", Sensor 2000, Vol. 20 No. 2, pp. 98-105, 2000, doi: 10.1108/02602280010319169

[11] T. Rovensky, A. Pietrikova, I. Vehec, L. Livovsky, "Stability of miniaturized non-trimmed thick- and thin-film resistors," Microelectronics Reliability, Vol. 84, pp. 88-94, 2018, doi: 10.1016/j.microrel.2018.03.011.

[12] S. Rajala, S. Tuukkanen, J. Halttunen, "Characteristics of piezoelectric polymer film sensors with solution processable graphene based electrode materials," IEEE Sensors Journal, Vol. 15, Issue 6, pp. 3102 - 3109, June 2015, doi: 10.1109/JSEN.2014.2344132

Inteligentné nabíjanie elektromobilov

[1]*Ján KAŇUCH,* [2]*Peter GIROVSKÝ,* [3]*Filip ŠULEK*
[1,2] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 1/9, 04200 Košice-Sever, Slovenská republika

[1]jan.kanuch@tuke.sk, [2]peter.girovsky@tuke.sk, [3]filip.sulek@student.tuke.sk

***Abstrakt* — V tomto príspevku je prezentované riešenie problematiky nabíjania elektromobilov s využitím nových progresívnych technológií. V prvej časti príspevku sú popísané súčasné technológie nabíjania elektromobilov a rozbor súčasného stavu tejto problematiky. Následné je v príspevku spracovaný rozbor riešenia plánovaného inteligentného nabíjania elektromobilov pomocou nových technológii, ktoré dokážu zabezpečiť tok elektrickej energie nielen do batérie, ale aj z batérie elektromobilu v spolupráci s technológiami inteligentného a plánovaného nabíjania.**

***Kľúčové slová* — nabíjanie elektromobilov, technológie nabíjania, inteligentne nabíjanie**

I. Úvod

Rozvoj elektromobility je v súčasnosti v plnom prúde. Vozidlá s elektrickým pohonom neustále zvyšujú svoj podiel na trhu. S pribúdajúcim počtom elektromobilov vzniká problém, ako bude dopĺňať elektrickú energiu také veľké množstvo elektromobilov ak trh ovládnu úplne, čo je v pláne Európskej únie. Ak sa aj pribúdajúcim elektromobilom prispôsobí energetická sieť, obyvatelia rodinných domov v mimo mestských oblastiach svoj elektromobil nabijú doma. Najväčší problém je však u elektrických vozidiel v husto obývaných oblastiach.

V súčasnosti v mestách a najmä v husto obývaných oblastiach majú vodiči často problém nájsť voľné parkovacie miesto, teda s voľnou nabíjačkou to bude asi veľmi podobné. Preto jedným z najvhodnejších riešení by mohla byť výstavba parkovacích domov pre elektromobily, pripravených pre nabíjanie veľkého množstva elektromobilov. Tieto parkovacie domy by mohli byť vybavené „inteligentným nabíjaním“. Pre rovnomernejšie vyťaženie siete budú elektromobily nabíjané podľa priority. Ak vodič vopred vie, že so svojím elektromobilom vyrazí ďalší deň až popoludní pričom ďalší vodič odchádza už ráno je zbytočné, aby sa tieto elektromobily nabíjali súčasne rovnakým výkonom.

II. Technológie nabíjania elektromobilov

Na rozdiel od spaľovacích vozidiel do ktorých dopĺňame palivo výhradne na čerpacích staniciach elektrické vozidlá ponúkajú širokú škálu možností doplnenia energie. Spôsoby, ako nabiť batériu do elektromobilu začínajú už doma, keďže v súčasnosti takmer každá domácnosť disponuje elektrickou prípojkou. Tieto možnosti ďalej pokračujú cez verejné nabíjacie stanice, ktoré sa opäť líšia ich nabíjacím výkonom, alebo či na nabíjanie využívajú jednosmerný alebo striedavý elektrický prúd. Alternatívnymi možnosťami, ktoré ale umožňuje len úzka škála elektromobilov, sú napríklad batérie s vymeniteľnými modulmi alebo technológia induktívneho nabíjania. Teda užívateľ elektromobilu si môže vybrať akým spôsobom doplní elektrickú energiu do svojho elektromobilu od čoho závisí najmä čas a tomu úmerne aj cena.

A. Verejné nabíjacie stanice

Jedným z najčastejším spôsobov nabíjania elektromobilov sú verejné nabíjacie stanice, ktoré dokáže využiť takmer úplne každý elektromobil (Obr. 1). Ich výhodou je, že sa nemusia nachádzať len pri frekventovaných ťahoch ale vďaka ich kompaktnosti a dostupnosti elektrickej energie ich nájdeme už aj v súčasnosti bežne na miestach akými sú parkoviská pri obchodných centrách, biznis centrách alebo fabrikách. Teda túto nabíjaciu stanicu môže elektromobil využiť aj počas toho, ak jeho majiteľ pracuje, nakupuje alebo využíva iné služby. Nabíjacie stanice pre elektromobil delíme hlavne podľa toho, či na nabíjanie využívajú jednosmerný alebo striedavý elektrický prúd.



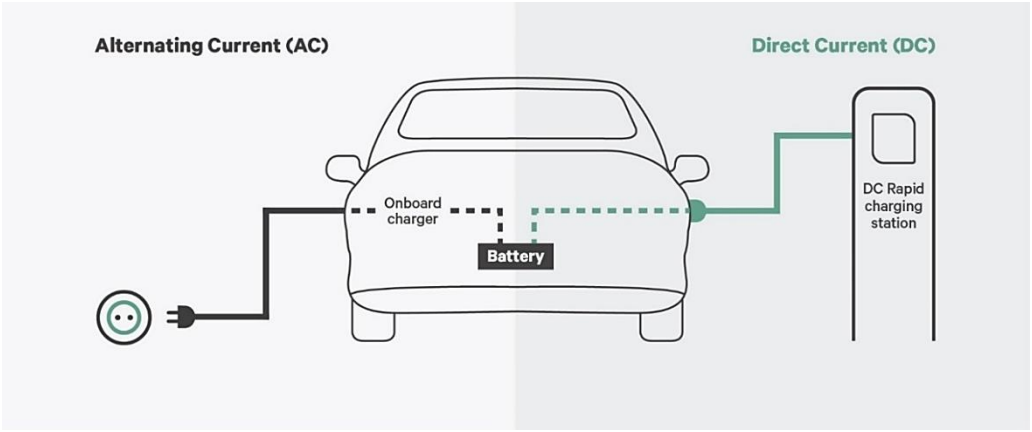
Obr. 1 Verejná nabíjacia stanica pre elektromobil [1].

Nabíjacie stanice využívajúce jednosmerný prúd

Najčastejšie sú práve DC nabíjacie stanice považované za tie, ktoré elektromobil nabijú takzvaným rýchlonabíjaním. Je to z dôvodu, že táto stanica už usmerní striedavý prúd z elektrickej siete na jednosmerný a týmto prúdom dobíja batériu. To ma za následok, že nabíjanie nie je obmedzené výkonom palubnej nabíjačky, keďže výkonová časť je uložená mimo vozidla v nabíjacej stanici. Vďaka tejto výhode vieme nabíjací výkon zvýšiť až niekoľko násobne oproti striedavej nabíjacej stanici bez ohľadu na parametre konkrétneho elektromobilu. Rýchle nabíjanie tak vie využiť široké spektrum elektromobilov [2]. Takáto nabíjacia stanica dokáže nabiť batériu elektromobilu už behom niekoľkých desiatok minút. V súčasnosti sa považuje pri rýchlom nabíjaní za štandard nabitie batérie zo stavu 10% kapacity na 80% kapacity za čas približne 30 minút. Výkon rýchlonabíjacích staníc začína na 50 kW pričom v dnešnej dobe sú bežné stanice s nabíjacím výkonom už 175 kW. Najnovšia generácia elektromobilov ponúka už aj možnosť použitia takzvaných ultrarýchlonabíjacích staníc s nabíjacím výkonom až 350 kW [3].

Nabíjacie stanice využívajúce striedavý prúd

Ako bolo spomenuté, tak nabíjanie striedavým prúdom je síce pomalšie, ale ponúka iné výhody. Jednou z nich je, že ho podporujú všetky elektromobily a taktiež aj plug-in hybridy.



Obr. 2 Nabíjanie elektromobilu striedavým a jednosmerným prúdom [4].

Pri tomto nabíjaní mení striedavý prúd na jednosmerný až palubná nabíjačka, ktorá je zabudovaná v elektromobile (Obr. 2). Nabíjací výkon je obmedzený parametrami tejto palubnej nabíjačky, teda parametrami konkrétneho elektromobilu. Pri tomto nabíjaní je často potrebné používať aj vlastný nabíjací kábel, pričom v rýchlonabíjačkách je už zvyčajne zabudovaný. Avšak

nabíjanie pomocou AC nabíjačky je zvyčajne lacnejšie a taktiež aj šetrnejšie voči životnosti batérie [4]. Ak teda najmä pri dlhšej trase potrebuje užívateľ rýchlo doplniť energiu do vozidla, tak je vhodnejšie využiť DC nabíjanie. Ale ak nie je limitovaný časom, tak lacnejšie a šetrnejšie sú nabíjačky využívajúce striedavý prúd.

B. Domáce nabíjacie stanice

Jednou z najväčších výhod vlastnenia elektromobilu môže byť pre užívateľa práve to, že si elektromobil môže nabíjať aj v pohodlí domova. Avšak táto výhoda sa týka len majiteľov rodinných domov. Majitelia elektromobilov žijúcich v husto obývaných oblastiach túto výhodu využívať nemôžu, nakoľko nie je možné pripojiť elektromobil k elektrickej sieti napríklad v panelákovom byte.

Elektromobil môžeme nabíjať aj priamo z bežnej elektrickej siete, ktorú máme v domácnosti a to jednofázovým prípadne dvoj alebo trojfázovým prúdom. Ďalšou možnosťou je zakúpiť si zariadenie s názvom „wallbox“ (Obr. 3). Je to zariadenie, ktoré plní funkciu domácej nabíjačky a dosahuje vyšší nabíjací výkon. Ten môže byť až 22 kW, pričom pri nabíjaní bez wallboxu je maximálny nabíjací výkon len 2,3 kW. Okrem toho je obrovskou výhodou, že súčasťou wallboxov sú aj elektrické ochrany, ktoré môžu pri poruchách ochrániť ako vozidlo, tak aj elektrickú sieť [5]. Pri zapojení na priamo sa musíme spoľahnúť na kvalitu našej domácej elektroinštalácie. Pri nedostatočne dimenzovaných vodičoch môže dôjsť pri preťažení ku skratu medzi prehriatými vodičmi a v najhoršom prípade aj k požiaru. Dá sa predpokladať, že s vývojom technológie budú čoraz viac dostupné aj výkonnejšie nabíjačky a elektromobil si dokážeme nabiť za relatívne krátky čas aj doma. Výhodou je opäť možnosť voľby majiteľa vozidla akú nabíjačku si zakúpi a bude využívať. Nevýhodou sú ale ďalšie náklady na zakúpenie, inštaláciu a servis takéhoto zariadenia. Okrem toho moderné domáce nabíjacie stanice ponúkajú aj rôzne funkcie online monitoringu a kontroly pred preťažením. Majitelia rodinných domov so solárnymi panelmi môžu ušetriť ďalšie financie. Avšak pri nabíjaní z energie so solárnych panelov narážame na problém s uchovaním elektrickej energie nakoľko slnko nemusí svietiť práve vo chvíli keď potrebujeme nabiť naše vozidlo.



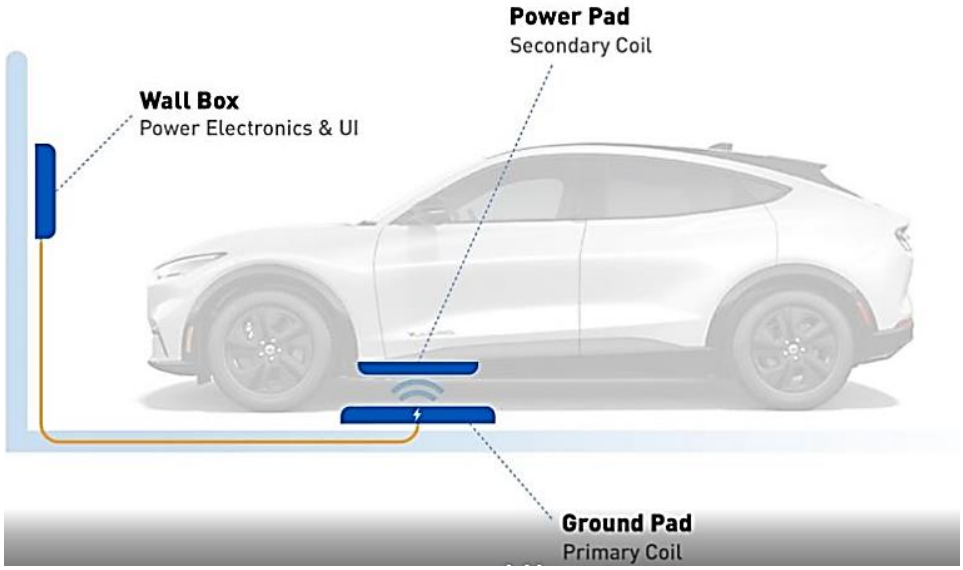
Obr. 3 Domáca nabíjacia stanica [6]

C. Indukčné nabíjanie

Za najviac pohodlný spôsob dopĺňania energie by sa dalo považovať bezkontaktné alebo teda indukčné nabíjanie. Najideálnejšie by bolo, ak by vodič nabíjal svoj elektromobil už počas jazdy. Nemusel by tak nikde zastavovať ani na dlhších trasách a nabíjanie by ho tak prakticky vôbec neobťažovalo. Druhou alternatívou by bolo, ak by sa elektromobil nabíjal státím na parkovacom mieste, bez použitia akéhokoľvek konektora a káblov. Obe tieto možnosti ponúka indukčné nabíjanie, ktoré je založené na princípe prenosu energie pomocou rezonančnej elektromagnetickej indukcie. Auto obsahuje cievku, ktorá prijíma elektrickú energiu z elektromagnetickej cievky v nabíjacej podložke [7].

Statické indukčné nabíjanie

Vodič zaparkuje elektromobil v parkovacom dome alebo parkovisku vybavenou touto technológiu nad nabíjaciu podložku. Po zapojení a aktivovaní by nabíjanie fungovalo tak, ako vidíme na Obr. 4 [8].



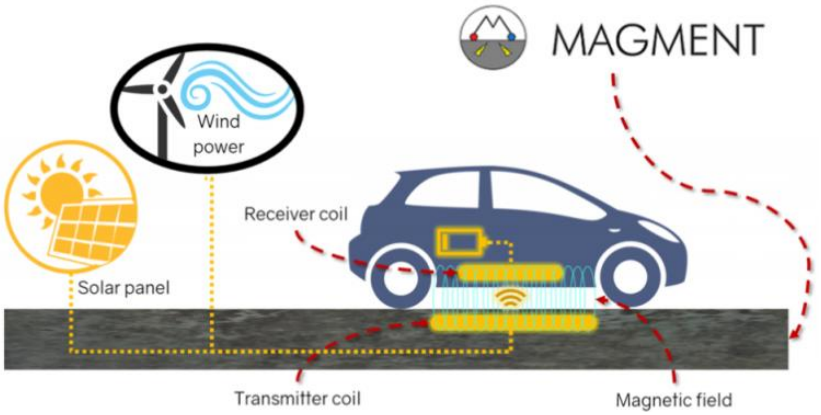
Obr. 4 Statické indukčné nabíjanie [8]

V súčasnosti túto technológiu ponúka už viacero spoločností. Napríklad automobilka BMW ponúka indukčné nabíjanie s nabíjacím výkonom 3,2 kW a účinnosťou 85%. Značka Hyundai predstavila už modernejšiu technológiu s výkonom 11 kW a účinnosťou až 90%. Dokáže tak nabiť elektromobil Hyundai Genesis za 6 hodín [8]. Existujú ale už aj prototypy, ako napríklad od vedcov zo Švédskej Chalmer University, ktorý vyvinuli systém indukčného nabíjania s výkonom až 500 kW za 10 minút [9].

Medzi najväčšie výhody tohto systému patri určite najmä pohodlie, keďže užívateľ nemusí pripájať žiaden konektor. Takéto nabíjanie funguje aj s pomerne vysokou účinnosťou a to až cez 90% pri relatívne vysokých nabíjacích výkonoch. Dá sa považovať aj za ekologické nakoľko k nemu nepotrebujeme žiadne káble ani konektory, čo znižuje aj riziko úrazu elektrickým prúdom. Nevýhodou je však stále veľmi vysoká obstarávacia cena oproti klasickým spôsobom nabíjania. Problémom môže byť aj kompatibilita medzi rôznymi značkami elektromobilov. Keďže vieme, že len nabíjacích konektorov existuje väčšie množstvo a v tomto prípade by sa musel používať jednotný nabíjací štandard naprieč všetkými značkami automobilov.

Dynamické indukčné nabíjanie

Koncovým a najideálnejším riešením by bolo dosiahnutie bezkontaktného nabíjania počas jazdy. Princíp je v podstate úplne rovnaký ako pri statickom indukčnom nabíjaní. V ceste, tesne pod asfaltom, sú zabudované cievky, ktoré generujú elektromagnetické pole počas celého úseku cesty (Obr. 5).



Obr. 5 Dynamické indukčné nabíjanie [9]

Elektromobil, ktorý je vybavený vhodným prijímačom, tak dokáže odoberať energiu z týchto cievok a dobíjať batériu. Spoločnosť Stellantis sa už pokúsila v Taliansku vytvoriť úsek dlhý približne jeden kilometer, kde potvrdila funkčnosť tohto konceptu. Výsledkom bola porovnateľná účinnosť systému, ako pri bežnom „káblovom“ jednosmernom nabíjaní. Spoločnosť Motor City,

sa dokonca túto technológiu pokúsi vyskúšať aj na verejnej komunikácii v Detroite. Pomôcť by mohla aj nemecká firma Magment GmbH, ktorá pracuje na vývoji magnetizovateľného betónu, ktorý by ešte zvýšil účinnosť takéhoto nabíjania [9].

Táto technológia má ale vysokú obstarávaciu cenu a komplexný problém vybudovania a servisovania takejto infraštruktúry. Okrem toho je opäť potrebné, aby boli všetky elektromobily vhodne vybavené a kompatibilné pre takéto nabíjanie. Problémom by mohlo byť aj zabrániť tomu, aby sa takáto nabíjacia sieť nepreťažila najmä pri hustej doprave a nabíjaní nákladných vozidiel, autobusov a ďalších elektromobilov so zvýšenou spotrebou elektrickej energie.

D. Cenové porovnanie nabíjania rôznymi spôsobmi

Okrem voľby miesta a rýchlosti nabíjania, má majiteľ elektromobilu možnosť si vybrať spôsob nabíjania aj podľa ceny. Túto voľbu majiteľ spaľovacieho vozidla nemá, pretože ceny na čerpacích staniciach sú zväčša veľmi podobné.

Najlacnejším spôsobom nabíjania elektromobilu je nabíjanie domácou nabíjačkou. Podľa dát spoločnosti ZSE (pri cenách elektrickej energie v roku 2023) zaplatil užívateľ, ktorý nabíjal elektromobil doma, za 100 km jazdy len 2,88 €, ak uvažujeme priemernú spotrebu elektromobilu 18 kWh/100 km. Zákazník má aj možnosť nabíjať v noci za zníženú tarifu, kde môže klesnúť cena až na 1,62 € za 100 km jazdy [10].

Avšak veľká časť obyvateľov žije v bytoch, kde nemajú možnosť nabíjania elektromobilu doma a sú odkázaný len na verejné nabíjacie stanice. Tú sa ceny líšia nie len od spôsobu nabíjania, ale aj poskytovateľa a hlavne od toho, či máme zaplatený nabíjací program pri ktorom vychádza pri určitom odbere výhodnejšia cena. Ako vidíme na Obr. 6, najlacnejšie nabijeme elektromobil na verejnej nabíjacej stanici pri pomalšom AC nabíjaní.

POROVNANIE NABÍJACÍCH STANÍC

	Typ nabíjania	Cena za kWh	Náklady na 100 km jazdy	Úspora nákladov oproti spaľovaciemu vozidlu
Program ZSE Drive Eco (0€ mesačne)	AC (pomalé) nabíjanie	0,39 €	7,02 €	28 %
	DC nabíjanie výkonom 50 kW	0,49 €	8,82 €	9 %
	Ultra nabíjanie 150 kW	0,59 €	10,62 €	-9 %
Program ZSE Drive Partner (40 kWh za 11,90€ mesačne)	AC (pomalé) nabíjanie	0,29 €	5,22 €	46 %
	DC nabíjanie výkonom 50 kW	0,29 €	5,22 €	46 %
	Ultra nabíjanie 150 kW	0,39 €	7,02 €	28 %
Program ZSE Drive Flat (voľných 400 kWh za 89€ mesačne)	AC (pomalé) nabíjanie	0,29 €	5,22 €	46 %
	DC nabíjanie výkonom 50 kW	0,29 €	5,22 €	46 %
	Ultra nabíjanie 150 kW	0,29 €	5,22 €	46 %
Program GreenWay Energia Standard (0€ mesačne)	AC (pomalé) nabíjanie	0,42 €	7,56 €	22 %
	DC nabíjanie výkonom 50 kW	0,65 €	11,70 €	-20 %
	Ultra nabíjanie 150 kW	0,78 €	14,04 €	-44 %
Program GreenWay Energia PLUS (30 kWh za 9,90€ mesačne)	AC (pomalé) nabíjanie	0,32 €	5,76 €	41 %
	DC nabíjanie výkonom 50 kW	0,54 €	9,72 €	0 %
	Ultra nabíjanie 150 kW	0,66 €	11,88 €	-22 %
ProgramGreenWay Energia MAX (voľných 100 kWh za 29,90 € mesačne)	AC (pomalé) nabíjanie	0,26 €	4,68 €	52 %
	DC nabíjanie výkonom 50 kW	0,46 €	8,28 €	15 %
	Ultra nabíjanie 150 kW	0,52 €	9,36 €	4 %
Program ejoin GO (0 € mesačne)	AC (pomalé) nabíjanie	0,29 €	5,22 €	46 %
	DC nabíjanie výkonom 50 kW	0,39 €	7,02 €	28 %
	Ultra nabíjanie 150 kW	-	-	-
Slovnaft Plugee (registrovaný užívateľ)	AC (pomalé) nabíjanie	0,29 €	5,22 €	46 %
	DC nabíjanie výkonom 50 kW	0,54 €	9,72 €	0 %
	Ultra nabíjanie 150 kW	0,54 €	9,72 €	0 %
Tesla Supercharger	Nižšia tarifa	0,55 €	9,90 €	-2 %
	Vyššia tarifa	0,66 €	11,88 €	-22 %
Lidl	AC (pomalé) nabíjanie	0,24 €	4,32 €	56 %
	DC nabíjanie výkonom 50 kW	0,39 €	7,02 €	28 %
Nabíjanie doma ZSE Energia	Bežná tarifa D2	0,16 €	2,88 €	70 %
	Nízka tarifa D4 (+zvýhodnená tarifa Elektrina pre Elektromobil)	0,09 €	1,62 €	83 %
Náklady na 100 km diesel (cena 1,62 eur/l, spotreba 6 l/100 km)			9,72 €	
Náklady na 100 km benzín (cena 1,5 eur/l, spotreba 7 l/100 km)			10,50 €	

Obr. 6 Porovnanie cien za nabíjanie v rôznych nabíjacích staniciach [10]

V závislosti od toho, či máme zakúpený program, sa ceny u rôznych spoločností pohybujú v rozmedzí 4,32 € až do 7,02 €. Najčastejšia cena je ale cena okolo 5 € na 100 km jazdy. Pri rýchlejšom DC nabíjaní začínajú ceny už pri 5,22 €, avšak je potrebné mať predplatený nabíjací program. Ako vidíme na Obr. 6, pri tomto nabíjaní sa už ceny líšia omnoho viac v porovnaní od poskytovateľov. Najdrahšia je firma GreenWay, ktorá ponúka nabitie elektromobilu za cenu 11,7 € na 100 km. Za ultrarýchle nabitie elektromobilu s výkonom až 150 kW si musí užívateľ priplatiť ešte viac, nakoľko ceny siahajú až do 14,04 € na 100 km. Pri programe ZSE Drive Flat sa ceny za rôzne spôsoby nabíjania nelíšia, avšak tento program stojí 89 € mesačne a teda oplatí

sa vodičom s väčším nájazdom kilometrov a tak aj väčšou spotrebou elektrickej energie. Zaujímavosťou je, že takmer vo všetkých prípadoch vidíme ekonomickú úsporu oproti jazdeniu na spaľovacom vozidle. Pri nafte zaplatíme 9,72 € na 100 km a pri benzíne 10,5 €. Teda v porovnaní s domácim nabíjaním elektromobilu, kde zaplatíme za 100 km jazdy len 2,88 €, môže byť úspora aj viac ako 70% [10].

Okrem spôsobu nabíjania vplýva na relatívne veľký rozdiel v cene aj krajina v ktorej nabíjame elektromobil. Napríklad v Írsku je možné nabiť elektromobil ako za 6 € tak aj za 30 €. Teda rozdiely v cenách sú omnoho väčšie ako na Slovensku. Taktiež je zaujímavé, že najdrahšie si nabijeme elektromobil v krajinách ako Dánsko alebo Belgicko, ktoré sa považujú za ekonomicky aj ekologicky veľmi vyspelé. Je to spôsobené tým, že v týchto krajinách je na cestách väčší podiel elektromobilov a tomu aj lepšie prispôsobená infraštruktúra. Naopak najlacnejšie nabijeme elektromobil v krajinách ako Kosovo, Gruzínsko, Srbsko, kde je elektromobilita na veľmi nízkej úrovni. Z tohto trendu sa dá teda predpokladať, že so stúpajúcim počtom elektromobilov a lepšou nabíjacou infraštruktúrou, sa bude postupne zvyšovať aj cena nabíjania elektromobilov. Na Obr. 7 vidíme, že Slovensko sa umiestnilo cenami v priemere, čo približne odpovedá aj rozvinutosti elektromobility u nás [11].



Obr. 7 Porovnanie cien za nabíjanie elektromobilov v Európe [12]

III. INTELIGENTNÉ NABÍJANIE ELEKTROMOBILOV

Najväčším problémov počas nabíjania elektromobilov pri ich masívnom rozšírení je možné preťaženie elektrickej siete. To môže nastať, ak sa začne nabíjať plným výkonom súčasne príliš veľké množstvo elektromobilov napr. počas energetickej špičky.

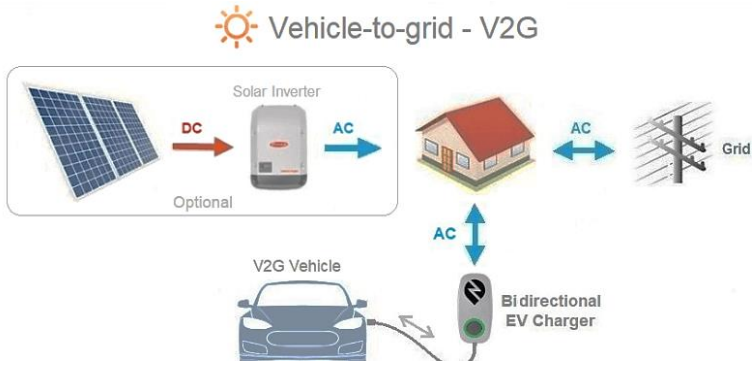
Jedným z riešení je prechod väčšieho počtu majiteľov elektromobilov na denné nabíjanie a nabíjanie vozidiel v práci alebo na verejných nabíjačkách, teda mimo energetickej špičky. Ak sa budú elektromobily nabíjať v neskorých ranných a skorých popoludňajších hodinách, keď má sieť prebytok solárnej kapacity, ktorá sa nevyužíva, bude zaťaženie elektrickej sústavy menšie a

nebude tak potrebné skladovanie prebytočnej solárnej elektrickej energie. Už v súčasnosti ponúka napríklad Kalifornská spoločnosť PG&E majiteľom elektromobilov elektrinu za 25 centov za noc, teda od polnoci do 7.00 h a za 36 centov od 7.00 h do 14.00 h [13].

Problém je, že pre majiteľa elektromobilu je nekomfortné nabíjať elektromobil v presne definovaný čas, kedy to je pre elektrickú sieť výhodné. Mohol by síce ušetriť finančné prostriedky, ale musel by v definovaný čas prísť k elektromobilu a zapojiť ho do nabíjačky. Taktiež by mohol nastať problém, kedy sa rozhodne väčší počet majiteľov elektromobilov zapojiť elektromobil do nabíjačky v približne podobný čas, čím tak opäť dôjde k veľkým nárokom na elektrickú sieť. Riešenie teoreticky ponúkajú technológie, ktoré dokážu zabezpečiť tok elektrickej energie aj z batérie elektromobilu spolu s technológiami inteligentného a plánovaného nabíjania.

A. Technológia V2G

Skratka „V2G“ znamená „Vehicle to grid“, teda v preklade „z vozidla do siete“. Názov tejto technológie výstižne popisuje aj jej funkciu a to práve možnosť prenášať energiu aj z batérie elektromobilu späť do elektrickej siete. Ako vidíme na Obr. 8, ďalšou možnosťou využitia je v prípade slnečného počasia nabíjanie elektromobilu zo solárnych panelov. Tak dokážeme efektívnejšie využiť elektrickú energiu získanú zo solárnych panelov, keďže práve vo chvíli slnečného počasia nemusíme mať pripojenú dostatočnú záťaž v domácnosti a v danej chvíli nebude možnosť ani odovzdať energiu do siete. Vtedy môžeme využiť batériu elektromobilu a nabiť ju. Následne v čase energetickej špičky vieme pomocou nabíjačky s technológiou V2G odovzdať energiu do siete a elektromobil zase naspäť nabiť napríklad v noci, v čase nižšieho zaťaženia [14].

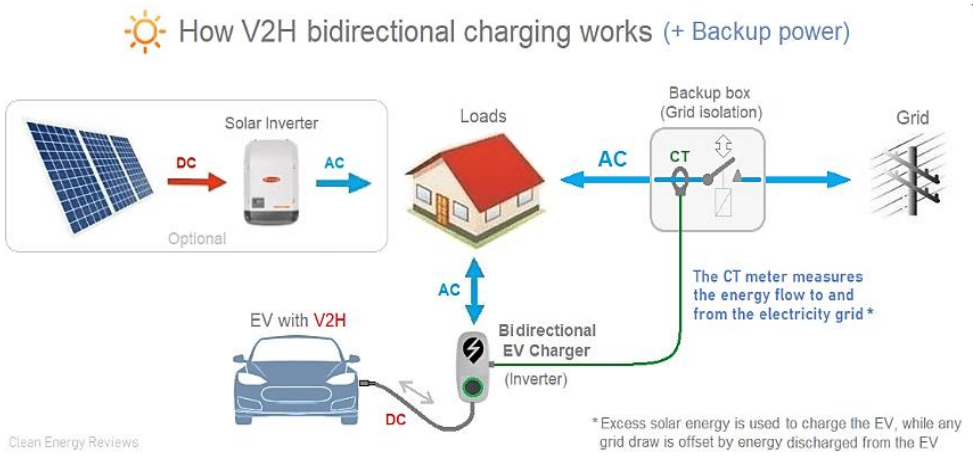


Obr. 8 Technológia V2G [14].

Jedným z dôvodov prečo sa technológia Vehicle-to-grid zavádza veľmi pomaly je skutočnosť, že normy týkajúce sa V2G sú náročné a zložité, pretože zahŕňajú reguláciu výkonu, bezpečnosti a elektrických požiadaviek pri dodávaní energie späť do siete.

B. Technológia V2H

Technológia V2H (Obr. 9) je veľmi podobná predošlej, avšak energia zo solárnych panelov a batérie elektromobilu sa využíva len lokálne na napájanie domácnosti. Práve preto sa nazýva „Vehicle to home“, teda z vozidla do domácnosti.

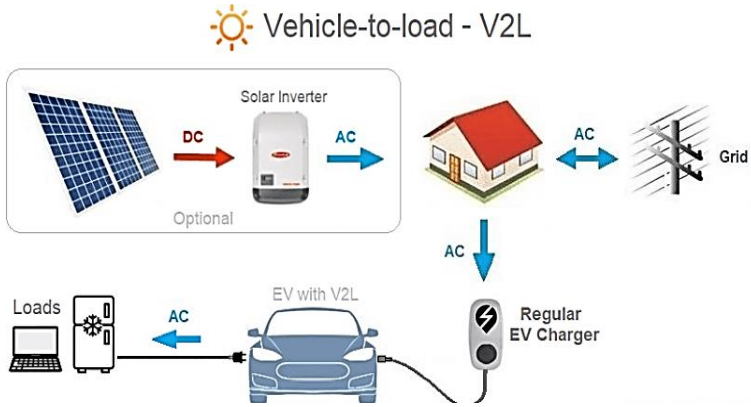


Obr. 9 Technológia V2H [14].

Elektromobil v podstate funguje ako bežný batériový systém v domácnosti a pomáha zvyšovať energetickú sebestačnosť, spolu so solárnou energiou. Okrem kompatibilnej obojsmernej nabíjačky potrebujeme aj zariadenie, ktoré dokáže merať parametre elektrickej siete. V prípade, že záťaž v domácnosti je malá, energia zo solárnych panelov nabíja batériu elektromobilu. Naopak v prípade väčšej záťaže, ako dokážu vyrobiť solárne panely sa energia využije z batérie vozidla. Okrem toho v prípade zistenia poruchy elektrickej siete ju vieme cez stýkač odpojiť a následne napájať domácnosť len z batérie a solárnych panelov [14].

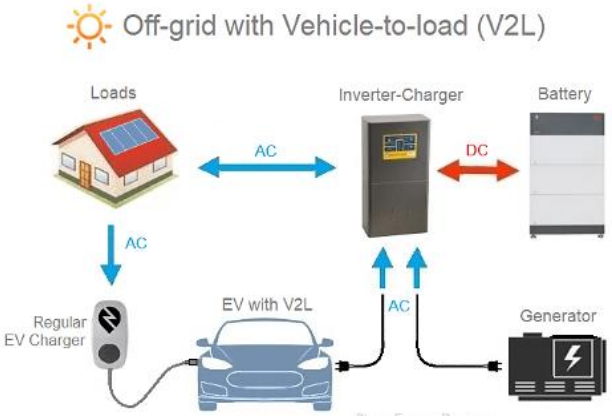
C. Technológia V2L

Najjednoduchším riešením je technológia s názvom „Vehicle to Load“ (Obr. 10), v preklade z vozidla do záťaže. Elektromobil, ktorý disponuje touto technológiou, dokáže pomocou špeciálneho konektora napájať cez zabudovanú obojsmernú nabíjačku bežné domáce spotrebiče. Pri tejto technológií nepotrebujeme disponovať nabíjacou stanicou s možnosťou obojsmerného toku výkonu, avšak takúto možnosť musí ponúkať samotný elektromobil [14].



Obr. 10 Technológia V2L [14].

V súčasnosti to je napríklad Ford-F150 ktorý má možnosť V2L až prostredníctvom štyroch 2,4 kW zásuviek s celkovým výkonom 9,6 kW. Medzi ďalšie vozidlá s technológiou V2L patria aj modely Hyundai IONIQ 5 a Kia EV6, ktoré majú interné a externé zásuvky striedavého prúdu a rad elektrických vozidiel BYD vrátane modelu BYD Atto 3. Vozidlá s V2L môžu pomocou predlžovacích káblov poskytovať záložné napájanie na prevádzku vybraných spotrebičov (Obr. 11).



Obr. 11 Technológia V2L ako záložný zdroj [14].

Napájanie V2L je však možné pripojiť aj priamo k záložnému rozvádzaču alebo dokonca k hlavnému rozvádzaču pomocou špecializovaného prepínača striedavého prúdu. Elektromobil s touto technológiou je možné zapojiť aj do systému solárneho napájania mimo siete, aby sa znížila alebo dokonca odstránila potreba záložného generátora. Väčšina off-grid solárnych systémov obsahuje obojsmerný menič, ktorý môže technicky využívať energiu z akéhokoľvek zdroja striedavého prúdu vrátane vozidla s V2L.

D. Vozidlá disponujúce technológiou V2G

Technológiu V2G má v súčasnosti len niekoľko elektromobilov, medzi ktoré patria novšie modely Nissan Leaf (ZE1), Nissan e-NV200, Mitsubishi Outlander a Mitsubishi Eclipse plug-in

hybrid (Tab. Hoci sú tieto vozidlá k dispozícii už niekoľko rokov, vo väčšine krajín je táto technológia stále v skúšobnej fáze. [26]

S prichádzajúcim prechodom na decentralizovanejší energetický systém založený na obnoviteľných zdrojoch si výrobcovia automobilov začínajú uvedomovať mnohé výhody V2G a implementujú technológiu obojsmerného nabíjania do budúcich elektrických vozidiel (Tabuľka 1). Je známe, že viacerí známi výrobcovia vrátane európskych automobiliek Volkswagen a Volvo vyvíjajú elektrické vozidlá so schopnosťou V2G. Nedávno Hyundai nasadil 25 upravených vozidiel IONIQ 5 na skúšobnú prevádzku Vehicle-to-Grid v holandskom meste Utrecht. Toto mesto sa spolu s poskytovateľom mobility We Drive Solar usiluje o rozsiahle nasadenie technológie V2G v snahe stať sa prvým obojsmerne napájaným regiónom na svete.

Tabuľka 1
Vozidlá podporujúce obojsmerné nabíjanie

Model vozidla	Typ konektora	Podpora V2G	Podpora V2H	Podpora V2L
Nissan Leaf ZE1	Chademo	Áno	Áno	Nie
Mitsubishi Outlander PHEV	Chademo	Áno	Áno	Nie
Hyundai Ioniq 5	CCS	Nie	Nie	Áno (3,6 kW)
KIA EV6	CCS	Nie	Nie	Áno (3,6 kW)
Genesis GB60	CCS	Nie	Nie	Áno (3,6 kW)
BYD Atto 3	CCS	Nie	Nie	Áno (2,4 kW)
BYD Seal	CCS	Nie	Nie	Áno (2,4 kW)
Ford F-150	CSS	Nie	Áno	Áno (9,6 kW)
MG ZS EV (2022)	CSS	Nie	Nie	Áno (2,2 kW)
Volkswagen ID	CSS	Áno	Áno	Nie
Cupra Born EV	CSS	Áno	Áno	Nie

Začiatkom roka 2023 viceprezident spoločnosti Tesla spomenul, že spoločnosť bude integrovať možnosť obojsmerného nabíjania do svojich vozidiel novej generácie, čo naznačuje, že by mohla byť k dispozícii do roku 2025. Stále však neexistuje žiadne oficiálne potvrdenie, že spoločnosť Tesla zahrnie technológie V2G, V2H alebo V2L a to aj napriek prudkému nárastu záujmu a zásadnej úlohe, ktorú tieto technológie nepochybne zohrajú [14].

IV. ZÁVER

Aktuálne predpoklady hovoria, že v roku 2030 bude na Slovensku v pesimistickom predpoklade dvadsaťnásobný počet elektromobilov a v optimistickom až tridsaťpäťnásobný, pričom aj najoptimistickejšie predpoklady nárastu nabíjacích bodov nie sú dostatočné pre dosiahnutie vnímania elektromobilov ako bezproblémovej alternatívy k spaľovacím motorom. Najväčším problémom je nabíjanie elektromobilov počas energetickej špičky, kde je jedným z možných riešení rozdelenie nabíjacieho výkonu elektromobilov mimo energetickú špičku a zároveň využitie kapacity batérie elektromobilov práve v energetických špičkách. Napríklad priemerný elektromobil má kapacitu batérie 60 kWh, čo je šesťkrát viac ako typická domáca solárna batéria a približne trikrát viac energie, ako spotrebuje priemerná domácnosť za deň, čo je približne 20 kWh. Takže len jeden elektromobil by mohol potenciálne napájať domácnosť niekoľko dní. Takéto technológie už v súčasnosti existujú. Základnou je technológia so skratkou „V2G“ teda „Vehicle to Grid“. Tak ako napovedá názov, táto technológia dokáže zabezpečiť okrem nabíjania elektromobilu zo siete aj spätný tok energie z batérie elektromobilu do siete. Vďaka tejto technológii dokážeme efektívnejšie využiť elektrickú energiu získanú zo solárnych panelov, keďže práve vo chvíli slnečného počasia nemusíme mať pripojenú dostatočnú záťaž v domácnosti a v danej chvíli nebude možnosť ani odovzdať energiu do siete. Následne vieme využiť batériu elektromobilu a nabiť ju. Potom v čase energetickej špičky vieme pomocou nabíjačky s technológiou V2G odovzdať energiu do siete a elektromobil zase naspäť nabiť napríklad v noci, v čase nižšieho zaťaženia. Podobne vieme využiť aj technológiu „Vehicle to home“, ktorá sa využíva ale len lokálne pre napájanie domácnosti alebo „Vehicle to Load“ kde batéria elektromobilu slúži ako záložný zdroj pre vybrané spotrebiče.

POĎAKOVANIE

Táto práca bola podporená projektom vedeckej agentúry KEGA podľa zmluvy číslo: 032TUKE-4/2024.

LITERATÚRA

[1] Shannon Osaka, the Washington Post [online]. 22. September 2023. [Dátum: 25. Október 2023]. Dostupné na internete: <https://www.washingtonpost.com/climate-environment/2022/09/22/its-common-charge-electric-vehicles-night-that-will-be-problem/>.
[2] ACnabíjačka.sk [online]. [Dátum: 9. September. 2023]. Dostupné na internete: <https://www.acnabijacka.sk/blog/nabijacie-stanice >.
[3] Elektromobilita. Ročenka 2023 [online] 18. Apríl. 2023. [Dátum: 9. September. 2023]. Dostupné na internete: <https://www.nextech.sk/library/PDF%20ARCHIV/Elektromobilita/Elektromobilita%202023.pdf>.
[4] Wallbox.com [online]. [Dátum: 25. September. 2023]. Dostupné na internete: <https://wallbox.com/media/graphics-FAQs-05.jpg>
[5] KLAVESNICA [online]. [Dátum: 20. September. 2023]. Dostupné na internete: <https://www.klavesnica.sk/blog/wallbox-co-to-je-a-ako-to-funguje#:~:text=Wallbox%20je%20vlastne%20ak%C3%A1si%20s%C3%BAkromn%C3%A1,sa%20jeho%20bat%C3%A9ria%20%C3%BAplne%20nabila.>.
[6] Gareth Roberts, Fleetnews [online]. 25. Máj 2023. [Dátum: 12. November 2023]. Dostupné na internete: <https://www.fleetnews.co.uk/news/latest-fleet-news/electric-fleet-news/2023/05/25/average-battery-range-for-new-electric-vehicles-almost-300-miles#:~:text=The%20average%20distance%20a%20new,300%20miles%2C%20new%20analysis%20suggests.
[7] EV Industry Blog [online]. [Dátum: 15. September. 2023]. Dostupné na internete: <https://evchargingsummit.com/blog/everything-you-need-to-know-about-wireless-ev-charging/>.
[8] Vivek Shah. CarExpert [online]. 21. Jún. 2023 [Dátum: 15. September. 2023]. Dostupné na internete: <https://www.carexpert.com.au/car-news/inductive-car-charging-what-is-it
[9] Prom Wad [online]. 6. Apríl. 2023 [Dátum: 16. September. 2023]. Dostupné na internete: < https://promwad.com/news/wireless-charging-technology#:~:text=Wireless%20charging%20for%20automobiles%20works,charging%20receiver%20inside%20your%20vehicle.>
[10] Rastislav Žembery. Blog ZSE [online]. 17. Február 2024 [Dátum: 19. marec 2024]. Dostupné na internete: <https://www.zse.sk/blog/elektromobilita/je-nabijanie-stale-vyhodnejsie-ako-tankovanie>
[11] Filip Spevar. Môjelektromobil.sk [online]. 25. Augusta 2023 [Dátum: 19. September 2023]. Dostupné na internete: <https://www.mojelektromobil.sk/ake-drahe-je-nabijanie-elektromobilov-na-slovensku-v-porovnani-so-zvyskom-europy/>
[12] Audi Media Center [online]. 9. December 2023 [Dátum: 17. Marec 2024] Dostupné na internete: <https://www.audi-mediacenter.com/en/press-releases/new-audi-charging-hub-opens-in-munich-15570>
[13] Shannon Osaka, the Washington Post [online]. 22. September 2023. [Dátum: 25. Október 2023]. Dostupné na internete: <https://www.washingtonpost.com/climate-environment/2022/09/22/its-common-charge-electric-vehicles - night-that-will-be-problem/>
[14] Jason Svarc, Clean energy reviews [online]. 2. Október 2023. [Dátum: 28. Február 2024]. Dostupné na internete: <https://www.cleanenergyreviews.info/blog/bidirectional-ev-charging-v2g-v2h-v2l>

Test Stand for 3 Phase Asynchronous Electric Motor

Juraj ĎURIŠIN

[1] Department of Technologies in Electronics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

juraj.durisin@tuke.sk

***Abstract* — A test stand for a 3 phase asynchronous electric motor was constructed for educational and test purposes. The stand enables to control the asynchronous motor and to measure its various parameters. It allows to simulate operation of an asynchronous motor in a real automobile.**

***Keywords* — electric vehicle, asynchronous motor, slip, torque, *V/f* ratio**

I. INTRODUCTION

Nowadays, amid rising environmental concerns, there is a growing pressure in Europe or USA to exclude vehicles with combustion engines from transport. Combustion engines, fueled by petrol, gasoline or LPG, pollute the environment. The typical content of an exhaust gas is the following: O_2, H_2O, CO_2, CO, HC and NO_X. Except O_2 and H_2O, all other components harm human health, directly or indirectly. During the past decades, with the advent of catalytic converters and improved combustion process, the content of CO, HC and NO_X in exhaust gases has been noticeably suppressed at negligible level. However, despite the CO_2 content has been suppressed too (Fig. 1), concerns remain. The level of CO_2 emissions, from passenger cars, is estimated to be around 3 billion tons worldwide. CO_2 is a typical greenhouse gas causing rising the surface temperature of the Earth. With the growing temperature, there can be expected significant climate changes, e.g. rising sea level, expansion of deserts or drying out of rivers, all leading to shrinking of farmlands in today's populated areas. Thus, the climate change can cause worldwide famine and poverty of billions of people.

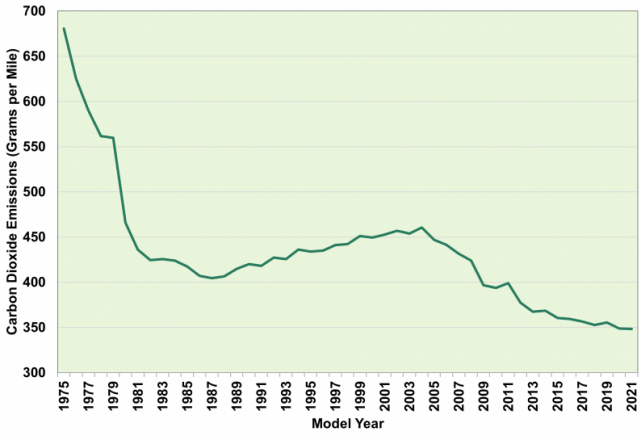


Fig. 1 Average CO_2 emissions from a new vehicle by model year [1].

One of the possible ways to reduce CO_2 emissions is to substitute traditional fuels by new ones, e.g. hydrogen. Hydrogen can be burnt in a combustion engine, after some modifications of its construction are done. However, hydrogen production and storage is very expensive, compared to today's traditional processes. Especially if hydrogen is stored and transported in the liquid form – hydrogen is a very explosive substance if mixed with air.

Another option might be using electric vehicles. The invention of an electric car goes back to the 19th century. The electric cars, such as Flocken Elektrowagen or La Jamais Contente, proved

the concept of an electric vehicle to be viable. The main drawback of these or similar vehicles, designed also later during the 1960's or 1970's, was their limited power and electric range. Therefore, cars powered by combustion engines quickly prevailed and the concept was almost forgotten.

Today, with the above mentioned issue of CO_2 emissions, there is a renewed interest in electric vehicles. With the advent of new materials and technologies during the past decades, the problem of limited power and electric range can be solved. New batteries, typically lithium-ion batteries with high energy and power characteristics, and new more effective electric motors together with improved aerodynamics and lowered friction forces significantly extend power (× 100 kW) and electric range (× 100 km) of these vehicles.

In electric cars, there are used 3 phase alternating current (AC) synchronous or asynchronous electric motors. Synchronous motors exhibit higher torque at lower speeds and are lighter than asynchronous motors. Asynchronous motors show more power than synchronous motors, therefore they are more preferred in electric vehicles.

3 electric phase asynchronous (induction) motors are used to drive various electric vehicles, e.g. Mercedes-Benz EQC 400 4MATIC, Tesla model X or Audi e-tron. The performance of an induction motor does not depend only on its construction, but also how it is controlled. Therefore, to understand fundamental principles of an asynchronous motor, and how to control it (a proper combination of revolutions, torque and power), a test stand for a 3 phase induction motor was built.

II. Asynchronous Electric Motor

An asynchronous or induction motor is a type of electric motor powered by alternating current. A motor can be single electric phase or multiphase, usually it is powered by 3 electric phases.

The principle of an induction motor lies in the rotating magnetic field $\boldsymbol{B}$ created by coils of the stator. The rotating magnetic field $\boldsymbol{B}$ results in an induction of the voltage U_i in the windings or squirrel cage of the rotor, according to the Faraday's law of induction:

$$U_i = -(\boldsymbol{B} \times \boldsymbol{v})l = -Bvlsin\beta \tag{1}$$

, where $\boldsymbol{v}$ is the net rotational velocity of $\boldsymbol{B}$ relative to the rotational velocity of the rotor, l is the length of the rotor wires (bars) exposed to $\boldsymbol{B}$ and β is the angle between the direction of $\boldsymbol{B}$ and $\boldsymbol{v}$.

The voltage across the windings U_i causes a flow of the electric current I across the rotor. Now, with the windings exposed to the stator's magnetic field $\boldsymbol{B}$, the magnetic force $\boldsymbol{F_m}$ is formed and moves, rotates the rotor. The magnitude of the magnetic force $\boldsymbol{F_m}$ is defined as:

$$F_m = BIlsin\alpha \tag{2}$$

, where α the angle between the direction of I (l) and $\boldsymbol{B}$. The direction of $\boldsymbol{F_m}$ is determined by the Fleming's left hand rule.

The rotational speed of the stator (or $\boldsymbol{B}$) is approximately up to 5 % larger than the rotational speed of the rotor. The difference between the speed of the stator (n_{stator}) and rotor (n_{rotor}) is called the slip, where the relative slip is defined as:

$$s = \frac{n_{stator} - n_{rotor}}{n_{stator}} \times 100\ \% \tag{3}$$

Under constant conditions, the length (magnitude) of $\boldsymbol{B}$ does not change. Therefore, to induce U_i in the motor windings at any condition (in the motoring region), $\boldsymbol{B}$ has to travel around the stator faster (asynchronously) than the rotor ($v \neq 0$ and $s \neq 0$). Thus, the relative slip s is the fundamental operational condition of an induction motor [2].

The induced torque of an asynchronous motor τ_{ind} results from its induced mechanical power P_{mech} and angular frequency of the rotor ω_{rotor} according to:

$$\tau_{ind} = \frac{P_{mech}}{\omega_{rotor}} \tag{4}$$

, where $\omega_{rotor}=(1-s)\omega_{stator}$. The output mechanical power of the motor P_{mech_output} on its shaft is defined as:

$$P_{mech\_output} = P_{mech} - P_{fw} - P_{misc} \tag{5}$$

, where P_{fw} are friction and windage losses and P_{misc} are other miscellaneous losses.

In induction motors, the stator consists of multiple Cu coils and the rotor usually consists of a squirrel cage made of Cu or Al. The cage has skewed rotor bars to improve motor properties. There are 2 equivalent circuits of an induction motor. To calculate electrical properties, the transformer model is used. To calculate the demanded induced torque τ_{ind} (mechanical property),

the transformer model is modified by the Thevenin's theorem [2].

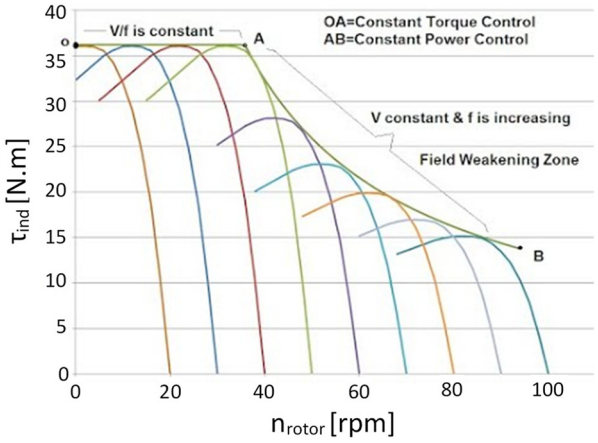


Fig. 2 The induced torque τ_{ind} vs. the speed of the rotor n_{rotor} for various *V/f* ratios [3].

There are various ways how to control the speed and torque of an induction motor. In practice, especially where frequent or sudden speed and torque control is necessary, like in vehicles, the motor construction modification is excluded. Therefore, only input electric parameters can be changed here to vary the motor's behaviour. To do this, the input voltage or a combination of the input voltage and the voltage frequency is modified (*V/f* ratio, Fig. 2). Controlling the speed and torque of an induction motor in a car is necessary, because the demanded value of τ_{ind} vs. n_{rotor} (~vehicle speed) is changed according to the driving conditions: acceleration, deceleration, uphill or downhill ride [2]. In electric automobiles, induction motors show some construction modifications (liquid stator or rotor cooling, etc.) compared to stand alone induction motors, but the basic principle of an asynchronous motor is the same.

To obtain electrical parameters of an asynchronous motor needed to predict (to calculate) and thus to control its performance, the following tests must be performed first: direct current test, no load test and locked (or blocked) rotor test.

III. TEST STAND

The built test stand consists of the following components: main switch, circuit breaker, power meter, 2 two way switches, delta/star switch, Hall sensor, small magnet, induction motor, frequency inverter, cables and a measuring point to measure voltage and current in the motor (Fig. 3). The stand is powered by 3 phase AC.

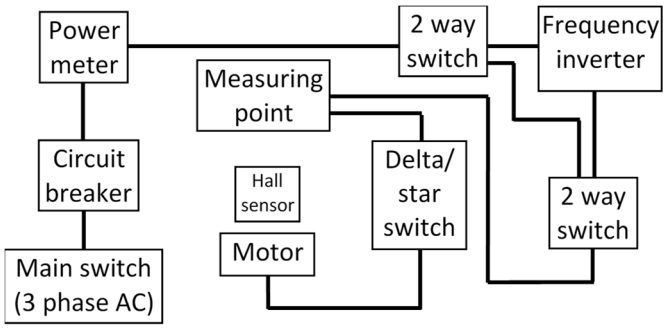


Fig. 3 Layout of the test stand.

The main switch is used to turn on/off the stand. The circuit breaker (OMU system, 1 A, 400 V, type C) breaks the stand circuit if needed. The digital power meter (Eastron SDM72DR) measures the power consumed. It is used for initial calibration; it shows correct values (that can be applied in calculations) only if the frequency inverter is excluded from the circuit. The 2 two way switches are used to include or exclude the frequency inverter from the test. The delta/star switch allows to power the motor in the delta or star configuration. This option is useful especially during the motor's start. The Hall sensor detects revolutions of the motor via a magnet being attached to the motor's shaft.

The 3 phase VYBO Electric 1-AL712-8 motor has, under typical conditions, the following properties: input voltage 230/400 V AC, input voltage frequency 50 Hz, maximal power 120 W, maximally 680 rpm, 8 poles, delta/star configuration and working environment in the range of -35-+50°C [4].

The frequency inverter VYBO Electric A550 Plus 4T enables modification of the input voltage and frequency for the induction motor (Fig. 4). Basic data of the inverter: output voltage 0-400 V AC, output frequency range 0.1-999.9 Hz, output frequency resolution 0.01 Hz. The inverter can be controlled by buttons or by a computer via RS485 [5].



Fig. 4 Front panel of the frequency inverter VYBO Electric A550 Plus 4T [5].

The test stand is depicted in the following figure (Fig. 5).



Fig. 5 Test stand for a 3 phase asynchronous motor.

The built test stand enables to test and to control a 3 phase induction motor thus simulating the operation of an induction motor in a real vehicle.

IV. Conclusion

The constructed test stand is a useful tool for educational and test purposes. It allows students to control and to measure various characteristics of a 3 phase VYBO Electric 1-AL712-8 motor or a similar asynchronous motor of a comparable size, power, etc. The test stand enables to comprehend in practice various methods how to set the speed and torque of an induction motor. Input electric parameters (input voltage and *V/f* ratio) can be modified to vary the motor's performance. In this way one can simulate the action of an asynchronous motor in an actual automobile.

References

[1] https://www.energy.gov/eere/vehicles/articles/fotw-1223-january-31-2022-average-carbon-dioxide-emissions-2021-model-year.
[2] https://abs.cu.edu.tr/Dokumanlar/2017/EEE318/666957712_3_induction_machines.pdf.
[3] https://www.youtube.com/watch?v=t0mXLK_1E5g.
[4] https://vyboelectric.sk/obchod/elektromotor-0-12kw-1al712-8-vybo-electric/.
[5] https://vyboelectric.sk/obchod/frekvencny-menic-standard-e550-4t0007/.

Neurónový regulátor rýchlosti pre jednosmerný cudzobudený motor

[1]*Peter GIROVSKÝ,* [2]*Ján KAŇUCH,* [3]*Jaroslava ŽILKOVÁ*

[1,2,3] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

[1]peter.girovsky@tuke.sk, [2]jan.kanuch@tuke.sk, [3]jaroslava.zilkova@tuke.sk,

***Abstrakt* — V tejto práci je prezentovaný návrh neurónového regulátora pre jednosmerný motor. Cieľom príspevku bolo navrhnúť neurónový regulátor rýchlosti jednosmerného cudzobudeného motora a porovnať ho s klasickým riadením pomocou PI regulátora. V článku je prezentovaný skrátený návrh PI regulátora a neurónového regulátora rýchlosti jednosmerného cudzobudeného motora. Navrhnuté regulátory boli simulačne overené a testované a zároveň prebehla aj ich verifikácia na reálnom zariadení. Verifikácia na reálnom zariadení prebiehala pomocou HIL simulácie.**

***Keywords* — jednosmerný motor, neurónová sieť, riadenie, simulácia, PI regulátor**

I. Úvod

Umelá inteligencia, resp. umelé neurónové siete (UNS), sa v súčasnosti tlačia do popredia pre ich široký rozsah použitia, napríklad v rozpoznávaní vzoriek, triedení vzoriek, navigácii v neznámom teréne, modelovaní, a taktiež v oblasti riadenia. UNS je univerzálny aproximátor, tj. vie s dostatočnou presnosťou opísať vzorový systém. Jej vlastnosti sú vhodné pre modelovanie, ale taktiež riadenie nelineárnych systémov. Veľkou výhodou využitia riadenia pomocou UNS v porovnaní s klasickou metódou návrhu regulácie pohonov je fakt, že UNS nepotrebuje poznať presný matematický popis regulovaného systému. Ak vychádzame z tohto faktu, tak pri návrhu neurónových regulátorov nepotrebujeme poznať presný matematický popis. Pre návrh regulátorov sú potrebné len vzory vstupných a žiadaných výstupných dát.

II. Umelé neurónové siete

Umelé neurónové siete predstavujú matematickú reprezentáciu biologických neurónových sietí, ktorá sa ich snaží napodobniť vo všetkých aspektoch fungovania. Rovnako ako pri biologických neurónových sieťach aj pri ANN, neuróny týchto sietí slúžia pre prijímanie, spracovávanie, uchovávanie, prenášanie a využívanie informácií. Z hľadiska podobnosti funkcie s ľudským mozgom majú ANN schopnosť učiť sa a získané informácie následne uchovávať. Na rozdiel od konvenčných systémov, ktoré pracujú sekvenčne na základe presne určeného algoritmu, ANN spracovávajú informácie paralelne ako závislosti medzi vektorom vstupných a výstupných údajov. Podobnosť ANN s fungovaním ľudského mozgu môžeme zhrnúť do nasledujúcich bodov:

- znalosti sú získavané v procese učenia s učiteľom, alebo bez učiteľa
- naučené znalosti sú uchovávané v medzi neurónových prepojeniach
- informácie sú spracovávané paralelne

Medzi hlavné výhody umelých neurónových sietí môžeme zahrnúť:

- schopnosť učiť sa, t. j. schopnosť adaptovať sa pri riešení konkrétnej úlohy z hľadiska aktuálneho stavu vonkajšieho prostredia
- univerzálnosť, nakoľko pomocou umelých neurónových sietí môžeme modelovať správanie sa sústavy
- robustnosť z hľadiska eliminácie šumu a zmien parametrov systému
- sú veľmi blízke ľudskému vnímaniu okolia a sú vhodné pre prácu so systémami typu „black box“ s neznámymi alebo neúplnými parametrami.

III. Návrh riadenia jednosmerného motora

Neurónový regulátor, ktorý bol porovnávaný s PI regulátorom bol navrhovaný pre

jednosmerný cudzobudený motor.

A. Jednosmerný cudzobudený motor

Pri riešení elektrických a mechanických stavov JCBM vychádzame zo sústavy diferenciálnych rovníc prvého rádu:

$$u_a = R_a i_a + L_a \frac{di_a}{dt} + C\emptyset\omega \quad (1)$$

$$u_b = R_b i_b + L_b \frac{di_b}{dt} \quad (2)$$

$$u_i = C\emptyset\omega \quad (3)$$

$$m_{em} = C\emptyset i_a \quad (4)$$

$$m_{em} = M_Z + J\frac{d\omega}{dt} \quad (5)$$

Konštanta *C* predstavuje konštrukčnú konštantu jednosmerného motora:

$$C = \frac{p' N_2}{2\pi a} \quad (6)$$

kde: *p'* - je počet pólových párov
N_2 - je počet vodičov kotvy
a - je počet paralelných vetiev vinutia kotvy

Následne pre ustálený stav jednosmerného cudzobudeného motora, kedy sa prúd kotvy, budiaci prúd ani uhlová rýchlosť motora nemenia, môžeme písať rovnice v tvare:

$$U_a = R_a I_a + C\emptyset\omega \quad (7)$$

$$U_b = R_b I_b \quad (8)$$

$$U_i = C\emptyset\omega \quad (9)$$

$$M_{em} = C\emptyset I_a \quad (10)$$

$$M_{em} = M_Z \quad (11)$$

Matematický model jednosmerného cudzobudeného motora vychádza z diferenciálnych rovníc (1)-(5), ktoré popisujú správanie sa motora počas dynamických dejov. Z uvedených rovníc vyplýva, že jednosmerný cudzobudený motor predstavuje z hľadiska regulácie lineárnu sústavu s dvoma riadiacimi veličinami (napätie kotvy a napätie budenia). Výstupnou veličinou je uhlová rýchlosť, prípadne moment alebo prúd motora. Úpravou uvedených rovníc pomocou Laplaceovej transformácie dostávame rovnice JCBM v tvare:

$$U_a = R_a I_a + L_a I_a . s + C\emptyset\omega \quad (12)$$

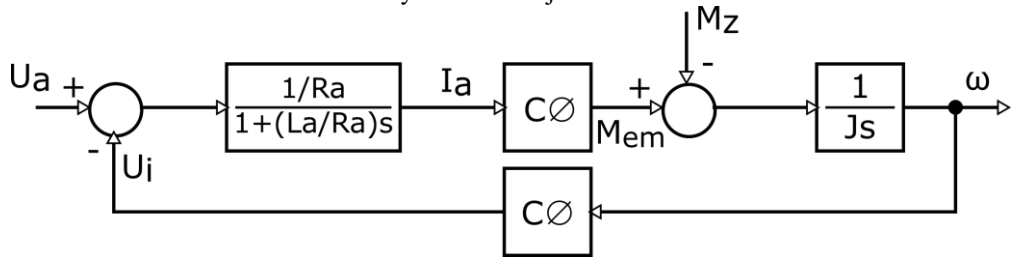
$$U_b = R_b I_b + L_b I_b . s \quad (13)$$

$$U_i = C\emptyset\omega \quad (14)$$

$$M_{em} = C\emptyset I_a \quad (15)$$

$$M_{em} = M_Z + J\omega . s \quad (16)$$

Za predpokladu, že uvažujeme konštantné budenie motora, výsledná simulačná schéma môže byť zjednodušená, pričom jednosmerný cudzobudený motor bude riadený iba napätím kotvy. Simulačná schéme JCBM s konštantným budením je zobrazená na Obr. 1.



Obr. 1 Simulačný model jednosmerného cudzobudeného motora s konštantným budením.

Parametre jednosmerného cudzobudeného motora:

$U_a = 42V$ - *Napätie kotvy*
$I_a = 6A$ - *Prúd kotvy*
$n_{max} = 4000rot/min$ - *Maximálne otáčky motora*
$R_a = 0.678R$ - *Odpor kotvy*

$L_a = 0.002H$ - *Indukčnosť kotvy*
$J = 0.000436kg.m2$ - *Moment zotrvačnosti motora*
$P = 250W$ - *Výkon*
$K_e = 0.0979Vs/rad$ - *Magnetizačná konštanta*

B. Návrh PI regulátora pre JCBM

Pri návrhu riadenia rýchlosti JCBM motora sme postupovali klasickými metódami, a to pomocou kritéria optimálneho modulu a kritéria symetrického optima, ktoré sa využívajú na návrh regulátorov rýchlosti a prúdu. V našom prípade pôjde o kaskádne zapojenie regulátorov.

Návrh spojitého regulátora prúdu pomocou kritéria optimálneho modulu.
Návrh podľa KOM možno popísať v niekoľkých krokoch:

1. Kompenzujeme väčšiu časovú konštantu T_2, čo je prvé pravidlo KOM. Vo väčšine prípadov pri elektrických pohonoch je to časová konštanta T_a.
$$T_{Ri} = T_a = 0{,}0029 \tag{17}$$

2. Výpočet zosilnenia PI regulátora, pre ktorý platí všeobecný vzťah.
$$K_{Ri} = \frac{T_2}{2T_1K} = \frac{T_a}{T_nK_a} = 4 \tag{18}$$

Zavedená náhradná časová konštanta T_n je konštanta prúdového obvodu. Jej hodnotou určíme, za aký čas dosiahne žiadaný prúd hodnotu 0,63-násobku referenčnej hodnoty prúdu. Konštantu volíme tak, aby nebola rýchlejšia ako vzorkovanie vzhľadom na možnosti hardvéru.

3. Výpočet zosilnenia PI regulátora, pre ktorý platí všeobecný vzťah.
$$T_{ii} = \frac{T_{Ri}}{K_{Ri}} = 0{,}00074 \tag{19}$$

Návrh regulátora podľa KOM zabezpečí prekmit výstupnej hodnoty prúdu menší ako 5% a dobu ustálenia rovnú cca 9-násobku menšej časovej konštanty.

Návrh spojitého regulátora rýchlosti pomocou kritéria symetrického optima.
Keďže ide o kaskádne zapojenie regulátorov rýchlosti a prúdu vychádzame z faktu, že regulátor rýchlosti je nadradený regulátoru prúdu, resp. regulátor prúdu je podradeným regulátorom. Táto štruktúra je v praxi najpoužívanejšia. Návrh podľa KSO možno popísať v niekoľkých krokoch, pričom uvažujeme, že menšia časová konštanta $T_1 = J$ a väčšia časová konštanta $T_2 = T_n$.:

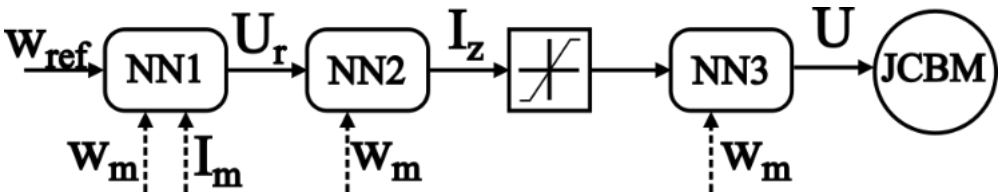
1. Kompenzácia 4-násobku menšej časovej konštanty konštantou T_R.
$$T_{R\omega} = 4T_1 = 4J = 0{,}0017 \tag{20}$$

2. Zosilnenie regulátora rýchlosti je dané:
$$K_{R\omega} = \frac{T_n}{2JK_e} = 5{,}857 \tag{21}$$

3. Integračná časová konštanta je potom daná ako:
$$T_{i\omega} = \frac{T_{R\omega}}{K_{R\omega}} = 0{,}0002978 \tag{22}$$

C. Návrh neurónového regulátora pre JCBM

Princíp riadenia rýchlosti je veľmi podobný ako u klasickej štruktúry s PI regulátormi. Do riadenia vstupuje želaná hodnota uhlovej rýchlosti. Pre regulátor NN1 platí, že jeho vstupom je želaná hodnota uhlovej rýchlosti, keďže je spätno-väzobne zaväzbený do vstupu $\omega(k-1)$ je privádzaná aktuálna hodnota uhlovej rýchlosti z regulovaného PMDC motora. Výstupom NN1 je potom hodnota riadiaceho napätia U_r, ktoré zodpovedá žiadosti uhlovej rýchlosti. Toto riadiace napätie U_r by vyvolalo bez ďalšej regulácie v motore nedovolený prúd. Keďže ho je potrebné obmedziť, musíme ako podradený regulátor zaradiť regulátor prúdu NN2.Tento regulátor na základe vstupu U_r z regulátora NN1 a taktiež vstupu uhlovej rýchlosti vygeneruje žiadanú hodnotu prúdu I_z.



Obr. 2 Principiálna štruktúra nevrhnutého neurónového regulátora.

Neurónová sieť NN1 predstavuje inverzný model JCBM motora. Výstupom inverzného modelu JCBM je napätie kotvy U_r, pričom vstupom je uhlová rýchlosť a jej oneskorená vzorka, a taktiež prúd kotvy I_m a jeho oneskorené vzorky. Pripojením skutočnej rýchlosti na vstup oneskorenej vzorky rýchlosti *w(k-1)* dosiahneme spätno-väzobné zapojenie regulátora rýchlosti. Štruktúra siete NN1 je [6 6 1].

Neurónová sieť NN2 predstavuje prúdový model JCBM. Výstupom tejto siete je želaná hodnota prúdu I_z. Táto hodnota je vygenerovaná sieťou NN2 na základe žiadosti uhlovej rýchlosti z NN1. To znamená, že nie je obmedzená a pre ochranu motora je potrebné túto hodnotu obmedziť na prípustnú hodnotu. V prípade JSCB motorov je táto hodnota prúdu rovná maximálne $2I_N$.Štruktúra siete NN2 je [6 6 1].

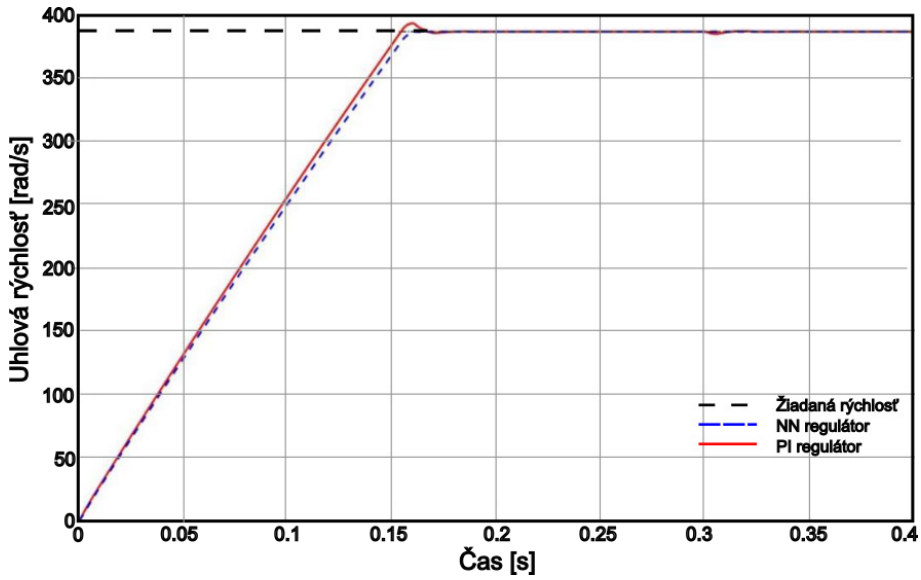
Posledná neurónová sieť NN3 predstavuje rovnako inverzný model JCBM motora. Výstup z neurónovej siete NN3 je hodnota napätia *U*, t.j. hodnota riadiaceho napätia pre napájanie motora po obmedzení prúdu motora. Štruktúra NN3 je [6 9 1].

IV. Testovanie nevrhnutých regulátorov

Navrhnutý PI a neurónový regulátor boli testované najskôr simulačne a následne boli návrhy regulátorov verifikované aj na reálnom motore.

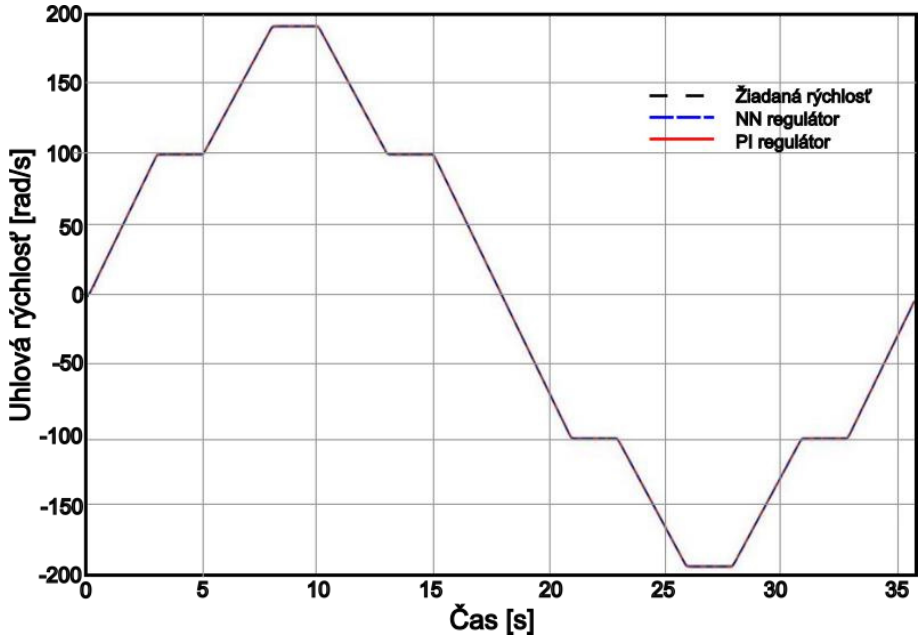
A. Simulačné overenie regulácie JCBM

Simulačné overenie navrhnutých regulátorov prebiehalo v programe MATLAB/Simulink, kde boli porovnávanie výsledky regulácie s PI regulátorom a neurónovým regulátorom. Výsledky porovnania jednotlivých regulátorov sú zobrazené na nasledujúcich obrázkoch.



Obr. 3 Odozva regulácie na žiadanú skokovú zmenu rýchlosti.

Na Obr. 3 je odozva regulátorov v kaskádnom zapojení na skokovú zmenu na nominálnu uhlovú rýchlosť. Ako je vidieť z priebehu, neurónový regulátor narastá rýchlejšie, ale má väčší prekmit, ako PI regulácia. Ak považujeme ustálený stav v rozmedzí ±2%, tak sa neurónová regulácia ustáli v čase 0.151_s a PI regulácia v čase 0,1545s. V čase 0,3s bol motor zaťažený momentom $0{,}2.M_N$.



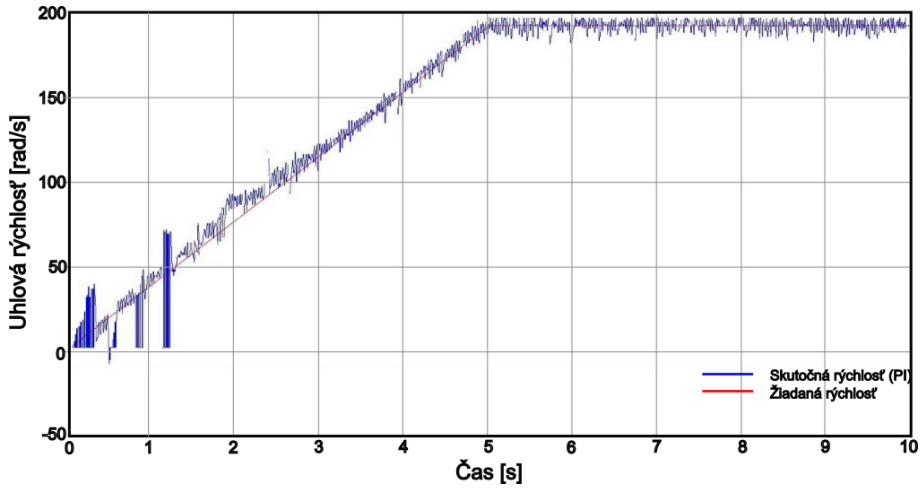
Obr. 4 Odozva regulácie na zmenu žiadanej rýchlosti.

Na Obr. 4 je zobrazený priebeh otáčok po rampe s reverzáciou riadený kaskádnou štruktúrou. Ako je vidieť z obrázkov 3 a 4, priebehy oboch rýchlostí sa zhodujú.

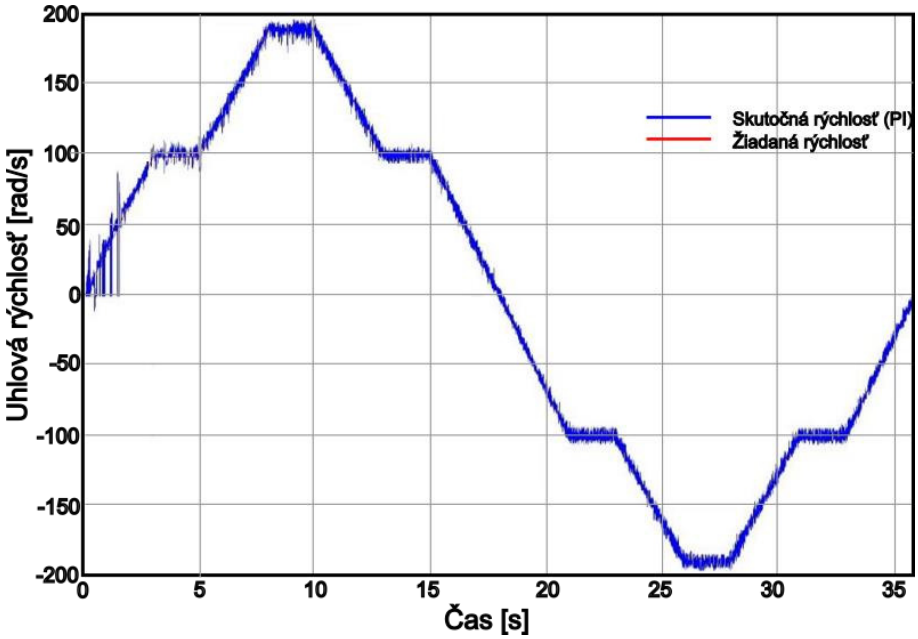
B. Praktická verifikácia navrhnutých regulátorov

Praktická verifikácia bola robená pomocou HIL simulácie s využitím toolboxu Real-Time v MATLABe pomocou karty HUMUSFOT MF624. Na Obr. 5 a Obr.6 sú zobrazené priebehy rýchlosti pri ktorých bol použitý PI regulátor.

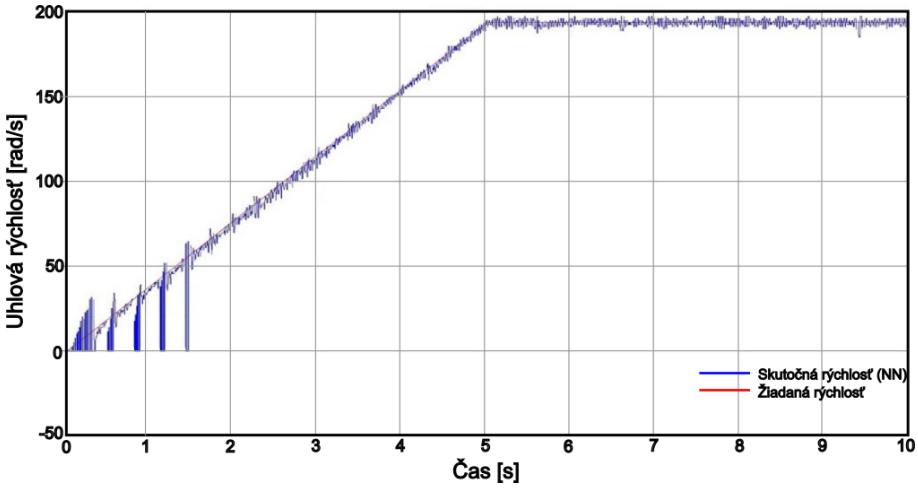
Na Obr. 7 a Obr.8 sú zobrazené priebehy rýchlosti pri ktorých bol použitý neurónový regulátor regulátor.



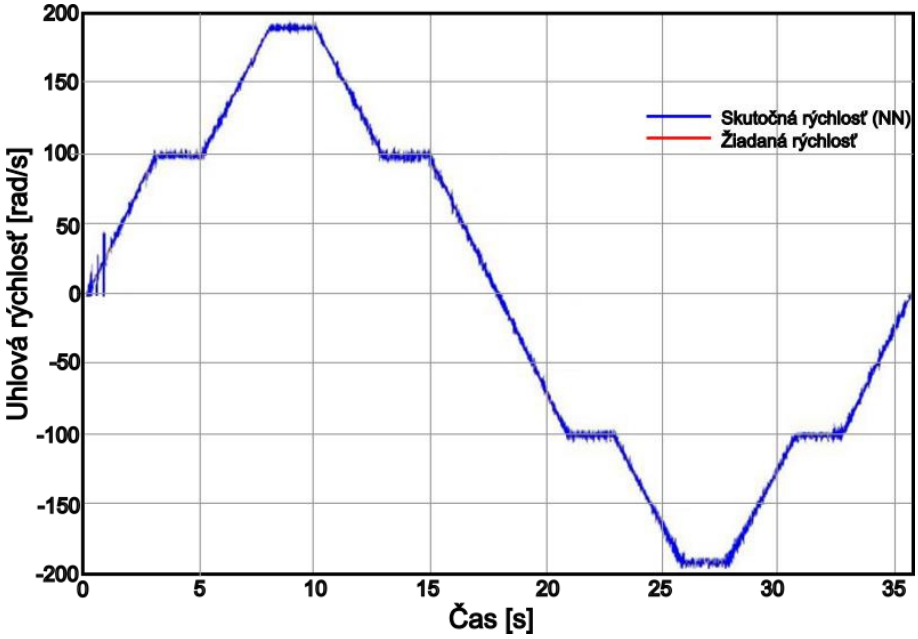
Obr. 5 Rozbeh motora po rampe s PI regulátorom.



Obr. 6 Priebeh otáčok JCBM s PI regulátorom



Obr. 7 Rozbeh motora po rampe s neurónovým regulátorom.



Obr. 8 Priebeh otáčok JCBM s neurónovým regulátorom

V. ZÁVER

Cieľom práce bolo navrhnúť reguláciu rýchlosti jednosmerného motora s permanentnými magnetmi neurónovou sieťou a porovnať ju s klasickou PI reguláciou. Porovnanie jednotlivých regulátorov bolo nielen simulačné, ale aj praktické a realizované pomocou H-mostíka a vstupno-výstupnej výpočtovej karty HUMUSOFT MF624 prepojenej s Real-Time Desktop doplnkom v prostredí SIMULINK. Počas realizácie práce sme sa museli vysporiadať s problémami pri návrhu neurónovej regulácie, a tiež pri praktickej realizácii regulácie. Zo simulovaných priebehov sme zistili, že neurónová sieť ma dobrý potenciál k lepšej dynamike regulovaného systému ako je to pri PI regulácii, avšak za predpokladu, že bude správne natrénovaná a hardvér na jej otestovanie bude dostatočne výkonný. V našom prípade sa podarilo simulačne dokázať, že kaskádna neurónová regulácia má lepšiu dynamiku pri skokovej žiadosti otáčok. V praktickej realizácii nebolo možné overiť skokovú odozvu systému vzhľadom na hardvérové obmedzenia, a tak porovnanie bolo realizované len s reguláciou otáčok bez regulácie prúdu. Vzhľadom na tento fakt sme mohli prakticky reguláciu odskúšať len na rozbeh po rampe tak, aby sme neprekročili dovolenú hodnotu záberového prúdu. Z merania sme usúdili, že obe metódy regulácie sú porovnateľné. Celkovo navrhnuté a prakticky realizované regulátory sú plne funkčné a spĺňajú všetky požiadavky, ktoré boli na nich kladené.

POĎAKOVANIE

Tento príspevok vznikol s podporou Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied v rámci projektu VEGA 1/0363/23.

Táto práca bola podporená projektom KEGA 032TUKE-4/2024.

LITERATÚRA

[1] Priddy, L. Kevin; Keller, E. Paul. Artificial Neural Networks – An Introduction. SPIE Bellingham, Washington USA, 2005. ISBN 0-8194-5987-9
[2] G.E. Hinton and J.A. Anderson. Parallel Models of Associative Memory. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1989.
[3] M. Minsky and S. Papert. Perceptrons. An Introduction to Computational Geometry. The MIT Press, Cambridge, MA, 1969.
[4] C.L. Giles and T. Maxwell. Learning, invariance, and generalization in high-order neural networks. Applied Optics 26:4972-4978, 1987.
[5] J. Hertz, A. Krogh, and R.G. Palmer. Introduction to The Theory of Neural Computation. Addison-Wesley Publishing Company, Redwood City, California, 1991.
[6] J.L.Elman and D. Zipser. Learning the hidden structure of speech. Journal of the Acoustical Society of America, 83: 1615-1626, 1988.
[7] M.W. Hirsh. Saturation at High Gain in Discrete Time Recurrent Networks. Neural Networks, 3: 449-453, 1994.
[8] H.-U. Bauer and K.R. Pawelzik. Quantifying the neighborhood preservation of selforganizing feature maps. IEEE Transactions on Neural Networks, 3(4):570-579, 1992.
[9] C. Bouton, M. Cottrell, J.C. Fort, and G. Pagés. Self-organization and convergence of the Kohonen algorithm. In: Probabilités Numériques (Eds. N. Bouleau and D. Talay), INRIA, Paris, France, 163-180, 1991.
[10] P. Vas. Artificial-Intelligence-based electrical machines and drives, Oxford University Press, 1999.
[11] W. Leonhard. Control of Electrical Drives. Springer Berlin Heidelberg, 2001. ISBN 3-540-41820-2
[12] W. Piotr. Dynamics and Control of Electrical Drives. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-20221-6.

Application for design and creation of graphics

[1]*Stanislav GUSKOV,* [2]*Miroslav MURIN ,* [3]*Ondrej KAINZ,*
[4]*Miroslav MICHALKO*

[1,2,3,4]Department of Computers and Informatics, FEEI, Technical University of Košice, Slovak Republic

[1]stanislav.guskov@student.tuke.sk, [2]miroslav.murin@tuke.sk, [3]ondrej.kainz@tuke.sk, [4]miroslav.michalko@tuke.sk

Abstract **– Nowadays, a large number of applications for creating bitmap graphics are complex or paid, which is a disadvantage for the average user. The task of this work is to create an application for the design and creation of graphics, which will be easy to use and will have the possibility of easy expansion of functionalities by users for working with graphics. Before the actual implementation, there was a phase of analysis of already existing solutions, their positives and negatives. Subsequently, the proposal of the future application was made, where we selected the best technologies for our application. The next phase was the implementation itself, which mainly focused on adding all the necessary functionalities and also developing a simple and easy-to-use application. From the functionality side of the system, it has functional options for working with images and also drawing. The result is a user-friendly application for creating graphics.**

Keywords **– Design, graphics editor, drawing, image editing, web application, JavaScript, Java, PostgreSQL, security**

I. Introduction

In today's digital age, every company that has created an application for design and creation of graphics is trying to make as much money as possible from users. The more popular and advanced a design application from any company, the more it charges to use it. But not every user needs a wide range of tools for working with graphics, and therefore does not want to pay for the full version of the application. In response to this situation, there are also free versions of design tools, but they often have a minimal number of features or try to force users to pay for a graphics tool by inserting the company's brand or logo into the graphic solution. Sometimes they can also limit the maximum resolution of work, even denying the user the ability to download their solution or simply limit the number of available features of the application.

Moreover, these graphics tools often have only a version that you have to install on your computer, which takes quite a bit of time. And if you have an old computer or other old device, using the tool can be unpleasant and time-consuming because the device's performance is not sufficient. Even this graphical tool may only support a specific operating system, which will not allow the user to install the tool on their computer if they do not have a supported operating system installed by this application. This is precisely why companies prefer to create applications that can be used through an Internet browser, and users use such applications.

Another rather important problem is that applications for the design and creation of graphics are often aimed at professionals who work with graphics. Therefore, for a person who has never encountered tools for design and work with graphics, it is confusing when they open such an application. Because these design applications have a large number of features, tools, and plugins, which causes an increased amount of time that a person has to spend working on their task. Therefore, a friendly user interface is a vital component for such design applications.

By creating an application for design and creation of graphics, these problems should be solved. The most important goal of implementing this system should be that the system will have a completely open-source base, where any of its functionality can be created by users of the system and added to the framework as a plugin that anyone can add to their project. This will be ensured by the application having its plugin database where users will be able to add their created plugins and share them with all users. One of the other goals of creating the application would be that

the application will be available on any device that supports working on the Internet. In this way, this application will achieve the greatest possible usability and support of a large number of devices, as well as the greatest possible reach among users, because it will be available from any device, anywhere. In addition, the implementation of the application should be aimed at clients who do not need to do complex projects in the graphic sense and do not have experience with such systems. And therefore the application must be as user-friendly as possible. Also, the open-source code will allow any user to create a user interface that they like the most. Or use any interface created by other users. The most basic function of the created design tool would be that the user will be able to insert any number of images and make the composition they need with them without having to pay. In addition, it is important that the application supports different file types related to graphics. An equally important function of the application would be the ability to format embedded images or created graphic units, adjust their dimensions, set their parameters, crop, color the area and more.

II. Design applications

This chapter primarily describes the differences between various design and graphics applications. Each of the mentioned applications aims to introduce innovations in its respective field, but it fails to address many problems, as it does not perceive them as necessary to solve within the given domain. The main issue these apps face is their limited range of options.

A. Photoshop

It is one of the most popular tools for working with graphics or creating a design. The main goal, in our opinion, that Adobe wants to achieve with this product is to provide the most advanced tools possible for professionals, which has earned it an almost flawless reputation among professionals focusing on graphics creation. This is according to the authors of the document [1], which characterizes for whom this tool is intended and also describes its functionalities, advantages, and disadvantages.

This application is installed on your device, which must have one of the compatible operating systems installed, including Windows, macOS, iPadOS, and Android. Considering the complexity of this application and its necessary installation on the device, in our opinion, it is too complicated for the average user. Its user interface is saturated with various tools that Photoshop has accumulated during its development, and therefore the average user will be completely confused by the number of tabs with functions that they will not even know how to use.

Our design solution would need to have a user interface so that the user doesn't feel lost when turning it on for the first time and knows exactly what each function does. For a common user, this design application is expensive, and there is no version available for free download.

B. Microsoft Designer

Another application for designing and creating graphics is Microsoft Designer. This application is part of the wider Microsoft ecosystem and provides a large number of tools for professionals as well as novice designers.

Using this system for design is easy because the application is accessible on the website, thus reaching the widest possible audience. The application is also convenient for users who use Microsoft systems thanks to its inclusion in the ecosystem of this company. In our opinion, it is ideal from the perspective of ordinary users, rather than experts. This ensures the widest possible scope for our application.

Another positive aspect is that this app makes it easy to create your own designs by implementing a large number of pre-made templates and designs. Additionally, the integration of an intelligent system that creates templates based on user-entered text improves the user-friendliness of this system.

We could apply a similar approach to the development of our application. For example, by incorporating existing templates to meet users' current design requirements. Simultaneously, it would be beneficial to create an open database of templates where anyone could share their solutions.

C. Canva

Canva, an online graphic design and creation application, is created by a company with the same name as the app. It offers the possibility to use existing templates and also enables more advanced editing of graphic solutions.

The application provides a wide variety of available templates, but in some cases, it can be a disadvantage that some of them are paid, which limits its usability. On the contrary, the fact that the application provides a large number of ready-made templates does not mean that users won't miss the missing basic tools and functions.

D. Gimp

The Gimp application, created by The Gimp Team, is an example of a system that is gradually developed in collaboration with a community of enthusiasts. The application provides extensibility through integration with a large number of programming languages, including Scheme, Python, Perl, and others. The result is a high level of customization, as evidenced by the large number of scripts and plugins created by the community.

An important aspect of the given system is its open-source code, which opens many opportunities to improve the system, even for those who are not direct creators of the application but are interested in improving it.

The direction of our solution will be oriented towards a similar approach because, in our opinion, it will reach a larger user base. At the same time, we must be careful about the security of our solution. Even with such large corporations as the one that created the Canva tool, incidents occur, as for example, the authors of the document [2] describe. By gaining access to the GitHub of one of Canva's developers, the attacker gained full access to the user base. We will have to pay attention to such situations when developing our application.

III. Technology for web application

Our application for the design and creation of graphics must be developed using modern technologies that comply with current standards. This will allow us to save a lot of time during implementation, solve security challenges, and achieve the longest possible support for selected technologies.

A. Database technology

It is important for every application to store data in a quickly accessible and safe place. Database systems serve to fulfill this task. Currently, there are two paths: relational (SQL) and non-relational (NoSQL) database servers. The authors of the document [3] claim that the difference between relational and non-relational database systems lies in the way they scale. Based on this property and the specific characteristics of given database systems, the authors concluded that non-relational database systems work faster than relational database systems. We decided to stick with relational database systems.

However, now we need to choose a relational database system that would suit our needs. The Oracle database system is not suitable for our solution. There is a free version, but for commercial use, we have to purchase a license. MySQL, on the other hand, has a free version for commercial use, but users have to pay for some additional functions. However, that doesn't bother us because using MySQL is free as long as we don't modify the source code of MySQL itself. Microsoft SQL Server, like Oracle, is not available for free for commercial use. PostgreSQL, like MySQL, is free, but unlike MySQL, we can modify it, rewrite its source code, and more.

We selected the PostgreSQL database. PostgreSQL implements more modern technologies and is therefore better suited for use in a newly created application.

B. Back-end technology

Back-end is part of the three-layer architecture, serving as the middle layer responsible for calculations. Our application will need this layer to compute changes when working with graphics, format images, add new elements and similar tasks. Currently there is a wide range of languages used in implementing the application layer, with the most popular being JavaScript, PHP, Python and Java.

JavaScript is a versatile scripting language that can be used for both the application and presentation layers. When used in the application layer, JavaScript is combined with Node.js. PHP, like JavaScript, is widely used and known for its scripting capabilities. Python, on the other hand, is renowned for its simplicity and code readability, also serving as a scripting language. Java being platform-independent aims to run on any operating system facilitated by the Java Virtual Machine with its source code available freely.

To compare these languages and select the most suitable one for our needs it's essential to consider various factors. PHP and Java have the advantage of being runnable on any operating system. JavaScript also enjoys this advantage but its application layer requires Node.js. Performance-wise,

PHP and Java are similar with Java being considered more stable. JavaScript particularly when combined with Node.js is preferable for websites with a large number of users and increased input-output operations. However, Python may not be suitable for web pages with a high number of requests that perform input/output operations. Java offers a wide range of existing libraries for various functionalities which can be advantageous.

We have chosen Java as the language for our application layer. Java offers extensive possibilities for application design and high-performance calculations, such as graphic image editing, as suggested by the authors of [4].

C. Front-end technology

The user interface is a crucial part of the system, without which the system would be entirely unmanageable. To simplify the implementation process we will utilize pre-existing web frameworks that offer a wide array of tools for web interface development such as React, Angular and Vue.js.

Now, we intend to compare these three frameworks. Each of them is open source and free for commercial use. However, it's worth noting that while we refer to them as frameworks, React is actually a UI library, Vue.js is a progressive framework and only Angular is a full-fledged framework. According to the authors [5], React and Vue.js have minimal differences, with React being praised for its flexibility and responsiveness, while Vue.js boasts a customizable architecture. Additionally, the authors of [6] note performance differences between these frameworks, stating that React has the smallest interaction time, while Angular and Vue.js are comparable. This observation is corroborated by the authors of [7], who also highlight that Vue.js performs exceptionally well when handling static elements.

We have chosen React as our preferred framework. In our view, it is the most suitable choice due to its ease of learning and application development.

IV. Web application security technologies

We need to pay close attention to the security of the application. Without a secure system we won't earn the trust of our users. Our primary focus should be on preventing attackers from accessing user data including passwords and private information.

A. SQL injections

According to the authors of [8], SQL injections are among the most dangerous attacks on web applications. The same authors also provide us with defense methods. One of the simplest methods is to add SQL expressions to a blacklist. However the best defense is to use parameterized queries. This involves not immediately inserting user input as a query into the database but first filtering out any injected SQL code. This is why we do not intend to develop our application solely in Java without utilizing any frameworks. For instance as mentioned by the authors of [9], the Spring Boot framework which we will use for our system already incorporates methods to protect against SQL injections.

B. XSS attacks

XSS is a security flaw that can affect web applications allowing attackers to inject their malicious code into your application. According to the authors of [10] the most common methods of defense against XSS attacks are input and output validation. As mentioned, we will use Spring Boot. The authors of [9] claim that Spring Boot provides protection against XSS attacks by default.

C. CSRF attacks

Cross-site request forgery or CSRF attack is cited by [11] as one of the most common attacks used by intruders. In essence, this type of attack deceives the victim by, for example, using a malicious link, allowing attackers to send or receive data on behalf of the victim. Similar to XSS attacks the authors of [9] state that the Spring Security framework offers options for protection against CSRF attacks.

D. Monitoring system

The monitoring system includes the recording of any event in the system, which according to the authors [12] provides opportunities for analysis and detection of security threats. In the future, using this system, we will be able to detect or prevent newly emerging threats.

V. Application design for graphic design and creation

Without proper design and description of its construction further implementation will likely be complex and time-consuming. In this chapter, we will outline the structure of the database, the user interface, and the methods of communication between the server and the application. We'll also address data processing on two fronts: within the web application and within our Java component.

A. Database design

Our application requires data storage, primarily for users and their associated projects and components. It doesn't necessitate a very complex database structure. As mentioned earlier, we've chosen the PostgreSQL database server. For our application, we'll need to create storage for a few system entities, such as users, projects, and components. We will implement these entities with the necessary attributes. Large files, such as project files, will not be stored in the database; instead, they will be stored on the local machine's file system.

B. User interface design

The user interface is a critical component of the application as it is what the user interacts with. To expedite development we'll utilize existing solutions for certain aspects, such as the Konva library for object manipulation and Bootstrap for pre-made buttons. Our interface is segmented into several key sections each responsible for displaying specific content to ensure user navigation clarity.

The primary section is the login page where users input their credentials to access the application. Upon successful login users are directed to a page displaying their projects where they can create new projects or navigate to existing ones. When a user selects a project they are presented with the project screen containing all the necessary tools for drawing, image editing and other graphic operations. Finally, the component page allows users to manage components within their projects, enabling addition or removal as needed. In figure 1 we see the component shop.

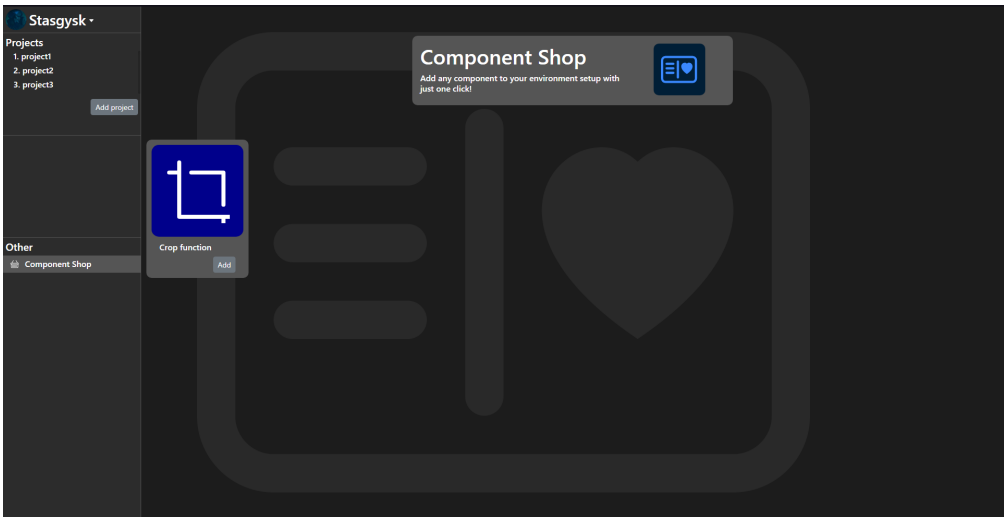


Fig. 1 Component Shop

C. Application structure design

We will use the common web application structure comprising a back-end server which will communicate with the database and the front-end application with the user interacting primarily with the front-end system. This structure is typical for web servers.

The back-end will be built using a Java based solution with Spring Boot providing robust functionality and easy integration. For the front-end we will develop a React application supplemented with additional libraries for enhanced features and usability. As for the database, we have opted for PostgreSQL a reliable and versatile choice that aligns well with our application's needs.

VI. Graphic processing design

To handle the graphic solutions created by users through the interface, we must design a system for transferring these designs into file formats like PNG or JPEG. Given that the user interface isn't suitable for processing these solutions, our task is to establish communication between the user interface and the Java component enabling the transmission of data to be processed.

As a solution we've opted to implement the graphic solution processing on the Java component side. This approach allows us to leverage the capabilities of Java for image processing. Additionally we'll require a web component to facilitate data transmission from the user interface to the Java component for processing.

A. Design of data processing on the web interface side

The main task of the web interface is to help the user visualize his graphic solution, arrange all the elements as he wants and use his chosen effects. It will also deal with the forwarding of all this data to the Java component. One of its several functions is obtaining the coordinates of objects that are shown on the canvas.

To work with a graphic solution on the web component side, we will use the Konva library, it provides great options for working with graphic objects, it is available for React. It is a very simple tool for our needs, it includes a large number of functions. In figure 2 we see the editor.

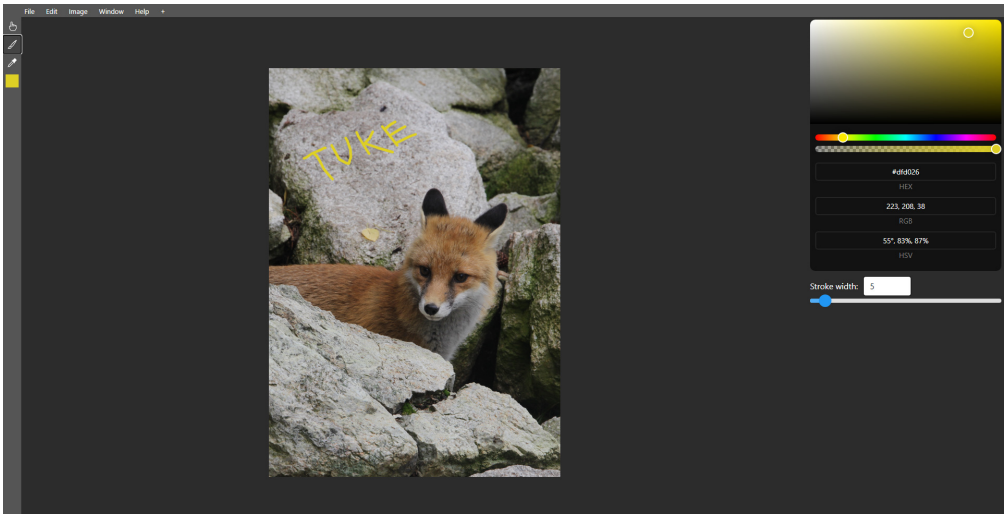


Fig. 2 Image Editor

B. Java component-side data processing design

The Java component will be responsible for processing received data from the web interface. We will ensure this by using already created libraries for processing images and graphic solutions. One such library is, for example, ImageJ, which provides all the functions we need. Also, according to the description of the library's capabilities by the authors, it is the fastest Java library for image processing.

VII. Implementation of an application for creating graphics

The way web applications are developed nowadays is very diverse. We chose Java as the language for the application layer, but Java also provides many different libraries to work with any part of our system. In this chapter, we will describe our selection of system libraries for working with the database, the use of REST services, and also the selection of the library for working with images.

A. Implementation of communication with the database

Communication with the database is a crucial aspect of application development as choosing the right method ensures the safety of user data. Given that we're using Java for the database communication layer we have several options available.

In the absence of frameworks like Spring Boot, Java offers JDBC (Java Database Connectivity). However JDBC requires writing SQL commands directly which poses security risks.

Alternatively, Spring Boot provides enhanced database communication interfaces. Spring Data JPA, for instance, incorporates SQL injection prevention mechanisms such as prepared queries. This approach automatically guards against SQL injection by filtering user input. Given these advantages, we will employ Spring Data JPA in our application.

B. Implementation of REST services

Our application will adopt a standard method of communication known as REST services. To implement this we will utilize the Spring Web MVC Framework on the Java component side, enabling us to create REST endpoints. For the React component side we will employ React Axios which not only provides protection against CSRF attacks but also offers a fast and memory-efficient approach to creating REST services.

C. Implementation of the image creation mechanism

Image data will be processed by the ImageJ library which applies effects to each received object and combines them into a graphic solution. The graphic solution is constructed gradually by adding each object to a byte of storage space, which is then converted into a file and saved on the user's computer.

The processing of the graphic solution is managed using the ImageJ library. It begins by creating a 16-bit image with dimensions provided by the web interface along with the background color specified by the user. Next each object is read and added to the image one by one while being transformed according to the applied effects. This is facilitated by the capabilities of the ImageJ library.

Each modified image is stored in the user's project folder. When the user wishes to view this image the program retrieves the data from the project folder for display.

VIII. Conclusion

The aim of this paper is to analyze existing graphic design applications and propose a solution that addresses their limitations followed by the implementation of the proposed solution. While there are numerous graphic design applications available not all of them meet the user's needs either due to lacking features or being paid services. Our application seeks to resolve these issues by offering a comprehensive set of features including those found in existing solutions and addressing their shortcomings.

Prior to the design phase we conducted an analysis of existing applications focusing on their strengths and weaknesses. Based on this analysis we devised a solution that aimed to address the shortcomings of other applications while incorporating their positive aspects. During the design process we considered potential solutions to common problems and improvements to existing functionalities.

The resulting application for design and graphics creation prioritizes user-friendliness functionality and performance. It is divided into distinct parts each responsible for specific functionalities. The application enables users to create graphic solutions with ease offering tools for drawing image editing and project management. Users can save their creations either locally or on the application server allowing for easy access and retrieval.

User accounts are an integral part of the application. Additionally users can customize their profiles and preferences within the system. A component store is provided offering users a selection of functional components to enhance their graphic design capabilities.

Looking ahead the application has significant potential for further development. Future iterations could include features such as easy integration of user-generated components external plugins and public availability. Continuous improvement and enhancement of existing functionalities should also be prioritized.

Acknowledgment

This work was supported by Cultural and Educational Grant Agency (KEGA) of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the project No. 060TUKE-4/2022.

REFERENCES

[1] P. Bowen and L. Hua, "Application exploration of photoshop computer design software in contemporary art design," in *2021 7th International Symposium on Mechatronics and Industrial Informatics (ISMII)*, 2021, pp. 218–221.

[2] M. H. Nguyen Ba, J. Bennett, M. Gallagher, and S. Bhunia, "A case study of credential stuffing attack: Canva data breach," in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2021, pp. 735–740.

[3] T. Capris, P. Melo, N. M. Garcia, I. M. Pires, and E. Zdravevski, "Comparison of sql and nosql databases with different workloads: Mongodb vs mysql evaluation," in *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, 2022, pp. 214–218.

[4] F. Devai and M. Rahman, "The suitability of the java language for computer-graphics and visualisation research," in *Proceedings Sixth International Conference on Information Visualisation*, 2002, pp. 273–278.

[5] C. M. Novac, O. C. Novac, R. M. Sferle, M. I. Gordan, G. BUJDOSó, and C. M. Dindelegan, "Comparative study of some applications made in the vue.js and react.js frameworks," in *2021 16th International Conference on Engineering of Modern Electric Systems (EMES)*, 2021, pp. 1–4.

[6] R. N. Diniz-Junior, C. C. L. Figueiredo, G. De S.Russo, M. R. G. Bahiense-Junior, M. V. Arbex, L. M. Dos Santos, R. F. Da Rocha, R. R. Bezerra, and F. T. Giuntini, "Evaluating the performance of web rendering technologies based on javascript: Angular, react, and vue," in *2022 XVLIII Latin American Computer Conference (CLEI)*, 2022, pp. 1–9.

[7] R. Ollila, N. Mäkitalo, and T. Mikkonen, "Modern web frameworks: A comparison of rendering performance," *Journal of Web Engineering*, Mar. 2022.

[8] A. Rai, M. M. I. Miraz, D. Das, H. Kaur, and Swati, "Sql injection: Classification and prevention," in *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, 2021, pp. 367–372.

[9] A. Aborujilah, J. Adamu, S. M. Shariff, and Z. Awang Long, "Descriptive analysis of built-in security features in web development frameworks," in *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2022, pp. 1–8.

[10] T. Wang, D. Zhao, and J. Qi, "Research on cross-site scripting vulnerability of xss based on international student website," in *2022 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, 2022, pp. 154–158.

[11] M. S. Rao, B. Kalyani, B. Vathsalya, K. Dhanunjay, and A. L. Narayana, "Cross-site request forgery as an example of machine learning for web vulnerability detection," in *2023 3rd International Conference on Smart Data Intelligence (ICSMDI)*, 2023, pp. 422–426.

[12] D. V. Stepanenko, K. L. Stoychin, and D. V. Shevchenko, "Analysis of operating system event logs when investigating information security incidents," in *2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, 2023, pp. 313–315.

Súčasný stav možnosti nabíjania elektromobilov na Slovensku

[1]*Ján KAŇUCH,* [2]*Peter GIROVSKÝ,* [3]*Jaroslava ŽILKOVÁ*
[1,2,3] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 1/9, 04200 Košice-Sever, Slovenská republika

[1]jan.kanuch@tuke.sk, [2]peter.girovsky@tuke.sk, [3]jaroslava.zilkova@tuke.sk

***Abstrakt* — V tomto príspevku je prezentované riešenie problematiky nabíjania elektromobilov na Slovensku. V prvej časti príspevku je popísané v akom stave sú v súčasnosti možnosti nabíjania elektromobilov, teda analýza aktuálneho počtu nabíjacích staníc, čo veľmi úzko súvisí s rozvojom predaja elektromobilov a teda aj vypracovaním scenára rozvoja nabíjacej infraštruktúry do budúcnosti. Podstatný je aj vhodný výber lokalít, kde sa bude nabíjacia infraštruktúra budovať. V poslednej časti príspevku spracovaný rozbor problémov počas nabíjania elektromobilov pri ich masívnom rozšírení a s tým aj súvisiace aj možné zvýšené nároky na výrobu a rozvod elektrickej energie.**

***Kľúčové slová* — nabíjanie elektromobilov, nabíjacie stanice, nabíjacia infraštruktúra**

I. Úvod

Elektromobilita v doprave na Slovensku je relatívne stále začínajúci a rozvíjajúci sa segment a preto na veľa problémov, ktoré môže jej rozvoj ešte priniesť, nepoznáme riešenia. Jedným z nich je aj problematika nabíjania elektrických vozidiel v husto obývaných oblastiach.

Donedávna bol pre potencionálnych záujemcov jedným z najväčších problémov dojazd elektrických áut. Avšak už v súčasnosti nie je problematické nájsť elektromobily s dojazdom väčším ako 500 kilometrov, čo ponúka relatívne pohodlný dojazd aj pre ľudí, ktorí jazdia s autom dlhšie trasy. Pokiaľ vlastnia garáž alebo rodinný dom, kde môžu elektromobil pohodlne nabiť, nemajú už zásadný problém, ktorý by ich odrádzal od kúpy elektromobilu. Dokazujú to aj dáta Slovenskej asociácie elektrických vozidiel, podľa ktorých je v súčasnosti až 80% nabíjaní realizovaných pomocou domácej nabíjačky a len 20% tvoria verejné nabíjacie body [1]. To potvrdzuje, že po elektromobile v súčasnosti siahajú najmä ľudia, ktorí majú možnosť nabíjať doma. Pre človeka bývajúceho na sídlisku je nabíjanie zásadný problém, ktorý ho môže odradiť od kúpy elektromobilu, keďže v týchto oblastiach je často problém nájsť voľné parkovacie miesto, nie to ešte voľnú nabíjaciu stanicu. Obyvatelia miest tak často nabíjajú svoj elektromobil mimo sídlisk, na rýchlonabíjacích staniciach, aby sa následne rýchlo mohli dopraviť k svojmu cieľu alebo domovu [2].

Podľa Medzinárodnej agentúry pre energiu, v roku 2014 žilo 54% globálneho obyvateľstva v mestských oblastiach. Očakáva sa, že tento počet do roku 2050 vzrastie na 66%. Aj na Slovensku sú tieto čísla veľmi podobné. Dáta z posledného sčítania obyvateľov v roku 2021 ukazujú, že v mestách žije na Slovensku až 53,2% obyvateľov [3]. Vzhľadom na neustále rastúce populácie v meste, je pravdepodobné, že dopyt po ekologickejších alternatívach dopravy, ako sú elektrické vozidlá, bude rásť. Práve v mestách a husto obývaných oblastiach sú najviac citeľné emisie z dopravy pre bežných ľudí, ktorým tak znižujú komfort pre život. Okrem emisií by sa rozšírením elektromobilov zvýšil životný komfort obyvateľov aj vďaka nižšej hlučnosti. Keďže vozidiel v doprave neustále pribúda, je zvlášť dôležité pre mestá a sídliská, aby v nich premával väčší počet elektromobilov na úkor spaľovacích vozidiel. K tomu je ale nevyhnuté vybudovať nabíjaciu infraštruktúru, ktorá zabezpečí pohodlné užívanie elektromobilu aj pre ľudí žijúcich v bytových domoch bez garáže a možnosti nabíjania elektromobilu v domácom prostredí.

Do tejto problematiky sa radia aj turistické destinácie s vysokým prílevom návštevníkov. Zabezpečenie dostatočného počtu verejných nabíjacích staníc na populárnych turistických miestach je preto tiež kľúčové pre uspokojenie potrieb vlastníkov elektromobilov. Tých bude neustále pribúdať a práve možnosť nabitia elektromobilu, môže byť jedným z faktorov, ktoré ovplyvnia výber turistického miesta ktoré navštívia. V súčasnosti sa nabíjacie stanice nachádzajú najčastejšie pri hoteloch. Tieto nabíjačky zvyknú byť často určené ale len pre hotelových hostí a bežní vodiči ich využívať nemôžu.

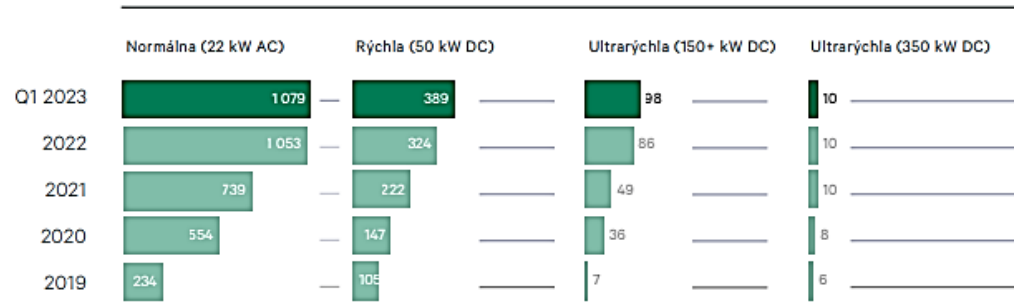
Problém s nabitím elektromobilu môže byť aj na akciách, kde sa zhromažďuje veľa ľudí, ako sú napríklad festivaly, koncerty alebo športové akcie. Na štadiónoch, ktoré majú veľké parkoviská, stačí teoreticky osadiť nabíjacie stanice. Problémom môžu byť napríklad festivaly, ktoré sú na odľahlejších miestach a parkovanie majú riešené len provizórne, na lúkach a podobne. Tam bude zvlášť problematické riešiť nabíjaciu infraštruktúru pre veľké množstvo elektromobilov, ktoré sa budú potrebovať nabíjať súčasne.

II. Možnosti nabíjania elektromobilov na Slovensku v súčasnosti

V akom stave sú v súčasnosti možnosti nabíjania elektromobilov na Slovensku, to najlepšie definuje aktuálny počet nabíjacích bodov a trend akým tempom pribúdajú.

A. Aktuálny počet nabíjacích staníc na Slovensku

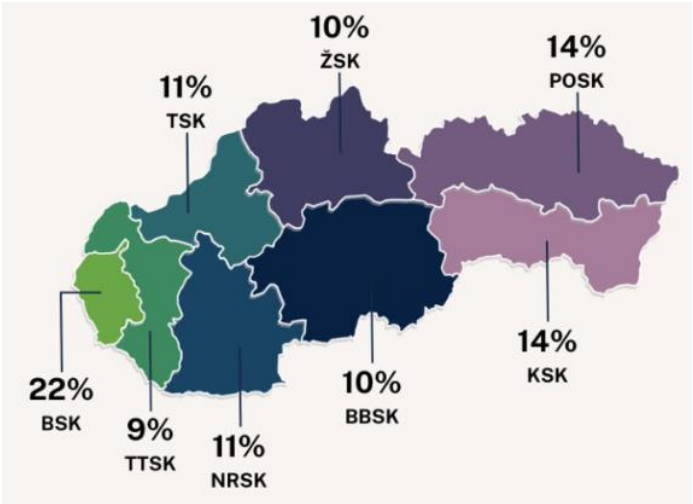
Tak ako celý svet, aj Slovensko bolo v rokoch 2019 až 2021 ovplyvnené pandémiou, čo sa negatívne prejavilo aj na cenách nabíjania elektromobilov. Ale nárast počtu nabíjacích bodov ovplyvnený nebol, ako vidíme na Obr. 1. Aktuálne sa na Slovensku nachádza už 1576 nabíjacích bodov, pričom ich medziročný rast je až 55%. Ako vidíme, tak najväčšiu prevahu majú stále AC nabíjačky s výkonom 22 kW [2].



Obr. 1 Trend nárastu nabíjacích bodov na Slovensku [2].

Pozitívny je trend nárastu nabíjacích bodov s rýchlym alebo ultrarýchlym nabíjaním. Z 629 lokalít s nabíjacou stanicou, ponúka až 331 aj rýchle alebo ultrarýchle nabíjanie. Teda na Slovensku si nabijeme elektromobil s výkonom 50 – 350 kW už na viac ako polovici nabíjacích lokalít. Dôležité je, že čoraz bežnejším štandardom sa stávajú nabíjačky s výkonom 100 – 150 kW. Trend nárastu výkonu nabíjacích staníc je dôležitý aj z hľadiska riešenia problematiky nabíjania v husto obývaných oblastiach, nakoľko rýchlo nabitý elektromobil dokáže uvoľniť miesto na nabíjanie pre ďalší.

Nové nabíjačky stále pribúdajú najčastejšie vo väčších mestách a frekventovaných dopravných koridoroch, pričom v menších mestách a na vidieku je rozvoj nabíjacej infraštruktúry výrazne pomalší. Na tieto miesta sa častejšie osadzujú nabíjačky s nižším výkonom. Dokazuje to aj aktuálna mapa rozloženia nabíjacích bodov na Obr. 2.



Obr. 2 Mapa rozloženia nabíjacej infraštruktúry na Slovensku [4].

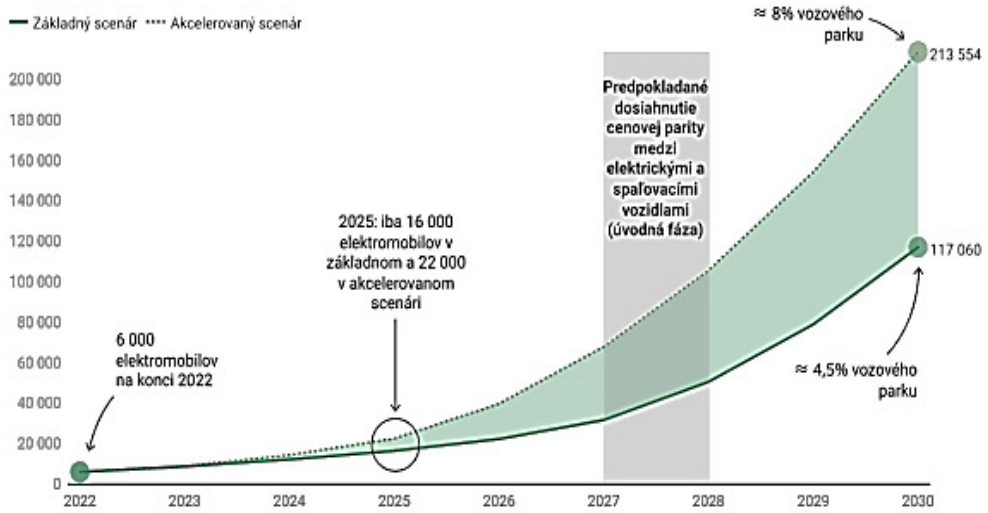
Vidíme. že najviac nabíjacích bodov sa percentuálne nachádza v Bratislavskom, Košickom a Prešovskom kraji, kde sa zároveň nachádzajú aj tri najväčšie slovenské mestá, pričom jednoznačne najväčší podiel má najväčšie slovenské mesto Bratislava [5].

Ďalším dôležitým faktorom pri posudzovaní rozvoja verejnej nabíjacej infraštruktúry je aj pomer inštalovaného výkonu všetkých bodov v pomere k počtu elektrických vozidiel vo vozovom parku. Podľa súčasnej metodiky zavedenej v Nariadení o zavádzaní infraštruktúry pre alternatívne palivá (AFIR), by mala byť vybudovaná nabíjacia infraštruktúra pre každý elektromobil s výkonom 1,3 kW. Pri dátach zo začiatku roku 2023, by sme na Slovensku mali mať nabíjaciu sieť s výkonom na úrovni 7 800 kW. Keďže na Slovensku máme k dispozícii 1576 nabíjacích bodov s výkonom 61 000 kW, tento parameter bez problému spĺňame. Avšak je to spôsobené hlavne stále relatívne nízkym počtom elektromobilov a nie pokročilou nabíjacou infraštruktúrou [2]. Avšak ak by sme súčasnú nabíjaciu infraštruktúru posudzovali len podľa dostupného nabíjacieho výkonu, aktuálne bez problému spĺňa požiadavky a je dokonca výrazne predimenzovaná [5].

Pre zaujímavosť, ak sa pozrieme na stav mimo Slovenskej republiky, podľa dát z prvého štvrťroka 2023, všetky krajiny Európskej Únie spolu ponúkajú nabíjaciu infraštruktúru o počte 428 000 klasických AC nabíjacích bodov a 52 000 rýchlych DC nabíjacích bodov,. Zaujímavosťou je, že až 70% týchto nabíjacích bodov sa nachádza v krajinách – Holandsko, Nemecko, Francúzsko, Veľká Británia a Taliansko. Teda tiež mimo Slovenska je rozvoj elektromobility veľmi rôznorodý aj z pohľadu samotných krajín a regiónov. Ale Slovensko za týmito krajinami výrazne zaostáva, dokonca aj pri medziročnom náraste 55% nabíjacích bodov.

B. Rozvoj predaja elektromobilov na Slovensku

Slovenská asociácia elektrických vozidiel namodelovala dva možné scenáre nárastu elektromobilov na Slovensku (Obr. 3). Základný scenár, ktorý nepočíta so zásadnou zmenou politiky štátu v podpore predaja elektromobilov, počíta s nárastom elektromobilov hlavne kvôli postupne znižujúcej sa ponuke spaľovacích vozidiel. Druhý scenár, je takzvaný akcelerovaný, ktorý vychádza z predpokladu, že štát viazaný svojimi záväzkami voči EU vstúpi do procesu rozvoja elektromobilov na Slovensku, rôznymi podpornými opatreniami. Spôsoby, ako môže štát urýchliť nárast počtu predaja elektromobilov, sú najmä vo forme priamej finančnej podpory nákupu elektrických áut, podpory rozvoja nabíjacej infraštruktúry alebo vo forme rôznych úľav v daňovej a účtovnej oblasti.



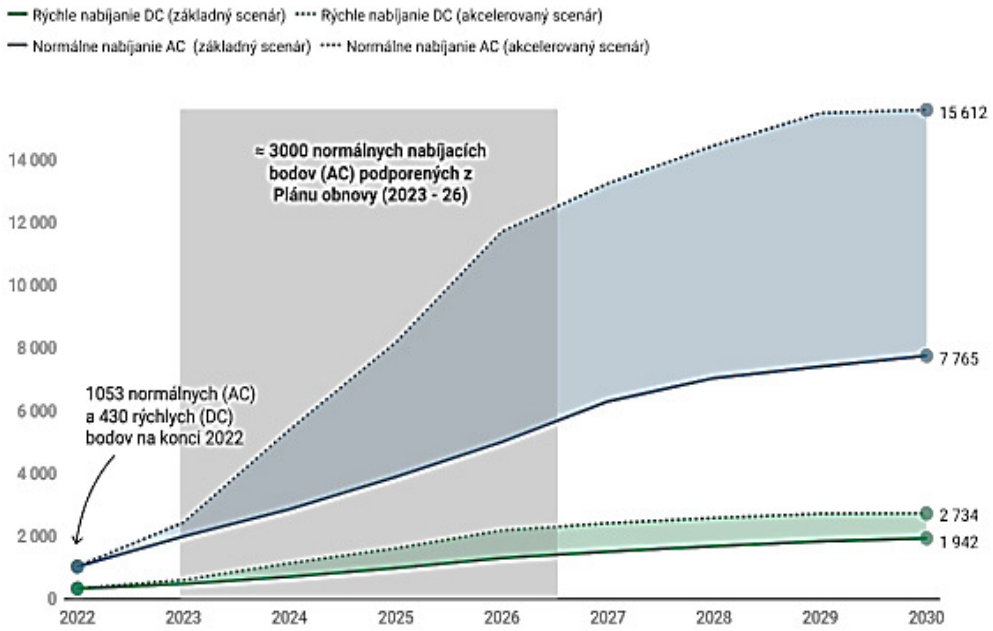
Obr. 3 Scenáre rozvoja predaja elektromobilov na Slovensku [2].

Ako vidíme na Obr. 3, bez zásahu štátu sa predpokladá pre najbližšie roky stagnácia na trhu s elektrickými vozidlami a k výraznejšiemu nárastu predaja elektromobilov príde až po dosiahnutí cenovej parity medzi elektrickými a spaľovacími vozidlami spolu s postupne sa znižujúcou ponukou spaľovacích vozidiel, kvôli čoraz prísnejším emisným kvótam. Pri akcelerovanom scenári by sme sa mohli dostať k takmer dvojnásobnému predaju elektrických vozidiel, čím by sme sa dostali na počet 213 554 elektromobilov, čo by ale stále tvorilo len 8% vozového parku. Teda predpokladá sa, že na Slovensku nebudú ani v roku 2030 tvoriť elektromobily majoritný podiel medzi všetkými vozidlami. Avšak súčasný počet, približne 6 000 elektromobilov, sa podľa predpokladov do roku 2030 v horšom prípade zväčší takmer dvadsaťnásobne a v optimistickejšom prípade až tridsaťpäťnásobne. Tomu teda musí odpovedať aj rozvoj nabíjacej

infraštruktúry a najmä vyriešenie problematiky nabíjania v husto obývaných oblastiach, ktoré môže výrazne zvýšiť predajnosť elektromobilov u obyvateľov týchto oblastí.

C. Rozvoj nabíjacej infraštruktúry na Slovensku

Rovnakou metodikou bol vypracovaný aj scenár rozvoja nabíjacej infraštruktúry do roku 2030 (Obr. 4). Tu je dôležité podotknúť, že nezahŕňa súkromné nabíjacie body, medzi ktoré patria okrem domáceho nabíjania aj nabíjačky v administratívnych alebo obytných budovách, hoteloch a ďalších miestach kde nemá prístup široká verejnosť.



Obr. 4 Scenáre rozvoja verejnej nabíjacej infraštruktúry na Slovensku [2].

Základný scenár opäť uvažuje s rozvojom bez výraznej podpory štátu a nárast nabíjacej infraštruktúry bude kopírovať len dopyt pre súkromný sektor, ktorý bude úmerný nárastu elektromobilov. Ďalším bodom je predpoklad, že sa nepodarí využiť zdroje z Plánu obnovy, ktoré predstavujú sumu 46 miliónov eur. Keďže musia byť vyčerpané do roku 2026, je predpoklad, že tieto financie nebudú kvôli problémom s meškaním výziev alebo oneskoreným začiatkom čerpania využité v plnej výške. Pri akcelerovanom scenári sa počíta s využitím týchto finančných prostriedkov a ďalšou podporou štátu pri rozvoji nabíjacej infraštruktúry. Tento scenár počíta už aj s podporou miest vo výstavbe normálnych nabíjacích AC bodov na sídliskách, čo by mohlo pomôcť rozvoju elektromobility v mestách. Okrem toho zahŕňa aj výraznejšiu akceleráciu výstavby ultrarýchlych nabíjacích bodov pri diaľniciach a väčšiu investíciu súkromného sektora. Súkromné investície sú ale veľmi závislé od rýchlosti rozvoja predaja elektrických vozidiel, hlavne kvôli potrebe ich následnej návratnosti a zisku.

Ako na Obr. 4 môžeme vidieť, tak akcelerovaný scenár výrazne ovplyvní najmä nárast klasických AC nabíjacích bodov, ktoré by sa vybudovali výraznejším podielom v mestách. V pesimistickejšom scenári sa vybuduje do roku 2030 len polovica týchto nabíjacích bodov, čo môže veľmi negatívne ovplyvniť rozvoj elektromobilov práve v husto obývaných oblastiach. V prípade rýchleho DC nabíjania vidíme menší rozdiel medzi oboma scenármi, avšak rozdiel je stále takmer 800 nabíjacích bodov. Dôležité je podotknúť, že ani v akcelerovanom scenári sa nám podľa dát SEVA nepodarí dosiahnuť dostatok nabíjacích bodov na to, aby sa elektromobily začali vnímať ako bezproblémová alternatíva k vozidlám so spaľovacím motorom. Teda nabíjacia infraštruktúra potrebuje veľmi výrazný impulz, ktorý pomôže dosiahnuť Slovensku cieľ k uhlíkovej neutralite.

D. Stratégia umiestňovania nabíjacích staníc

Efektívne riešenie problematiky nabíjania elektromobilov spočíva aj vo vhodnom výbere lokalít, kde sa bude nabíjacia infraštruktúra budovať. Pre správny výber lokality na ktorej by mohol byť umiestnený nabíjací bod, je najdôležitejšie posúdiť hustotu obyvateľstva v danej lokalite. Hustota obyvateľstva je definovaná ako počet obyvateľov na jednotku plochy. Medzi husto obývané oblasti patria najčastejšie mestá a sídliská. Podľa dát z roku 2011, žilo na desiatich najviac obývaných kilometroch štvorcových až 150 000 ľudí, čo predstavuje takmer 3 %

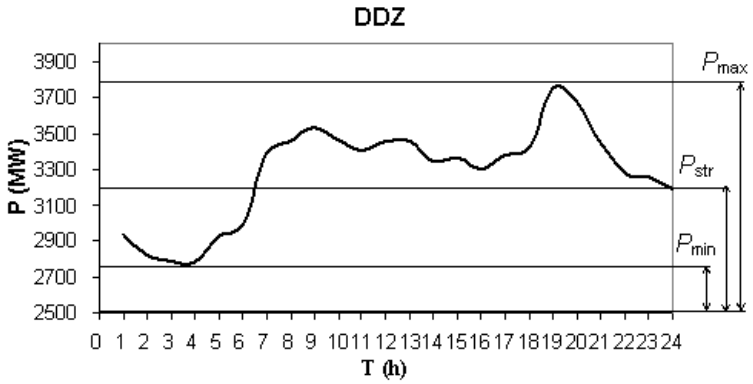
obyvateľov Slovenska. Len pre zaujímavosť, tak najľudnatejšou oblasťou na Slovensku, je Sídlisko Ťahanovce v Košiciach, kde pripadá na jeden kilometer štvorcový až 2 462 obyvateľov [6]. Tieto oblasti sa vďaka veľkej hustote obyvateľstva vyznačujú aj vyššou koncentráciou infraštruktúry, vybavenosti, služieb a teda aj celkovým nárokom na energetiku.

Veľmi veľa starších sídlisk v mestách má problém s nedostatkom parkovacích miest. Pri ich projektovaní nebolo počítané s toľkým počtom vozidiel. V minulosti sa nepočítalo často ani s jedným vozidlom na rodinu, pričom v súčasnosti je bežné, že rodina vlastní dve aj viac motorových vozidiel. Toto silno ovplyvňuje aj rozvoj elektromobilov. Keďže v husto obývaných oblastiach žije veľké množstvo ľudí na malej ploche, zvyšuje sa aj potencionálny počet ľudí, ktorí môžu vlastniť elektromobil a potrebujú ho nabíjať. Obyvatelia husto obývaných oblastí sú vo veľkej miere odkázaní výhradne na verejné parkovacie miesta a preto je zvlášť potrebné umiestňovať nabíjacie body v komfortnej blízkosti a dostupnosti. Ďalším špecifikom miest s vysokou hustotou obyvateľov je, že väčšina z nich na týchto miestach síce žije, ale pracuje alebo študuje mimo týchto oblastí. Teda väčšina obyvateľov ráno opustí svoj domov a mestskú časť a vráti sa domov v poobedňajších hodinách, po práci, štúdiu. Tento fakt bude dôležitý v budúcnosti, keď sa domov vráti väčšina majiteľov elektromobilov v približne rovnaký čas a zapojí elektromobil do nabíjačky. To výrazne zvýši požiadavky na aktuálny výkon elektrární a prenosovej siete.

III. Zaťaženie elektrickej siete nabíjaním elektromobilov

Najväčším problémov počas nabíjania elektromobilov pri ich masívnom rozšírení je možné zvýšenie nárokov na elektrickú sieť. Jedným problémom pri budovaní nabíjacej infraštruktúry, nových nabíjacích bodov s dostatočne výkonnými nabíjačkami je ich samotné napájanie. Elektrická sieť má svoje obmedzenia a s narastajúcou záťažou, ktorú elektromobily zásadne zvýšia, môže naraziť na svoje limity. Keďže v minulosti sa nepočítalo s rozvojom elektromobilov, elektrárne a rozvodné siete boli dimenzované na úplne inú záťaž.

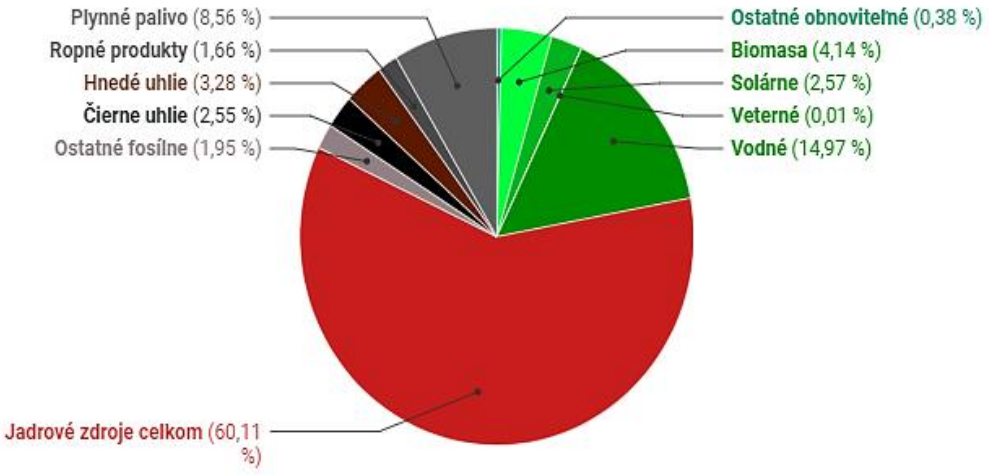
Veľkým problémom môže byť nárast spotreby práve v určitú hodinu. Ako vidíme na Obr. 5, ktorý zobrazuje denný diagram zaťaženia elektrickej siete, najväčšiu špičku vidíme v čase medzi 18 a 20 hodinou večer.



Obr. 5 Denný diagram zaťaženia elektrickej siete [7].

Je to presne čas kedy sa ľudia vracajú z práce a škôl, začínajú tak v domácnostiach svietiť, kúriť a využívať domáce spotrebiče. Je teda pravdepodobné, že majitelia elektromobilov ten svoj pripoja do nabíjačky práve po príchode z práce, niekedy v týchto hodinách. To spôsobí ďalšie zvýšenie tejto špičky, čo môže byť pre stabilitu energetickej siete problém [8].

Z hľadiska problematiky nabíjania elektromobilov v energetickej špičke, môžu mať problém najmä krajiny, ktoré využívajú z veľkej časti elektrickú energiu z obnoviteľných zdrojov, keďže v týchto elektrárňach je náročné regulovať okamžitý výkon. Napríklad veterné alebo solárne elektrárne nemusia dodávať dostatočný výkon práve v najväčšom zaťažení počas energetickej špičky. Výhodou Slovenskej republiky je, že väčšinu elektrickej energie stále vyrábame z jadrových zdrojov. Ako vidíme na Obr. 6, podľa dát z roku 2022 sa jedná o viac ako 60,11 % z celkovej vyrobenej elektrickej energie na Slovensku. Druhým najvýznamnejším zdrojom sú vodné elektrárne s podielom takmer 15 %. Medzi významné zdroje patria aj plynné palivo z ktorého pochádza 8,56 % elektrickej energie a čierne a hnedé uhlie spolu tvoria 5,83 % [7]. Problémom je nízke percento elektrickej energie z obnoviteľných zdrojov okrem vodných elektrární, čo môže byť v budúcnosti z hľadiska ekológie problém. V západných európskych krajinách v súčasnosti dochádza k zatváraniu jadrových elektrární, keďže ani tento zdroj sa nepovažuje za úplne ekologický, čo môže v budúcnosti zasiahnuť aj Slovensko.



Obr. 8 Zdroje elektrickej energie na Slovensku v r.2022 [7].

Podľa analýzy Ministerstva hospodárstva SR, by pri očakávanom počte 116 000 elektromobilov v roku 2030, vzrástla spotreba elektrickej energie na Slovensku o približne 2%. Problémom je, že síce nárast o 2% by pravdepodobne naša energetická sieť zvládla, avšak 116 000 elektromobilov by na Slovensku tvorilo len približne 5% zo všetkých vozidiel. Ak by sme teda uvažovali, že vozový park na Slovensku sa plne elektrifikuje, nároky na výrobu a distribúciu stúpnu až o približne 40%, respektíve 12 TWh ročne. To predstavuje ročnú výrobu troch jadrových blokov.

Okrem výroby samotnej elektrickej energie je potrebné zabezpečiť aj jej distribúciu do samotných nabíjačiek. Ak by sme aj dokázali vyrobiť dostatok elektrickej energie pre nabíjanie všetkých elektromobilov, veľký problém môže nastať pri prenosovej sústave, ktorá pri projektovaní v minulosti nebola dimenzovaná na prenášanie takých výkonov, ktoré by stačili na narastajúce nároky v spojení s nabíjaním elektromobilov. Podľa experta Miroslava Gejdoša zo spoločnosti Stredoslovenská distribučná, nie je naša distribučná sústava pripravená na plánovaný nárast počtu elektromobilov. Základom bude využitie finančných prostriedkov z plánu obnovy na posilnenie a zvýšenie odolnosti distribučnej sústavy [9].

V súčasnosti v EU platí zákaz predaja spaľovacích vozidiel od roku 2035. Dá sa očakávať že na Slovenských cestách by takmer úplnú väčšinu elektromobily mohli dosiahnuť až niekedy po roku 2050, nakoľko priemerný vek vozidla na Slovensku je približne 14,3 roka [10]. To znamená, že nárast požiadaviek na energetiku bude postupný a energetická infraštruktúra sa musí rozvíjať spolu s nárastom počtu elektromobilov.

IV. ZÁVER

Viac ako 54 % globálneho obyvateľstva žije v mestských oblastiach, pričom na Slovensku sú tieto dáta takmer rovnaké. Pri plánovanej úplnej elektrifikácii dopravy do roku 2050 je potrebné vyriešiť aj dostupnosť nabíjania pre viac ako polovicu obyvateľstva, ktorý nemajú k dispozícii domáce nabíjanie. Pozitívnym trendom je medziročný nárast až o 55% nových nabíjacích bodov na Slovensku, pričom viac ako polovica ponúka rýchle nabíjanie s výkonom 50 kW a viac. Ak vezmeme do úvahy pomer inštalovaného výkonu všetkých bodov v pomere k počtu elektrických vozidiel, podľa súčasných metodík má Slovensko momentálne dostatočnú nabíjaciu infraštruktúru, avšak faktom je v súčasnosti stále veľmi nízky počet elektromobilov. Aktuálne predpoklady hovoria že v roku 2030 bude na Slovensku v pesimistickom predpoklade dvadsaťnásobný počet elektromobilov a v optimistickom až tridsaťpäťnásobný, pričom aj najoptimistickejšie predpoklady nárastu nabíjacích bodov nie sú dostatočné, pre dosiahnutie vnímania elektromobilov ako bezproblémovej alternatívy k spaľovacím motorom. Ďalším problémom nabíjania elektromobilov je vyrobiť dostatočné množstvo elektrickej energie. Podľa dát Ministerstva hospodárstva SR, by pri očakávanom počte 116 000 elektromobilov v roku 2030, vzrástla spotreba elektrickej energie na Slovensku o približne 2%, avšak tento počet elektromobilov by tvoril len 5% z celkového počtu automobilov. Pri plnej elektrifikácii by stúpli nároky na výrobu a distribúciu elektrickej energie až o 40%, čo predstavuje ročnú výrobu troch jadrových blokov.

POĎAKOVANIE

Táto práca bola podporená projektom vedeckej agentúry KEGA podľa zmluvy číslo: 032TUKE-4/2024.

LITERATÚRA

[1] Miroslav Mudroň. Môj elektromobil.sk 23. Január 2024 [Dátum: 5. Október 2023]. Dostupné na internete: <https://www.mojelektromobil.sk/nabijacie-siete-slovensko-zhrnutie>
[2] SEVA [online]. [Dátum: 5. Október 2023]. Dostupné na internete: < https://seva.sk/report23/>
[3] Trend [online]. 20. December 2021 [Dátum: 5. Október 2023]. Dostupné na internete: <https://www.trend.sk/spravy/statisticky-urad-zverejnil-prve-cisla-tohtorocneho-scitania-obyvatelov>
[4] Rastislav Žembery. Blog ZSE [online]. 17. Február 2024 [Dátum: 19. marec 2024]. Dostupné na internete: <https://www.zse.sk/blog/elektromobilita/je-nabijanie-stale-vyhodnejsie-ako-tankovanie>
[5] Elektromobilita. Ročenka 2023 [online] 18. Apríl. 2023. [Dátum: 9. September. 2023]. Dostupné na internete: <https://www.nextech.sk/library/PDF%20ARCHIV/Elektromobilita/Elektromobilita%202023.pdf>.
[6] Pavol Škápik, Viktor Robert Bederka. DenníkN [online]. 14. Máj 2018 [Dátum: 12 Október 2023]. Dostupné na internete: < https://dennikn.sk/blog/1119465/klebetime-ze-husto-je-nielen-v-kosiciach/>
[7] Radovan Potočár. Energie-potral. [online] 2. Jún 2023 [Dátum: 25. Október 2023]. Dostupné na internete: <https://www.energie-portal.sk/Dokument/vyroba-elektriny-slovensko-elektrarne>
[8] Martin Domček, Pravda, TASR [online]. 9. Marec 2023. [Dátum: 25. Október 2023] Dostupné na internete: <https://auto.pravda.sk/magazin/clanok/659604-odbornici-distribucna-siet-na-slovensku-nezvladne-narast-elektromobility/>
[9] Plán obnovy – cestovná mapa k lepšiemu Slovensku, [Dátum: 3. 4. 2024] Dostupné na internete: <https://www.planobnovy.sk/site/assets/files/1069/01_kapitola_1_vseobecne-ciele-sudrznost-planu_1.pdf/>
[10] Martin Domček, Čoraz staršie: Takto starnú autá na Slovensku a v EÚ, Pravda, TASR [online]. 13. október 2023. [Dátum: 5. 5. 2023] Dostupné na internete: < https://auto.pravda.sk/magazin/clanok/684778-coraz-starsie-pozrite-sa-aky-je-priemerny-vek-aut-na-slovensku-a-v-eu/>

Implementácia SD-MANET siete do simulátora Mininet-WiFi

[1]*Hai Tran NGOC* [2]*Ján PAPAJ*

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

[1]hai.tran.ngoc@student.tuke.sk, [2]jan.papaj@tuke.sk

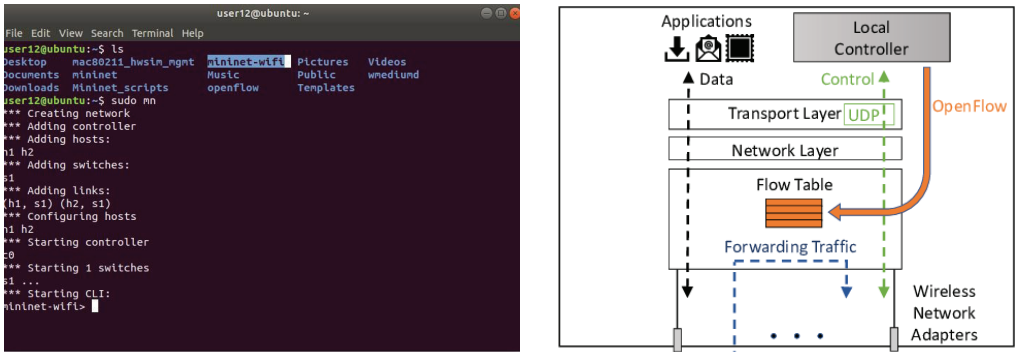
Abstract **– Mobilné Adhoc siete (MANET) predstavujú dynamické a autonómne siete, ktoré nevyžadujú pevnú infraštruktúru a umožňujú bezdrôtovým zariadeniam komunikovať priamo medzi sebou. S príchodom 6G technológií sa otvárajú nové možnosti a výzvy pre implementáciu týchto sietí. Softvérovo definované siete (SDN) prinášajú flexibilné, efektívne a centralizované riadenie, čo môže výrazne vylepšiť efektivitu, výkonnosť a adaptabilitu MANET v prostredí 6G. Táto práca skúma koncept softvérovo definovaných mobilných Adhoc sietí (SD-MANET) v prostredí 6G, analyzujúc ich architektúru a výhody. Implementácia SDN do MANET umožňuje lepšiu škálovateľnosť, zníženie latencie a efektívnejšie využitie spektra, čo je kritické pre podporu nových 6G aplikácií, ako sú rozsiahle IoT systémy, inteligentné mestá a autonómne vozidlá. Na základe simulácií a experimentov demonštrujeme, ako môžu SD-MANET v 6G prostredí riešiť tradičné výzvy v MANET, ako je nestabilita spojení a nízka priepustnosť, pričom zdôrazňuje potrebu ďalšieho výskumu v oblasti štandardizácie a bezpečnostných protokolov. Celkovo tento prístup naznačuje, že integrácia SDN a 6G technológií môže priniesť významné pokroky v oblasti mobilných Adhoc sietí, čím prispeje k rozvoju novej generácie bezdrôtových komunikácií.**

Keywords **– MANET, SDN, Mininet-WiFi, SD-MANET**

I. Úvod

Mobilné Adhoc siete (MANET) sa stávajú čoraz dôležitejšou súčasťou moderných bezdrôtových komunikácií, najmä v tam, kde nie je dostupná pevná infraštruktúra. Tieto siete umožňujú mobilným zariadeniam komunikovať priamo medzi sebou, čím poskytujú flexibilitu a odolnosť voči rôznym druhom výpadkov. S príchodom 6G technológií, ktoré sľubujú výrazné zlepšenie rýchlosti, kapacity a spoľahlivosti bezdrôtových sietí, sa otvárajú nové príležitosti pre implementáciu a optimalizáciu MANET [1]. MANET je bezdrôtová sieť, ktorá pozostáva zo skupiny mobilných uzlov pripojených bezdrôtovo v samostatne nakonfigurovanej a samoopravnej sieti bez fixnej štruktúry. Uzly MANET sa môžu voľne pohybovať náhodne, pretože topológia siete sa často mení. Každý tento uzol sa správa ako zdroj, smerovač a cieľový uzol, pretože posiela dátovú komunikáciu do iných uzlov siete. Topológia siete MANET je dynamická, pretože konektivita medzi uzlami sa môže časom meniť v dôsledku mobility uzlov. Tieto uzly by mali byť schopné detegovať prítomnosť iných takýchto zariadení, aby umožnili komunikáciu a zdieľanie informácií. Keďže sa počet uzlov mení za chodu, menia sa aj informácie o smerovaní. Každý uzol komunikuje so susednými uzlami v rámci svojho dosahu bezdrôtovej komunikácie. Uzly, ktoré sú súčasťou siete MANET ale mimo dosah ostatných uzlov komunikujú pomocou viacskokovej trasy (multihop route) prechádzajúce cez iné uzly v sieti. Tieto trasy sa menia s topológiou siete a sú určené pomocou smerovacích protokolov (napr. DSDV, DSR, AODV, TORA, ZRP...).

Softvérovo definované siete (SDN) predstavujú revolučný prístup v správe a riadení sietí, umožňujúci oddelenie riadiacej vrstvy od vrstvy dátovej [2]. Tento koncept prináša výhody ako centralizované riadenie, jednoduchšie aktualizácie a konfigurácie siete, a lepšiu optimalizáciu zdrojov. Integrácia SDN do MANET v prostredí 6G môže viesť k vytvoreniu efektívnych a dynamických sietí, ktoré budú schopné zvládnuť náročné požiadavky budúcich aplikácií. SDN je prístup k sieti využívajúci softvérové ovládače alebo aplikačné programové rozhrania (API) na komunikáciu so základnou hardvérovou infraštruktúrou a na priamy prenos v sieti [2]. V SDN sa riadiaca rovina (control plane) odoberie prepínaču a priradí sa centralizovanej jednotke nazývanej SDN kontrolér. Správca siete tak môže riadiť sieťovú prevádzku prostredníctvom SDN kontroléra bez toho, aby zasahoval do jednotlivých prepínačov. Dátová rovina sa stále nachádza v prepínači a keď paket vstúpi do prepínača, o jeho presmerovaní sa rozhoduje na základe záznamov tabuľky



(a) Ukážka prostredia Mininet-WiFi (b) OpenFlow architektúra pre SD-MANET

Obr. 1: SD-MANET v simulačnom prostredí

toku (Flow Table), ktoré sú vopred priradené SDN kontrolérom [3]. Typická SDN architektúra sa skladá z troch vrstiev:

- ***Aplikačná vrstva –*** obsahuje typické sieťové aplikácie (napr. Firewall, Load Balancing).
- ***Riadiaca vrstva –*** pozostáva z SDN kontroléra, ktorý umožňuje abstrakciu aplikácií nad ním.
- ***Vrstva infraštruktúry –*** pozostáva z fyzických prepínačov, ktoré tvoria dátovú rovinu a vykonávajú pohyb dátových paketov.

Tieto vrstvy komunikujú prostredníctvom množiny rozhraní [3]:

1) ***Northbound rozhrania API (medzi aplikačnou a riadiacou vrstvou)***
2) ***Southbound rozhrania API (medzi riadiacou a vrstvou infraštruktúry)***

Model SDN sa považuje za novú generáciu vytvárania sietí. Preukázala svoju účinnosť predovšetkým v káblových sieťach, konkrétne dátové centrá.

Mobilné Adhoc siete (MANET) a softvérovo definované siete (SDN) predstavujú dva významné koncepty v oblasti moderných bezdrôtových komunikácií. S príchodom 6G technológií sa očakáva, že kombinácia týchto dvoch prístupov prinesie nové možnosti, aplikácie a výzvy [4].

A. Softvérovo definované MANET siete

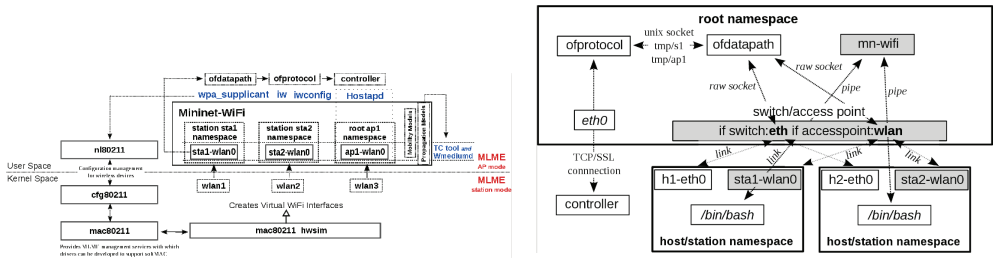
Softvérovo definovaná MANET (SD-MANET) predstavuje skupinu mobilných bezdrôtových uzlov, ktoré dynamicky vytvárajú dočasnú sieť na komunikáciu bez potreby fixnej infraštruktúry využívajúce koncepty SDN [5]. SD-MANET kombinuje vlastnosti softvérovo definovanej siete s charakteristikami MANET, čo umožňuje flexibilnú a dynamickú správu siete [6].

V porovnaní s distribuovanými smerovacími protokolmi (napr. AODV a OLSR) môže prístup založený na SDN pre MANET siete poskytnúť vyššiu priepustnosť, rýchlejšiu odozvu na poruchy a predĺženie životnosti siete [7]. Na komunikáciu medzi dátovou a riadiacou rovinou sieťových zariadení sa využíva SDN protokol OpenFlow, ktorý umožňuje vziať jednu implementáciu riadiacej roviny a použiť ju na riadenie implementácie dátovej roviny. Jedným z hlavných faktorov OpenFlow je jeho zjednodušenie. Zjednodušuje sieť takým spôsobom, že oddeľuje komplexný softvér z uzlov a nasadí do siete externé servery nazývané SDN kontroléry [6].

SDN kontrolér je softvér, ktorý sa správa ako kontrolný bod na riadenie celej siete. Riadenie a manažment MANET sietí ponúka väčšiu flexibilitu, nové funkcie a služby sieti. Model SDN poskytuje centralizovanú sieť, kde všetky funkcie riadiacej roviny a roviny manažmentu sa posúvajú do centralizovanej jednotky – SDN kontroléra [3]. SDN umožňuje sieťovým administrátorom definovať zásady alebo správanie v kontroléri prijímaním rozhodnutí pre každý nový tok. Môžeme uvažovať, že SDN kontrolér je mozgom siete a definuje kolaboratívne správanie siete a riadi zúčastnené uzly a ich správanie v sieti. Každý uzol musí byť nakonfigurovaný tak, aby podporoval OpenFlow, teda bude obsahovať OpenFlow prepínač a bude sa správať ako smerovač a zároveň koncový hostiteľ [8].

B. Openflow protokol pre SD-MANET

Medzi najdôležitejšiu funkciu protokolu OpenFlow je oddelenie riadiacej a dátovej roviny (Obr. 1). Sieť SDN na báze OpenFlow tvoria prepínače, ktoré komunikujú s jedným alebo viacerými kontrolérmi a preposielajú pakety pomocou OpenFlow protokolu [9]. SDN kontrolér vytvára a konfiguruje správanie preposielania dát nastavením pravidiel v tokovej tabuľke (Flow Table) a ich odoslaním ku každému prepínaču v sieti. Tieto pravidlá pozostávajú zo zhody kritérií a opatrení. OpenFlow prepínač môže obsahovať jednu alebo viaceré týchto tokových tabuliek [8].



(a) Architektúra nástroja Mininet-WiFi (b) Komponenty nástroja Mininet-WiFi

Obr. 2: Mininet-WiFi pre SD-MANET

Kontrola zhody sa začína pri prvej tokovej tabuľke a môže pokračovať do ďalších tabuliek sériovo. Keď na prepínač prichádza paket, začnú sa kontrolovať záznamy (entries) v prvej tabuľke a podľa priority. Paket s najvyššou prioritou sa zhoduje ako prvý. Akcie určujú to, ako tento paket spravuje prepínač, napr. modifikuje paket, prepošle ho cez určitý port/do kontroléra alebo ho zaplaví [10].

Protokol OpenFlow sa taktiež používa na povolenie kontrolérovi stiahnutie, aktualizovanie a zmazanie záznamov v tokovej tabuľke prepínača. V tejto situácií môže fungovať v reaktívnom alebo proaktívnom spôsobe:

- *Reaktívny* – v tomto spôsobe prepínač nerozhoduje o preposielaní, ak prichádzajúci paket sa nezhoduje žiadnej zo záznamov v tokovej tabuľke, pošle ho do kontroléra ako packet-in správu
- *Proaktívny* – kontrolér vykoná akciu podľa jeho konfigurácie, napr. vypočíta cestu pre paket a pošle aktualizovaný tokový záznam prepínaču ako packet-out správu

C. Simulátor Mininet-WiFi

Mininet-WiFi je softvérový nástroj, ktorý slúži na simuláciu bezdrôtových sietí. Je to rozšírenie pôvodného nástroja Mininet, ktorý je určený pre simuláciu klasických (káblových) počítačových sietí. Mininet-WiFi pridáva bezdrôtovú funkcionalitu, čo umožňuje vytvárať a testovať bezdrôtové siete v prostredí, ktoré simuluje reálne bezdrôtové siete [4]. Celý proces virtualizácie nástroja Mininet-WiFi je založený na báze procesov bežiacich na sieťových komponentoch, zvaných namespaces a virtuálnych sieťových rozhraniach operačného systému Linux (viď Obr. 2a). Tieto menné priestory sú v podstate kópiou sieťového zásobníka OS Linux, v ktorom sú ukladané trasy, pravidlá Firewall a sieťové zariadenia. Fungujú ako keby to boli skutočné počítače s rovnakými sieťovými vlastnosťami.

Z Obr.2b) vidíme, že medzikomponentovú komunikáciu zahajuje v inicializácii modul nazývaný $mac802.11\_hwsim$, zodpovedný za virtualizáciu Wi-Fi sieťových kariet, obsahuje istý počet virtuálnych bezdrôtových rozhraní pre všetky uzly, ktoré boli preddefinované používateľom.

v jadre (Kernel Space) OS Linux všetky funkcie podporované modulom $mac80121\_hwsim$ pochádzajú z frameworku $mac802.11$ založenom na SoftMAC, ktorý je používaný vývojármi na naprogramovanie ovládačov pre bezdrôtové zariadenia. Nachádza sa tu aj API konfigurácia haldy $cfg80211$, pričom pri jeho spustení sa o interakciu medzi jadrom a užívateľským priestorom stará konfigurácia $nl80211$.

Hlavné sieťové aplikácie využívané nástrojom Mininet-WiFi sa vykonávajú v užívateľskom priestore. Medzi tieto aplikácie patrí napríklad hostapd (Host Access Point Daemon), čo je softvér na úrovni užívateľa schopný spustiť rozhranie bezdrôtovej siete na prístupových bodoch a autentifikačných serveroch.

II. Implementácia SD-MANET do prostredia Mininet-WiFi

Integrácia SDN a MANET v prostredí 6G prináša nový rozmer do oblasti bezdrôtových sietí. 6G technológie sľubujú zlepšenia v rýchlosti, kapacite, spoľahlivosti a latencii, čo môže výrazne ovplyvniť výkonnosť SDN MANET. Medzi hlavné parametre, ktoré sú potrebné brať do úvahy pre bezpečnosť v MANET sieti sú:

- *Sieťová réžia* – tento parameter sa vzťahuje na počet paketov vygenerovaných zabezpečeným prístupom; vysoká réžia siete predstavuje vyššiu stratu paketov,
- *Doba spracovania* – každý zabezpečený prístup potrebuje čas na to, aby vedel detegovať zlomyseľné uzly; vzhľadom na dynamickú topológiu siete MANET, je vysoká šanca, že cesta medzi dvoma uzlami sa môže prerušiť z dôvodu ich mobility,

Tabuľka 1: SD-MANET v simulátore Mininet-WiFi - softvérové požiadavky

Softvér	Funkcia	Verzia
Ubuntu	Operačný systém VM	22.04 LTS
Mininet-WiFi	Simulačný nástroj	v2.4.3
Floodlight	OpenFlow kontrolér	v1.2S
OLSR	Smerovací protokol	v1S
BATMAN	Smerovací protokol	v2024.1S
OpenFlow	Smerovací protokol	1.3S
Iperf	Nástroj na meranie parametrov	v1S
VirtualBox	Monitor VM	v7.0.18

- *Spotreba energie* – v sieti MANET uzly majú obmedzený zdroj energií; takže optimalizácia spotreby energie môže byť v nich náročná.

Aj keď integrácia SD-MANET v prostredí sľubuje mnoho výhod, existuje niekoľko výziev a problémov, ktoré je potrebné riešiť:

- Náročnosť – integrácia SDN do dynamických a rýchlo sa meniacich MANET môže byť komplexná,
- Bezpečnosť a spoľahlivosť – zaistenie bezpečnosti a aj spoľahlivosti s vysokou mobilitou je náročné,
- Obmedzené zdroje – uzly majú zvyčajne obmedzenú šírku pásma a výpočtový výkon.

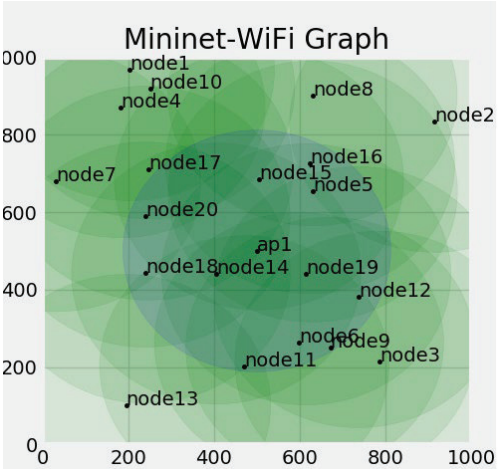
Simulačné prostredie využíva VirtualBox ako virtualizačné prostredie s obrazom Ubuntu 22.04 LTS vo virtuálnom stroji. Týmto virtuálnym strojom bola nakonfigurovaná úložná kapacita 20 GB, 8192 GB RAM a 4-jadrový procesor. Tab. 1 a Tab. 2 obsahujú softvérové komponenty a sieťové parametre použité pri implementácií a simulácií siete SD-MANET.

Topológie siete budeme vytvárať pomocou skriptov v objektovo-orientovanom jazyku Python. Nástroj Mininet-WiFi disponuje aj programom s užívateľským rozhraním MiniEdit alebo Virtual Network Descriptor na vytváranie topológií, avšak tieto možnosti majú v súčasnosti pomerne veľa limitácií. Postup pri vytváraní navrhnutej siete zahŕňa tieto kroky:

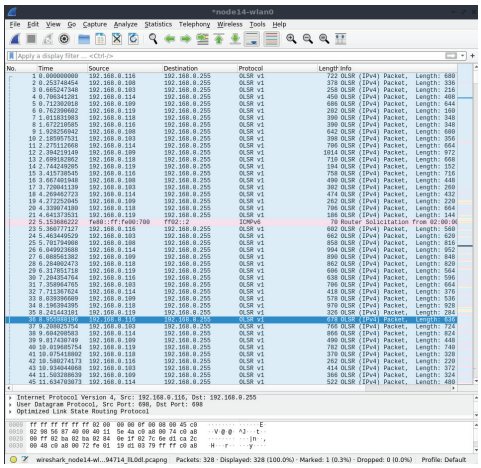
- **Importovanie knižníc -** objekt *RemoteController* predstavuje vzdialený ovládač a *OVSKernelSwitch* prepínač bežiaci v priestore jadra (kernel space), ktorý je zároveň prepínačom typu Open vSwitch. V priestore jadra sú kontrolér a prepínač jednoducho procesmi. Knižnice setLogLevel a info nám umožňujú vytvárať rôzne informácie v čase spustenia príkazového riadku topológie, pričom tento príkazový riadok je vytvorený pomocou knižnice CLI a možno v ňom spustiť akýkoľvek príkaz spustiteľný v systéme Linux. Knižnica $Mininet\_wifi$ obsahuje objekt, ktorý inicializuje celú našu topológiu podľa príslušných parametrov. O abstrakciu elektromagnetického rušenia (EMI) v bezdrôtových sieťach sa starajú knižnice interference a wmediumd. Okrem toho, knižnica wmediumd slúži na virtualizáciu sieťových kariet v prostredí Mininet-WiFi. Wmediumd je odporúčané používať aj vtedy, ak bezdrôtové rozhrania musia byť od seba izolované. Na vytváranie liniek v režime Ad-hoc medzi zariadeniami využijeme knižnicu adhoc. Všetky vyššie spomínané knižnice možno upraviť podľa potrieb.
- **Inicializácia siete a vytvorenie zariadení -** vytvoríme funkciu s názvom $create\_manet()$, ktorá bude zavolaná v hlavnej funkcii main. Budú teda vytvorené všetky objekty, čiže zariadenia a spojenia vytvorené medzi nimi. Na začiatku inicializujeme našu sieť s bezdrôtovými linkami, ktorá bude využívať vzdialený kontrolér a prepínač typu OVSKernelSwitch, riadkom kódu $net = Mininet_wifi(controller = RemoteController, link = wmediumd, wmediumd_mode = interference, switch = OVSKernelSwitch)$.
- **Konfigurácia modulu šírenia bezdrôtového komunikačného systému -** matematický model používaný simulátormi na napodobnenie správania bezdrôtového média.
- **Konfigurácia mobilných staníc -** vytváranie bezdrôtových komunikačných liniek medzi samotnými mobilnými stanicami a prístupovým bodom a ich následné pripojenie do Wi-Fi siete. Tento krok bol vykonany pomocou funkcie *addLink()* s argumentmi: *objekt (self), režim asociácie (cls), názov rozhrania (intf), názov siete (ssid), režim alebo štandard 802.11 (g), kanál (channel), a názov MANET smerovacieho protokolu (proto).*
- **Konfigurácia modelu mobility -** model mobility slúži na napodobňovanie pohybu uzla z jedného bodu do druhého. Nástroj MininetWiFi v súčasnosti podporuje 7 modelov mobility: *Random Walk, Truncated Levy Walk, Random Direction, Random Waypoint, Gauss Markov, Reference Point a Time Variant Community*.

Tabuľka 2: SD-MANET v simulátore Mininet-WiFi - softvérové požiadavky

Parameter	Hodnota
Počet uzlov	20, 40, 60, 80, 100
Rýchlosť pohybu uzla	5 – 10 m/s
Model mobility	Random Waypoint
Režim paketu	TCP a UDP
Protokoly	OLSR, B.A.T.M.A.N a OpenFlow
WiFi štandard	802.11g
Režim spojenia	Ad-hoc
Plocha simulácie	1000m x 1000m



(a) Ukážka topológie siete



(b) Ukážka komunikácie v simulátore Mininet-WiFi a WireShark na uzle 14.

Obr. 3: Simulačné prostredie Mininet-WiFi

Tabuľka 3: SD-MANET a sieťové parametre pre protokol OLSR

Počet uzlov	Prenosová rýchlosť $[kb/s]$	Šírka pásma $[kb/s]$	Oneskorenie $[ms]$	Úspešnosť doručenia $[\%]$
20	809	1002	21,260	62,1
40	559	635	98,575	48,2
60	423	564	1925,762	22,07
80	221	289	31238,004	5,02
100	107	142	50688,293	1,83

III. Dosiahnuté výsledky

Cieľom simulácii bolo overenie si možnosti nasadenia SD-MANET do prostredia Mininet-WiFi (Obr. 3). Pri analýze sme použili nástroj *iperf*, ktorý sa využíva meranie sieťových parametrov ako napr. šírka pásma, kolísanie oneskorenia (jitter). Tento nástroj funguje na princípe klient/server, pričom sa najprv zapne server na jednom uzle a následne na druhom uzle. Okrem nástroja iperf sme využili aj nástroj WireShark na zistenie správania pri nadviazaní komunikácie medzi dvoma uzlami (Obr. 4).

Komunikáciu sme vykonávali medzi uzlami **node14** a **node17** (Obr. 3a). Pomocou WireShark sme zistili, že protokol OLSR neustále vysiela pakety na hľadanie ciest ku každému uzlu v sieti (viď Obr. 3b). Smerovaciu tabuľku uzla node14 je možné zobraziť pomocou príkazu route, pričom táto tabuľka sa dynamicky mení každú chvíľu. Pri nadviazaní komunikácie pomocou príkazu ping, kde uzol **node14** bol najprv mimo dosahu uzla a postupne sa k nemu približoval bola priemerná doba prenosu dát s veľkosťou 64B 25,419 ms.

Pre sieť s 40 uzlami bolo poslaných 1540 ICMP správ, kde ich úspešnosť dosahovala 48,2%. Počet odoslaných ICMP správ v sieti s 60 uzlami bol 3540 a ich úspešnosť doručenia bola 22,07%. V sieti s 80 uzlami bolo preposlaných 6320 ICMP správ s úspešnosťou doručenia 5,02% a sieť so 100 uzlami vykazovala úspešnosť doručenia správ iba 1,83% pri 9900 celkovo odoslaných ICMP správ. Z Tab.3 je zrejmé, že postupným zvyšovaním počtu uzlov v sieti sa nám znižuje prenosová rýchlosť a aj šírka pásma, teda možno skonštatovať, že počet uzlov v sieti má priamy vplyv na výkon siete.

Z dosiahnutých výsledkov z Tab. 4 vyplýva, že s narastajúcim počtom mobilných uzlov v sieti

Tabuľka 4: SD-MANET a sieťové parametre pre protokol **BATMAN**

Počet uzlov	Prenosová rýchlosť $[kb/s]$	Šírka pásma $[kb/s]$	Oneskorenie $[ms]$	Úspešnosť doručenia $[\%]$
20	809	1002	21,260	62,1
40	559	635	98,575	48,2
60	423	564	1925,762	22,07
80	221	289	31238,004	5,02
100	107	142	50688,293	1,83

```
mininet-wifi> dpctl dump-flows -O OpenFlow13
*** sw0 ------------------------------------------------------------------------
 cookie=0x0, duration=729.921s, table=0, n_packets=62, n_bytes=4553, priority=0 actions=CONTROLLER:65535
 cookie=0x0, duration=729.921s, table=1, n_packets=0, n_bytes=0, priority=0 actions=CONTROLLER:65535
 cookie=0x0, duration=729.921s, table=2, n_packets=0, n_bytes=0, priority=0 actions=CONTROLLER:65535
 cookie=0x0, duration=729.921s, table=3, n_packets=0, n_bytes=0, priority=0 actions=CONTROLLER:65535
 cookie=0x0, duration=729.921s, table=4, n_packets=0, n_bytes=0, priority=0 actions=CONTROLLER:65535
```

Obr. 4: Toky v simulátore Mininet-WiFi.

klesá šírka pásma aj prenosová rýchlosť. Protokol BATMAN dosahuje v priemere nižšie prenosové rýchlosti a šírku pásma ako OLSR a poskytuje nižšiu úspešnosť doručenia ICMP správ pri počte uzlov 20 – 60 oproti protokolu OLSR, na druhej strane lepšie zvláda doručenie správ pri vyššom počte (80 až 100 uzlov), ale za cenu nižšej prenosovej rýchlosti. Počas simulácie sme taktiež zistili, že priemerné oneskorenie prijatých ICMP správ pri použitom protokole BATMAN vykazuje nižšie hodnoty ako protokol OLSR.

IV. Záver

V článku sme venovali problematike integrácie softvérovo definovaných sietí (SDN) a mobilných adhoc sietí (MANET) v kontexte budúcich 6G technológií.

Simulácia a následná analýza ukázala v simulátore Mininet-Wifi, že počet uzlov a vzdialenosť medzi uzlami v súčasnosti vo veľkej miere vplýva na celkový výkon siete. Z dosiahnutých výsledkov sme taktiež zistili, že smerovací protokol OLSR je vhodnejší pre siete, kde vzdialenosť medzi uzlami je veľká a protokol BATMAN sa viac hodí v sieťach s viacerými uzlami.

Na záver možno konštatovať, že integrácia SDN do MANET v prostredí 6G predstavuje perspektívne riešenie pre budúce bezdrôtové komunikácie. Tento článok predstavuje úvodný krok integrácie do simulačného prostredia Mininet-Wifi. Dúfame, že postupným vývojom simulačných nástrojov a integrácie 6G do nich táto práca poslúži ako základ pre ďalší výskum a vývoj v oblasti SDN a MANET.

Poďakovanie

Tento výskum bol financovaný agentúrou na podporu výskumu a vývoja SR, výskumný grant č. APVV-17-0208 a VEGA 1/0260/23.

Literatúra

[1] G. Kaur and P. Thakur, "Routing protocols in manet: An overview," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, vol. 1, 2019, pp. 935–941.

[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] F. A. Yaseen, N. A. Alkhalidi, and H. S. Al-Raweshidy, "Itor-sdn: Intelligent tor networks-based sdn for data forwarding management," *IEEE Access*, vol. 12, pp. 4792–4800, 2024.

[4] 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.

[5] H. Iesar, W. Iqbal, Y. Abbas, M. Y. Umair, A. Wakeel, F. Illahi, B. Saleem, and Z. Muhammad, "Revolutionizing data center networks: Dynamic load balancing via floodlight in sdn environment," in *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*, 2024, pp. 1–8.

[6] X. Chen, T. Wu, G. Sun, and H. Yu, "Software-defined manet swarm for mobile monitoring in hydropower plants," *IEEE Access*, vol. 7, pp. 152 243–152 257, 2019.

[7] 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.

[8] S. Asadollahi and B. Goswami, "Experimenting with scalability of floodlight controller in software defined networks," in *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, 2017, pp. 288–292.

[9] S. Rabia, S. I. I, G. Lilia, and K. Benjamin, "Sdmanet: Enhancing manets with hybrid protocols through sdn integration," in *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, 2024, pp. 1–8.

[10] M. Abdollahi, W. Ni, M. Abolhasan, and S. Li, "Software-defined networking-based adaptive routing for multi-hop multi-frequency wireless mesh," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 13 073–13 086, 2021.

Automated Blocking of Elements Overlapping Web Content

[1]*Jozef AVUK*, [2]*Matúš SULÍR*

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovakia
[1]jozef.avuk@student.tuke.sk, [2]matus.sulir@tuke.sk

Abstract **– A large portion of websites contain unwanted visual elements overlapping their meaningful content, such as cookie notices, discount offers, or other notifications. Although there exist multiple solutions for web content blocking, they are usually based on manually curated and periodically maintained lists. We present a machine learning model, based on the XGBoost algorithm, that was trained on a large number of elements to fully automatically distinguish unwanted overlays on web pages. In an evaluation on real-world websites, it achieved a 73% recall and 100% precision. We also developed a browser extension that utilizes this model to block unwanted overlays during web browsing.**

Keywords **– Websites, overlays, blocking, machine learning, browser extension**

I. Introduction

Many websites contain non-essential visual elements that overlap their primary content, distract the user, and thus hinder user experience. Examples of such elements include cookie notices, discount offers, news, notifications about related events, assistance chatbots, and many more. In the past, pop-ups were represented by separate browser windows; therefore, blocking the function call "window.open" was sufficient to eliminate them [1]. As web development progressed, new methods emerged. This includes the use of iframes, CSS properties such as z-index, and a vast range of JavaScript methods. Consequently, detecting unwanted visual elements is more complicated, and more attention has been devoted to finding ways to block these types of elements.

Currently, there are several commercial tools available that focus on blocking unwanted content. AdBlock and I Don't Care about Cookies are among the most popular tools used by users. However, these tools employ static blocking techniques, relying on content lists that require regular maintenance and updates. Our goal is to create a machine learning model to detect unwanted web overlays without the necessity to manually update a list of rules for every specific website. The model should be then utilized in a browser extension blocking unwanted elements. The datasets, models, and the browser extension is available online[1].

II. Related Work

As the problem grew, so did the interest in finding a solution. The use of blockers is widespread among web users, with up to 200 million desktop and 300 million mobile devices having them, and these numbers continue to grow [2]. Now we will review machine learning approaches for web content blocking, selected specific lists, and element blocking software.

A. Machine Learning Approaches

There exist multiple machine learning approaches for blocking advertisements on websites. For example, AdRemover [3] tries to block ads using random forests solely based on URL addresses of elements. Percival [4] analyzes the visual representation of websites and detects ads using deep learning, specifically a customized SqueezeNet architecture. NoMoAds [5] focuses on mobile advertisement detection, utilizing the C4.5 decision tree algorithm.

Another similar problem is the blocking of tracking elements that record the user's behavior, including sensitive information, on multiple websites to provide more targeted advertisements. DMTrackerDetector [6] is a machine learning approach to detecting tracking elements. The authors included multiple classifiers in the evaluation, J48 being the most efficient based on the results.

[1]https://git.kpi.fei.tuke.sk/kpi-zp/2023/dp.jozef.avuk/dp-app

TrackerDetector [7] analyzes JavaScript files for the presence of various expressions and patterns related to tracking. In the evaluation, the BFTree classifier achieved the best accuracy.

Finally, some approaches combine the detection of advertisements and tracking code together. AdGraph [8] builds a graph representation of multiple aspects of a website, namely the HTML content, network requests and JavaScript execution. The approach utilizes a random forest model and achieved 95% accuracy when compared to human-generated filter lists. AdFlush [9] is a solution with improved accuracy and performance, making this approach a candidate for real-world deployment.

Although the aforementioned approaches are focused on blocking ads and tracking content, we are not aware of any approach utilizing machine learning specifically for blocking elements overlapping meaningful website content. While some of these elements might be advertisements, there exist also many non-ad overlays, such as cookie preference dialogs. This aspect makes our work unique.

B. Lists

1) Fanboy's Annoyances List: Fanboy's Annoyance List[2] specifically focuses on blocking social media content, pop-up windows, advertisements, and other actions that users may find annoying. It is a manually curated list that requires nontrivial maintenance effort. Lists of this type are regularly reviewed and updated, with contributions from developers as well as regular users [10].

2) EasyList: EasyList[3] offers a wide range of benefits for users. Its existence is closely related to the aforementioned browser extensions, as they are the ones that utilize it. EasyList allows users to actively contribute by submitting inappropriate and annoying site elements to its database. With this information, extensions can effectively block such content. Additionally, EasyList provides separate lists for certain countries, which contain blocked content specific to websites in their respective languages. It supports three basic types of rules [11]:

- Network rules – used to identify URLs associated with requests to block them.
- Element rules – used to identify HTML elements that should be hidden.
- Exemption rules – used to grant exemptions for URLs that should not be blocked.

C. Blocking Software

1) Poper Blocker: Poper Blocker is an extension focusing exclusively on pop-up windows. Its goal is to block windows that to a large extent prevent the normal functionality of the site and impair the flow of movement on the site. However, he can likewise block smaller annoying pop-ups (annoying content that appears in new windows), various overlays, and notifications for deactivating ad blockers [12].

2) AdBlock: AdBlock[4] provides a significant number of features that ensure a smooth browsing experience by blocking ads. Like other similar extensions, it utilizes content filtering, which is connected to the aforementioned EasyList, from which AdBlock retrieves data to block unwanted content.

The extension offers various functions, such as the ability for users to choose whether they allow certain ads categorized as intrusive. They can also customize which ads, if any, they want to see on platforms like YouTube or Twitch to support content creators. Additionally, users can customize general functions such as displaying the number of blocked ads on a site or enabling anonymous data collection. Advanced settings are available for experienced users.

The filter list is arguably the most crucial part and function offered by AdBlock. By adjusting the filters, it is possible to combat circumvention techniques, protect against cryptocurrency mining, and block pop-up notifications related to cookies that are automatically accepted for proper site functionality when necessary.

3) Stands Fair AdBlocker: This extension, similar to AdBlock, provides functions to filter out inappropriate content. It is specifically designed for the Google Chrome browser and offers the ability to customize the number of acceptable ads displayed on a page. Fair ads are considered those that do not contain offensive or inappropriate content and also contribute to the website's revenue [13]. The extension also includes features to block malware, pop-ups, and tracking.

[2] https://fanboy.co.nz/
[3] https://easylist.to/
[4] https://getadblock.com/

4) Ghostery: It primarily focuses on user data security [14]. This includes protection against malicious code that may appear in these windows and potentially allow criminals to steal data [15]. In terms of the elements themselves, their presence in large numbers on a page can cause browsing issues due to resource overload [16]. Despite increasing reports and subsequent blocking of scam attempts, the security measures are still insufficient, as the number of these attempts is significantly higher than what is reported [17].

As an open-source extension, it provides the opportunity for interested individuals to explore its inner workings. This allows users to understand how it functions and the data it accesses, and examine the entire source code.

5) I Don't Care about Cookies: This web extension[5] aims to remove website notifications related to cookies. The presence of cookie-related pop-ups can be unpleasant for users as they often remain on the screen until the user interacts with them. Additionally, in many cases, access to the website is restricted unless the user accepts cookies. This extension automatically blocks these pop-ups and notifications on nearly all web pages, significantly improving the user experience and interaction within websites.

III. Data Collection

In order to address this problem effectively, it was crucial to create a comprehensive dataset that includes a large number of informative elements capable of capturing key characteristics for distinguishing unwanted content.

The data collection was performed semi-automatically. One of the authors browsed a collection of websites, and for each HTML element (tag) on the websites, the following properties were automatically collected using a browser extension:

- URL - the URL of the web page where the element is located,
- Tag - the name of the HTML tag,
- Class - CSS element classes,
- ID - HTML element ID,
- Z-index - Z-index of the element,
- Position - the element's position type in CSS, e.g., relative or absolute,
- ChildElements - number of child elements,
- Location - the location (coordinates) of the element on the page
- Size - the size of the element,
- Color - the color of the element,
- FontSize - font size of the element,
- Visibility - determines whether the element is currently visible on the page,
- Display - CSS element display style,
- DOM-Index - position of the element in the Document Object Model,
- HTML - the full HTML code of the element,
- Text-Content - text content of the element.

There is one label for each element:

- Unwanted - specifies whether the element is unwanted and can be removed from the site.

Each time an unwanted overlay was displayed, the author manually clicked on it and our browser extension marked it as unwanted in the database. In total, more than 22 million elements were collected.

IV. Dataset Preprocessing

After creating the dataset, it was necessary to focus on its pre-processing. Editing our dataset consisted of the following steps:

- selection of appropriate attributes,
- data formatting,
- processing of categorical attributes,
- and processing of numerical attributes.

A. Selection of Appropriate Attributes

In the selection process of choosing the most relevant attributes, the following ones were removed:

- URL – it would contribute to bias, as the goal of the model is to predict elements on various websites.

[5] https://www.i-dont-care-about-cookies.eu/

- HTML – not easily parseable by machine learning models, and it is a bearer of a lot of duplicate information contained in other features.
- Tag – the attribute contained a huge number of unique values, and we could not determine the final set of its values.
- Class – the attribute also contained a large number of unique values.
- ID – the attribute contained a large number of unique values, some of which corresponded to one of the groups of values of the "Class" attribute.
- Text-Content – due to the multilingual context of the attribute values, which were sometimes misleading, the attribute was removed.

The majority of the removed features are long texts that are not suitable for direct inclusion in conventional machine learning models. Some of the information contained in these removed textual attributes (e.g., HTML) is indirectly present in the preserved features (such as ChildElements). An analogous approach was used by AdGraph [8] that included only selected properties of the URLs (such as length) instead of the whole URLs in its feature set.

B. Data Formatting

Several modifications were made to the format of certain attributes within the dataset. These modifications were necessary because these attributes initially contained multiple values within a single cell. The modifications were specifically applied to the Location, Size, and Color attributes. For the Location attribute, new attributes were created to represent the X and Y components separately. The Size attribute was divided into Width and Height attributes to provide more detailed information about the size of each element. As for the Color attribute, which was originally in RGBA format, each color component was extracted individually. However, it was later determined that the color attribute did not significantly contribute to the accuracy of the model's predictions. Therefore, only the Alpha component, representing the transparency of the element, was retained. Additionally, the FontSize attribute was adjusted. The original values were non-integer pixel values, which were then rounded to whole numbers and converted into numerical form. These modifications were made to enhance the structure and usability of the dataset for further analysis and model training purposes.

C. Processing of Categorical Attributes

We include non-numeric, categorical attributes in our dataset. In the case of the created dataset, it was necessary to process the categorical attributes like Position, Visibility, and Display.

The process of handling the categorical values in the dataset involved using the OneHotEncoder method. This method is primarily used for nominal data, which means data that are not directly comparable. Since our case involved nominal data, this method was selected.

The technique of OneHotEncoder involves converting each categorical value into a binary vector with values 0 and 1. In this process, new columns were created, each representing a unique value that the attribute took on. For each unique value, corresponding numerical columns with values of 0 and 1 were generated.

By applying the OneHotEncoder method, the categorical attributes were transformed into a suitable format that could be effectively utilized in further analysis and machine learning models.

D. Processing of Numerical Attributes

The process involved adjusting the deviations of numerical values within the dataset. Certain key attributes exhibited significant variations in magnitude, which needed to be addressed. Additionally, some of these attributes contained text values that were replaced with numerical equivalents (such as Z-index).

To address these issues, the MinMaxScaler function from the scikit-learn Python library was utilized. This function scaled the numerical values to a common range, resolving the problem of varying magnitudes and ensuring consistency within the dataset.

The following attributes were subject to scaling of numerical values: Z-index, ChildElements, Location_X, Location_Y, Width, Height, Color, FontSize, and DOM-Index.

V. Machine Learning Model

After creating and subsequently modifying the dataset, it was possible to proceed with further steps in the work. These steps involved selecting the appropriate algorithm, implementing it, creating the actual web extension, and deploying it for evaluation.

In the process of creating and preparing the dataset, the next step was to select an optimal algorithm for the problem at hand. During the training of these models, the data was divided into an 80:20 ratio, where 80 percent represented the training data and the remaining 20 percent was used for testing.

A. Algorithm

The selection of the algorithm was largely influenced by the size of the dataset, the type of input data, the type of desired output values, and also by the fact that this model will need to be deployed in a real environment. This means that the produced results need to be as optimal and accurate as possible. We selected from the following algorithms: RandomForestClassifier, DecisionTreeClassifier, GradientBoostingClassifier, GaussianNB, XGBClassifier. The selection was based mainly on the classification algorithms used in existing approaches mentioned in section II-A, which often successfully utilized tree-based approaches (e.g., random forest in AdRemover [3] and AdGraph [8]).

Table 1
Performance of Individual Machine Learning Algorithms

Algorithm	**False Positive Rate**	**Recall**
Logistic Regression	0.0004	0.1799
Random Forest	0.0001	0.7633
Decision Tree	0.0004	0.8426
Gradient Boost	0.0011	0.5325
Gaussian Naive Bayes	0.6541	0.9976
XGBoost	0.0001	0.8107

Multiple algorithms achieved very good results. Although Decision Tree had a slightly better recall, XGBoost achieved the lowest false alarm rate, which is a crucial indicator because blocking essential elements on a webpage could result in errors and hinder the user experience even more than not using any blocker at all. Therefore, it was chosen as the most suitable algorithm for this specific problem.

B. Hyperparameters

To select the best hyperparameters for the XGBoost algorithm, we used the TuneSearchCV method from the Tune_sklearn[6] library. It performs cross-validation training until termination criteria are fulfilled and returns the set of the best hyperparameters. In our case, the selected hyperparameters were: n_estimators=200, learning_rate=0.1, max_depth=25, subsample=1.0, colsample_bytree=0.7, min_child_weight=3, gamma=0.3, reg_alpha=1, reg_lambda=1.

VI. Browser Extension

Next, we will describe the design and implementation of the browser extension for blocking unwanted elements. Since we used Python machine learning libraries and browser extensions are written using JavaScript, we decided to use a client-server architecture connecting code in these two programming languages. The client, which is the web browser extension, sends data about the identified elements to the server's API endpoint. The server then processes the data based on the trained model. After processing the data, the server generates a prediction and sends the result back to the client. The client receives information about whether the element should be removed or kept based on the prediction.

A. Server

The server side was programmed using the Flask library as an API interface. This API interface contains only one endpoint called "predict", to which the data about the elements on the web page is sent using the POST method. Subsequently, these received data are processed in the same way as during the model training. This is made possible because all the necessary preprocessing functions, such as MinMaxScaler or OneHotEncoder, are executed on the server. After that, the trained model, which is also stored on the server, is loaded. Once the model is loaded, it can create the final prediction of whether specific elements are desired or undesired. The resulting prediction contains values ranging from 0 to 1, where a value closer to 1 indicates that the model is more confident that the element is undesired.

[6]https://github.com/ray-project/tune-sklearn

B. Client

The client-side part takes the form of a browser extension for various types of browsers. Its primary function is to collect all elements, whether visible or invisible, retrieve their properties, and then send them to the server side. A simple user interface was created for this extension. It consists of a button that enables or disables the extension. Additionally, the extension displays the domain the user is currently on. During the implementation of the client-side solution, the most important aspect was to solve the method of sending data about elements and ensure their sending even in cases where elements are created on the page with a delay. We addressed this issue by setting a timer to 800 milliseconds. This ensured the initial page load and sending of all elements to the server. In the case of elements that appeared later on the page, we utilized the MutationObserver interface. It triggers a specified function when changes occur on the page. This interface allows us to obtain a list of newly added elements, which in our case are used to send additionally displayed elements to the server.

VII. Evaluation

We evaluated the effectiveness of our approach through practical testing on twenty real web pages. In this testing, a web page was loaded without any blocking mechanism. The total number of elements present on the page was recorded. After examining the web page, the number of elements that hindered the usability or overlapped the content was determined. The actual testing process involved running our extension, which blocked the elements identified as unwanted by the model. By obtaining the count and list of specific blocked elements, it was possible to determine the number of correctly and incorrectly blocked elements. The results are presented in Table 2.

Table 2
Results of the evaluation on selected real-world websites

Website	**Total elements**	**Unwanted**	**True positives**	**False positives**	**Recall**	**Precision**
www.bbc.com	1453	2	2	0	100%	100%
www.nytimes.com	2294	1	1	0	100%	100%
www.cnn.com	2641	2	2	0	100%	100%
www.wsj.com	3703	2	1	0	50%	100%
www.nbcnews.com	2489	1	1	0	100%	100%
www.theguardian.com	3553	1	0	0	0%	100%
www.bloomberg.com	1896	1	1	0	100%	100%
www.auto.cz	3224	1	1	0	100%	100%
www.huffpost.com	2498	1	1	0	100%	100%
www.foxnews.com	3621	0	0	0	100%	100%
www.blesk.cz	3884	1	1	0	100%	100%
www.idnes.cz	1552	1	1	0	100%	100%
www.ta3.com	3583	1	1	0	100%	100%
www.tvnoviny.sk	1418	2	1	0	50%	100%
www.aktuality.sk	1391	1	1	0	100%	100%
www.sme.sk	16009	2	1	0	50%	100%
www.novinky.cz	1878	2	0	0	0%	100%
www.denik.cz	1203	2	1	0	50%	100%
www.lidovky.cz	946	1	1	0	100%	100%
www.topky.sk	2210	1	1	0	100%	100%
Total	***61446***	***26***	***19***	***0***	***73%***	**100%**

The practical evaluation of the extension led to the conclusion that the implemented model successfully identifies unwanted elements that overlap the content with a 73% recall. Out of the 26 unwanted elements, 19 were accurately detected and removed. Furthermore, no false positives were recorded, ensuring that the meaningful content of the page remained untouched. This does not indicate overfitting, since due to the nature of the problem, we intentionally prioritized a lower false positive rate during the algorithm selection, as we already mentioned.

VIII. Conclusion

In this paper, we presented a machine learning model and a browser extension to block unwanted elements covering the meaningful content of the websites. Based on our practical testing, we can state that our solution for blocking unwanted content does not interfere with the functionality of websites during the removal of elements. In the future, we could improve the model by using more advanced machine learning approaches, such as neural networks.

ACKNOWLEDGMENT

This work was supported by project VEGA No. 1/0630/22 Lowering Programmers' Cognitive Load Using Context-Dependent Dialogs.

REFERENCES

[1] T. Petricek, "Pop-up from hell: On the growing opacity of web programs," 10 2021. [Online]. Available: https://tomasp.net/blog/2021/popup-from-hell/
[2] K. Garimella, O. Kostakis, and M. Mathioudakis, "Ad-blocking: A study on performance, privacy and counter-measures," in *Proceedings of the 2017 ACM on Web Science Conference*, 2017, pp. 259–262.
[3] T. Vo and C. Jaiswal, "AdRemover: The improved machine learning approach for blocking ads," in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2019, pp. 1–4.
[4] Z. A. Din, P. Tigas, S. T. King, and B. Livshits, "PERCIVAL: Making In-Browser perceptual ad blocking practical with deep learning," in *2020 USENIX Annual Technical Conference (USENIX ATC 20)*. USENIX Association, Jul. 2020, pp. 387–400.
[5] A. Shuba, A. Markopoulou, and Z. Shafiq, "NoMoAds: Effective and efficient cross-app mobile ad-blocking," *Proceedings on Privacy Enhancing Technologies*, vol. 4, pp. 125–140, 2018.
[6] Q. Wu, Q. Liu, Y. Zhang, P. Liu, and G. Wen, "A machine learning approach for detecting third-party trackers on the web," in *Computer Security – ESORICS 2016*, I. Askoxylakis, S. Ioannidis, S. Katsikas, and C. Meadows, Eds. Cham: Springer International Publishing, 2016, pp. 238–258.
[7] Q. Wu, Q. Liu, Y. Zhang, and G. Wen, "TrackerDetector: A system to detect third-party trackers through machine learning," *Computer Networks*, vol. 91, pp. 164–173, 2015.
[8] U. Iqbal, P. Snyder, S. Zhu, B. Livshits, Z. Qian, and Z. Shafiq, "AdGraph: A graph-based approach to ad and tracker blocking," in *2020 IEEE Symposium on Security and Privacy (SP)*, 2020, pp. 763–776.
[9] K. Lee, C. Lim, B. Jin, T. Kim, and H. Kim, "Adflush: A real-world deployable machine learning solution for effective advertisement and web tracker prevention," in *Proceedings of the ACM on Web Conference 2024*, ser. WWW '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 1902–1913.
[10] S. S. Hashmi, M. Ikram, and S. Smith, "On optimization of ad-blocking lists for mobile devices," in *Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2019, pp. 220–227.
[11] P. Snyder, A. Vastel, and B. Livshits, "Who filters the filters: Understanding the growth, usefulness and efficiency of crowdsourced ad blocking," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 4, no. 2, 2020.
[12] "Pop up blocker for chrome™." [Online]. Available: https://poperblocker.com/
[13] Y. Karo, "Stands fair adblocker." [Online]. Available: https://www.standsapp.org/
[14] D. Bouhnik and G. Carmi, "Interface application comprehensive analysis of Ghostery," *International Journal of Computer Systems*, vol. 5, no. 3, p. 03, 2018.
[15] M. G. Alkhozae and O. A. Batarfi, "Phishing websites detection based on phishing characteristics in the webpage source code," *International Journal of Information and Communication Technology Research*, vol. 1, no. 6, 2011.
[16] R. Bilton, "Ghostery: A web tracking blocker that actually helps the ad industry," 2022. [Online]. Available: https://venturebeat.com/2012/07/31/ghostery-a-web-tracking-blocker-that-actually-helps-the-ad-industry/
[17] S. Al-Rousan, A. Abuhussein, F. Alsubaei, L. Collen, and S. Shiva, "Ads-guard: Detecting scammers in online classified ads," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2020, pp. 1492–1498.

Návrh hierarchickej dátovej štruktúry pre ukladanie geometrie časovo-variabilných 3D scén

[1]*Kristián STUTIAK,* [2]*Branislav MADOŠ,* [3]*Heidar KHORSHIDIYEH*

[1,2,3] Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]kristian.stutiak@student.tuke.sk, [2]branislav.mados@tuke.sk,
[3]heidar.khorshidiyeh @tuke.sk

***Abstrakt* — Článok sa zoberá problematikou reprezentácie geometrie voxelizovaných troj-dimenzionálnych scén prostredníctvom hierarchických dátových štruktúr založených na oktantových stromoch resp. orientovaných acyklických grafoch, ktorých hlavnými reprezentantmi sú dátová štruktúra sparse voxel octrees. resp. sparse voxel directed acyclic graphs. Z týchto dátových štruktúr boli odvodené ďalšie, pridávaním ďalších vlastností, ktoré umožňujú vyššiu kompresiu dát. Veľká pozornosť je v tejto súvislosti venovaná bezstratovej reprezentácii geometrie, aj keď sa objavujú aj dátové štruktúry a súvisiace algoritmy ich zostavovania, ktoré sú založené na stratovej reprezentácii geometrie. Väčšina týchto dátových štruktúr bola zameraná na ukladanie statických troj-dimenzionálnych scén. Pozornosť však začala byť venovaná aj takým dátovým štruktúram, ktoré umožňujú počas aktívneho používania dátovej štruktúry zmenu kódovanej informácie, teda zmenu geometrie scény, bez nutnosti rozsiahlej rekompresie dátovej štruktúry. Osobitnou kategóriou je potom snaha o zakódovanie geometrie časovo-variabilnej troj-dimenzionálnej scény, teda o akési zachytenie troj-dimenzionálnej videosekvencie. Práve tejto problematike sa venuje tento článok a prináša návrh hierarchickej dátovej štruktúry založenej na sparse voxel octrees, ktorá ale pridáva aj kódovanie štvrtej dimenzie, do ktorej sa zaznamenáva informácia o zmenách geometrie troj-dimenzionálnej scény v čase.**

***Kľúčové slová* — geometria troj-dimenzionálnej scény, bezstratová kompresia dát, časovo-variabilné hierarchické dátové štruktúry, sparse voxel octrees, sparse voxel directed acyclic graphs,**

I. Úvod

Reprezentácia informácií vo forme multi-dimenzionálnych pravidelných mriežok dát je bežnou súčasťou rôznych aplikácií v oblasti vedy, výskumu, aplikovaného výskumu a vývoja v priemysle, ale aj medicíne a ďalších, skrátka všade, kde je vhodné dáta takto usporiadať. Následne je bežným prístupom tieto dáta vizualizovať, aby bolo možné lepšie pochopiť súvislosti, ktoré sú týmito dátami reprezentované.

Medzi bežne vizualizované patria troj-dimenzionálne dáta, ktoré môžu byť umelo vytvárané, alebo sú získavané senzormi, ako je to napríklad v prípade medicínskych zobrazovacích zariadení (počítačová tomografia) alebo v prípade priemyselných zariadení (priemyselná počítačová tomografia). Tak, ako sa zvyšoval výkon a kapacita počítačov, vizualizácia troj-dimenzionálnych dát sa stala súčasťou počítačovej grafiky v zábavnom priemysle, teda v oblasti počítačových hier, ako aj virtuálnej reality (virtual reality) a rozšírenej reality (augmented reality).

Ak je v oblasti počítačovej grafiky využívaná troj-dimenzionálna scéna, tvorená voxelmi, usporiadanými v troj-dimenzionálnej pravidelnej mriežke, je ukladanie dát predstavujúcich túto scénu náročný problém, kvôli vyžadovanej kapacite sekundárnych úložných zariadení počas archivovania týchto dát, ako aj operačnej pamäte počítača, resp. pamäte grafickej karty, počas vizualizácie týchto dát, alebo realizácie rôznych analýz.

Obvykle sú v prípade voxelizovaných scén v oblasti počítačovej grafiky priradené jednotlivým voxelom atribúty ako je farba alebo rôzne materiálové vlastnosti a medzi nimi je aj atribút, ktorý umožňuje definovať geometriu scény. Každý voxel môže byť z tohto hľadiska pasívny a vtedy sa mu nastaví príslušný atribút na hodnotu „0“ alebo je aktívny a vtedy sa tento atribút nastaví na hodnotu „1“. Aj keď objem dát predstavuje iba 1b na voxel, je tento objem dát pre troj-dimenzionálne scény neprakticky veľký a uloženie informácie o geometrii scény si vyžaduje lepšie riešenie. Medzi populárne riešenia tohto problému patria hierarchické dátové štruktúry,

založené na oktantových stromoch alebo orientovaných acyklických grafoch. Bližšie sa im venujú sekcie II a III tohto článku.

Hierarchické dátové štruktúry sú vyvíjané aktuálne hlavne pre reprezentáciu statických scén, ktoré sa po ich zostavení a po zostavení príslušných HDS, ktoré ich reprezentujú, nemenia. Môže však vznikať požiadavka na ich ad-hoc úpravu, čo si však v prípade väčšiny týchto HDS vyžaduje ich dekompresiu, úpravu a následnú rekompresiu, čo môže byť náročné na pamäť, výpočtový výkon a čas.

Ešte náročnejšia je požiadavka zakódovať informáciu o troj-dimenzionálnej scéne, ktorá je časovo-variabilná, zjednodušene povedané, ak má byť zaznamenaná akási videosekvencia meniacej sa troj-dimenzionálnej scény. Ilustráciou náročnosti môže byť riešenie otázky, aký objem dát môže predstavovať informácia o geometrii takejto časovo-variabilnej troj-dimenzionálnej scény. Ak je potrebné v nekomprimovanom, jednoduchom formáte pravidelnej mriežky reprezentovať geometriu statickej scény s rozmerom $1024 \times 1024 \times 1024$ voxelov, kde je potrebné zaznamenať 1b/vox, je pre záznam geometrie takejto scény potrebné vyčleniť až 1Gb alebo 128MB priestoru v operačnej pamäti resp. na sekundárnom úložnom zariadení. Ak je potrebné zaznamenať videosekvenciu, každé časové políčko bude vyžadovať na zaznamenanie spomínaných 128MB. Ak je teda potrebné vytvoriť videosekvenciu s 1024 časovými políčkami, čo pri 25 políčkach na sekundu predstavuje videosekvenciu s dĺžkou trvania 40,96 sekundy, bude potrebných celkovo 1Tb alebo 128GB priestoru.

Vyššie spomínané nároky na pamäť resp. úložný priestor preto zakladajú potrebu úspornejšej reprezentácie geometrie takýchto časovo-variabilných scén. Preto je skúmaná otázka, či by bolo možné vytvoriť stromovú štruktúru analogickú SVO, ktorá by zachytávala nielen priestorové informácie ale aj informácie o plynutí času a umožnila by tak úspornejšiu reprezentáciu dát. V tomto článku je skúmané, či takto zostavená dátová štruktúra môže byť na binárnej úrovni kompaktnejšia ako ukladanie geometrie scény do SVO, keď sa takáto dátová štruktúra vytvorí osobitne pre každé políčko videosekvencie.

II. Súvisiace Práce

Základnou štruktúrou pre reprezentáciu geometrie troj-dimenzionálnych scén sa stala Sparse Voxel Octrees (SVO), ktorá vo svojej štruktúre implementuje ukazovatele na dcérske uzly pre možnosť rýchleho traverzovania. Aby bolo možné vytvoriť kompaktnejšiu dátovú štruktúru, vznikli Pointerless Sparse Voxel Octrees (PSVO), ktoré nemajú vo svojej konštrukcii ukazovatele na dcérske uzly, čo však na druhej strane znemožňuje ich jednoduché a rýchle traverzovanie. Na SVO boli založené ďalšie dátové štruktúry, medzi ktoré patria napr. Clustered Spare Voxel Octrees (CSVO), ktoré v roku 2022 navrhol Madoš a kol. [1].

Baert a kol. navrhli v roku 2013 v [2,3] dvojkrokový algoritmus, kde v prvom kroku dochádza k voxelizácii polygonálneho povrchového modelu zostaveného z trojuholníkov a v druhom kroku sa vytvára SVO. Tento algoritmus navrhli ako out-of-core, to znamená, že tak vstupné dáta, ako aj medziprodukt a výstupné dáta môžu svojim objemom ďaleko presahovať veľkosť operačnej pamäte počítača, ktorá je pre spustenie programovej implementácie tohto algoritmu potrebná. Pätzold a kol. potom v roku 2015 navrhli algoritmus [4], ktorý zostavuje SVO priamo, bez potreby medziproduktu.

V roku 2013 navrhol Kämpe a kol. v [5] dátovú štruktúru Sparse Voxel Directed Acyclic Graphs (SVDAG), ktorá jednoduchou úpravou mení SVO na orientovaný acyklický graf – Directed Acyclic Graph (DAG) a umožňuje zvýšiť kompaktnosť tejto dátovej štruktúry oproti SVO. Zatiaľ čo SVO neumožňuje, aby mal dcérsky uzol viac ako jeden materský uzol, SVDAG toto umožňuje. Vzniká tak možnosť vyhľadať zhodné podstromy, nachádzajúce sa v dátovej štruktúre a úsporne ich reprezentovať tak, že sa uvedie iba jeden takýto podstrom, na ktorý sa ale bude viacnásobne odkazovať z vyššie umiestnených uzlov. V roku 2016 Villanueva a kol. v [6,7] predstavili Symmetry-aware Sparse Voxel Directed Acyclic Graphs (SSVDAG), ktorá umožňuje nájsť zhodu aj v prípade, ak je pre jej nájdenie nutné aplikovať reflektívnu symetriu v jednej, dvoch alebo vo všetkých troch osiach scény. V roku 2020 navrhol Vokorokos a kol. dátovú štruktúru Pointerless Sparse Voxel Directed Acyclic Graphs (PSVDAG), ktorá, napriek tomu, že neobsahuje ukazovatele na dcérske uzly, umožňuje spájanie spoločných podstromov [8]. Následne predstavili Madoš a Ádám v roku 2021 v [9] algoritmus pre rýchlu transformáciu PSVDAG dátovej štruktúry do SVDAG dátovej štruktúry.

Pre možnosť rýchlej zmeny uloženej geometrie scény v dátovej štruktúre predstavil Careil a kol. dátovú štruktúru HashDAG v roku 2020 v [10]. Možnosťou úspornej reprezentácie časovo-variabilných dát sa zaoberalo viacero príspevkov, medzi nimi [11-15].

III. Časovo-Variabilné Sparse Voxel Octrees

Klasická Sparse Voxel Octrees (SVO) hierarchická dátová štruktúra dekomponuje troj-dimenzionálny priestor vo všetkých troch osiach priestoru x, y, z vždy na polovicu, čím vytvára osem oktantov priestoru. Pre každý oktant sa potom posudzuje, či je daný oktant pasívny, teda neobsahuje ani jeden aktívy voxel, alebo je daný oktant aktívny, teda obsahuje aspoň jeden aktívny voxel. To sa potom pretaví do zostavenia masky dcérskych uzlov, ktorá je zložená z 8 Header Tagov (HT), kde každý zastupuje jeden takýto oktant a v prípade, ak je oktant pasívny, má daný HT hodnotu 0, ak je aktívny, má HT hodnotu 1. Poradie spracovávania oktantov ako aj poradie zoradenia HT je závislé od rozhodnutia a v počítačovej grafike sa preferuje zoradenie podľa Mortonovho poradia (Morton order).

V prípade časovo-variabilných troj-dimenzionálnych scén je potrebné zaznamenať geometriu troj-dimenzionálnej scény pre každé časové políčko. To je možné realizovať tak, že sa považuje každé časové políčko za samostatnú statickú troj-dimenzionálnu scénu a jej geometria sa teda zaznamená vždy do samostatnej SVO hierarchickej dátovej štruktúry. Ak je teda potrebné zapísať 256 časových políčok, je vytvorených 256 samostatných SVO dátových štruktúr a pri prehrávaní tejto videosekvencie sú používané jedna po druhej. Ak je realizovaný takýto postup, je možné použiť klasické SVO spolu s klasickými dobre preskúmanými kompresnými algoritmami.

V prípade časovo-variabilných troj-dimenzionálnych scén je možné pre kompresiu dát využiť časovú koherenciu dát medzi snímkami. Je teda možné zvýšiť kompresný pomer tak, že sa dáta viacerých časových políčok v sekvencii a potenciálne všetkých políčok, komprimujú do jednej hierarchickej dátovej štruktúry. Ak je teda scéna zložená z $N \times N \times N = N^3$ voxelov, kde N sa dá zapísať ako 2^m, je možné zobrať do úvahy sekvenciu N časových políčok a vytvoriť tak štvor-rozmernú mriežku voxelov, kde tri rozmery tvoria geometrické rozmery v ose x, y a z a štvrtú os t tvorí čas. Potom je možné zostaviť 4D ekvivalentnú štruktúru k SVO. Pre jej zostavenie budú platiť analogické pravidlá ako je to v prípade SVO Ak je teda počet voxelov na hrane scény a počet časových snímok rovný N, pričom $N = 2^m$, potom je vytvorených m vrstiev uzlov, pričom vrstvy sú očíslované od 0 do $m - 1$. Tie uzly, ktoré sú vo vrstvách s poradovými číslami od 0 do $m - 2$ sú označené ako interné uzly – Internal Node (INODE), uzly uložené v poslednej vrstve, teda vrstve s poradovým číslom $m - 1$ sú listové uzly – Leaf Node (LNODE). Vzhľadom k tomu, že scéna je štvorrozmerná, je delená v každej ose na polovicu a vytvára sa tak 16 podpriestorov. Preto aj maska dcérskych uzlov CHNM má 16 Header Tagov a za ňou je uložené pole ukazovateľov na aktívne dcérske uzly. Tých je variabilné množstvo, od 1 do 16. Listové uzly obsahujú iba 16b CHNM, bez ďalších ukazovateľov na dcérske uzly, pretože tie sa už v dátovej štruktúre nenachádzajú. Pre správne kódovanie obsahu tejto dátovej štruktúry je potrebné, aby počet časových snímok bol taký istý, ako je rozmer v geometrických osiach v počte voxelov. Ak je snímok menej, je možné doplniť ďalšie tak, že všetky voxely v danej snímke sú nastavené ako pasívne. Ak je snímok viac, je možné vytvoriť usporiadanú množinu 4D dátových štruktúr, keď každá z nich kóduje príslušnú sadu časových snímok.

Na Obr. 1 je možné vidieť Backus-Naurovu formu tak pre klasickú SVO hierarchickú dátovú štruktúru (Obr. 1a), ako aj pre ST SVO verziu (Obr. 1b).

```
SVO ::= (n)<NODE>
NODE ::= <INODE>|<LNODE>
INODE ::= <CHNM>(24)<BIT><PTS>
LNODE ::= <CHNM>(24)<BIT><BIT>
CHNM ::= (8)<HT>
PTS ::= (1)*(8)<PT>
PT ::= <BIT>
HT ::= <BIT>
BIT ::= "0" | "1"
```

a)

```
ST SVO ::= (n)<NODE>
NODE ::= <INODE>|<LNODE>
INODE ::= <CHNM>(16)<BIT><PTS>
LNODE ::= <CHNM>(16)<BIT><BIT>
CHNM ::= (16)<HT>
PTS ::= (1)*(16)<PT>
PT ::= <BIT>
HT ::= <BIT>
BIT ::= "0" | "1"
```

b)

kde platí:

<SYM> - je neterminálny symbol SYM
"sym" - je terminálny symbol sym
(n)<SYM> - je symbol SYM použitý n-krát
(n)*(m)<SYM> - je symbol SYM, použitý od n do m krát
| - je alternatíva

Obr.1 Backus-Naurova forma pre a) SVO a b) ST SVO

IV. VÝSLEDKY TESTOVANIA

Testovanie prebehlo s využitím troch modelov, Angel Lucy, Porsche a Skull (Obr. 2), ktoré boli pôvodne reprezentované ako polygonálne povrchové modely, zložené zo sietí trojuholníkov. Následne boli voxelizované do dvoch rôznych rozlíšení 128^3 a 256^3. Potom boli vygenerované jednotlivé časové políčka, ktorých bolo 128 resp. 256 a to tak, že v prvom prípade dochádzalo k posunu modelu v scéne o jeden voxel v ose x scény a v druhom prípade bol posun modelu zabezpečený vždy o jeden pixel v ose y scény v každom časovom políčku. Boli vygenerované zoznamy aktívnych voxelov scén, reprezentovaných ich Morton adresou, ktoré boli usporiadané podľa tejto adresy vzostupne a počas zostavovania dátových štruktúr boli v poradí od najmenšej Morton adresy po najväčšiu Morton adresu postupne načítavané na spracovanie.



Obr. 2 Vizualizácie modelov a) Angel Lucy, b) Skull a c) Porsche po ich voxelizácii

Najskôr boli pre každý model a jeho rozlíšenie vygenerované SVO pre samostatné časové políčka, teda bolo vygenerovaných vždy 128 resp. 256 SVO, ktorých veľkosť binárnej reprezentácie bola sčítaná. Následne bol pre každý model a jeho rozlíšenie vygenerovaný jeden ST SVO. Výsledné veľkosti binárnej reprezentácie množiny SVO resp. ST SVO pre každý model a rozlíšenie možno vidieť v Tabuľke 1 resp. Tabuľke 2.

Ako vidno z výsledkov zistených počas testovania, bolo v prípade použitia ST SVO v porovnaní s množinou SVO dosiahnuté skompaktnenie uloženia dát od 1.91 násobku v prípade modelu Skull voxelizovaného do rozlíšenia 256^3 až po 2.04 násobok, v prípade modelu Angel Lucy voxelizovaného do rozlíšenia 256^3.

Tabuľka 1 Výsledok kompresie časovo-variabilných scén Angel Lucy, Porsche a Skull s rozlíšeniami 128^3 a 256^3 v SVO a ST SVO verzii

	Posun modelu po osi x					
	Angel Lucy		Porsche		Skull	
Rozlíšenie	128^3	256^3	128^3	256^3	128^3	256^3
Aktívnych voxelov	22 481	91 517	54 203	233 040	74 103	298 845
3D SVO [KB]	5 769	49 133	15 912	130 371	11 562	107 563
ST SVO [KB]	2 946	25 234	8 223	67 570	5 714	53 556
Pomer	1.96	1.95	1.94	1.93	2.02	2.01

Tabuľka 2 Výsledok kompresie časovo-variabilných scén Angel Lucy, Porsche a Skull s rozlíšeniami 128^3 a 256^3 v SVO a ST SVO verzii

	Posun modelu po osi y					
	Angel Lucy		Porsche		Skull	
Rozlíšenie	128^3	256^3	128^3	256^3	128^3	256^3
Aktívnych voxelov	22 481	91 517	54 203	233 040	74 103	298 845
3D SVO [KB]	3 704	31 506	13 339	110 350	13 443	123 043
ST SVO [KB]	1 814	15 412	6 590	54 447	6 810	64 528
Pomer	2.04	2.04	2.03	2.03	1.97	1.91

V. ZÁVER

Tento článok bol venovaný problematike reprezentácie geometrie voxelizovaných troj-dimenzionálnych scén v oblasti počítačovej grafiky, pričom sa zameral na skúmanie možností kompaktnej binárnej reprezentácie tejto geometrie prostredníctvom hierarchických dátových štruktúr. Posúva problematiku od reprezentácie statických troj-dimenzionálnych scén k časovo-variabilným troj-dimenzionálnym scénam, teda zjednodušene povedané k videosekvenciám troj-dimenzionálnych scén, pričom sa snaží aplikovať pre uloženie geometrie týchto scén hierarchické dátové štruktúry založené na stromoch.

V rámci tejto práce bola skúmaná možnosť ukladania informácií o jednotlivých časových políčkach samostatne, prostredníctvom klasických SVO, tak, ako by boli zaznamenávané klasické troj-dimenzionálne statické scény, a súčasne bola skúmaná možnosť interpretovať plynutie času ako ďalší rozmer a teda možnosť vytvoriť ST SVO dátovú štruktúru. Následne boli porovnané veľkosti binárnych reprezentácií vzniknutých hierarchických dátových štruktúr.

Z testov vyplynulo, že bolo možné dosiahnuť, pri použití ST SVO, skompaktnenie uloženia dát od 1.91 násobku v prípade modelu Skull voxelizovaného do rozlíšenia 256^3 až po 2.04 násobok, v prípade modelu Angel Lucy voxelizovaného do rozlíšenia 256^3.

V ďalšom výskume bude vyhodnocované, aký vplyv na toto skompaktnenie reprezentácie dát bude mať zložitejšia dynamika scény a využitie iných hierarchických dátových štruktúr ako je napr. SVDAG, SSVDAG alebo CSVO.

POĎAKOVANIE

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

REFERENCIE

[1] Madoš, B.; Chovancová, E.; Chovanec, M.; Ádám, N. CSVO: Clustered Sparse Voxel Octrees—A Hierarchical Data Structure for Geometry Representation of Voxelized 3D Scenes. Symmetry 2022, 14, 2114. https://doi.org/10.3390/sym14102114

[2] Baert, J.; Lagae, A.; Dutré, Ph. Out-of-Core Construction of Sparse Voxel Octrees. In Proceedings of the 5th High-Performance Graphics Conference (HPG '13), Anaheim, CA, USA, 19–21 July, 2013, pp. 27–32. https://doi.org/10.1145/2492045.2492048

[3] Baert, J.; Lagae, A.; Dutré, Ph. Out-of-Core Construction of Sparse Voxel Octrees. Computer Graphics Forum 2014, 33, pp. 220–227. ISSN 0167-7055. https://doi.org/10.1111/cgf.12345

[4] Pätzold, M.; Kolb, A. Grid-free out-of-core voxelization to sparse voxel octrees on GPU. In Proceedings of the 7th Conference on High-Performance Graphics (HPG '15), Los Angeles, CA, USA, 7–9 August 2015, pp. 95–103. ISBN 9781450337076. https://doi.org/10.1145/2790060.2790067

[5] Kämpe, V.; Sintorn, E.; Assarson, U. High Resolution Sparse Voxel DAGs. ACM Transactions on Graphics 2013, 32, pp. 1–13. ISSN 0730-0301. https://doi.org/10.1145/2461912.2462024

[6] Villanueva, A.J.; Marton, F.; Gobetti, E. Symmetry-aware Sparse Voxel DAGs. In Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D '16), Redmond, WA, USA, 27–28 February, 2016; pp. 7–14. ISBN 978-1-4503-4043-4. https://doi.org/10.1145/2856400.2856420

[7] Villanueva, A.J.; Marton, F.; Gobetti, E. Symmetry-aware Sparse Voxel DAGs (SSVDAGs) for compression-domain tracing of high-resolution geometric scene. J. Comput. Graph. Tech. (JCGT), 2017, 6, pp. 30. http://jcgt.org/ published/0006/02/01

[8] Vokorokos, L.; Madoš, B.; Bilanová, Z. PSVDAG: Compact Voxelized Representation of 3D Scenes Using Pointerless Sparse Voxel Directed Acyclic Graphs. Computing and Informatics, 2020, 39, pp. 587-616. ISSN 1335-9150 (print); 2585-8807 (online). https://doi.org/10.31577/cai_2020_3_587

[9] Madoš, B.; Ádám, N. Transforming Hierarchical Data Structures—A PSVDAG—SVDAG Conversion Algorithm. Acta Polytechnica Hungarica, 2021, 18, pp.47-66. ISSN 1785-8860. https://doi.org/10.12700/APH.18.8.2021.8.3

[10] Careil, V.; Billeter, M.; Eisemann, E. Interactively Modifying Compressed Sparse Voxel Representations. In Computer Graphics Forum, 2020, 39, pp. 111–119. ISSN 0167-7055. https://doi.org/10.1111/cgf.13916

[11] Kämpe, V.; Rasmuson, S.; Billeter, M.; Sintorn, E.; Assarsson, U. Exploiting Coherence in Time-Varying Voxel Data. In Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, Redmond, WA, USA, 27–28 February 2016; pp. 15–21. ISBN 9781450340434. https://doi.org/10.1145/2856400.2856413

[12] Martinek, M.; Thiemann, P.; Stamminger, M. Spatio-temporal filtered motion DAGs for path-tracing. In Computers & Graphics, 2021, 99, pp. 224-233. ISSN 0097-8493. https://doi.org/10.1016/j.cag.2021.07.008

[13] Zhang, J.; Owen, C.B. Octree-Based Animated Geometry Compression. In Proceedings of the Conference on Data Compression,Washington, DC, USA, 23–25 March 2004; p. 508. ISBN 978-0-7695-2082-7

[14] Zhang, J.; Xu, J. Optimizing Octree Motion Representation for 3D Animation ACM-SE 44. In Proceedings of the 44th annual Southeast Regional Conference, Melbourne, FL, USA, 10–12 March 2006; pp. 50-55. ISBN 978-1-59593-315-7. https://doi.org/10.1145/1185448.1185461

[15] Zhang, J.; Xu, J.; Yu, H. Octree-Based 3D Animation Compression with Motion Vector Sharing. In Proceedings of the 2007 4th International Conference on Information Technology New Generations, Las Vegas, NV, USA, 2–4 April 2007; pp. 202–207. https://doi.ieeecomputersociety.org/10.1109/ITNG.2007.138

Gathering Information about Events in Košice

[1]*Ivan YATSKIV,* [2]*Miroslav MURIN,* [3]*Ondrej KAINZ,*
[4]*Miroslav MICHALKO*

[1,2,3,4]Department of Informatics, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Slovak Republic

[1]ivan.yatskiv@student.tuke.sk, [2]miroslav.murin@tuke.sk, [3]ondrej.kainz@tuke.sk,
[4]miroslav.michalko@tuke.sk

Abstract **– Event information gathering systems have become a useful component of modern life, due to the fact that they offer a way to discover more about cultural, social, and educational happenings. A few existing solutions were analysed before the design phase to identify their characteristics and limitations. This piece of writing is dedicated to show how our system for gathering information about events was designed and implemented. The developed system is capable of the addition, modification, and removal of event details, and also incorporates advanced features. Moreover, it makes use of advanced security services for authentication and authorization. The system was evaluated with a target group and successfully met all defined objectives, including additional functionalities.**

Keywords **– event filtering, events, JWT, security, programming, web application**

I. Introduction

Košice, the cultural centre of eastern Slovakia, is bursting with life and hosts a great amount of cultural, sporting and social events that offer unforgettable experiences to both locals and visitors. Košice is well known for its active event management, with annual international film festivals, music concerts, exhibitions, sporting events and local fairs reflecting the cultural diversity of the region.

Despite the enormous number of events available, there is a significant gap in the centralisation and accessibility of event information, which often makes it a challenge for residents and visitors to take full advantage of what the city has to offer. Current channels for collecting event information are scattered all across the bulletin boards, campus boundaries, and various social media platforms. This fragmented approach often turns out into incomplete, outdated or poorly presented information. The consequence is a population that is poorly informed about local events, which prevents them from participating and fully engaging in community events.

The problem of information overload is especially visible in the city centre and among students, who are more like others constantly bombarded with notices and posters within the campus and dormitories. This phenomenon, known as the "active user paradox", described by author in [1], suggests that the more information is provided, the less effectively it is used by the target audience. Thus, although Košice is bursting with activity, the community's ability to navigate this sea of information and find relevant events is quite limited.

To address the problem, this paper proposes the creation of a centralized information system that aims to simplify the process of finding and retrieving information about events in and around Košice. The system will allow users to filter events by interest or date, thus increasing the clarity and accessibility of information.

In addition, the system is intended to create a platform for event organisers to promote their activities more effectively, reach a wider audience and encourage richer interaction in the community. By providing a centralised hub of event information, the system will be a medium between event organisers and attendees, improving the overall user experience.

II. Analysis of existing event gathering systems

Event gathering systems are designed to simplify the process of planning, organising and promoting different events, but very often they are not clear enough for the end user. As the author argues in [2], what makes the system convenient is the ease of use. If users can switch from intent to actions without having to think much, the design has succeeded and is convenient.

A. Project Eventmie

One of the large selection of solutions to the problem of collecting events is Eventmie. As the author argues in [3], sometimes less is more. But not in the above context. The functionality of the service contains only basic elements: a small list of events, a search bar, a calendar to select dates, and a drop-down menu with event categories.

At first glance, the web-page appears minimalistic and intuitive, but during testing and attempting to use its functionality, several major shortcomings were discovered.

The search bar suffers from a lack of responsiveness - it does not respond to user input using suggestions (autocomplete). This practice, as Donald A. Norman argues in his work [4], provokes the user frustration because it does not match expectations and needs in use and also represents a violation of the principles of user-friendliness. If a search for an event or city according to the criteria entered in the dashboard yields results, it is easily recognizable that the input data has been entered correctly. However, in a situation where the database does not contain any event information about the city or event, the user is not notified about the absence of data, which can lead to unnecessary review and doubts as, to whether user has entered the correct information.

The calendar is well done. It offers the possibility of a choice of events for a specific day,“today”, “this/next week”, “this/next month”, and “this weekend”. There is also an option to select the time interval in which the corresponding events will be displayed. The design is intuitive, and as the author insists in [3], when an object is easy to interpret and understand, it means that it is also well-designed.

As regards the overall appearance and functionality of this planning service, it is evident that improvements are needed to increase user-friendliness, responsiveness, and event information relevancy.

B. Do312

Another service called Do3122 is aimed at supporting the residents of the city of Chicago in planning their upcoming free weekends.

Looking at the main page we can see a friendly color palette, which appears at first glance to contain all the information the user needs. Do312’s announcement centralisation solution has an extended functional scope and presents itself in a more maintainable form. Nevertheless, even this interface is not perfect and would certainly benefit from further improvements.

As stated by the author in [5], elements that are related should be grouped together, while unrelated elements should be separated to avoid confusion. In the present case, the navigation icons are visible on the panel and the user does not make it difficult for the user to find the calendar, filters, or search bar. A nice bonus is also the ability to view the time sequence in which they published new announcements of available activities, alongside articles and posts related to the topics of those events.

The search tool is fairly well implemented, but not in the best possible way. In a single search menu for venues, the events themselves, and even for individual performers or organisations, the results are displayed in a single table. Although the solution attempts to separate individual specifications and help the user distinguish between them at a logical level through icons, the interface is still not completely intuitive and the meaning of individual icons is not easily understood. A similar type of behavior of websites violates one of the basic principles of user experience - intuitive mapping.

The calendar has very limited functionality compared to the previous solution. It lacks options such as “today/weekend/next week” and the like. Also what missing is the date range selection, the calendar can only select a specific day, nothing more.

It can be said that the Do312 service offers Chicago residents a convenient way to discover and plan their leisure activities. Even though the user interface and some features could be more intuitive and even better optimized, the overall structure and features such as an updated event list and user promotion of events make Do312 a useful tool for the community. To improve the overall user experience it would be worth paying attention to details such as search and calendar functions to ensure ease and efficiency in their use.

III. Choosing architecture and technologies

In the development of our project, we had to choose an optimal architectural pattern and appropriate technologies for the back-end and front-end of our application. After careful consideration, we decided on the following technology solutions that fit our needs the best. As the author states in [6], clean and well-structured software architecture is the basis for efficient and sustainable development.

A. Choosing an Architectural Pattern

We chose MVC (Model-View-Controller) because of its clear separation of responsibilities between components, which improves code maintainability and readability. As the author points out in [7], MVC divides an application into three main components: the Model (data), the View (user interface), and the Controller (logic control). This pattern allows developers to separate the application logic from the user interface, making it easier to test and maintain. This pattern supports efficient component testing and offers high flexibility and extensibility, which is ideal for our application. Patterns such as MVP (Model-View-Presenter) and MVVM (Model-View-View-Model) were discarded because they introduce more complex state management and are less intuitive for our particular project structure and needs. Also, as the author in [8] claims that MVP can increase the code complexity.

B. Choosing Technologies for the back-end

We selected Java with Spring framework, which proved to be the most suitable due to its robustness, and ability to efficiently handle extending of the web application. This framework provides the necessary flexibility and stability for the long-term development and maintenance of the system. JavaScript was also concidered. Author in [9] emphasizes Javascript as the language that features broad support for front-end development and, thanks to Node.js, has become relevant for the back-end as well. Although this language also provides high performance and efficiency for asynchronous operations, we did not select it for the back-end, because it would be unnecessarily complex and hard to learn. Python was considered as well, but its level of scalability and robustness would not meet our requirements for the system.

C. Choosing Technologies for the front-end

For the front-end, we chose React.js as according to [10], React introduced the concept of a virtual DOM, which allows for efficient updating and rendering components. React also promotes modular and sustainable development. It guarantees high-performance thanks to the virtual DOM and offers a seamless user experience. We didn't choose alternatives like Angular and Vue, because Angular is suitable for large, complex applications, and Vue, is ideal for projects with an emphasis on simplicity and scalability. React was preferred for its flexibility and ease of integration with other technologies, which is key to our needs. This choice of technologies and architectural patterns allowed us to develop a robust and scalable application that met all the technical and functional requirements of our project.

IV. Design of the system

In this part of the paper, we will deal with the design of a web application that provides users in Košice with a centralized, efficient, and user-friendly platform for discovering local events. The key goal is to create a system that will allow users not only to easily find events that interest them but also to do this efficiently.

A. Database

For our project of collecting information about events in Košice, we decided to use PostgreSQL. PSQL was chosen because of its performance, support for complex SQL queries, and decent scalability. PostgreSQL is also affordable and secure.

Database design involves several key entities:

- User: Represents the users of the system with attributes such as ID, email, and phone number. Each user is assigned a role that defines their privileges in the system.
- Event: Includes event details such as ID, description, location, date, etc. Events can be categorized based on the type of leisure, for example, parties, education, food, etc.
- Category: allows categorization of events, making it easy to filter and search them.

B. User interface

The conceptual model in Figure 1 illustrates a simple deployment process of our application.

- Common scenario
  - The user searches for an event or selects it from the menu on the main page
  - The user goes to the detailed information of the event
  - The user learns more about the event

The main purpose of our application is to navigate the user to events, which were mentioned in the common scenario. The system also possesses a report that contains information for an organizer about the number of interested people.

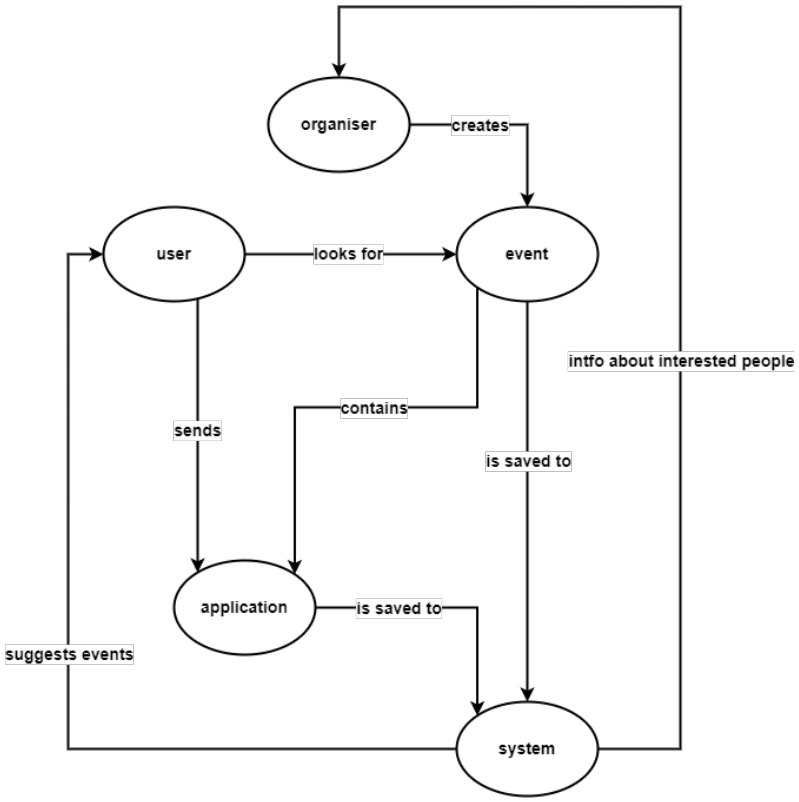


Fig. 1 Conceptual model

V. Implementation

While developing the web application, we used Java with the Spring Boot framework and JavaScript with the React.js framework. This combination is usually suitable for large and large-scale projects, but we also chose it because of the possibilities of extending it in the future and improving our skills in these technologies. We chose a PostgreSQL database to store our data. The whole application is adapted to different devices thanks to the use of Bootstrap, which ensures that it displays correctly from mobile phones to large monitors.

A. Event filtering

We have enhanced our platform by expanding the event filtering options, showed in the Figure 2, which include searching by event type, minimum and maximum price, as well as event date. These filters are now integrated into individual sections such as “Recent Events”, making it easier for users to find relevant events without the need for manual filtering. The integration of the category filter into the extended filter has greatly simplified the application code, centralizing filtering operations and improving code clarity and maintainability. We also introduced a simpler way to filter out free events, by setting “minPrice” and “maxPrice” to 0, which makes the developers’ job easier and improves the user experience. In addition, using the Axios library to manage HTTP requests improves the efficiency and reliability of data processing in our web application.

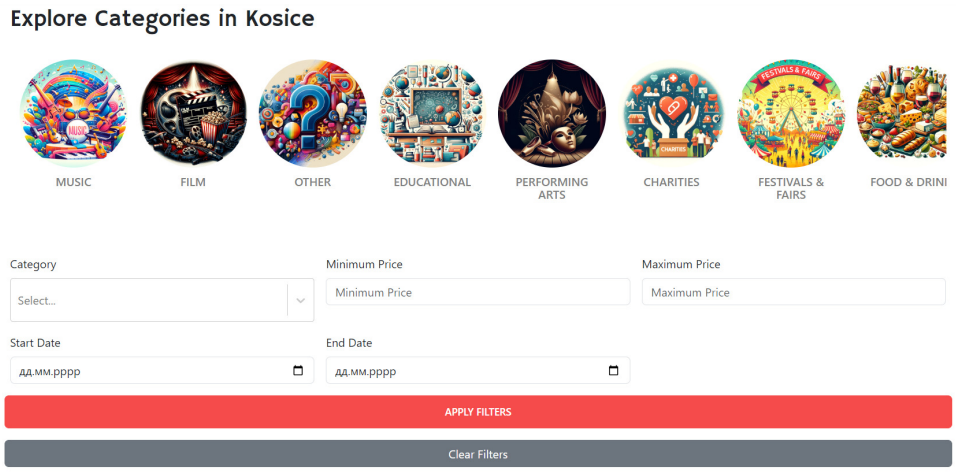


Fig. 2 Event filtering

B. Detailed event view

When we got to the implementation of the detailed event view, we emphasized providing important information such as event name, time, date, entry fee, and exact location. The information is displayed prominently in the page for the user to quickly notice. To improve the user experience, we decided to keep the original formatting of the text event descriptions, which includes indenting, paragraphing, and highlighting the text, using the Linkify library to display the links correctly.

The site also includes an interactive map using the Leaflet library that shows the exact location of the event, making it easier to plan a visit. To complement the maps, we have added a featured events slider to the site, which acts as a marketing tool to promote less popular events and increase interest in featured events. The overall implementation of the detailed event view was aimed at increasing user convenience and satisfaction with a clever structure of information and intuitive design.

C. Security

To secure our event information collection system, we have used two main security mechanisms – OAuth2 and JWT authentication. OAuth2 allows users to log in through their existing accounts (e.g., Google, Github), thus simplifying the process of registration and login. On the other hand, JWT provides safe authentication tokens that carry all necessary information used for verifying the identity of a user that is valid within a certain time frame. With these two technologies, we will be able to protect data while at the same time authenticating and authorizing users effectively.

The system has two different user roles including “USER” who has the least amount of privileges, “ORGANIZER” responsible for adding and managing all events as well as “ADMIN” with full access rights within all users over the system among other things. The roles are used to make sure that permissions are managed effectively and that every person logging into the account is authorized accordingly.

VI. Conclusion

The work aimed to create a platform that would centralize information about different cultural, social, and educational events and make event planning easier. The system was designed with an emphasis on flexibility, user-friendliness, and extensive functionality that would include not only an overview of available events but also the addition of new ones, as well as tools to manage them. The developed system includes functionalities that make it much easier to search for events according to different criteria. Authorized users can add, edit, and delete events, allowing dynamic content management. Furthermore, the system provides tools for quick access to free events or those which are the most up-to-date. Advanced search allows the filtering of events by categories, dates, prices, and other parameters, increasing efficiency and search accuracy. To increase the usefulness for users, the system has also been integrated with a map function that shows the exact locations of events, which is particularly valuable for participation planning. All of this makes the event search process easier for end users.

Development of the system was carried out using current technologies and approaches, which ensured its stability, security, and scalability. As a result, the system provides a reliable solution for organisations and offers a modern interface for users.

The demonstration of security practices using OAuth2 and JWT authentication mechanisms provides a robust foundation for educating students on modern security practices in software development.

The system has been implemented to be user-friendly, which was confirmed by the positive feedback from the target group, which was mainly students and young adults. User-side testing was carried out without any prior explanation of the functionality, which allowed us to determine the extent to which the interface is intuitive and easy to understand. The application was tested not only by members of the focus group but also by students of technical disciplines, especially those specialising in software engineering. To get feedback, we provided users with a questionnaire containing a total of 12 questions. Of these, 10 questions focused on evaluating the usability of the system using the System Usability Scale (SUS), and the remaining 2 questions focused on evaluating the system's functionality and interface friendliness. In total, we obtained feedback from 47 users. The results of the questionnaire showed that we achieved a SUS score of 81.78, which is an excellent result, indicating a high level of satisfaction with the usability of the system.

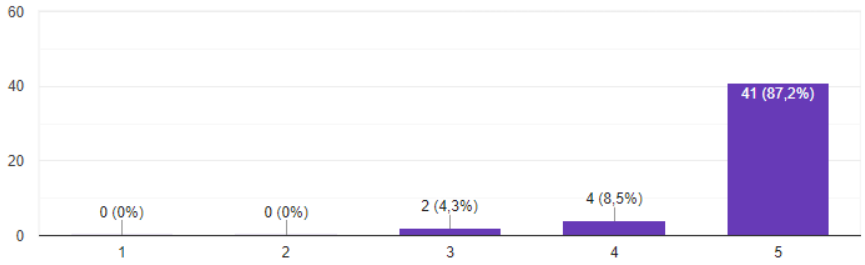


Fig. 3 Results of system functionality evaluation

Figure 3 shows that the majority of users were satisfied with the functionality of the system. Users were allowed to rate their satisfaction on a scale of 1 to 5, where 1 meant "Very dissatisfied" and 5 meant "Very satisfied".

Figure 4 shows the results of the evaluation of the design of the application according to the same principle. We asked users to rate how much they considered the design of the app as friendly and intuitive. The results showed that users rated the design of the system as convenient and intuitive, confirming the success of our design and implementation decisions. These results provided valuable information and confirmed that the system is ready for further expansion and deployment to real operation.

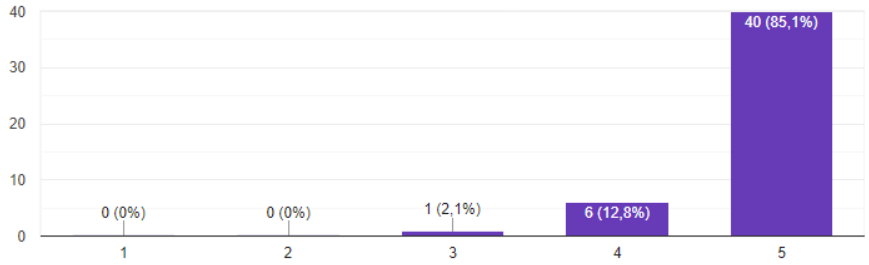


Fig. 4 Results of system functionality evaluation

In summary, the developed system can be considered beneficial to the residents and visitors of Košice. Its implementation has shown that it is capable of effectively solving the problems related to the collection and accessibility of event information, leading to better organisation and user satisfaction. The project has therefore fulfilled its objectives and confirmed its practical usefulness.

ACKNOWLEDGMENT

This work was supported by Cultural and Educational Grant Agency (KEGA) of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the project No. 004ŽU-4/2024.

REFERENCES

[1] H. L. Jakob Nielsen, *Prioritizing Web Usability*, 1st ed. New Riders Press, 2006.
[2] S. Krug, *Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability*, third edition ed., ser. Voices That Matter. New Riders, 2014.
[3] L. R. Peter Morville, *Information Architecture for the World Wide Web: Designing Large-Scale Web Sites*, third edition ed. O'Reilly Media, 2006.
[4] D. A. Norman, *The Design of Everyday Things*. Basic Books, 2002.
[5] J. B. William Lidwell, Kritina Holden, *Universal principles of design: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design*, second edition, revised and updated ed. Rockport Publishers, 2010.
[6] R. C. Martin, *Clean Architecture: A Craftsman's Guide to Software Structure and Design*, 1st ed., ser. Robert C. Martin Series. Prentice Hall, 2017.
[7] E. Gamma, R. Helm, R. Johnson, and J. M. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*, 1st ed. Addison-Wesley Professional, 1994.
[8] A. Syromiatnikov and D. Weyns, "A journey through the land of model-view-design patterns," 04 2014, pp. 21–30.
[9] M. Haverbeke, *Eloquent JavaScript*, 1st ed. No Starch Press, 2011.
[10] A. Banks and E. Porcello, *Learning React Modern Patterns for Developing React Apps*, 2nd ed. O'Reilly, 2020.

Kompaktný senzorový uzol s TCP/IP konektivitou a protokolom TLS 1.3

[1]Martin JANITOR, [2]Miloš DRUTAROVSKÝ

[1]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]martin.janitor@student.tuke.sk, [2]milos.drutarovsky@tuke.sk

Abstrakt – **Článok opisuje možnosti implementácie protokolu TLS 1.3 do senzorového uzla na báze vývojovej dosky Wiznet W5100S-EVB-Pico. Zabezpečenie spojenia na báze klient-server v IoT sieti je realizované s využitím protokolu TLS 1.3. V článku sú opísané kroky na integráciu protokolu TLS 1.3 pre použitú cieľovú hardvérovú platformu na báze procesora RP2040 s dvoma jadrami ARM Cortex M0+. Pozornosť je venovaná aj opisu dostupných generátorov náhodných čísel, ktoré cieľová platforma ponúka a ich testovaniu pomocou štatistických testov FIPS 140-2. V experimentálnej časti sú uvedené experimentálne zistené doby trvania fázy nadviazania spojenia v protokole TLS 1.3 pre neoptimalizovanú a optimalizovanú verziu firmvéru.**

Kľúčové slová – **ARM Cortex M0+, IoT, programovací jazyk C/C++, RNG, TLS 1.3, vstavaný systém, W5100S-EVB-Pico**

I. Úvod

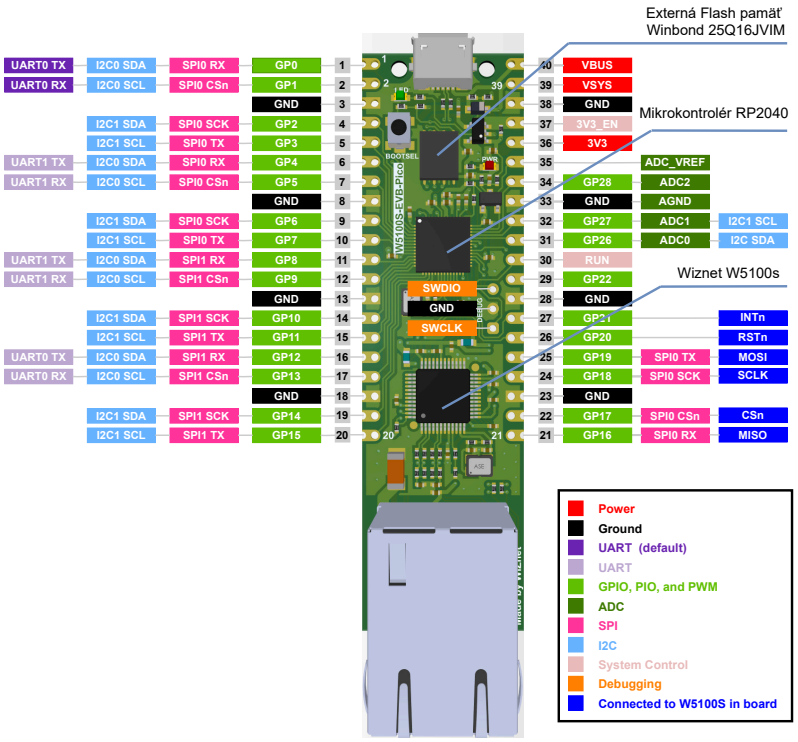
Vznik Internetu vecí (IoT – Internet of Things) priniesol spôsob, akým integrujeme inteligentné zariadenia do nášho každodenného života, od domácej automatizácie až po priemyselné aplikácie. Zariadenia môžu medzi sebou komunikovať, zdieľať dáta a byť vzdialene ovládané užívateľom pre získanie a spracovávanie dát zo senzorových uzlov. Rýchla integrácia zariadení do IoT prináša výzvu ako dáta zabezpečiť pred nežiaducimi užívateľmi, ktorí vstúpia nepovolane do siete. Zabezpečenie komunikácie medzi zariadeniami IoT a servermi sa stáva kľúčové pre ochranu citlivých údajov a zachovanie integrity systémov.

Aplikovanie senzorového uzla do IoT siete vyžaduje zvoliť vhodné metódy na zabezpečenie komunikácie. Pre senzorové uzly, ktoré sú realizované na báze vstavaných systémov sa môžu vyskytnúť potencionálne problémy s využitím potrebných bezpečnostných metód. Vstavané systémy zvyčajne poskytujú obmedzené zdroje pre kryptografické účely, napríklad nedisponujú kvalitným generátorom náhodných čísel alebo majú nižší výpočtový výkon oproti desktopovým počítačom (PC – Personal Computer). Vývojová doska Wiznet W5100S-Evb-Pico je kompaktná doska, ktorá integruje základné funkcie pre efektívne nasadenie IoT uzla do TCP/IP (Transmission Control Protocol/Internet Protocol) siete, vrátane sieťového pripojenia, čím sa stáva vhodným kandidátom pre rôzne IoT aplikácie. Vývojová doska disponuje dodatočným rozhraním RJ45, ktorý poskytuje sieťové káblové spojenie, čo v praxi môže mať viacero výhod oproti bezdrôtovému prenosu. Primárnymi výhodami káblového prenosu oproti bezdrôtovému sú rýchlejší a spoľahlivejší prenos dát, väčšia bezpečnosť a nižšia náchylnosť na okolitý šum a rušenie.

Článok obsahuje stručný opis hardvérovej aj softvérovej časti použitej vývojovej dosky. V softvérovej časti sú opísané vývojové balíky určené pre vývojovú dosku a možnosti nastavenia sieťových parametrov, ktoré vývojová doska ponúka. Súčasťou softvérovej časti je aj opis procesu použitia dostupných kryptografických knižníc MIRACL a TIIGER TLS 1.3 na realizáciu zabezpečeného spojenia s využitím protokolu na zabezpečenie v transportnej vrstve (TLS – Transport Layer Security) verzie 1.3. Hardvérová časť obsahuje opis vývojovej dosky a jej periférií a rozhraniami, ktorými disponuje. V článku je upriamená pozornosť aj na dostupné generátory náhodných čísel, ktorými doska disponuje. Jednotlivé generátory sú následne experimentálne testované pomocou základných štatistických testov (FIPS – Federal Information Processing Standards) 140-2. Experimentálnym výsledkom článku je overenie TLS 1.3 spojenia medzi senzorovým uzlom a PC platformou.

II. Hardvérové komponenty senzorového uzla

Použitá vývojová doska W5100S-EVB-Pico [1] je založená na doske Raspberry Pi Pico a dodatočného ethernetového čipu W5100S [2] od firmy Wiznet, ktorý ponúka sieťovú konektivitu. Základom dosky je mikrokontrolér RP2040 [3], ktorý má k dispozícii rôzne typy rozhraní. K procesoru je pripojená externá Flash pamäť pomocou 4-vývodového sériového periférneho rozhrania (QSPI – Quad Serial Peripheral Interface). Externá Flash pamäť Winbond 25Q16JVIM [4] disponuje veľkosťou 2 MB a slúži na uloženie programu. Sieťová konektivita je realizovaná pomocou káblového rozhrania s využitím dostupného rozhrania RJ45 a čipu Wiznet W5100S, ktorý je pripojený k RP2040 pomocou sériového periférneho rozhrania (SPI – Serial Peripheral Interface). RP2040 obsahuje aj zavádzací program (bootloader), ktorý umožňuje RP2040 emulovať univerzálnu sériovú zbernicu (USB – Universal Serial Bus) ako Flash disk. Nahrávanie kódu do RP2040 je možné realizovať pomocou USB Flash disk rozhrania. Na Obr. 1 je znázornená vývojová doska, ktorá je využitá ako senzorový uzol pre IoT siete.



Obr. 1 Vývojová doska Wiznet W5100S-EVB-Pico s dodatočným rozhraním RJ45 pre sieťovú konektivitu.

A. *Mikrokontrolér RP2040*

RP2040 je 32-bitový mikrokontrolér postavený na architektúre ARMv6-M s využitím redukovanej inštrukčnej sady (RISC – Reduced Instruction Set Computer). Disponuje dvomi jadrami ARM Cortex M0+, ktoré môžu dosahovať taktovaciu frekvenciu až 133 MHz s využitím externého 12 MHz kryštálu a slučky fázového závesu (PLL – Phase-Locked Loop). RP2040 využíva architektúru malého endiánu. Poskytuje rôzne periférie ako napríklad UART (Universal Asynchronous Receiver-Transmitter), SPI, I2C (Inter-Integrated Circuit) zbernice, analógovo digitálny prevodník (ADC – Analog-to-Digital Converter), PWM (Pulse Width Modulation) a USB. RP2040 nedisponuje internou Flash pamäťou na uloženie programu, ponúka možnosť pripojenia externej Flash pamäte pomocou QSPI rozhrania. RP2040 obsahuje internú statickú pamäť s ľubovolným prístupom (RAM -Random Access Memory) s veľkosťou 264 kB a pamäť určenú na čítanie (ROM – Read Only Memory) s veľkosťou 16 kB. RP2040 bol navrhnutý s cieľom minimalizovať spotrebu, čo ho robí vhodným kandidátom pre vstavané systémy, ktoré vyžadujú energetickú efektivitu [3].

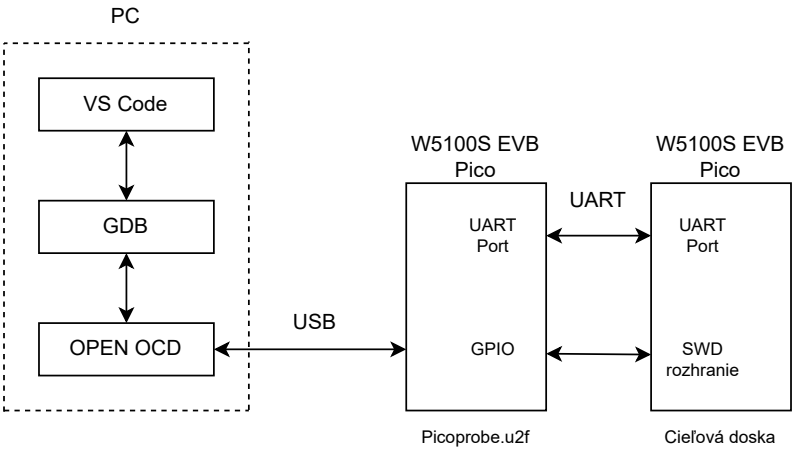
B. *Ethernetový čip Wiznet W5100S*

Čip Wiznet W5100S je určený pre integráciu vstavaných systémov do siete internet. Čip ponúka hardvérovú podporu pre rôzne internetové protokoly ako napríklad TCP (Transmission Control Protocol), UDP (User Datagram Protocol), ICMP (Internet Control Message

Protocol), IPv4 (Internet Protocol version 4), ARP (Address Resolution Protocol), IGMP (Internet Group Management Protocol) a PPPoE (Point-to-Point Protocol over Ethernet). Spracovanie sieťových protokolov je realizované priamo v čipe, dôsledkom čoho je menšia vyťaženosť hlavného procesora. Čip má podporu pripojenia k procesoru pomocou sériového SPI rozhrania alebo paralelnej zbernice. Čip disponuje hardvérovou podporou pre rozhranie RJ45 [2].

C. Ladiaci systém

Ladiaci systém na vývojovej doske poskytuje sériové ladiace rozhranie (SWD – Serial Wire Debug) pre komunikáciu s externým ladiacim nástrojom. Ako ladiaci nástroj je možné využiť externý ladiaci nastroj Picoprobe alebo ďalšiu dodatočnú vývojovú dosku. Ladenie s využitím dodatočnej dosky vyžaduje firmvér picoprobe.u2f, ktorý emuluje ladiaci nástroj. Dodatočná vývojová doska komunikuje s cieľovou doskou cez vstupno/výstupné vývody (GPIO – General Purpose Input/Output), ktoré sú pripojené k rozhraniu SWD. Ladiaci systém podporuje 4 body zastavenia (breakpoint) a 2 sledovacie body (watchpoint). Na Obr. 2 je znázornená bloková schéma ako je možné realizovať ladiaci systém s dedikovanou doskou a aké nástroje je potrebné mať zahrnuté vo vývojovom prostredí VS (Visual Studio) Code. Na stránke [5] je uvedený tutoriál ako realizovať ladenie s využitím dodatočnej vývojovej dosky.



Obr. 2 Štruktúra ladiaceho systému s využitím dodatočnej vývojovej dosky a nástrojov pre ladenie.

III. Softvérová podpora pre vývojovú dosku

Vývojový balík (SDK – Software Development Kit) poskytuje aplikačné rozhranie (API – Application Programming Interface) pre prácu s RP2040 a čipom Wiznet W5100S. Pre vývojovú dosku Wiznet W5100S-EVB-Pico je dostupných viacero verzií SDK, ktoré si užívateľ môže zvoliť pre vytvorenie vlastných projektov. Dostupné sú verzie, ktoré sú vytvorené v programovacích jazykoch C a C++ s využitím asemblerovských (ASM) inštrukcií alebo v programovacom jazyku MicroPython. Zaujímavou verziou SDK je aj verzia, ktorá integruje operačný systém FreeRTOS. V experimentoch bolo využité Wiznet SDK verzie 2.3.0.

A. Vývojový balík – SDK

SDK poskytuje rozhranie pre vývoj aplikácií na doske W5100S-EVB-Pico. Wiznet SDK je založené na SDK pre Raspberry Pi Pico a knižnici na obsluhu ethernetového čipu Wiznet W5100S pre sieťovú konektivitu. Vývojový balík je postavený na hierarchickej štruktúre CMake súborov, ktoré sú určené pre zostavovavenie projektov. SDK poskytuje rôzne API rozhrania pre manipuláciu s dostupnými perifériami. Rozhranie SDK umožňuje vygenerovanie viacerých výstupných súborov v rôznych formátoch ako napríklad u2f, hex, bin a elf.

API pre sieťové rozhranie

SDK disponuje API rozhraním, ktoré podporuje rôzne funkcionality sieťových protokolov. Pre nastavenie sieťových parametrov sú dostupné dve metódy, ktoré poskytujú užívateľovi základnú sieťovú konfiguráciu. Jednotlivé metódy sú uvedené v nasledujúcom zozname.

Možnosti nastavenia sieťových parametrov:

1) Statické nastavenie – API rozhranie poskytuje užívateľovi štruktúru, kde je možné nastaviť sieťové parametre. Nevýhodou tejto metódy je potreba získať sieťové parametre a manuálne nastaviť statickú IP adresu na smerovači.
2) Dynamické nastavenie – API rozhranie ponúka funkcie pre komunikáciu so serverom na dynamickú konfiguráciu sieťových parametrov pomocou protokolu DHCP (Dynamic Host Configuration Protocol). DHCP server poskytuje užívateľovi automatické pridelenie sieťových parametrov, čo eliminuje potrebu manuálneho zadávania.

Pre prevod doménového názvu na IP adresu sú dostupné aj funkcie pre prácu so systémom názvu domén DNS (Domain Name System). Sieťové rozhranie implementuje BSD (Berkley Sockets) funkcie pre vytvorenie sieťovej komunikácie medzi klientom a serverom.

B. Kryptografické knižnice MIRACL a TIIGER TLS 1.3

Pre aplikovanie TLS 1.3 protokolu do vývojovej dosky je potrebné zvoliť vhodnú kryptografickú knižnicu, ktorá ponúka implementáciu protokolu TLS 1.3 a možnosti integrácie do vstavaného systému. Medzi najpoužívanejšie patria napríklad štandardné kryptografické knižnice WolfSSL a MbedTLS, ktoré poskytujú implementáciu TLS protokolu aj pre vstavané systémy. Počas realizácie projektu sme však zvolili kryptografické knižnice MIRACL [6] a TIIGER TLS 1.3 [7] pre vytvorenie aplikácie, ktorá poskytuje TLS 1.3 spojenie v roli klienta. Hlavným dôvodom bola ich vysoká miera škálovateľnosti a prehľadnosť implementácie. Typický firmvér senzorového uzla pozostáva z vlastného aplikačného kódu, ktorý je napísaný v programovacom jazyku C alebo C++ a konfigurácie sieťových nastavení, ktorý je linkovaný s kryptografickými knižnicami.

Kryptografická knižnica MIRACL poskytuje implementácie rôznych kryptografických algoritmov pre realizovanie šifrovania, autentizácie, hašovacích funkcií a rozhrania pre generovanie náhodných čísel. Knižnica je prenositeľná na rôzne platformy, keďže neobsahuje žiadne ASM inštrukcie, ktoré sú závislé na architektúre. Knižnica je dostupná vo viacerých programovacích jazykoch ako napríklad C, C++, Go, Rust, Python, Java, Javascript a Swift.

Kryptografická knižnica TIIGER TLS 1.3 obsahuje implementáciu protokolu TLS 1.3, ktorý slúži na realizáciu zabezpečenej komunikácie medzi klientom a serverom cez nezabezpečenú sieť. Silnou stránkou knižnice je jej modulárnosť vďaka ktorej poskytuje možnosť užívateľovi modifikovať algoritmy, ktoré si dokáže prispôsobiť podľa vlastných preferencií. Modifikáciu kryptografických algoritmov je možné realizovať v SAL (Security Abstraction Layer) vrstve, kde je možné pridať vlastné implementácie algoritmov. SAL vrstva obsahuje funkcie pre inicializáciu kryptograficky bezpečného pseudo-náhodného generátora čísel (CSPRNG – Cryptographically Secure Pseudo Random Number Generator) a kryptografické metódy, ktoré sa využívajú počas TLS 1.3 spojenia.

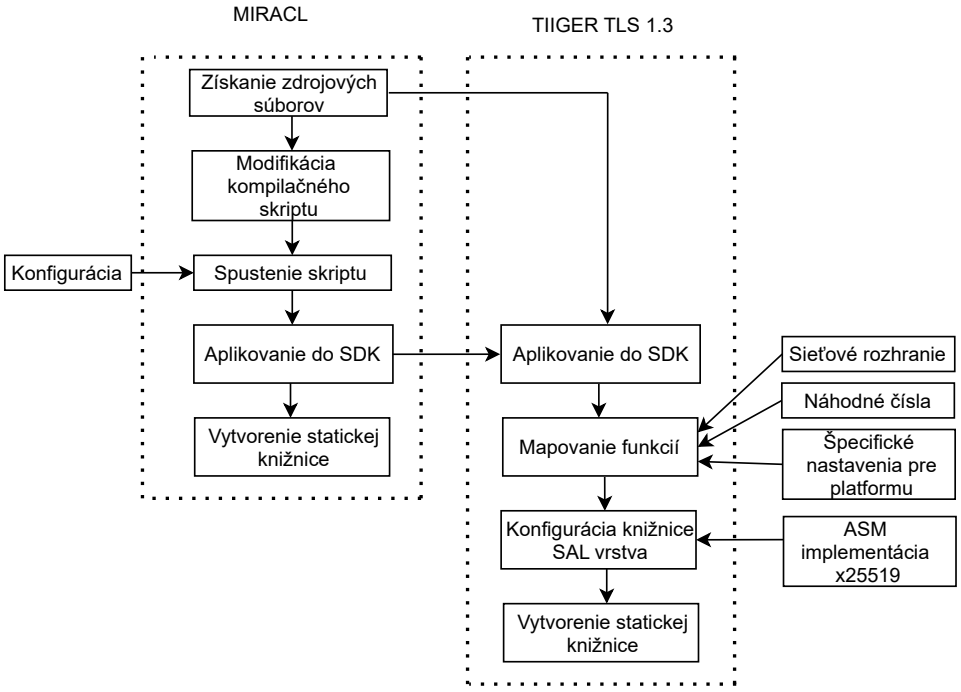
Úpravou SAL vrstvy dokážeme docieliť aj zefektívnenie výpočtov pre cieľovú platformu senzorového uzla. Zaujímavou vlastnosťou je možnosť využitia ASM implementácie eliptickej krivky x25519 v konštantnom čase pre mikrokontrolér s jadrom ARM Cortex-M0/M0+ [8]. Implementácia je realizovaná iba s využitím ASM inštrukcií, čo môže výrazne zrýchliť výpočty na realizovanie TLS 1.3 spojenia v cieľovom senzorovom uzle.

C. Integrácia kryptografických knižníc do Wiznet SDK

Aplikovanie kryptografických knižníc do SDK pre vstavané systémy vyžaduje nastavenie a konfiguráciu knižnice podla špecifických požiadavok, ktoré závisia od cieľového systému a preferencií užívateľa. Pridanie a nastavenie kryptografických knižníc MIRACL a TIIGER TLS 1.3 vyžaduje viacero potrebných krokov pre nasadenie. Na Obr. 3 je zobrazená schéma procesu integrácie kryptografických knižníc MIRACL a TIIGER TLS 1.3 do Wiznet SDK.

V prvom kroku je potrebné získať zdrojové súbory kryptografických knižníc, ktoré sú dostupné na stránke https://github.com/, pričom je potrebné sledovať aj verzie jednotlivých kryptografických knižníc.

Kompilácia knižnice MIRACL s využitím programovacieho jazyka C++ pozostáva z predpripravených Python skriptov, ktoré realizujú kompiláciu pre operačné systémy založené na linuxe. Po spustení skriptu si užívateľ zvolí konfiguráciu kryptografických algoritmov, ktoré plánuje využívať v projekte a na základe toho sa vygeneruje statická knižnica. Zdrojové kódy knižnice poskytujú vysokú modulárnosť. Príkladom modulárneho zdrojového súboru je *rsa.cpp*, na základe ktorého je možné vytvoriť modifikované súbory pre šifrovací algoritmus RSA s veľkosťou 2048 alebo 4096 bitov. Po zadaní konfigurácie užívateľom do Python skriptu sa modifikujú modulárne zdrojové súbory a vytvorí sa statická knižnica.



Obr. 3 Proces integrácie knižníc MIRACL a TIIGER TLS 1.3 do vývojového prostredia SDK.

Pre aplikáciu MIRACL knižnice do Wiznet SDK je potrebne modifikovať Python skript aby zaznamenal do textového súboru, ktoré súbory boli modifikované. Túto informáciu užívateľ využije pri manuálnej integrácií do SDK. Po úspešnom spracovaní Python skriptu sa vytvorí textový súbor v ktorom má užívateľ názvy súborov, podľa zadanej konfigurácie. Jednotlivé modifikované súbory sa premiestnia do Wiznet SDK a pomocou nástroja CMake sa vytvorí statická knižnica MIRACL.

Po kompilácii knižnice MIRACL je potrebné vytvoriť statickú knižnicu TIIGER TLS 1.3. Kryptografická knižnica TIIGER TLS 1.3 vyžaduje algoritmy obsiahnuté v knižnici MIRACL. Pre využitie možností RP2040 a čipu Wiznet v senzorovom uzle je potrebné nastaviť mapovanie do cieľovej platformy.

Mapovanie do cieľovej platformy:

- Sieťové nastavenia – Knižnica poskytuje súbor *tls_sockets.h* pre nastavenia rozhrania sieťových parametrov. Základná implementácia je určená pre platformy s operačným systémom linux. Vhodnou modifikáciou súboru s využitím Wiznet API funkcií dokážeme aplikovať sieťové nastavenia, ktoré poskytuje Wiznet W5100S čip pre vývojovú dosku na realizovanie požadovaného sieťového pripojenia.
- Náhodné čísla – Knižnica poskytuje súbor *tls_sal_mhp.cpp*, ktorý implementuje SAL vrstvu. V SAL vrstve je možné nastaviť generátor náhodných čísel podľa vlastných preferencií. Primárne je vhodné využiť dostupné zdroje entropie, ktoré obsahuje RP2040. Jednotlivé generátory sú opísané v nasledujúcej sekcii. Náhodné dáta z dostupných generátorov sú využité na inicializáciu interného CSPRNG, ktorý je založený na hašovacej funkcii (SHA – Secure Hash Algorithm) SHA-256. Interný CSPRNG vyžaduje minimálne 128 bajtov náhodných dát.
- Špecifické nastavenia pre platformu – Užívateľ dokáže pozmeniť konfiguráciu SAL vrstvy pre využitie vlastných implementácii kryptografických algoritmov, prípadne ich môže redukovať aby získal menšiu veľkosť výsledného firmvéru. Typickým príkladom je využitie ASM implementácie x25519, ktorá umožňuje zrýchliť nadviazanie spojenia v protokole TLS 1.3.

IV. Generovanie náhodných čísel v RP2040

RP2040 umožňuje využiť viacero dostupných zdrojov entropie, ktoré je možné využiť na generovanie náhodných čísel. Jednotlivé zdroje entropie samostatne nezabezpečujú generovanie kryptografických bezpečných náhodných čísel. Pre zaistenie požadovanej kryptografickej bezpečnosti je vhodné jednotlivé zdroje entropie zlúčiť pomocou operácie exkluzívny súčet (XOR – Exclusive OR) a následne výstup využiť na inicializáciu CSPRNG.

Dostupné zdroje entropie v RP2040:

1) ADC prevodník – RP2040 disponuje rozhraním ADC, z ktorého je možné získať náhodné dáta. Náhodné dáta sa načítavajú z ADC vývodu na ktorý nie je pripojený žiaden vstup (floating-point pin). Zosnímaný šum môžeme využiť ako zdroj entropie.
2) Kruhový oscilátor (ROSC – Ring Oscillator) – Typ oscilačného obvodu bežne používaného v digitálnych systémoch na generovanie hodinového signálu pre procesor. RP2040 obsahuje ROSC register z ktorého je možné vyčítať jeden náhodný bit. Ak sa ROSC využíva na generovanie hodinového signálu, nie je možné načítavať náhodný bit z registra.
3) Časová vzorka – Funkcia *time_us_*64() vráti 64 bitov dlhú vzorku hardvérového časovača od začiatku vykonávania programu.
4) Hašovaná hodnota RAM pamäte – Vopred vymedzená oblasť RAM pamäte je hašovaná. Počiatočný obsah RAM je po zapnutí nedefinovaný, čo môže poskytnúť zdroj entropie.
5) Počítadlo prístupov k zbernici – RP2040 obsahuje súbor hardvérových počítadiel (Bus Performance Counters), ktoré sú primárne určené na monitorovanie aktivity a výkonu zberníc využívaných v RP2040. Procesor obsahuje hlavnú AHB-Lite sieť zberníc, ktorá presmerováva adresy a dáta na jednotlivé periférie. Počítadlá výkonu zberníc automaticky počítajú prístupy k AHB-Lite zbernici.
6) Jedinečné identifikačné číslo (ID – Identification Number) vývojovej dosky – RP2040 obsahuje dodatočnú externú pamäť Flash, ktorá obsahuje program a unikátne ID čipu s dĺžkou 64 bitov vložený fixne do pamäte počas výroby. Identifikátor je zvyčajne uložený v oblasti nevolatilnej pamäte (NOR Flash) a je jedinečný pre každú jednotlivú vyrobenú dosku.

V. Experimentálne výsledky

V rámci experimentálnych výsledkov bolo overené testovanie dostupných generátorov náhodných dát, ktoré má RP2040 k dispozícii a TLS 1.3 implementácia klienta s využitím kryptografie na báze eliptických kriviek (ECC – Elliptic Curve Cryptography).

A. Testovanie kryptografickej knižnice MIRACL a TIIGER TLS 1.3

V procese testovania TLS 1.3 spojenia s využitím kryptografických knižníc MIRACL a TIIGER TLS 1.3 sme realizovali porovnanie časových nárokov na dĺžku nadviazania spojenia (tzv. handshake procesu) v realizácii TLS s využitím rozdielnych implementácií eliptickej krivky x25519. Využité boli základná implementácia poskytnutá MIRACL knižnicou a ASM implementácia eliptickej krivky x25519 určená pre mikrokontrolér s jadrom ARM Cortex-M0/M0+ [8]. Testovanie bolo realizované s využitím ECC certifikátov na báze eliptickej krivky NIST-256. Server na platforme PC bol spustený s využitím aplikácie OpenSSL a klient bol spustený na senzorovom uzle. Server využíval certifikát Certifikačnej autority (CA – Certification Authority) a certifikát servera. Certifikát CA na strane klienta bol zadaný v súbore *tls_cacerts.cpp*. Klient overoval certifikát servera s využitím certifikátu CA, ktorý mal k dispozícii. Z uvedených výsledkov v Tabuľke 1 je vidno, že metóda s využitím ASM implementácie dosahuje kratší čas na realizovanie TLS 1.3 handshake procesu.

Tabuľka 1
Porovnanie TLS 1.3 implementácie s využitím ASM inštrukcií pre realizáciu handshake procesu

Časové nároky na realizáciu handshake procesu		
	TIIGER TLS 1.3	TIIGER TLS 1.3 + ASM kód [8]
TLS 1.3	1583,7 ms	1405,45 ms

B. Testovanie zdrojov entropie na RP2040

Overenie kvality zdrojov entropie bolo realizované pomocou základných štatistických testov FIPS 140-2. Dôležité je poznamenať účel FIPS 140-2 testov, ktoré sa využívajú primárne na všeobecné testovanie generátorov náhodných čísel (RNG – Random Number Generator) a poskytujú len základné informácie o kvalite náhodných čísel. Pre náročnejšie testovanie RNG pre kryptografické účely je možné využiť napríklad balík testov zo štandardu AIS-31, ktorý poskytuje širší rozsah testov oproti FIPS 140-2. Z výsledkov uvedených v Tabuľke 2 je možné skonštatovať, že načítanie náhodných dát z ADC vývodu neposkytuje dostatočnú entropiu pre využitie na generovanie kryptograficky bezpečných náhodných čísel. Generátor Počítadlo prístupu k zbernici taktiež neposkytuje dostatočnú entropiu pre využitie ako kryptografický bezpečný generátor. Generátory ROSC a Časová vzorka prešli všetkými dostupnými testami FIPS 140-2. Pre vykonanie testov FIPS 140-2 bolo potrebné získať vzorku náhodných dát s dĺžkou 20 000 bitov.

Tabuľka 2
Analýza dostupných zdrojov entropie náhodných čísel na platforme RP2040 s využitím testov FIPS 140-2

Výsledky testov FIPS 140-2			
	Pokerový test	Frekvenčný test	Test rovnakých reťazcov
ADC	NEÚSPEŠNÝ	NEÚSPEŠNÝ	NEÚSPEŠNÝ
ROSC	ÚSPEŠNÝ	ÚSPEŠNÝ	ÚSPEŠNÝ
Časová vzorka	ÚSPEŠNÝ	ÚSPEŠNÝ	ÚSPEŠNÝ
Počítadlo prístupu k zbernici	NEÚSPEŠNÝ	NEÚSPEŠNÝ	NEÚSPEŠNÝ
ADC + Zdroje entropie	ÚSPEŠNÝ	ÚSPEŠNÝ	ÚSPEŠNÝ

VI. Záver

Článok opisuje proces integrácie kryptografických knižníc MIRACL a TIIGER TLS 1.3 do senzorového uzla na báze vývojovej dosky Wiznet W5100S-EVB-Pico. Bolo opísané softvérové aj hardvérové vybavenie senzorového uzla. Vytvorený senzorový uzol umožňuje zabezpečenú sieťovú konektivtu na externé servery s využitím protokolu TLS 1.3. Vytvorený firmvér senzorového uzla podporuje konektivitu s využitím DHCP a DNS serverov pre uľahčenie konfigurácie sieťových nastavení. S využitím štatistických testov FIPS 140-2 sme experimentálne overili vhodnosť dostupných zdrojov entropie na platforme RP2040.

Výsledkom experimentálnej časti bola plnohodnotná aplikácia napísaná v programovacom jazyku C++ s využitím kryptografických knižníc, ktorá realizovala zabezpečenú komunikáciu medzi klientom a serverom s využitím protokolu TLS 1.3 a ECC certifikátov. Získané výsledky jasne demonštrujú, že aj s pomerne obmedzenými hardvérovými zdrojmi, ktorými disponuje senzorový uzol, je možné realizovať zabezpečenú komunikáciu medzi klientom a serverom na báze protokolu TLS 1.3 s prakticky využiteľným časovaním v sieťach IoT. Detailné výsledky sú dostupné v [9].

Poďakovanie

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-22-0400.

Literatúra

[1] "W5100s-evb-pico overview," 2024. [Online]. Available: https://docs.wiznet.io/Product/iEthernet/W5100S/w5100s-evb-pico
[2] "W5100s," 2024. [Online]. Available: https://docs.wiznet.io/Product/iEthernet/W5100S/overview
[3] "Rp2040 datasheet," 2024. [Online]. Available: https://datasheets.raspberrypi.com/rp2040/rp2040-datasheet.pdf
[4] "Winbond W25Q16JV - 3V 16M-BITSERIAL FLASH MEMORY WITH DUAL/QUAD SPI." [Online]. Available: https://docs.rs-online.com/19e0/0900766b81622f8b.pdf
[5] S. Hymel, "Raspberry pi pico and rp2040 - c/c++ part 2 debugging with vs code." [Online]. Available: https://www.digikey.be/en/maker/projects/raspberry-pi-pico-and-rp2040-cc-part-2-debugging-with-vs-code/470abc7efb07432b82c95f6f67f184c0
[6] M. Scott, "The miracl core library," multi-lingual crypto library, specifically designed to support the Internet of Things. [Online]. Available: https://miracl.com/
[7] ——, "On TLS for the internet of things, in a post quantum world," 2023, https://eprint.iacr.org/2023/095. [Online]. Available: https://eprint.iacr.org/2023/095
[8] T. Pornin, "Faster Modular Inversion and Legendre Symbol, and an X25519 Speed Record," Sept. 2020. [Online]. Available: https://research.nccgroup.com/2020/09/28/faster-modular-inversion-and-legendre-symbol-and-an-x25519-speed-record/
[9] M. Janitor, "Zabezpečená komunikácia senzorového uzla s využitím TLS protokolu a certifikátov," *Diplomová práca, FEI TUKE, Košice*, May 2024.

Virtual reality environment for support of art education

[1]*Branislav SOBOTA,* [2]*Štefan KOREČKO,* [3]*Nikoleta DIMUNOVÁ*

[1,2,3] Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical university of Košice, Letná 9, 04200 Košice, Slovak republic

[1]branislav.sobota@tuke.sk

***Abstract* — The aim of this article is the mutual connection of virtual reality and art education in the context of addressing the specific needs of children with neurological disorders. We developed a virtual reality application and integrated it into the environment of an art education class within a school setting. We evaluated the application in practice. In a conclusion, virtual reality can enhance education by providing unforgettable and immersive experiences. Additionally, virtual reality has a unique ability to inspire and engage pupils.**

***Keywords* — virtual reality, eXtended reality, disabled people, art education**

I. INTRODUCTION

Virtual reality (VR) became a vital aspect of our lives today. In addition to VR, related technologies such as mixed reality (MR) or augmented reality (AR) to associated in the XR (eXtented reality) application platform are being raised [1]. We can encounter it, for example, in healthcare, industry, or even entertainment. It can be used to create almost anything with cost-efficiency and without limiting to real world. One part of research with XR that is gaining popularity is education. For XR purposes more and more hardware and devices are being made. The most common virtual environments at this level are virtual CAVE (Cave automatic virtual environment)-based systems and VR headset-based systems (HMD – Head Mounted Display). So XR platform can be excellent tool not only for standard pupils' education but also for education of disabled pupils. Various applications and systems were developed, and possible solutions are being sought that may help disabled people to facilitate communication, movement, and thus ensure the conditions for a fuller life [2]. One of the ways is also the development of specialized centers and auxiliary schools, including the special boarding school of Pavel Sabadoš in Prešov. This school is a partner school within the project solved in the LIRKIS laboratory (authors' home laboratory) and the examples presented in this article are from the process of solving this project. Children born with disabilities can learn to live with their handicaps from a young age and learn in all available ways. One of the feasible ways is the use of artistic procedures (e.g. drawing) in addition to the application of the aforementioned modern XR technologies.

The main focus of this article is to bring closer usage of XR and related technologies in the educational process, which are also addressed within the projects solved in the authors' home laboratory LIRKIS (Laboratory of Intelligent inteRfaces of Communication and Information Systems). The introductory part contains the description of the disabled people education. Second, main part describes developed application based on XR technologies using some art methods including the description of some detailed scenarios (tasks) for pupils education.

II. DISABLED PEOPLE AND EDUCATIONAL PROCESS

When we are talking about disabled people then we are talking not only about people with physical or sensory disabilities, but also with cognitive disorders (for example, people with learning disabilities or with a concentration disorder, e.g. ADHD). About 15% of the population suffers from some form of disability. These people have generally significant problems integrating into society. Many of them need specialized tools and other assistive technologies for their daily life [3]. Therefore, XR technology is also suitable for creating a more natural user interface with better accessibility and usability. To create such an environment, it is necessary to understand how these people can access information and whether they have specific problems and needs. However, these problems are specific to a particular type of disability, so they can be divided into the following categories: sensory disability such as visual, olfactory, or auditory disability, physical disability, limited mobility, mental disability, speech disorders, learning disorders such as dyslexia, dysgraphia, dyscalculia, or other types such as senility and different

dependencies. Disabled people need to be supported by supportive technology, especially alternatives to user interfaces (text readers, screen readers, etc.). New technologies including XR can help to gain more information, help in recovery, everyday activities and so on. Among the most supported groups are blind and the hearing-impaired people.

Many aspects of disabled people handicap, except blindness, can be eliminated using XR technologies. By creating a virtual reality environment that simulates the presence of people, they can move freely in the virtual space and look around. Virtual reality opens the door to new opportunities, such as running, swimming, climbing, or skiing for a group of people with physical disabilities. For example, people with phobias can face their fear, all in a safe simulated and controlled virtual reality environment. People with mental or learning disabilities, such as autism, can be present in stressful situations, such as a busy road or a new unknown space. This process could reduce anxiety again in a safe virtual environment [4]. But of course, virtual reality can also be used the other way around, to give healthy people an idea of what people with epilepsy, Alzheimer's disease or autism are experiencing. A very fascinating idea is the interconnection of art and virtual reality in individuals diagnosed with ADHD. XR is interactive, so users' actions can change the scene around them [5]. Another proof of development in digital technologies, is the inclusion of XR, in the curriculum of basic education subjects [6]. Art education develops skills, creativity, sense of aesthetics, as well as higher cognitive operations (analysis, synthesis, evaluation, critical thinking etc.)

One of the many fundamental theses of art education is also illustrated by the authors of the publication "Art and Design for Secondary School Children with SEN" [7], which outlines recommended teaching practices in the following sequence: drawing using lines and symbols, color theory and 3D shapes, circles and ellipses, drawing stones with pencil and paint, still life, compositional techniques, pastel portraits, glass painting, acrylic still life painting, watercolor painting, and silk painting. This sequence of assignments (according to difficulties) can serve as an indicator for creating reality. A similar "artistic" method with VR support also brings support in the therapy of dyslexia [8] and dysgraphia [9].

Drawing with a tablet provides more useful data. Its simple usage is likely more engaging for participants than the traditional approach based on pen and paper. Another common option is an iPad application called "Lets Draw". It allows free drawing on a blank canvas. The coordinates with timestamps can be exported for real-time interaction replay [10].

Romero-Ayuso et al. [11] published a systematic review aimed at evaluating the effectiveness of interventions based on virtual reality on cognitive deficits in children with Attention Deficit Hyperactivity Disorder (ADHD). Although the number of studies was not large, the results suggest, that XR-based interventions significantly improve the cognitive performance of children and adolescents with ADHD in tasks focusing on a short-term and sustained attention, hence reducing the number of missed tasks and increase the number of correct responses to target stimuli. This indicates that the final effect is more centered on attention and less focused on improving impulsivity or inhibitory control. These results are interesting, since they suggest, that interventions based on XR can improve symptoms of inattention, potentially making them very useful for children with ADHD.

Virtual reality, as an experimental platform used with children diagnosed with ADHD, was also utilized in the study by Savickaite et al. [10], where the authors intended to explore the practical effects of using drawing tools on perceptual experiences and investigate any perceptual differences brought by the three-dimensional nature of XR. Participants were tested via gaming computer Nvidia GTX 1080 I (Windows 10) with an HTC Vive Pro headset. The experiment was conducted through the TiltBrush application (Google, 2016) using SteamVR software (Steam, 2003). Results of the study revealed different individual preferences in drawing tasks, while using VR. This can lead to better understanding of individual differences in processing visual-spatial information for people diagnosed with ADHD. The study serves as an example of how classical psychological paradigms can be transferred into the virtual world.

III. System design

A. *Base software technology*

The *Unity* engine (`https://unity.com/`) was used as base software technology. *Unity* engine is an excellent choice for developing XR applications. It offers extensive support, rich documentation, and an active community - which facilitates the creation and sharing of XR projects. The Unity was also chosen on positive personal experience with it. The official Unity Asset Store website `/urlhttps://assetstore.unity.com/` was used to download of an individual assets. One of the used assets was "*Pictures frames with photos*" for realistic rendering of images. For realistic rendering of nature, was used the "*Lowpoly Environment*" plugin. Subsequently, to create the environment, such as statues, chairs, and a camera, was used the

assets: "*Discobolus Statue*", "*Environment Pack: Free Forest Sample*", "*Next-gen Camera*", "*Old Tables Chair*".

For the development of the application itself, a standard PC was used. Another device, that was necessary, was the Oculus (Meta) Quest. You can find more details on the official meta quest page `https://www.meta.com/quest/setup/`.

Simplified system concept is depicted in Fig. 1 . Pupil is working in virtual class (scenario). Virtual classroom is implemented by use of Unity engine and it is "running" on computer. Pupil (user) is immersed in the virtual classroom, and he can see virtual environment. Inside he interacts (performs tasks) with virtual classroom (virtual objects) by using VR headset. In the following chapters, the educational scenario (tasks) will be mainly described supported by examples of screenshots and implementation code (only Task1 in more detail).

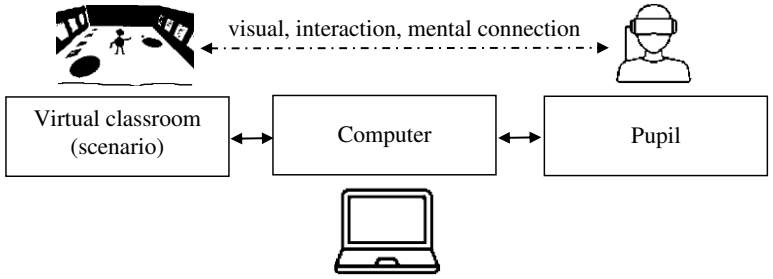


Fig. 1 Simplified system concept

B. Scenario

Based on existing findings, was designed, and created a 45-minute scenario of art education (lesson) in virtual reality (virtual classroom), which included three tasks for pupils. In the Unity program (C# base), was created a forest environment, that is both calming and engaging. The scenario and tasks sequence for pupils was developed (see Fig. 2).

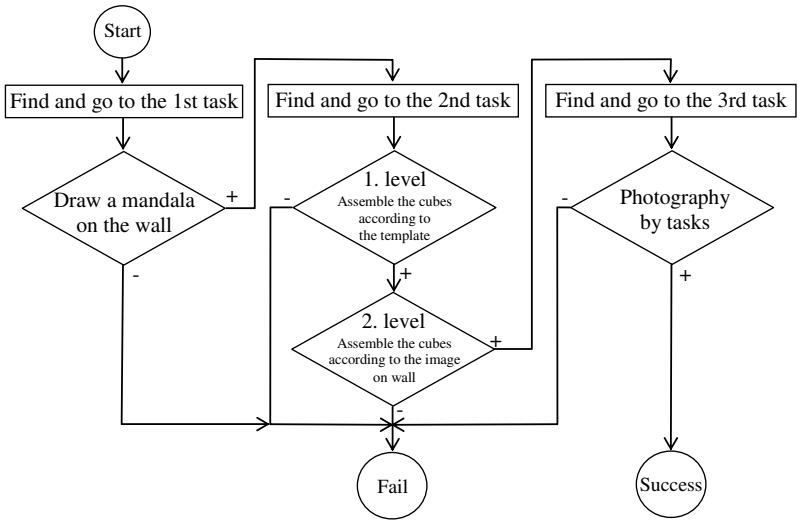


Fig. 2 The basic sequence of scenario tasks

- The goal of *the first task* is to focus on creativity and attention. It involves drawing a mandala. At the station for the first task, a drawing canvas is provided. In these surroundings, there are mandalas to inspire the pupils. This task will be time-limited to 15 minutes. This task will be described more detailed in the next chapter.
- The objectives of *the second task* are focused on cognitive skills, fine motor development, and imagination. The task is to build blocks (cubes). The first level involves building towers with blocks according to a template. If the pupil completes this task within the time limit of 5 minutes, he can proceed to the second level. In the second level, the task is to assemble blocks to match one of the images on the wall. This will be more challenging - as it requires more critical thinking. The pupil will have 10 minutes limit to complete the block assembly in the second level.
- The final part will be *the third task*, which involves photography. The pupil will be presented with a list of 5 objects to photograph. It is important, that the objects in the photograph are compositionally arranged correctly. The goal is to develop spatial orientation and attention to detail. The pupil will have 10 minutes limit to complete this task.

These two tasks (2 and 3) will not be described in more implementation detail. Their example views are shown in Fig. 3.

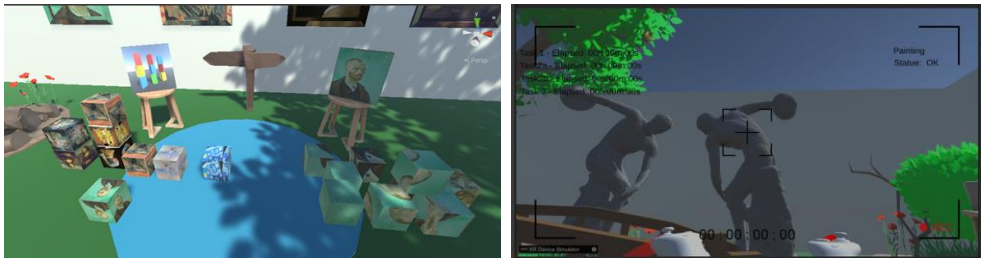


Fig. 3 Example views of Task2 and Task3

IV. TASK1 DESCRIPTION

For the first task, we have set up a purple station. It includes mandala templates on the wall for inspiration, as was already mentioned. There is also a pen available, that changes color (it also has an eraser), as well as a clean drawing canvas (Fig. 4).



Fig. 4 Example view of Task1

A. *Task 1 implementation*

The implementation starts with the drawing canvas. It was added to the environment as a 3D object called "*plane*". To this object, was assigned a script called "*Whiteboard*" (Fig. 5). Its implementation preview can be seen below. This script serves only to hold data about the texture. As we can see, the size of the canvas is set rigidly within *Vector2*. Given hardware limitations, this is an optimal solution. If this canvas should be to enlarge, it is certainly possible (the hardware limit must be respected). What phenomenon does this cause? We should view this as the resolution of an HMD display. If is larger display area, then higher resolution is possible.

```
1   public class Whiteboard : MonoBehaviour
2   {
3   public Texture2D texture;
4   public Vector2 textureSize = new Vector2(2048, 2048);
5
6           void Start()
7           {
8           var renderer = GetComponent<Renderer>();
9           texture = new Texture2D((int)textureSize.x, (int)textureSize.y);
10          renderer.material.mainTexture = texture;
11          }
12          }
```

Fig. 5 Whiteboard script code fragment



Fig. 6 Whiteboard with a pen

The pen creation was necessary next step (Fig. 6). The controller settings within VR headset were applied to the object *Pen*. The script "*XR Grab Interactable*" was attached for movement. The primary logic was implemented in the script "*WhiteboardMarker*". We set up color switching within the *ColorToggle()* function. The eraser is also included as part of the material. The *Draw()* function (Fig. 7 left) is where it is necessary to monitor whether the pupil (user) is drawing on the board or not, and then the touch of the pen itself. As a result, it is necessary to realize that it is drawn with a 3D object, i.e. pen, on a 2D object, i.e. on canvas.

```
private void Draw()
{
if(Physics.Raycast(tip.position, transform.up, out touch, tipHeight))
{
if (touch.transform.CompareTag("WhiteBoard"))
{
if(whiteboard == null)
{
whiteboard = touch.transform.GetComponent<Whiteboard>();
}
touchPos = new Vector2(touch.textureCoord.x, touch.
 textureCoord.y);
var x = (int) (touchPos.x * whiteboard.textureSize.x - ( penSize / 2));
var y = (int) (touchPos.y * whiteboard.textureSize.y - ( penSize / 2));
if(y < 0 || y > whiteboard.textureSize.y || x < 0 || x >
 whiteboard.textureSize.x)
{
return;
}
if (touchedLastFrame)
{
whiteboard.texture.SetPixels(x, y, penSize, penSize, colors);

        for( float f = 0.01f; f < 1.00f; f += 0.01f)
        {
        var lerpX = (int) Mathf.Lerp(lastTouchPos.x, x, f);
        var lerpY = (int) Mathf.Lerp(lastTouchPos.y, y, f);
        whiteboard.texture.SetPixels(lerpX, lerpY, penSize,
                penSize, colors);
        }
        transform.rotation = lastTouchRotation;
        whiteboard.texture.Apply();
        }
        lastTouchPos = new Vector2(x, y);
        lastTouchRotation = transform.rotation;
        touchedLastFrame = true;
        return;
        }
        }
        whiteboard = null;
        touchedLastFrame = false;
        }
```

```
void Update()
{
switch(taskState)
{
case TaskState.NOT_STARTED:

    if (Vector2.Distance(new Vector2(transform.position.x, transform.position.z), new
    Vector2(cameraTransform.
     position.x, cameraTransform.position.z)) < 0.5)
    {
    taskState = TaskState.STARTED;
    clock.timer = duration;
    clock.running = true;
    clock.message = "Elapsed: ";
    clock.endDuration = endDuration;
    r.material.color = new Color32(0, 255, 0, 255);
    }
    break;
    case TaskState.STARTED:
    if (Vector2.Distance(new Vector2(transform.position.x, transform.position.z), new
    Vector2(cameraTransform.
     position.x, cameraTransform.position.z)) > 1.25)
    {
    taskState = TaskState.COMPLETED;
    duration = clock.timer;
    clock.running = false;
    r.material.color = new Color32(255, 255, 255, 255);
    clock.message = "Finished: ";
    r.material.color = originalColor;
    }
    break;
    }
    }
```

Fig. 7 Draw script code fragment (left) and Taskstate checker code fragment (right)

The last position where drawing took place is taken. Then a new position is processed where the drawing ended. Subsequently, a straight line is created between these points. This line is a raster, created from points. The frequency of how often points need to be drawn is controlled within the for loop: *for (float f = 0.01f; f ¡ 1.00f; f += 0.01f)*. The variable "*f*" is of type *float*, so it is possible to define the precision much more sensitively. If the smaller value entered, the more points will be drawn, and the resulting image will be more accurate and sharper. The overall implementation is indicated in the code fragment depicted in (Fig. 7 right). The function tests two conditions. If the action (the draw) has not started, the input variables remain unchanged. If the action is started, individual parameters change. Action execution is assigned to the VR headset controller, and it is limited by the distance from the drawing plane.

V. APPLICATION EVALUATION

The respondent for evaluation was a pupil diagnosed with significant deficits in visuomotor performance, short-term visual memory, dysgraphia, and dyslalia (all in A category) with suspected ADHD. He was named as PupilXY for the purposes of verification study. VR headset for test was Oculus (Meta) Quest.

At the outset, we welcomed PupilXY and explained the purpose of our meeting, requesting his cooperation. We outlined the tasks he would be undertaking, and the time limit allotted for their completion. Subsequently, we asked if he understood the instructions and knew what to do. He was informed that if he felt uncomfortable during the tasks, he should verbally communicate this to us. Initially, we inquired whether he had any experience with virtual environments or virtual reality headsets. His response was negative, which was crucial information for us. Not everyone navigates the virtual world smoothly, so we first asked him to put on the virtual reality headset to see if this change was acceptable to him. After a brief trial, he confirmed that everything was fine. This initial conversation lasted for 10 minutes.

We conducted the evaluation immediately after completing the testing. For each task, we established the following assessment criteria:

- *Task1* - time taken to complete the task, histogram of colors used, extent of painted area (% canvas coverage);
- *Task2* - time taken to complete the task, coordination of movement, focus;
- *Task3* - time taken to complete the task, photographing 5 objects, spatial positioning of objects in the photograph.



Fig. 8 Example of a drawing test performing by the pupil.

A. *Evaluation of Task1*

PupilXY adhered to the 15-minute time limit. He created a mandala in the shape of a flower with petals. He used only one color - *black*, out of the possible three (*red*, *blue* and *black*). He placed the image in the center of the canvas, which represents 30% of the usable canvas area.

For the objective evaluation of the results, was utilized a histogram. This analysis was performed using a freely available application `https://sisik.eu/histo`. After providing the respondents with the created image, was generated a color histogram. The histogram itself shows low utilization rate of individual colors. This can be due to respondent's lack of experience with the option to change colors, or because the task did not specify, that they should also color the mandala. We note that the task was partially fulfilled because he did not use other colors and he did not cover the entire canvas area. An example of the histogram evaluation is illustrated in Fig. 9.

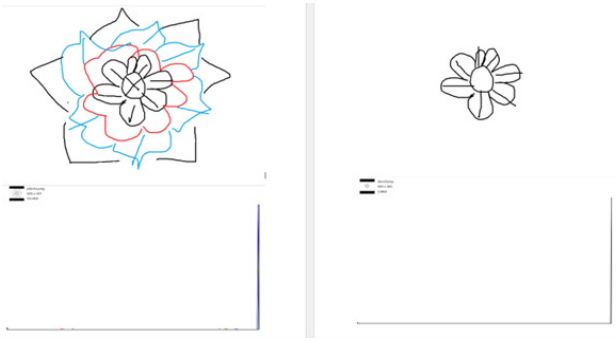


Fig. 9 Histogram

B. *Evaluation of Task2*

The task was divided into two parts. The time limit for the first part stacking blocks (cubes) into towers - was met within 5 minutes. However, the time limit for the second part of the task was not met. The task was completed 7 minutes later. Regarding the assessed criterion of movement coordination, we found, that the coordination was insufficient. This may be attributed to the student's lack of experience with joystick manipulation, which could have led to inadequate movement coordination. He focused his attention on stacking blocks according to the image template, but was insufficiently concentrated, as indicated by verbal expressions such as "*the block fell*", or "*I know where the block should go, but I can't do it*".

We conclude that the task was partially fulfilled because the time limit for the second part of the task was not adhered to and PupilXY was not fully focused on the task.

C. *Evaluation of Task3*

PupilXY adhered to the 10-minute time limit. He spent the most time searching for objects in the classroom space. He fulfilled the criterion of photographing 5 objects. The correct spatial positioning of the object in the photograph was achieved in 4 out of 5 photographs. In one instance, he only partially photographed the object. He verbally expressed opinions such as "*I'm stressed, how much time do I have left?*" and "*Do I have to photograph all 5?*"

We conclude that the task was partially fulfilled because one of the objects was not photographed correctly.

VI. Conclusion

As a part of the evaluation, we can state, that the selection of technologies, programs, environments and scenarios appears suitable for achieving the objectives set forth in our work. The integration of virtual reality, creativity, art and education is a dynamic field, that shapes our understanding of innovations and influences the personal development of individuals.

Virtual reality and related technologies represent one of the most progressive areas of computer science and information technology. It brings a change in communication with computing systems so that they are more and more natural, unambiguous and simple for people, including people with disabilities. The article presented our pilot solution in the form of an application based on XR technologies that can be used in the education of pupils with special educational needs but based on the artistic feeling of the user (pupils). Such an approach is quite innovative (not just a technocratic evaluation of knowledge) especially for disabled children, because every child deserves the same opportunities, whether they are handicapped or not. The possibility to work also in the field of artistic expression of possible fantasy ideas could be one of the ways. Thus, our work could be proof that virtual reality technologies have the potential to be used also in the environment of Slovak special education.

Acknowledgment

This work has been supported by the KEGA grant No. 048TUKE-4/2022: „Collaborative virtual reality technologies in the educational process "

References

[1] Sobota B., Hudák M., Pietriková E.: Collaborative XR Systems and Computer Games Development / In: Computer Game Development. - London (UK) : IntechOpen 2022, pp. 1-24 . ISBN 978-1-80355-564-5 available: http://dx.doi.org/10.5772/intechopen.105555.

[2] Rezaei G. H. Y. A., Heiden W.: "User Interface Design for Disabled People Under the Influence of Time, Efficiency and Costs," June 2014.

[3] Disabled World, "Disabilities: Definition, Types and Models of Disability," (online), cit. 2024-03, url: www.disabled-world.com

[4] Carl E., Stein A.T., Levihn-Coon A. et. al.: Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. J Anxiety Disord. 2019, 61:27–36. doi: 10.1016/j.janxdis.2018.08.003.S0887-6185(18)30246-9.

[5] Blog S. S., "Guide to virtual reality art: What it is and how to make VR art yourself," 2021, url: https://www.skillshare.com/en/blog/ guide-to-virtual-reality-art-what-it-is-and-how-to-make-vr-art-yourself/

[6] Pupala B., "Education for 21th century" Tech. Rep., 2022, url :´ https: //www.statpedu.sk/sk/aktuality/vzdelavanie-21-storocie-vychodiska-zmien-kurikule-zakladneho-vzdelavania.html

[7] Glynn M.: "Art and design for secondary school children with SEN," Tech. Rep., 2018, url: https://www.routledge.com/Art-and-Design-for-Secondary-School-Children-with-SEN-A-Resource-for-Inclusive-Teaching/Glynn/p/book/ 9781138559493

[8] Rodríguez-Cano S., Delgado-Benito V., et.al.: Design of a Virtual Reality Software to Promote the Learning of Students with Dyslexia. Sustainability. 2021. available: https : / / doi . org / 10 . 3390 /su13158425.

[9] Kalenjuk E. et. al.: Art-based research to explore children's lived experiences of dysgraphia. Children and Society. 2023. available: https://doi.org/10.1111/chso.12709.

[10] Savickaite S. et. al.: "The use of a tablet-based app for investigating the influence of autistic and adhd traits on performance in a complex drawing task," SpringerLink, Behavior research methods, pp. 54(5), 2479–2501, 2022. available: https://doi.org/10.3758/s13428-021-01746-8

[11] Romero-Ayuso Dulce et. al. : "Effectiveness of virtual reality-based interventions for children and adolescents with ADHD: A systematic review and meta-analysis," Sustainability, 2021. available: https://doi.org/10.3390/ children8020070

Učenie neurónových sietí typu MLP pomocou algoritmu simulovaného žíhania

[1]Daniel NOVÁK, [2]Marián MACH

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

[1]daniel.novak@student.tuke.sk, [2]marian.mach@tuke.sk

Abstrakt **– Článok sa zameriava na problematiku učenia váh dopredných neurónových sietí alternatívnymi metódami. Pozornosť je venovaná metaheuristickým prehľadávacím algoritmom, inšpirovaným prírodnými fenoménmi. Článok sa zaoberá využitím algoritmu simulovaného žíhania, používaného pre iteratívnu úpravu váh siete s cieľom minimalizovať klasifikačnú chybu učenej siete. V rámci experimentovania bol vyšetrovaný vplyv vybraných parametrov na dosahovanú správnosť klasifikácie. Experimenty boli realizované na štandardnej tabuľkovej dátovej množine pre klasifikáciu kultivarov (odrôd vína).**

Kľúčové slová **– klasifikačná úloha, metaheuristika, MLP, neurónová sieť, simulované žíhanie**

I. Úvod

Neurónové siete, či už hlboké alebo plytké, sú typicky učené algoritmami, založenými na gradientovom hľadaní minima chyby učenia, kvantifikujúcej rozdiel medzi predikovanými hodnotami a anotačnými hodnotami jednotlivých príkladov. V zásade, po návrhu štruktúry siete, je potrebné určiť hodnoty váhových koeficientov jednotlivých vrstiev siete, čo sa často označuje ako úloha optimalizácie reprezentačnej úrovne. Jedná sa vlastne o úlohu numerickej optimalizácie.

V tejto oblasti je však možné použiť aj iné, na gradiente nezaložené metódy. Jednou takou triedou použiteľných algoritmov sú metaheuristické algoritmy [1]. Jedná sa o širokú skupinu aproximačných algoritmov, reflektujúcich rôzne prírodné (fyzikálne, biologické, sociálne) fenomény. Na základe inšpirácie týmito fenoménmi sú vytvárané hľadacie operátory, pomocou ktorých algoritmy prehľadávajú priestor možných riešení s cieľom nájdenia optimálneho riešenia bez uviaznutia v lokálnom extréme.

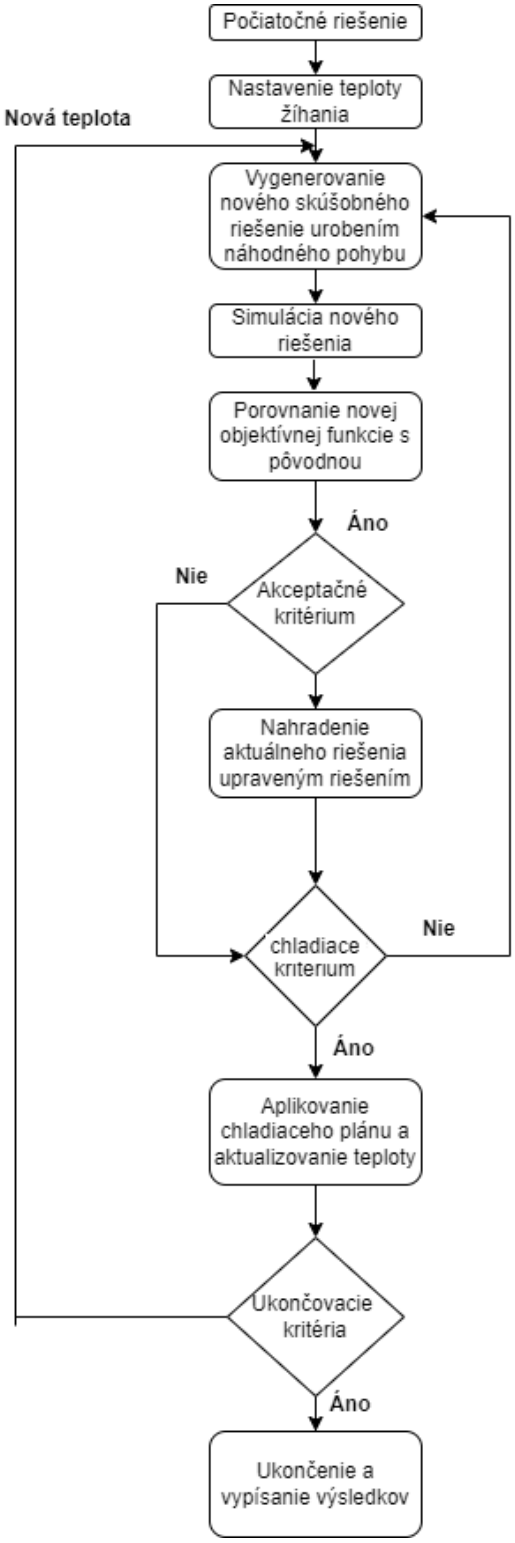
Neurónové siete boli trénované viacerými metaheuristickými algoritmami. V [2] bola použitá diferenciálna evolúcia a harmonické hľadanie v oblasti konvolučných sietí s experimentmi na dátových množinách MNIST a CIFAR-10. Autori v [3] využili pre konvolučné siete algoritmus optimalizácie rojom častíc. V [4] autor navrhol použitie distribuovanej verzie algoritmu kolónie umelých včiel a riešenie overil na dátovej množine MNIST. Algoritmus optimalizácie šedého vlka a sinus-kosínus algoritmus boli použité pre klasifikáciu medicínskych časových radov v [5].

Tento článok sa zameriava na využitie algoritmu simulovaného žíhania, pričom učenie obmedzuje iba na siete typu MLP (Multilayer perceptron). Štruktúra článku je nasledovná – v druhej sekcii sa venuje samotnému algoritmu simulovaného žíhania. V tretej sekcii je uvedená séria experimentov s daným algoritmom na úlohe hľadania riešenia globálneho minima chybovej funkcie vybranej dátovej množiny.

II. Algoritmus simulovaného žíhania

Simulované žíhanie je metaheuristický algoritmus, inšpirovaný fyzikálnym procesom žíhania – zahriatia materálu na vyššiu teplotu a veľmi pomalého ochladzovania (odoberania energie) s cieľom dosiahnuť optimálnu kryštálovú mriežku materiálu bez vnútorných štrukturálnych defektov (ukotvenie častíc v mriežke na najnižších energetických hladinách).

Jedná sa o individuálny prehľadávací algoritmus [6], ktorý v každom okamihu pracuje iba s jedným kandidátom na riešenie. V každej iterácii generuje nové skúšobné riešenie a rozhoduje sa, či toto nové riešenie nahradí aktuálne uvažované riešenie. Vývojový diagram algoritmu je na Obr. 1.



Obr. 1 Vývojový diagram algoritmu simulovaného žíhania.

Návrh algoritmu obsahuje niekoľko činností, ktoré daný algoritmus charakterizujú. Sú to tieto jeho súčasti:

- spôsob generovania nového skúšobného riešenia,
- akceptácia alebo odmietnutie vygenerovaného riešenia,
- kritérium pre zníženie teploty,
- spôsob znižovania teploty.

Skúšobné riešenie sa generuje náhodným spôsobom – výberom niektorého riešenia z okolia aktuálne uvažovaného riešenia. To znamená, že vzdialenosť medzi riešeniami musí byť menšia ako nejaký prah. Použité okolie musí byť také veľké, aby umožnilo únik z lokálneho extrému, a

zároveň také malé, aby sa pri generovaní využíval princíp exploatácie a hľadanie nemalo charakter čisto náhodného hľadania.

Akceptačné kritérium určuje pravdepodobnosť prechodu z aktuálneho riešenia na novo generované riešenie. Typicky je táto pravdepodobnosť určovaná podľa vzťahu (v tvare pre minimalizačnú úlohu)

$$prob = \begin{cases} 1.0 & \Delta \leq 0.0 \\ e^{\frac{-\Delta}{T}} & \Delta > 0.0 \end{cases} \tag{1}$$

kde Δ je rozdiel kvality nového a aktuálneho riešenia a T je aktuálna hodnota teploty. Ak teda novo generované riešenie je lepšie ako aktuálne, tak je akceptované s jednotkovou pravdepodobnosťou. Ak je horšie, tak môže byť akceptované, avšak šanca na akceptáciu klesá so stúpajúcim rozdielom medzi riešeniami. Tá šanca taktiež závisí od teploty, čím je teplota nižšia, tým je šanca horšieho riešenia tiež nižšia.

Teplota sa zvyčajne neznižuje v každej iterácii, ale iba každých niekoľko iterácií, aby sa systém dokázal stabilizovať pri aktuálnej teplote. O prechode na zmenu teploty rozhoduje chladiace kritérium. V prípade zmeny teploty sa táto teplota postupne znižuje podľa chladiaceho plánu, čo má za následok, že v skorších etapách algoritmu je pravdepodobnosť akceptácie horších riešení vyššia ale postupom času sa znižuje (znižuje sa miera explorácie na úkor posilnenia miery exploatácie).

III. Vybrané experimenty s učením pomocou simulovaného žíhania

A. Testovací problém

Použitý dataset obsahuje výsledky chemického skúmania vín, ktoré boli vyrobené z troch rôznych kultivarov, pričom všetky z nich boli pestované v rovnakej talianskej lokalite. Tento dataset obsahuje 13 atribútov, ktoré popisujú rôzne chemické aspekty vzoriek vína a obsahuje 178 záznamov, pričom každý z nich predstavuje samostatnú vzorku vína. Dátová množina je prístupná pomocou knižnice `scikit-learn`, kam bola prebraná z originálu v UCI Machine Learning Repository [7].

Všetkých 13 atribútov je numerických, pričom však sú použité rôzne rozsahy, z ktorých sú hodnoty atribútov vyberané. Všetky hodnoty sú kladné, keďže atribúty reprezentujú množstvá rôznych látok (napr. horčíka, fenolov, alkoholu, kyseliny jablčnej, ap.) nájdených vo vzorkách skúmaných vín. Dataset neobsahuje žiadne chýbajúce hodnoty, ktorých absenciu by bolo potrebné riešiť nejakou vhodnou metódou. Kategorizácia je robená do troch klasifikačných tried (kultivary Barolo, Grignolino a Barbera), pričom početnosť zastúpenia jednotlivých kategórií je 59, 71 a 48. V dátovej množine nie je výrazná nerovnováha výskytu niektorej z klasifikačných tried a teda nie je nutné realizovať niektorú z metód vyrovnávania početnosti zastúpenia klasifikačných tried.

B. Použitý klasifikačný model

Klasifikačný model v tvare doprednej neurónovej siete bol vytváraný v systéme `tensorflow`. Sieť mala jednoduchú štrukúru s dvomi skrytými, plne prepojenými vrstvami [8]. Bola definovaná nasledovným kódom.

```
inputs = Input(shape=(input_dim,))
x = tf.cast(inputs, dtype='float32')
x = Dense(nodes, activation='relu')(x)
x = Dropout(0.2)(x)
x = Dense(nodes, activation='relu')(x)
x = Dropout(0.2)(x)
x = Dense(output_dim)(x)
outputs = Softmax()(x)
model = Model(inputs=inputs, outputs=outputs)
```

Pre zjednodušenie štruktúry siete, obe skryté lineárne vrstvy mali rovnaký počet neurónov, definovaných parametrom `nodes`.

C. Výsledky experimentov

Experimenty, ktorých výsledky sú uvedené v tejto kapitole, boli realizované v [8] pri použití 100 iterácií algoritmu simulovaného žíhania s chladiacim plánom exponenciálneho tvaru

$$T = T_{init} e^{-\gamma t} \tag{2}$$

kde t reprezentuje číslo aktuálnej iterácie. Všetky použité konštanty boli experimentálne nastavené na hodnoty, pri ktorých algoritmus podával najlepší výkon.

Na začiatku učenia je generovanie návrhu počiatočného riešenia. V našom prípade bolo potrebné vygenerovať váhy počiatočného modelu. Použili sme náhodné generovanie, avšak pre voľbu pravdepodobnostnej distribúcie, podľa ktorej generovanie bolo realizované, existuje viacero možností. Vyskúšali sme osem možností, dosiahnuté výsledky sú uvedené v tab. 1.

Tabuľka 1
Porovnanie správnosti klasifikácie pre rozličné spôsoby inicializácie váh modelu

	Správnosť pred učením		Správnosť po učení	
Inicializácia	učiaca množina	testovacia množina	učiaca množina	testovacia množina
He normal	0.29576	0.31666	0.75350	0.76108
Glorot normal	0.36198	0.34998	0.76478	0.73890
Random normal	0.35776	0.35000	0.83098	0.87222
Truncated normal	0.32676	0.35556	0.86620	0.88334
He uniform	0.39438	0.43334	0.77324	0.76668
Glorot uniform	0.29860	0.30000	0.77182	0.76664
Random uniform	0.31548	0.30000	0.83662	0.82780
Orthogonal	0.31690	0.27776	0.75072	0.72222

V zásade boli použité tri druhy distribúcií:

- varianty normálnej distribúcie,
- varianty uniformnej distribúcie,
- ortogonálna inicializácia (ortogonálna dekompozícia matice náhodných čísel podľa normálnej distribúcie).

Jednotlivé varianty (He, Glorot, Random, Truncated) sa líšili hranicami intervalov, z ktorých boli náhodné hodnoty generované.

Z tabuľky vidno, že varianty He, Glorot a Orthogonal (špeciálne navrhnuté pre inicializáciu neurónových sietí) sa v prípade simulovaného žíhania osvedčili menej ako klasické distribúcie. Dôvodom môže byť ich generovanie hodnôt bližšie k nule kvôli limitovaniu rozsahu generovania, čo sa v prípade simulovaného žíhania ukázalo škodlivým.

Pri použití gradientových metód učenia sa ako jedna z regularizačných metód používa Dropout, keď dochádza k dočasnému nulovaniu váh s následkom ignorovania niektorých hodnôt pri určovaní nových váh. A keďže pri simulovanom žíhaní by sa nové riešenie malo vyberať z okolia aktuálneho riešenia (a okolie je často definované maximálnym počtom zmien), tak analogicky sme obmedzili počet váh, ktoré sa v jednej iterácii môžu meniť. Dochádza k náhodnému maskovaniu váh, pričom toto je riadené pravdepodobnosťou, že váha bude menená. Pravdepodobnosť maskovania môže byť dvojakého typu:

- statická - hodnota pravdepodobnosti $prob = p$ sa počas celého behu algoritmu nemení,
- dynamická - pravdepodobnosť sa z počiatočnej hodnoty p znižuje.

Dosiahnuté výsledky sú uvedené v tab. 2. Pri dynamickom nastavení sa hodnota znižovala lineárnym spôsobom, zohľadňujúcim pomer aktuálnej iterácie k celkovému počtu iterácií.

$$prob = p\left(1 - \frac{iter}{max\_iter}\right) \tag{3}$$

Tabuľka 2
Porovnanie správnosti klasifikácie pre rozličné pravdepodobnosti maskovania váh

Typ		Správnosť po učení	
hodnoty	p	učiaca množina	testovacia množina
stat	0.1	0.76478	0.73890
stat	0.3	0.71126	0.72222
stat	0.5	0.50702	0.52774
dyn	0.1	0.88450	0.91666
dyn	0.3	0.76460	0.79446
dyn	0.5	0.75120	0.71664

Z tabuľky vyplýva, že súčasná zmena väčšieho počtu váh súčasne pôsobí deštruktívnym spôsobom. Preto je vhodnejšie vyberať menší počet váh na zmenu v rámci jednej iterácie, a ešte lepšie je počas behu algoritmu tento počet zmenšovať. V dôsledku toho generované nové riešenie ostáva blízko aktuálneho riešenia a teda algoritmus takýmto spôsobom intenzifikuje svoje prehľadávanie podporovaním vyššej miery exploatácie.

D. Porovnanie s klasickými metódami

Pre zistenie, či má použitie simulovaného žíhania zmysel v kontexte existencie gradientových metód, sme experimentovali so sieťami rozličných veľkostí (menil sa počet uzlov v skrytých plne prepojených vrstvách). Pre zachovanie čo najpodobnejších podmienok pre oba prístupy, sieť pri učení simulovaným žíhaním bola inicializovaná rovnakým spôsobom ako pri gradientovom učení – odchýlky boli inicializované na nulu a váhy pomocou Glorot normal (čo sú implicitné inicializátory v použitej knižnici). Gradientová metóda (použitý optimalizátor Adam - Adaptive Moment Estimation) v každej epoche modifikuje všetky váhy. Na druhej strane simulované žíhanie použilo statickú pravdepodobnosť maskovania $p = 0.1$, spôsobujúcu, že v jednej iterácii sa modifikuje približne iba jedna desatina všetkých váh. Preto pre simulované žíhanie bolo použitých 1000 iterácií (s tisíc rôznymi hodnotami teploty T) a pre gradientovú metódu bolo použitých 100 epoch učenia. Výsledky sú v tab. 3 a tab. 4.

Tabuľka 3
Porovnanie dosiahnutých správností učenia - simulované žíhanie

Počet	Učiaca množina			Trénovacia množina		
uzlov	najlepšia	priemerná	smer. odchýlka	najlepšia	priemerná	smer. odchýlka
2	0.7113	0.66196	0.02521	0.6944	0.60552	0.04444
4	0.9014	0.89154	0.01055	0.9167	0.86666	0.02724
8	0.9507	0.93802	0.01569	0.9722	0.93888	0.02076
16	0.9648	0.92256	0.02314	0.9722	0.95000	0.04081

Tabuľka 4
Porovnanie dosiahnutých správností učenia - optimizér Adam

Počet	Učiaca množina			Trénovacia množina		
uzlov	najlepšia	priemerná	smer. odchýlka	najlepšia	priemerná	smer. odchýlka
2	0.6056	0.52958	0.08726	0.7500	0.58888	0.12716
4	0.6972	0.62816	0.07883	0.8333	0.68332	0.11196
8	0.8803	0.77888	0.07999	0.9722	0.93890	0.02719
16	0.9366	0.91830	0.01226	0.9752	0.96664	0.02080

Z tabuliek vidno, že pri zložitejších sieťach s 8 alebo 16 uzlami sú oba prístupy porovnateľné. Iná situácia je v prípade menších sietí s 2 alebo 4 uzlami. Výkon gradientového učenia je horší v priemerne dosahovaných hodnotách správnosti. O probléme svedčia aj rádovo väčšie hodnoty smerodajnej odchýlky na trénovacej množine, signalizujúcej rozkolísanosť výkonov, podávaných gradientovým prístupom.

IV. ZÁVER

Vykonané experimenty potvrdili, že učenie váh pomocou simulovaného žíhania dokáže poskytnúť výsledky, ktoré dokážu konkurovať výsledkom učenia tradičnými postupmi, založenými na gradientovom učení. Príťažlivá je najmä schopnosť simulovaného žíhania unikať z lokálnych extrémov a ďalej pokračovať v prehľadávaní priestoru váh. Do budúcna by sme chceli overiť použiteľnosť simulovaného žíhania aj v prípadoch učenia hlbších sietí a aj v kontexte rozsiahlejších dátových množín.

LITERATÚRA

[1] B. Akay, D. Karaboga and R.Akay, A comprehensive survey on optimizing deep learning models by metaheuristics, *Artificial Intelligence Review*, 2022, vol 55, 829-894.
[2] L.M.R. Rere, M.I. Fanany and A.M. Arymurthy, Metaheuristic algorithms for convolution neural network, *Computationa Intelligence and Neuroscience*, 2016, 1–13.
[3] A.R. Syulistyo, D.M.J. Purnomo, M.F. Rachmadi and A. Wibowo, Particle swarm optimization (PSO) for training optimization on convolutional neural networK (CNN). *Jurnal Ilmu Komputer dan Informasi*, 2016, vol 9, 52-58.
[4] A. Banharnsakun, Towards improving the convolutional neural networks for deep learning using the distributed artificial bee colony method. *Int. Journal of Machine Learning and Cybernetics*, 2018, vol. 10, 1301–1311.
[5] T.A. Rashid, P. Fattah and D.K. Awla DK, Using accuracy measure for improving the training of LSTM with metaheuristic algorithms, *Procedia Computer Science*, 2018, vol. 140, 324–333.
[6] Z. Michalewitz and D.B. Fogel, *How to solve it: modern heuristics*, 2nd ed., Springer, 2004.
[7] Wine - Using chemical analysis to determine the origin of wines. https://archive.ics.uci.edu/dataset/109/wine [prístupné 17.5.2024]
[8] D, Novák, *Učenie neurónových sietí prírodne inšpirovanými heuristickými algoritmami*, Bakalárska práca, Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach, 2023.

Návrh siete pre viac-pobočkovú firmu

[1]*Ľuboš OVSENÍK,* [2]*Jakub ORAVEC,* [3]*Štefan LAZORČÁK*

[1,2,3]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]lubos.ovsenik@tuke.sk, [2]jakub.oravec@tuke.sk, [3]stefan.lazorcak@student.tuke.sk

***Abstrakt* — Príspevok sa zaoberá návrhom počítačovej siete pre viac-pobočkovú firmu s postupným vylepšovaním funkčnosti a bezpečnosti medzi návrhmi s použitím Cisco zariadení. V poslednej 4. topológii sa nachádza Velocloud SD-WAN (Software Defined Wide Area Network) pre porovnanie s bežne používaným Ciscom.**

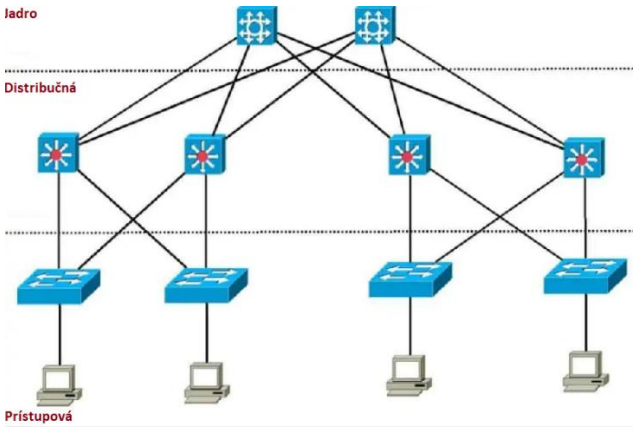
***Kľúčové slová* — Cisco zariadenia, počítačová sieť, prepínanie, SD-WAN siete, smerovanie, Velocloud**

I. Úvod

Firmy a spoločnosti si zakladajú na kvalitnej, rýchlej a bezpečnej počítačovej sieti pre svoje pobočky. Pri navrhovaní sieťovej topológie pre správne potreby firmy existuje veľa možností, pričom dôležitú úlohu zohráva aj finančný aspekt navrhnutého riešenia. Príspevok sa bude venovať návrhu topológií počítačovej siete a jej realizácii s ohľadom na finančnú stránku návrhu, tzn. bude snaha o minimalizáciu počtu sieťových zariadení s dôrazom na kvalitnú a bezpečnú sieť pre fiktívnu spoločnosť. Návrh a kompletná konfigurácia sieťovej topológie je realizovaná v prostrediach GNS3 (Graphical Network Simulator-3) a EVE-NG (Emulated Virtual Environment – Next Generation).

II. Trojúrovňový hierarchický model siete

Topológia siete sa používa na opis fyzickej a logickej štruktúry siete. Mapuje spôsob akým sú rôzne uzly v sieti umiestnené a prepojené, vrátane prepínačov a smerovačov, a taktiež aj spôsob toku údajov v sieti. Dobre naplánovaná topológia siete zlepšuje používateľskú skúsenosť a pomáha správcom siete maximalizovať jej výkon pri plnení obchodných potrieb. Pri správne zvolenej topológii je jednoduchšie lokalizovať chyby a nedostatky, odstraňovať problémy a zdieľať dáta cez siete. So správne riadenou topológiou siete môže firma zlepšiť efektivitu prenosu údajov, čo pomáha znižovať náklady na údržbu a prevádzku. Sieťové topológie sú vo všeobecnosti usporiadané dvomi spôsobmi, a to fyzickým (underlay) alebo logickým (overlay). Fyzický spôsob určuje skutočné prepojenia v sieti, ako kabeláž a umiestnenie rôznych komponentov, zatiaľ čo logický ukazuje, ako pretekajú údaje v rámci siete a od jedného zariadenia k druhému bez ohľadu na fyzické prepojenia medzi zariadeniami. Zatiaľ čo fyzická sieť využíva fyzické prepojenia na prenos údajov, skutočný tok dát je definovaný logikou a nie fyzickými prepojeniami [1].



Obr. 1 Trojúrovňový hierarchický model.

Vrstvová architektúra, ktorá sa bežne využíva v podnikoch a pobočkách, zvyčajne pozostávajúca z troch vrstiev, je široko používaný model na navrhovanie spoľahlivej, škálovateľnej a nákladovo efektívnej siete, ktorá slúži používateľom, čo sa k nej priamo pripájajú. Keďže siete môžu byť mimoriadne komplikované, s viacerými protokolmi a rôznymi technológiami, spoločnosť Cisco vyvinula vrstevný hierarchický model na navrhovanie spoľahlivej sieťovej infraštruktúry. Každá z vrstiev tohto modelu má svoje vlastné vlastnosti a funkcie, čo znižuje zložitosť siete. Sú kategorizované ako jadro (Core), distribučná vrstva (Distribution) a prístupová vrstva (Access) (Obr. 1) [2].

V LAN (Local Area Network) prostredí poskytuje prístupová vrstva koncovým zariadeniam prístup k sieti. Pre sieť malých podnikov vo všeobecnosti zahŕňa prepínače druhej vrstvy a prístupové body poskytujúce konektivitu medzi pracovnými stanicami a servermi. Táto vrstva plní množstvo funkcií, ako sú vysoká dostupnosť, zabezpečenie rozhraní, kontrolu protokolu na rozlíšenie adresy (ARP, Address Resolution Protocol), virtuálne záznamy riadenia prístupu, spanning-tree, PoE (Power over Ethernet), či pomocné siete VLAN (Virtual LAN) pre VoIP (Voice over Internet Protocol). Distribučná vrstva agreguje údaje prijaté od prepínačov prístupovej vrstvy predtým, ako sa prenesú do jadrovej vrstvy na smerovanie do konečného cieľa, kde sa najčastejšie využíva prepínač tretej vrstvy. Prepínač distribučnej vrstvy môže poskytovať upstream služby pre mnohé prepínače prístupovej vrstvy. Distribučná vrstva môže poskytovať agregáciu liniek, zabezpečenie založené na zásadách vo forme zoznamov riadenia prístupu (ACL, Access Control List) a filtrovania. Ďalšími službami sú smerovacie služby medzi LAN a VLAN a smerovacími doménami (napr. EIGRP a OSPF), redundancia a vyrovnávanie záťaže. Jadrová vrstva sa tiež označuje ako chrbtica siete. Pozostáva z vysokorýchlostných sieťových zariadení, ktoré sú navrhnuté tak, aby prepínali pakety čo najrýchlejšie a prepájali viaceré komponenty kampusu, ako sú distribučné moduly, servisné moduly, dátové centrum a WAN (Wide Area Network) sieť. Je kritická pre vzájomnú konektivitu medzi zariadeniami distribučnej vrstvy, resp. prepojenie distribučného bloku s rozhraním WAN a internetu. Jadro by malo byť vysoko dostupné a redundantné, agreguje prevádzku zo všetkých zariadení distribučnej vrstvy, takže musí byť schopné rýchlo posielať veľké množstvá údajov. Základnými vlastnosťami jadrovej vrstvy sú poskytovanie vysokorýchlostného prenpsu, spoľahlivosti a odolnosti voči poruchám, Taktiež to je vyhýbanie sa manipulácii s paketmi náročnými na CPU (Central Processing Unit) spôsobenej bezpečnosťou, kontrolou a klasifikáciou kvality služieb [3].

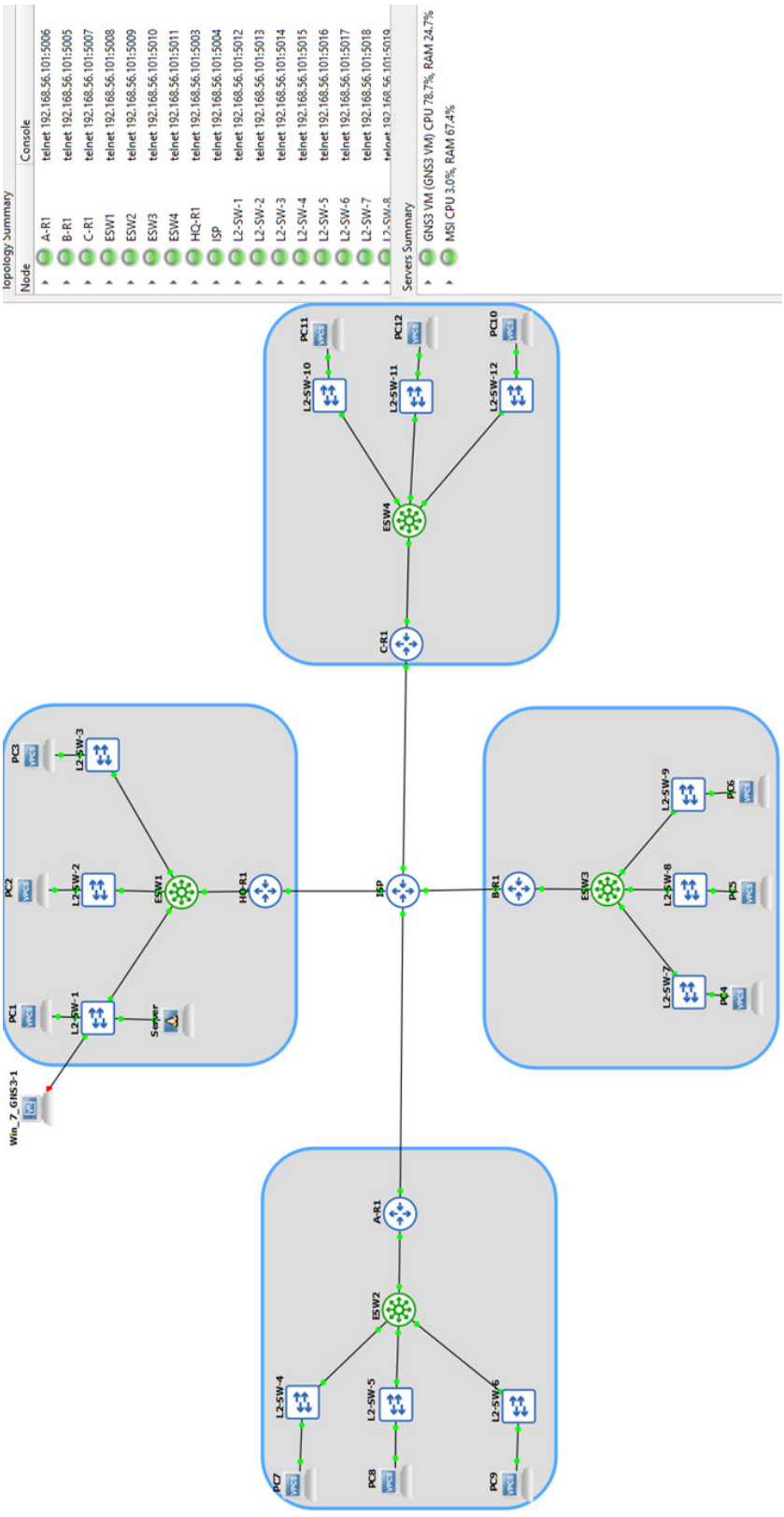
A. Fyzický návrh sieťovej topológie

Pri návrhu sieťovej topológie sa chceme čo najbližšie priblížiť realite, takže berieme do úvahy finančný aspekt riešenia. Nakoľko navrhujeme sieť pre 4 pobočkovú firmu, dá sa očakávať, že nebude mať veľký rozpočet na sieťové zariadenia, chod a spravovanie siete, no zároveň je pre firmu dôležité, aby mala spoľahlivú a kvalitnú sieť. Budeme tak prispôsobovať počet sieťových zariadení a typ internetového pripojenia, čomu prispôsobíme konfiguráciu. Predpokladáme, že si firma taktiež nebude chcieť platiť za drahé pripojenie, ako je napríklad MPLS (Multiprotocol Label Switching) a bude využívať klasickú internetovú linku.

Keďže dbáme na finančnú stránku riešenia, v prvej topológii bude mať každá pobočka jeden smerovač s jednou internetovou linkou k ISP (Internet Service Provider), jeden prepínač tretej vrstvy a tri prepínače druhej vrstvy. Počet prepínačov druhej vrstvy ako správca siete veľmi ovplyvniť nevieme, nakoľko to závisí od počtu pripojených koncových zariadení na daných pobočkách. V hlavnej pobočke HQ bude umiestnený server, ktorý bude reprezentovať firemný server na rôzne účely, či už na prístup k firemným súborom a dátam, alebo pre nás ako sieťového správcu na ukladanie konfigurácií sieťových zariadení (Obr. 2).

B. Logický návrh sieťovej topológie

Sieť je potrebné správne navrhnúť a rozložiť aj z logickej stránky. Nie každý koncový užívateľ by mal mať prístup ku každej časti sieti, takisto niektorí užívatelia by mali mať obmedzené možnosti využívania našej topológie. Rozdelili sme tak sieť jednotlivých pobočiek do podsietí pomocou VLAN. Vytvorili sme si VLAN10 a VLAN20 pre zamestnancov firmy, VLAN30 pre návštevníkov firmy a VLAN999 ako manažmentovú VLAN pre správcu siete. Základnou myšlienkou rozdelenia do podsietí je každej VLAN určiť, s kým môže a nemôže komunikovať. Ako správca siete máme prístup povolený všade, VLANy pre zamestnancov majú taktiež prístup všade okrem častí siete spravovaných správcom, sú napríklad prístup na sieťové zariadenia okrem servera. Najviac obmedzenou podsieťou je VLAN30 pre návštevníkov firmy, ktorí majú povolený len prístup na internet a majú tak blokovaný prístup k ostatným VLANám, čo budeme definovať na prepínači tretej vrstvy.



Obr. 2 Návrh topológie.

Súčasťou rozdelenia do podsietí je IP adresácia pre všetky zariadenia. V lokálnej sieti budeme pracovať s privátnymi rozsahmi IP adries, a to 192.168.X.X. Pre jednoduché rozdelenie jednotlivých pobočiek pri IP adresácii budeme meniť tretí oktet pre každú pobočku. HQ pobočka bude využívať rozsah 192.168.1X.X, pobočka A 192.168.2X.X, pobočka B 192.168.3X.X a pobočka C 192.168.4X.X. V poslednom oktete už prichádza na rad rozdelenie siete daných pobočiek do podsietí, kde treba brať do úvahy možný počet pripojených zariadení. Najväčší rozsah pridelíme VLAN30 pre zákazníkov, kde maska siete bude /24, takže pre možný počet 254 zariadení vrátane sieťových zariadení. Pre VLAN10 a VLAN20 bude maska siete takisto /24, ale

rozdelená na polovicu pre obe VLANy, takže každá VLAN bude mať masku /25 a možný počet zariadení 126. Pri poslednej manažmentovej VLANe takéto veľké rozsahy nepotrebujeme, nakoľko ju budeme využívať najmä na vzdialený prístup na sieťové zariadenia, a tak pridelíme masku /27, kde máme kapacitu 30 zariadení.

Z hľadiska funkčnosti siete a smerovania budeme využívať dynamické smerovacie protokoly, a to OSPF (Open Shortest Path First) na prepínači tretej vrstvy a smerovači a BGP (Border Gateway Protocol) na smerovači na komunikáciu medzi pobočkami. Pre prístup do internetu budeme využívať len jednoducho predvolenú trasu (Default Route) smerom k ISP. Bezpečnosť siete je takisto dôležitým faktorom, kde treba brať ohľad na riziko rôznych útokov, či už v rámci lokálnej siete na akejkoľvek pobočke alebo útokom z internetu. Jedným zo spôsobov zabezpečenia je už aj rozdelenie do podsietí a obmedzenie komunikácie určitým zariadeniam v lokálnej sieti. Ďalej budeme brať ohľad na fyzický útok, napríklad pripojenie vlastného počítača priamo na pobočke, ale aj logické útoky a odchytávanie našich preposielaných dát na internete.

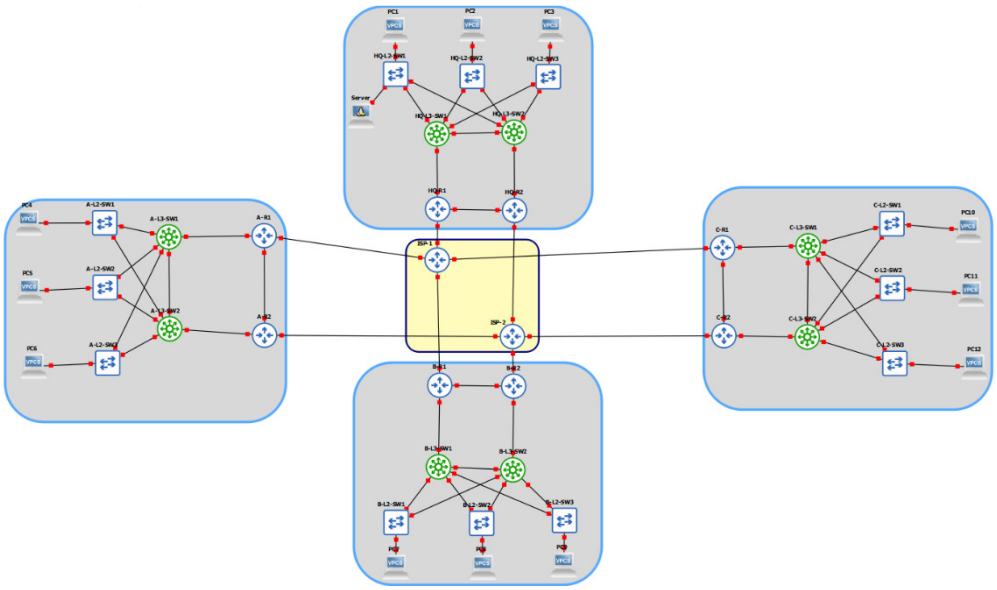
III. Konfigurácie jednotlivých topológií

A. *Konfigurácia 1. topológie*

Fyzická topológia je identická s obrázkom č. 2. Na smerovačoch sme sa venovali smerovaniu či už do internetu, ale aj medzi jednotlivými pobočkami a do LAN siete. Pripojenie do internetu sme riešili predvolenou trasou smerom k ISP smerovaču, no medzi pobočkami pomocou dynamického smerovacieho protokolu BGP s využitím bezpečnostného protokolu IPsec po odhalení bezpečnostných nedostatkov. Pri komunikácii s LAN sieťou sme zvolili protokol OSPF, nakoľko je štandardom a môže sa využiť. Z hľadiska funkčnosti siete sme sa ďalej venovali taktiež smerovaniu na prepínačoch tretej vrstvy pre spojenie so smerovačmi a oznamovanie lokálnych podsietí. Vďaka VLAN sme rozdelili siete jednotlivých pobočiek na podsiete pre návštevníkov, manažmentovú podsieť a dve podsiete pre zamestnancov.

B. *Konfigurácia 2. topológie*

Druhá topológia je vylepšenou verziou prvej sieťovej topológie. Hlavnú časť, ktorú treba zmeniť, je zabezpečenie pri prenose dát medzi pobočkami. V prvej topológii sme na smerovačoch zadefinovali preposielanie dát k ISP len predvolenou trasou. V druhej topológii tak implementujeme externý dynamický smerovací protokol BGP, pomocou ktorého budú pobočky medzi sebou prepojené. Komunikácia do internetu ostane zachovaná pomocou predvolenej trasy. Aby sme však zabezpečili aj prenos dát cez protokol BGP, spolu s ním implementujeme aj IPsec tunely práve pre BGP spojenia. Konfigurácia prepínačov druhej a tretej vrstvy ostane nezmenená, takisto sa nemení ani fyzický návrh topológie, ostáva rovnaký ako na Obr. 2.



Obr. 3 Návrh redundantnej topológie.

C. *Konfigurácia 3. topológie*

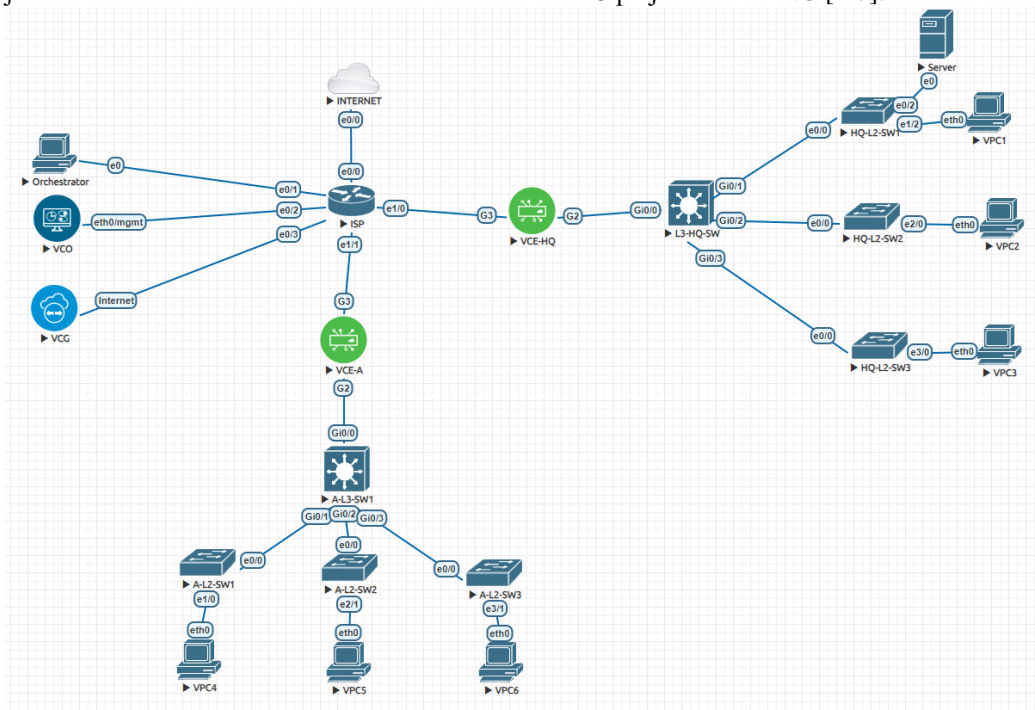
Veľkým nebezpečenstvom pri sieti sú výpadky zariadení alebo liniek, ktoré spôsobia nefunkčnosť celej siete alebo jej časti. Pri tretej topológii sa pozrieme na bezpečnosť siete aj z hľadiska dostupnosti. Od začiatku návrhu topológie berieme do úvahy finančný aspekt, tzn. že

firma nebude chcieť veľkú a komplikovanú sieť. Navrhujeme ale redundantnú topológiu, kde výrazne znížime riziko výpadku siete, keďže budeme mať záložné zariadenie, ktoré v prípade výpadku primárneho zariadenia prevezme jeho úlohu. Firma ako zákazník nemusí chcieť investovať do určitých častí siete, ale redundancia patrí medzi hlavné riešenia pri sieťach, ktoré si nemôžu dovoliť akýkoľvek výpadok.

Existuje niekoľko spôsobov, ako navrhnúť redundantnú topológiu. V našom prípade zvýšime redundanciu pridaním prepínača tretej vrstvy a smerovača pre každú pobočku. V predošlých topológiách používame len jeden smerovač a prepínač tretej vrstvy. Sú tak uzlami, ktoré ak vypadnú, znefunkční sa celá lokálna sieť. Pridaním ďalšieho smerovača a prepínača tretej vrstvy tak limitujeme výpadok celej siete, nakoľko druhé záložné zariadenie preberie funkcie primárneho zariadenia. Ďalší spôsob výpadku siete môže byť problém s pripojením k ISP. Zvýšime tak redundanciu pridaním novej internetovej linky k záložnému smerovaču, ale od iného poskytovateľa internetových služieb. Prepojenie jednotlivých zariadení je takisto dôležitou časťou dizajnu. Prepojili sme tak smerovače medzi sebou, primárny smerovač s primárnym prepínačom tretej vrstvy, záložný smerovač so záložným prepínačom tretej vrstvy a prepínače medzi sebou. Takisto zvyšujeme redundanciu aj pri prepínačoch druhej vrstvy. Každý z nich bude pripojený k obom prepínačom tretej vrstvy (Obr. 3).

D. Konfigurácia 4. topológie

Poslednou navrhovanou topológiou odbočíme od klasických Cisco zariadení. Rozhliadli sme sa po iných výrobcoch a novších technológiách, kde nás zaujalo SD-WAN. Výrobcov SD-WAN je takisto niekoľko, vrátane Cisca, no zvolili sme si Velocloud SD-WAN od spoločnosti VMware. Po potrebnom prieskume sme zistili, že Velocloud SD-WAN v emulačnom prostredí GNS3 nie je možné zrealizovať. Sme tak nútení z emulátoru GNS3 prejsť na EVE-NG [4-7].



Obr. 4 Fyzický návrh 4. topológie.

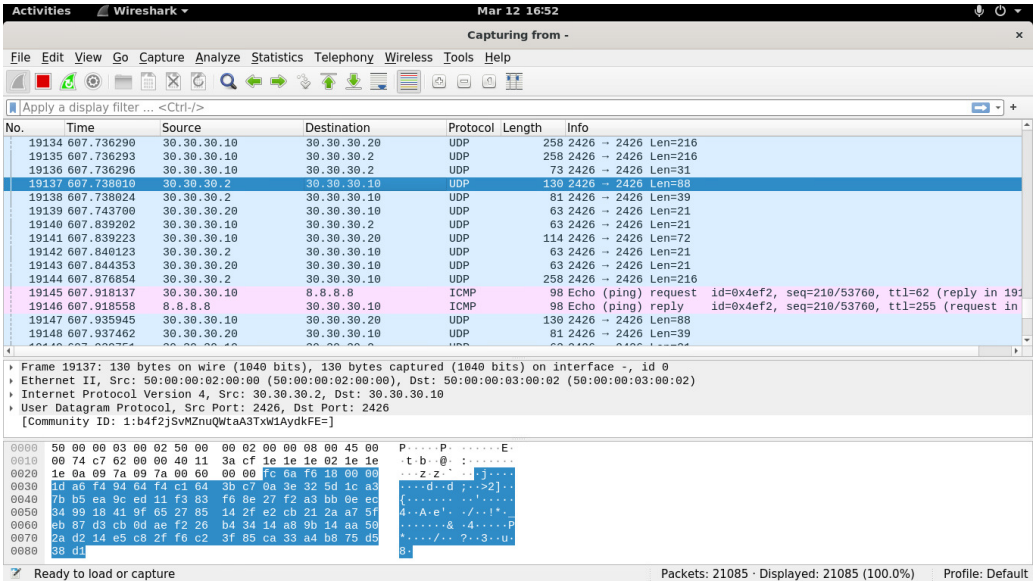
Po hlbšom prieskume sme taktiež prišlo na to, že Velocloud je hardvérovo veľmi náročný. Pre požadovanú funkčnosť siete s využitím Velocloud SD-WAN je potrebný Velocloud Orchestrátor, Velocloud brána a jednotlivé Velocloud Edges (smerovače). V praxi by sme sa orchestrátorom a bránou nezaoberali, nakoľko by nám ich poskytol priamo VMware alebo ISP, ktorý by túto technológiu sprostredkúval. V emulačnom prostredí sme však nútení si orchestrátor a bránu spojazdniť sami. Hardvérové požiadavky na orchestrátor v EVE-NG sú 4 vCPU a 16GB RAM pamäte, na bránu 4 vCPU a 8GB RAM a edge 2 vCPU a 4GB RAM. To je dohromady minimálne 12 vCPU a 32GB RAM pamäte, a to treba ešte brať do úvahy ostatné sieťové zariadenia ako prepínače, koncové zariadenia a že potrebujeme aspoň 2 edge smerovače. Pre tieto požiadavky sme oslovili univerzitu o sprístupnenie Ubuntu servera s týmito minimálnymi kritériami a poskytli nám server s 24 vCPU a 48GB RAM na realizáciu topológie pomocou EVE-NG. Nainštalovali sme si EVE-NG na sprístupnenom serveri, nahrali potrebné obrazy pre funkčnosť SD-WAN a ďalších sieťových zariadení. Ako sme už spomínali, pre Velocloud SD-WAN

potrebujeme orchestrátor, bránu a jednotlivé smerovače. Rozdielom oproti predošlým topológiám budú práve tieto smerovače, v LAN sieti ostane jeden prepínač tretej vrstvy a prepínače druhej vrstvy s koncovými zariadeniami. Pre konečné testovanie sme implementovali aj firemný server, na ktorom sme testovali rozdiely v konfigurácii Velocloud SD-WAN oproti klasickým Cisco smerovačom, pričom sa princíp konfigurácie snažíme zachovať rovnaký. Medzi pobočkami chceme zabezpečenie pomocou IPsec a smerom k prepínaču tretej vrstvy dynamický smerovací protokol OSPF.

Keďže orchestrátor a brána sú zariadenia spoločnosti VMWare a sú nám len sprostredkované na chod smerovačov, umiestnili sme ich mimo lokálnych sietí s tým, že cez internetovú linku a ISP k ním máme prístup. Všetka konfigurácia okrem prvotnej aktivácie SD-WAN zariadení prebiehala cez GUI orchestrátora, takže sme implementovali aj Windows koncové zariadenie, cez ktoré sa na orchestrátor pripájali a vytvárali potrebné konfigurácie (Obr. 4). Nad ISP sme umiestnili aj smerovač pre simuláciu prístupu na internet, kde sme na rozhranie k ISP definovali IP adresu 30.30.30.254 a loopback rozhranie s IP adresou 8.8.8.8.

Pre simuláciu odchytávania dát medzi pobočkami sme simulovali útočníka na internete a nahrávanie konfigurácie prepínača tretej vrstvy z pobočky A na firemný server v pobočke HQ. Tento prenos sme monitorovali pomocou programu Wireshark. Pri predošlých topológiách 2 a 3 sme pri prenose používali protokol ESP a všetky dáta boli šifrované. Pri Velocloude používame protokol UDP 2426. Na podporu Dynamic Multipath Optimization (DMPO) VMware zapuzdruje pakety do protokolu nazývaného Velocloud Multipath Protocol (VCMP). VCMP pridáva 31 bajtov pre používateľské pakety na podporu opätovného poradia, opravy chýb, analýzy siete a segmentácie siete v rámci jedného tunela. VCMP funguje na porte UDP 2426. Aby sa zabezpečilo konzistentné správanie vo všetkých potenciálnych scenároch, VCMP je šifrované pomocou transportného režimu IPsec. Pakety odoslané do internetu cez Edge smerovač nie sú štandardne zašifrované, pretože po opustení smerovača budú prechádzať priamo do otvoreného internetu.

Pri našej simulácii tak vidíme, že ping do internetu je odhaliteľný, ale preposielané dáta medzi pobočkami sú šifrované a nečitateľné, čo spĺňa náš cieľ (Obr. 5).



Obr. 5 Odchytávanie komunikácie z Edge smerovača HQ-VCE.

IV. ZÁVER

Kvalitná a bezpečná sieť je pre firmy základom komunikácie. Pri navrhovaní siete je niekoľko aspektov, ktoré je potrebné brať do úvahy. Medzi nich patria požiadavky prevádzky siete, zabezpečenie, segmentácia, škálovateľnosť a náklady na zariadenia či typ internetového pripojenia. Navrhli a nakonfigurovali sme 4 topológie pre 4-pobočkovú fiktívnu spoločnosť v prostrediach GNS3 a EVE-NG, ktoré sú založené na trojúrovňovom hierarchickom modeli. V topológiách sa nachádzajú smerovače, prepínače tretej a druhej vrstvy a koncové zariadenia pre testovanie siete. Keďže prihliadame na finančný aspekt riešení, snažili sme sa využiť čo najmenší potrebný počet sieťových zariadení a vyhnúť sa hardvéru ako firewall a bezpečnosť vyriešiť konfiguračne. Pri finančnom aspekte taktiež predpokladáme, že si spoločnosť ako typ pripojenia zvolí klasickú internetovú linku, namiesto napríklad MPLS pripojenia, ktoré je nákladné.

Na smerovačoch sme sa venovali smerovaniu či už do internetu, ale aj medzi jednotlivými pobočkami a do LAN siete. Pripojenie do internetu sme riešili predvolenou trasou smerom k ISP smerovaču, no medzi pobočkami pomocou dynamického smerovacieho protokolu BGP s využitím bezpečnostného protokolu IPsec po odhalení bezpečnostných nedostatkov. Pri komunikácii s LAN sieťou sme zvolili protokol OSPF, nakoľko je štandardom a môže sa využiť, aj keď nepracujeme výlučne s Cisco zariadeniami, čo sa nám osvedčilo pri štvrtej topológii s využitím SD-WAN smerovačov. Z hľadiska funkčnosti siete sme sa venovali smerovaniu na prepínačoch tretej vrstvy pre spojenie so smerovačmi a oznamovanie lokálnych podsietí. Vďaka VLAN sme rozdelili siete jednotlivých pobočiek na podsiete pre návštevníkov, manažmentovú podsieť a dve podsiete pre zamestnancov. Pri redundantnej topológii sme sa už museli venovať aj protokolom Spanning-Tree kvôli vzniknutým slučkám a pre rozdelenie záťaže siete a taktiež aj redundantnému protokolu HSRP. Z bezpečnostného hľadiska sme lokálnu sieť zabezpečili pomocou Port Security, access-listov na blokovanie komunikácie medzi určitými podsieťami a zabezpečenia vzdialeného prístupu na sieťové zariadenia.

Po nakonfigurovaní jednotlivých protokolov na sieťových zariadeniach sme topológie otestovali pripojením koncových zariadení, ktorým sme dynamicky pridelili IP adresy pomocou DHCP a jednoduchým pingom medzi zariadeniami a na internet, kde sme si overili správnu funkčnosť siete aj aplikácie bezpečnosti. Na overenie zabezpečenej komunikácie medzi jednotlivými pobočkami sme použili program Wireshark, pomocou ktorého sme si odsimulovali útok zo stredu odchytávaním paketov a dokázali, že prepojenie pobočiek pomocou BGP ponad IPsec je bezpečné, keďže útočník má odchytené dáta šifrované. Pri redundantnej topológii sme simulovali možné výpadky liniek a zariadení a potvrdili správnu funkčnosť siete cez záložné sieťové zariadenia.

Pri poslednej sieťovej topológii sme odbočili od zaužívaných Cisco zariadení a pozreli sme sa na SD-WAN. Aj spoločnosť Cisco vyvinulo vlastné SD-WAN zariadenia, no pre úplné odbočenie od spoločnosti Cisco sme si zvolili Velocloud SD-WAN od spoločnosti VMware. Mohli sme si tak vyskúšať a porovnať konfiguráciu Cisco zariadení a Velocloud SD-WAN, ktorá prebieha výlučne v grafickom rozhraní v orchestrátore. Vytvorili sme si zákazníka, konfigurácie jednotlivých smerovačov a profilov. Jednoduchou aktiváciou si SD-WAN smerovače konfiguráciu stiahli z orchestrátora, čím sme sa vyhli aplikácii konfigurácie riadok po riadku cez príkazový riadok. Využili sme takisto BGP ponad IPsec medzi pobočkami a overili úspešné šifrovanie odchytených prenášaných dát a smerom k LAN sieti protokol OSPF. Pomocou monitorovacej časti orchestrátora sme si mohli overiť stav internetovej linky v reálnom čase, ale aj typy komunikácií, ktoré smerovačom prechádzali, kde sme mohli prehľadne vidieť ICMP protokol, ktorý sme použili pri testovaní pripojenia na internet pomocou pingu a medzi pobočkami a taktiež FTP protokol pri nahrávaní konfigurácie na firemný server.

POĎAKOVANIE

Táto práca bola podporovaná nasledujúcimi výskumnými projektami: APVV-22-0400 – „Rozšírenie autonómnych aplikácií monitorovania, prostredníctvom viacpásmových UWB senzorov“ a VEGA 1/0260/23 – „Umelá inteligencia, robustné MANET a viacpreskokové D2D siete integrované do siete 6.generácie“.

LITERATÚRA

[1] 2024 Cisco Systems, Inc.: What is Network Topology? [online]. [cit. 2024-5-18]. Dostupné na internete: <https://www.cisco.com/c/en/us/solutions/automation/network-topology.html>.
[2] 2024 study-ccna.com.: Cisco three-layer hierarchical model [online]. [cit. 2024-5-18]. Dostupné na internete: <https://study-ccna.com/cisco-three-layer-hierarchical-model/>.
[3] 2024 Pearson Education, Cisco Press: Cisco Networking Academy Connecting Networks Companion Guide: Hierarchical Network Design [online]. [cit. 2024-5-19]. Dostupné na internete: <https://www.ciscopress.com/articles/article.asp?p=2202410&seqNum=4>.
[4] 2024 Hewlett Packard Enterprise Development LP: What is SD-WAN? [online]. [cit. 2024-5-21]. Dostupné na internete: https://www.arubanetworks.com/faq/what-is-sd-wan/.
[5] 2024 Cisco Systems, Inc.: What is SD-WAN? [online]. [cit. 2024-5-22]. Dostupné na internete: https://www.cisco.com/c/en/us/solutions/enterprise-networks/sd-wan/what-is-sd-wan.html.
[6] 2024 RSM US LLP.: 5 benefits of SD-WAN [online]. [cit. 2024-3-24]. Dostupné na internete: https://rsmus.com/insights/technology/cisco/5-benefits-of-sd-wan.html.
[7] 2024 IDG Communications, Inc.: The 3 types of SD-WAN architecture [online]. [cit. 2024-3-24]. Dostupné na internete: https://www.networkworld.com/article/964179/the-3-types-of-sd-wan-architecture.html.

Monitorovanie poveternostných podmienok pre FSO systémy s využitím drona

[1]*Ľuboš OVSENÍK,* [2]*Jakub ORAVEC,* [3]*David LECHNER*

[1,2,3]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]lubos.ovsenik@tuke.sk, [2]jakub.oravec@tuke.sk, [3]david.lechner@student.tuke.sk

***Abstrakt* — Príspevok sa zaoberá návrhom senzorového systému kompatibilného s komerčne dostupnými technológiami dronov a UAV (Unmanned Aerial Vehicle). Navrhnutý senzorový systém na základe teoretických znalostí z oblasti bezdrôtových prenosových systémov FSO (Free Space Optics) a RF (Radio Frequency) a odolnosti voči meteorologickým vplyvom, bude schopný efektívne merať viaceré veličiny a spracovávať namerané hodnoty na zjednodušenie ich analýzy.**

***Kľúčové slová* — analýza počasia, dron, FSO, optika voľného priestoru, RF, rádiofrekvenčný systém, senzorový systém, UAV**

I. Úvod

V súčasnom svete, kde rýchlosť a efektivita prenosu informácií definujú technologický pokrok, zohrávajú kľúčovú rolu inovatívne prístupy k telekomunikačným systémom. Jedným z perspektívnych smerov sú systémy FSO, ktoré umožňujú bezdrôtový prenos dát využitím optického signálu ako nositeľa informácií a voľného priestoru ako prenosového média. Tieto systémy ponúkajú vysokú prenosovú rýchlosť a sú vhodnou alternatívou v urbanizovaných alebo ťažko prístupných oblastiach, kde je inštalácia tradičných káblových sietí nákladná alebo technicky náročná. Napriek ich výhodám však efektívnosť FSO systémov výrazne ovplyvňujú poveternostné podmienky, ako sú hmla, dážď, alebo sneženie, ktoré môžu významne znižovať ich výkon a spoľahlivosť. V tejto súvislosti sa ponúka unikátna príležitosť pre využitie bezpilotných lietadiel, UAV a dronov, ktoré môžu byť nasadené na monitorovanie a optimalizáciu prenosových trás FSO systémov.

Drony, vybavené senzormi na meranie meteorologických údajov, môžu poskytnúť cenné informácie o aktuálnom stave prostredia a prispieť k dynamickému prispôsobeniu systémov tak, aby sa minimalizovali vplyvy nepriaznivého počasia. Využitie dronov pre tieto účely môže zefektívniť nasadenie FSO technológií, zvýšiť ich spoľahlivosť a otvoriť dvere pre ich širšie komerčné využitie. Výsledky tejto práce by mohli mať pozitívny dopad nielen na teoretické poznatky v oblasti telekomunikácií, ale aj na praktickú aplikáciu a automatizáciu týchto úloh [1-3].

II. Návrh senzorového systému využívajúci dron

Senzorový systém (SS) alebo bezdrôtové senzorové siete (WSN, Wireless Sensor Network) sú skupiny priestorovo rozptýlených senzorov na monitorovanie zmien prostredia vo vymedzenej oblasti. Senzorový systém sa skladá z jednotlivých uzlov, ktoré sa môžu pripojiť k rôznym snímačom a zhromažďovať údaje z prostredia o zmenách teploty, zvukoch, úrovni znečistenia, vlhkosti, vetre atď. Senzorové uzly komunikujú s ostatnými blízkymi uzlami v sieti, aby spoločne uľahčili prenos údajov zo senzorov cez sieť do hlavného uzla, kde s nimi ďalej pracujú operátori ktorými môžu byť osoby, alebo automatizovaný systém [4].

Zo získaných teoretických poznatkov jednoznačne vyplýva potenciál pre použitie spomínaných technológií Unmanned Aerial Vehicle (UAV) na vytvorenie senzorového systému pre meranie a optimalizáciu umiestnenia technológie Free Space Optics (FSO). Takýto senzorový systém spadá do kategórie mobilných bezdrôtových senzorových sietí, ktoré pozostávajú zo skupiny senzorových uzlov, ktoré sa môžu samostatne pohybovať, snímať fyzické prostredie a komunikovať. Mobilné bezdrôtové senzorové siete sú omnoho univerzálnejšie ako statické senzorové siete vzhľadom na ich mobilitu. Medzi výhody mobilných WSN v porovnaní so

statickými sieťami WSN patrí lepšie a kvalitnejšie pokrytie monitorovaného priestoru, vyššia kapacita kanálov a lepšia energetická účinnosť, ktoré plynú práve z ich mobility [5]. Informácie, ktoré mobilný uzol získa, môžu byť distribuované k ďalším mobilným uzlom vtedy, ak sa nachádzajú v ich dosahu. Sú využívané najmä pri živelných katastrofách, ako je napríklad tsunami, povodne, zemetrasenia, výbuchy sopiek alebo vojenské zásahy. Ich funkcionalita zahŕňa lokalizáciu, samo-organizáciu, kontrolu a navigáciu [4,5].

A. Obmedzenia a legislatívne požiadavky návrhu

Obmedzenia vyplývajú z fyzikálnej podstaty senzorového systému, ale taktiež z príslušnej legislatívy:

- **Výber senzorov** - Na meranie samotných dát je potrebné vybrať vhodné senzory. Bežne používané senzory v meteorológii zahŕňajú množstvo zariadení na meranie teploty vzduchu, atmosférického tlaku, vlhkosti vzduchu, rýchlosti vetra a senzory na meranie zrážok. Avšak ak by sme chceli získavať všetky tieto dáta pomocou drona, senzory môžu byť ťažké a veľké, to by vyžadovalo stále väčšieho drona schopného niesť všetky senzory, čím by sa zvýšila cena. Senzory by taktiež mali byť vybrané na základe ich presnosti, odolnosti voči poveternostným podmienkam, cieľu využitia a kompatibility s dronom.
- **Energetický manažment** - Drony majú obmedzenú výdrž batérie, čo môže obmedziť čas a rozsah meraní. Pri návrhu by sa mal klásť dôraz na optimalizáciu spotreby energie senzorov a drona, aby sa maximalizoval čas letu a efektivita meraní. Efektívne a ľahké systémy sú ideálne pre dlhodobejšie merania, avšak môžu trpieť nižšou presnosťou a menším rozsahom detekcie, zatiaľ čo systémy s maximálnym výkonom sú vhodnejšie pre krátke veľmi presné merania s potrebou detailných a čo najpresnejších dát.
- **Integrácia systému** - Senzory musia byť efektívne integrované s ostatnými systémami drona, ideálne ak sa senzory dokážu integrovať a nahrádzať. Príkladom môže byť odstránenie samostatného GPS modulu a využitie GPS dát zo systému drona, ktorý tieto údaje už integrovane zbiera. Je taktiež dôležité zabezpečiť, aby senzory neovplyvňovali aerodynamické vlastnosti drona.

Veľkými obmedzeniami dokáže byť aj legislatíva, ktorá určuje čo môže bezpilotný letecký systém vykonávať. Začiatkom roka 2024 vstúpili do platnosti nové pravidlá v súlade s nariadením Európskej únie, ktorá vyžaduje aby každý prevádzkovateľ UAS (Unmanned Aerial System) bol registrovaný na Dopravnom úrade SR, mal uzavreté platné poistenie na prevádzku drona [6].

Pôvodným cieľom tohto senzorového systému bolo vyhnúť sa týmto komplikáciám a preto bol zvolený dron Mavic Mini 2 ktorý svojou váhou 249g spadá do triedy C0 do 250g, avšak použitím senzorového systému spolu s dronom samotným váha prekročila tento limit, čím bolo nutné riadiť sa podmienkami kategórie A1/A3, ktorá si vyžadovala dodatočné opatrenia, ktoré bolo pred prevádzkou navrhnutého senzorového systému splniť:

- **Registrácia** na Dopravnom úrade SR a priradenie registračného čísla.
- **Absolvovanie výcvikového kurzu** na kategóriu UAS A1/A3 a získanie Európskeho certifikátu, nazývaný taktiež hovorovo "vodičák na drona".

B. Hardvérové komponenty senzorového systému

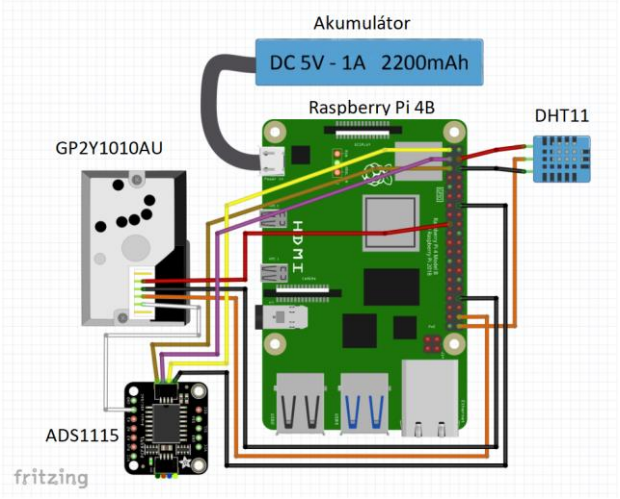
Funkčná verzia senzorového systému použitá na analýzu prostredia a zber výsledných dát sa skladá z dvoch základných častí:

- **Prvou je dron Mavick Mini 2** - Tento dron zahŕňa komponenty ako jeho napájacia batéria, vstavaný komunikačný modul obsahujúci GPS vysielač získavajúci dáta o aktuálnej polohe a je tiež používaný na stabilizáciu letu, prijímač a Wi-Fi modul určený na komunikáciu s operátorom.
- **Druhou časťou systému je senzorový systém** – Tento je umiestnený na ľahkej a vyváženej platforme. SS (Obr. 1) zbiera meteorologické údaje o vlhkosti, teplote a prašnosti monitorovaného prostredia.

Senzorový systém sa skladá z nasledovných komponentov:

- **Raspberry Pi** - Základ tejto platformy tvorí mikropočítač používajúci ARM systémové inštrukcie s názvom Raspberry Pi vo verzii 4B. Tento mikropočítač využíva operačný systém Raspbian OS založený na kerneli Linux. Má vstavený Wi-Fi a Bluetooth modul, ktorý sa používa na synchronizáciu dát drona a senzorového systému. Pomocou Wi-Fi vieme zariadenie používať aj v režime vzdialeného prístupu komfortne z nášho preferovaného zariadenia pomocou vstavaného protokolu VNC. Mikropočítač má k dispozícii rozhranie I2C na komunikáciu s analógovými senzormi, ktoré používa na komunikáciu s prevodníkom. Tento systém napája všetky ostatné senzory zobrazenej platformy pomocou 5V a 3.3V pinov. Všetky výpočty,

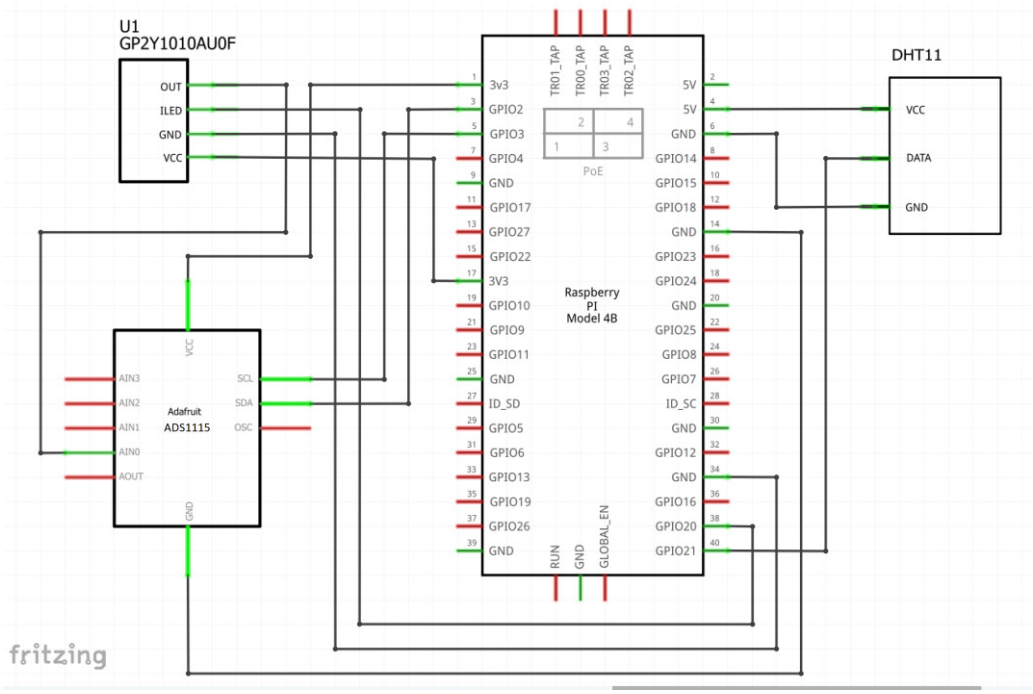
spracovanie a ukladanie dát v navrhnutom riešení sa vykonáva práve na Raspberry Pi, vrátane následného spracovania nazbieraných dát a ich vizualizácie.



Obr. 1 Vizualizácia výsledného senzorového systému.

• **Akumulátor** - Samotný mikropočítač je napájaný 2200mAh externým akumulátorom dodávajúcim prúd o veľkosti 1A pri napätí 5V. Na prepojenie je využitý kábel s USB-A portom zapojeným do akumulátora a USB-C portom do mikropočítača Raspberry Pi. Pri použití tejto batérie s takýmto prúdom je síce v operačnom systéme Raspbian zobrazené upozornenie na slabé napájanie, avšak tento fakt nemá vplyv na prevádzku alebo výsledky merania. Kapacita tohto akumulátora je plne dostačujúca na potreby zberu dát počas letu platformy, keďže kapacita a dolet drona samotného neprekračuje výdrž tohto akumulátora. Jeho nízka váha a štvorcový dizajn, ktorý poskytuje ľahké uchytenie boli dôvodmi pre jeho použitie.

• **DHT11** – Je senzor na meranie teploty a vlhkosti vzduchu. Ide o digitálny senzor, ktorý v sebe kombinuje termistor pre meranie teploty a kapacitný senzor pre meranie vlhkosti. Je relatívne lacný, ľahký, nevyžaduje dodatočné komponenty na priamu prevádzku s Raspberry Pi a čítanie jeho dát je veľmi jednoduché. Ako vidieť na schéme zapojenia senzorového systému (Obr. 2), je senzor pripojený na 5V pomocou pinu 4, uzemnený na pin 6 a Raspberry Pi z neho číta dáta pomocou pripojeného pinu 40. Senzor DHT11 dokáže bežne merať hodnoty teploty v rozsahu 0°C až 50°C a hodnoty vlhkosti od 20% do 80%.



Obr. 2 Schéma zapojenia výsledného senzorového systému.

• **GP2Y1010AU** – Je optický senzor prachových častíc, ktorý deteguje jemné prachové častice vo vzduchu. Využíva metódu rozptylu svetla, kde častice vo vzduchu rozptyľujú svetlo z LED diódy do fototranzistora, čím je možné odhadnúť množstvo častíc v ovzduší. Tento senzor je široko používaný a ľahký v porovnaní s konkurenčnými produktami. Jeho presnosť merania je horšia ako u iných senzorov, avšak je postačujúci na účely navrhnutého riešenia. V tomto prípade senzor meria častice PM10, ktoré sú merané v rozsahu 0.5V na 0.1 mg/m^3. Ako vidieť na schéme (Obr. 2), senzor je napájaný pomocou 3.3V z Raspberry Pi na pine 17, uzemnený na pine 34. Raspberry Pi ovláda jeho infračervenú LED na meranie pomocou pinu 38 a nakoniec odosiela svoje namerané hodnoty analógovým výstupom do ADC prevodníka.

• **ADS1115** - Tento 16 bitový prevodník analógového signálu na digitálny, je komponent potrebný na správne fungovanie senzora častíc GP2Y1010AU spolu s Raspberry Pi. Funguje na rovnakom napätí 3.3V, ako senzor častíc napojený na Raspberry pin 1, uzemnený na pin 14, číta analógový signál zo senzora častíc pomocou AIN0 a prevedené hodnoty komunikuje s Raspberry Pi pomocou pinov 3 a 5 fungujúcich na rozhraní I2C, ako môžeme vidieť na schéme Obr. 2.

III. VIZUÁLNA ANALÝZA SPRACOVANÝCH DÁT

Softvérové riešenie senzorového systému je založené na programovom systéme *process.py*, ktorý generuje KML súbor zo spracovaných senzorových informácií. Tento je vhodný pre vizualizáciu geografických dát v aplikáciách ako Google Earth, čím umožňuje intuitívnejšie pochopenie geopriestorových vzťahov a vzorcov v dátach. KML formát poskytuje bohaté možnosti pre zobrazenie a analýzu geopriestorových informácií v graficky príťažlivom a intuitívnom formáte (Obr. 3).



Obr. 3 Ukážka spracovaných dát letovej trasy.

Ako vidieť na obrázka (Obr. 3), táto vizualizácia zobrazuje body získané z drona, ktoré sú načítané do Google Earth. Každý bod označený ako bod s číselným identifikátorom reprezentuje špecifickú polohu s priradenými dátami, ktorými sú všetky spracované dáta z drona aj senzorového systému, ako sú geografické súradnice, čas, teplota, vlhkosť a údaje o prachových časticiach. Všetky tieto informácie sú dostupné na jednom mieste, body sa dajú jednoducho selektovať kliknutím na nich alebo výberom v zozname viditeľným naľavo. Po kliknutí na ľubovoľný bod sú všetky dáta asociované s daným bodom okamžite dostupné a viditeľné v okne na pravej strane obrazovky. KML súborov je možné nahrať ľubovoľný počet a dokážu byť individuálne skrývané a zobrazované na vytvorenie setu dát podľa potreby analýzy. KML súbory je možné v aplikácii Google Maps taktiež spájať do projektov, čím dokážeme dáta ešte lepšie oddeliť a spravovať podľa potrieb používateľa.

IV. OPTIMALIZÁCIA UMIESTNENIA FSO

Myšlienkou a dôvodom na praktické zostrojenie navrhnutého riešenia bolo reálne využitie na optimalizáciu prenosovej trasy FSO systémov v priestore (teréne). Ako meranú lokalitu sme sa rozhodli otestovať lokalitu novostavieb Zelená stráň. Dôvodom boli ideálne geografické podmienky, kde sa novostavby nachádzajú na okraji sídliska v blízkosti nákupného centra. Túto lokalitu predstavuje hornú vrchol trojuholníka viditeľný na Obr. 4. Pravý bod trojuholníka reprezentuje mestská časť Košická Nová Ves. Konkrétne bod tvorí veža miestneho kostola nachádzajúca sa v strednej časti obce. Nakoniec posledný bod na Obr. 4 tvoria novostavby

Heringeš, ktoré sú úplne oddelené od zvyšných budov a potrebujú teda na pripojenie samostatnú sieť.

Medzi všetkými týmito troma bodmi existuje prázdny vzdušný priestor s priamym vizuálnym kontaktom, ideálny na zavedenie FSO systému. Spôsob akým sú tieto lokality oddelené vytvára ideálnu príležitosť na implementáciu prenosu dát bezdrôtovým spôsobom bez naťahovania zbytočných káblov a na uspokojenie dopytu po pripojení, ktorý sa v novostavbách dá očakávať.

Vzhľadom na blízkosť tejto oblasti zhruba 200m od frekventovaného dopravného ťahu na jednej strane a obce vykurujúcej tuhým palivom na strane druhej, vieme predpokladať občasnú meteorologickú situáciu, ktorá by narušala našu prenosovú trasu prenosového FSO systému. Vzdialenosť samotných potenciálnych bodov na umiestnenie FSO hlavíc je viac ako 500m, čo vzhľadom na vlastnosti a váhovú kategóriu tohto drona používaného na zber dát predstavuje ideálny čas na odmeranie celej prenosovej trasy počas jedného letu.

Na optimalizáciu umiestnenia FSO hlavíc boli teda vykonané merania spôsobom každodenného merania po celej dĺžke potenciálnej prenosovej trasy počas doby jedného týždňa, pričom snaha bola o zachovanie približne rovnakého času merania, aby hodnoty odrážali čo najpresnejšie porovnania znečistenia počas dní v týždni. Výsledné hodnoty by mali ukázať, či týždňový medián prekračuje optimálnu hodnotu kvality ovzdušia a nedegraduje kvalitu nášho optického prenosu. Toto meranie prebehlo v aktívnom vykurovacom období počas dátumov 25.1.2024 - 30.1.2024, kedy sa očakával vplyv vykurovania tuhým palivom na kvalitu ovzdušia.



Obr. 4 Návrh umiestnenia FSO linky a jej trajektória.

V. Vyhodnotenie nameraných dát

Merania prebehli v stanovených dátumoch, samotný proces zberu dát bol bezproblémový a vzorku finálnych spojených dát pripravených programom *process.py* môžeme vidieť v Tabuľke 1.

Tabuľka 1
Ukážka spojených finálnych dát

UnixDron	Unix	Temperature	Humidity	Dust Density [microg/m3]
1706269215	1706269215.0	7.0	55.0	13.2
1706269216	1706269216.0	7.0	54.0	13.5
1706269216	1706269216.0	7.0	54.0	13.5
1706269216	1706269216.0	7.0	54.0	13.5
1706269216	1706269216.0	7.0	54.0	13.5

Ak by sme chceli zhodnotiť namerané dáta použitím vytvorenej vizuálnej analýzy, tak ako vidieť na Obr. 3 môžeme okamžite vyčítať, že vo väčšine bodov uskutočneného merania sú body zelené, čo predstavuje kvalitu ovzdušia v prijateľných hodnotách. Tento výsledok naznačuje, že lokalita vyzerá byť stabilný adept na inštaláciu FSO siete a jej efektivita by nebola pravidelne narušená. Samozrejme pre získanie presnejších dát by bolo potrebné získať väčšiu vzorku dát buď meraním v dlhšom období alebo zvýšením senzitivity merania pomocou zvýšenia počtu meracích letov za deň.

Avšak ak sa pozorne pozrieme na niektoré body vo vizuálnej analýze na Obr. 5, môžeme vidieť niekoľko žltých bodov indikujúcich mierne zvýšenú koncentráciu častíc v ovzduší na danom mieste. Na tomto obrázku môžeme v tabuľke na pravej strane vidieť konkrétnu hodnotu ktorú

tento žltý bod reprezentuje a to 62.4 μg/m^3.



Obr. 5 Znázornenie zvýšenej hodnoty koncentrácie častíc v ovzduší.

Pri hlbšom pohľade na Tabuľku 2, ktorá zobrazuje najvyššie namerané hodnoty za obdobie celého merania môžeme vidieť, že práve tento rovnaký čas odpovedá danému meraniu. Časovou podobnosťou tomuto jednorázovému nárastu meraných hodnôt môžeme dedukovať, že sa jednalo buď o lokálny prašný bod v mieste merania, alebo drobnú chybu meracieho prístroja, ktorým sa spôsobila dočasná odchýlka.

Tabuľka 2
Najvyššie hraničné hodnoty týždenného merania

UnixDron	Unix	Temperature	Humidity	Dust Density [microg/m3]
1706187733	1706187733.0	2.0	84.0	62.4
1706187733	1706187733.0	2.0	84.0	62.4
1706187733	1706187733.0	2.0	84.0	62.4
1706187733	1706187733.0	2.0	84.0	62.4
1706187935	1706187935.0	2.0	81.0	45.6

VI. Záver

V súčasnosti sa zvyšuje dopyt po pripojení na internet za pomoci bezdrôtových komunikačných technológií. Vplyvom rastúceho dopytu po kvalitnom internetovom pripojení sa poskytovatelia internetu snažia pokryť čoraz väčšie územie s čo najväčšou prenosovou rýchlosťou. V tejto práci bol navrhnutý mapovací systém poveternostných podmienok pre vhodné umiestnenie FSO liniek v skúmanom priestore s využitím UAV. Tento systém obsahoval kompaktný senzorový systém na meranie poveternostných podmienok a dron ako platformu pre upevnenie tohto systému. Tento systém umožňuje získať predstavu o kvalite ovzdušia v skúmanom prostredí pre umiestnenie FSO komunikačnej linky.

Poďakovanie

Táto práca bola podporovaná nasledujúcimi výskumnými projektami: APVV-22-0400 – „Rozšírenie autonómnych aplikácií monitorovania, prostredníctvom viacpásmových UWB senzorov“ a VEGA 1/0260/23 – „Umelá inteligencia, robustné MANET a viacpreskokové D2D siete integrované do siete 6.generácie“.

Literatúra

[1] CISCO PRESS. Wireless System Architecture: How Wireless Works. 2023. Dostupné na: https://www.ciscopress.com/articles/article.asp?p=344242.
[2] ELECTRONICS HUB. Wireless Communication: Introduction, Types and Applications. 2023. Dostupné na: https://www.electronicshub.org/wirele ss-communication-introduction-types-applications/.
[3] COMMUNICATIONS, FiberWorkOptical. Free SpaceOptics Products. 2023. Dostupné na: http://www.fiberwork.net/products/fso.html.
[4] WRITER, Staff. Sensor Networks: The Advantages and Disadvantages You Need To Know. 2019. Dostupné na: https://www.totalphase.com/blog/2019/04/sensor-networks-the-advantages-and-disadvantagesyou-need-to-know/.
[5] TEAM, Tesca. What is Wireless Sensor Network, and Types of WSN? 2021. Dostupné na: https://www.tescaglobal.com/blog/what-is-wireless-sensor-network-and-types-of-wsn/.
[6] SR, Dopravný úrad. Registrácia prevádzkovateľov UAS. 2023. Dostupné na: http://letectvo.nsat.sk/bezpilotne-letectvo/registracia-prevadzkovatelov-uas/.

Skúmanie vlastností optických filtrov v DWDM prenosovom systéme

[1]*Ľuboš OVSENÍK,* [2]*Jakub ORAVEC,* [3]*Slavomír ŠTOSEL*

[1,2,3]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

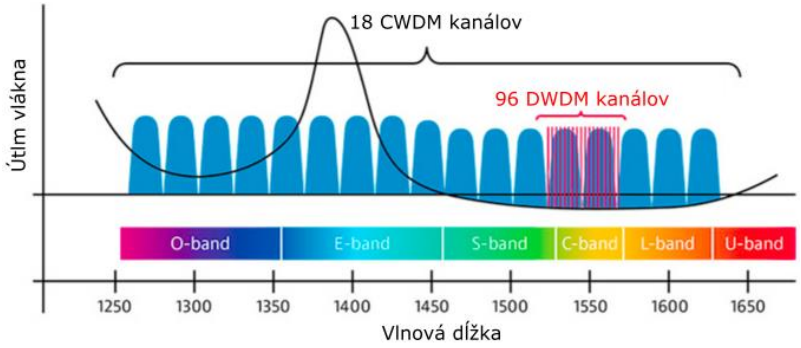
[1]lubos.ovsenik@tuke.sk, [2]jakub.oravec@tuke.sk, [3]slavomir.stosel@student.tuke.sk

***Abstrakt* — Príspevok sa zaoberá problematikou skúmania vplyvu optických filtrov vo WDM (Wavelength Division Multiplexing) prenosových systémoch. Zameriavame sa na rôzne nastavenia a vlastnosti optických filtrov pri WDM prenosových systémoch rôznej vzdialenosti. Optické filtre slúžia na rozdelenie multiplexovaných vlnových dĺžok. Posúdime FBG (Fiber Bragg Grating) a FP (Fabry-Perot) filtre, ktoré sú založené na rôznych princípoch a majú rôzne vlastnosti. V simulačnej časti práce sa zameriame na simulácie s demultiplexorom, FBG a FP filtrom pri vzdialenostiach komunikačného kanálu 50 a 80 km s prenosovou rýchlosťou 10 Gbit/s.**

***Kľúčové slová* — demultiplexor, Fabry-Perot filter, FBG filter, multilexor, optické filtre, WDM**

I. Úvod

V súčasnej ére neustáleho rozvoja telekomunikačných technológií sa optické siete stávajú kľúčovým pilierom pre efektívny prenos obrovských dátových tokov. V rámci týchto sietí zohrávajú optické filtre dôležitú úlohu pri spracovaní a riadení optických signálov podľa ich vlnovej dĺžky. V kontexte multiplexovania vlnovej dĺžky (WDM) sa vlastnosti a účinnosť filtrov stávajú kritickými pre dosiahnutie spoľahlivého a vysokovýkonného prenosu optických signálov [1,2,3]. Vlnové multiplexovanie WDM 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 Wavelenght 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.



Obr. 1 Kanálové porovnanie CWDM a DWDM.

DWDM je multiplexovanie s hustým delením vlnovej dĺžky a jeho vlnové dĺžky sa nachádzajú v úzkej oblasti 1525 nm až 1565 nm známej ako pásmo C a rozširujú sa na spektrum 1570 - 1610 nm pásma L. DWDM má rozstup vlnových dĺžok 0,8; 0,4; 0,2 alebo 0,1 nm. Vďaka úzkemu rozstupu vlnových dĺžok DWDM prenáša 40, 80, 96 alebo až 160 vlnových dĺžok s využitím oveľa užšieho rozstupu (Obr. 1). To umožňuje systémom DWDM prenášať obrovské množstvo údajov cez jedno vláknové spojenie, pretože umožňujú, aby sa na to isté vlákno zmestilo oveľa viac vlnových dĺžok [2]. DWDM je optimálny na komunikáciu na veľké vzdialenosti až do

120 km a viac vďaka schopnosti využívať optické zosilňovače, ktoré dokážu cenovo výhodne zosilniť celé spektrum 1550 nm alebo C-pásmo, ktoré sa bežne používa v aplikáciách DWDM. Tým sa prekonávajú dlhé úseky útlmu alebo vzdialenosti a pri zosilnení pomocou zosilňovačov EDFA (Erbium Doped Fiber Amplifier) majú systémy DWDM schopnosť prenášať veľké množstvo údajov na dlhé vzdialenosti siahajúce do stoviek alebo až tisícov kilometrov. Okrem schopnosti podporovať väčší počet vlnových dĺžok ako CWDM sú platformy DWDM schopné spracovať aj protokoly s vyššou rýchlosťou, keďže väčšina dodávateľov optických prenosových zariadení dnes bežne podporuje 100 Gb/s alebo 200 Gb/s na kanál, t.j. na jednu vlnovú dĺžku, zatiaľ čo nové technológie umožňujú prenos 400 Gb/s a viac [3].

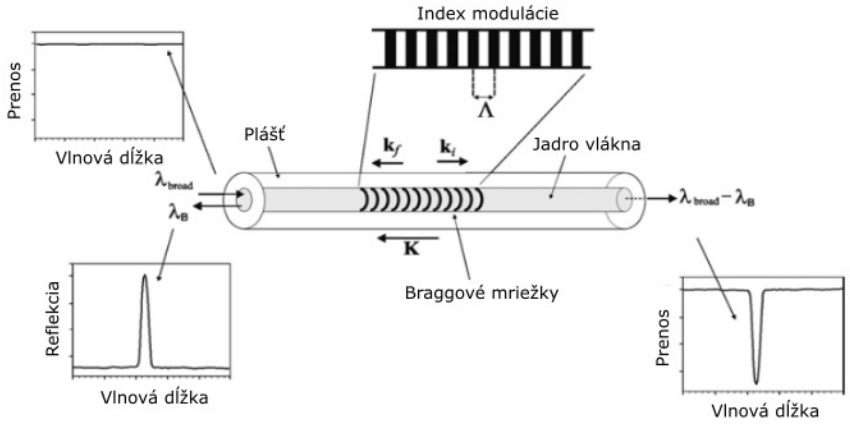
II. Optické filtre

Vlnové filtre v optických prenosových systémoch sú špeciálnou podskupinou fyzikálnych komponentov definovaných tak, že vyberajú alebo modifikujú časti spektra signálu. Optické vlnové filtre sú v skutočnosti definované vzhľadom na modifikácie, ktoré vyvolávajú vo frekvenčnom spektre. Naše úvahy obmedzíme na použitie optických filtrov v optických sieťach.

A. Vláknové Braggove mriežky (FBG)

Vláknová Braggova mriežka (Fiber Bragg Grating) je dôležitou súčasťou technológie optických vlákien, ktorá zohráva kľúčovú úlohu v rôznych aplikáciách, ako sú telekomunikácie, snímanie a spracovanie optických signálov. FBG sú v podstate periodické štruktúry v jadre optického vlákna, ktoré vytvárajú odraz alebo prenos svetla podľa vlnovej dĺžky (Obr. 2). Tieto mriežky sa vytvárajú vyvolaním periodickej zmeny indexu lomu jadra vlákna, zvyčajne pomocou pôsobenia ultrafialového (UV) svetla. Periodicita zmeny indexu pôsobí ako filter špecifický pre vlnovú dĺžku, ktorý umožňuje, aby sa určité vlnové dĺžky svetla odrážali späť, zatiaľ čo iné prepúšťa. Táto vlastnosť selektívneho odrazu je založená na princípe Braggovej difrakcie, kde sa dopadajúce svetlo s určitou vlnovou dĺžkou odráža späť v dôsledku konštruktívnej interferencie v štruktúre mriežky.

FBG mriežky sú známe svojou schopnosťou presného riadenia spektrálnych charakteristík svetla, vďaka čomu sú cenné v rôznych optických systémoch. Môžu byť navrhnuté tak, aby mali špecifické stredové vlnové dĺžky, šírky pásma a profily odrazivosti, čo umožňuje prispôsobenie na základe požiadaviek aplikácie. V optických komunikačných systémoch sa FBG používajú na úlohy, ako je smerovanie signálu, kompenzácia rozptylu a filtrovanie vlnovej dĺžky. Používajú sa aj v optických senzoroch na meranie parametrov, ako je teplota, napätie a tlak, a to vďaka ich citlivosti na vonkajšie rušivé vplyvy. FBG celkovo spôsobili revolúciu v oblasti technológie optických vlákien tým, že umožnili efektívnu a všestrannú manipuláciu so svetlom v optických vláknach, čo viedlo k pokroku v telekomunikáciách, snímaní a iných optických aplikáciách [4].



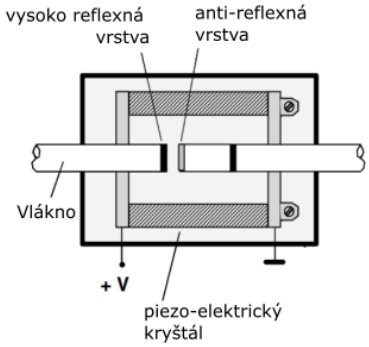
Obr. 2 Ilustrácia Braggovej mriežky s konštantným indexom modulácie amplitúdy a periódy.

B. Fabryho-Perotov (FP) interferenčný filter

Fabryho-Perotov (FP) interferometrický filter využíva na dosiahnutie selektivity vlnovej dĺžky proces interferencie viacerých lúčov. Filter má zvyčajne jeden vstupný a jeden výstupný port a využíva dve vysoko odrazové dosky, ktoré spolu tvoria rezonančnú dutinu vytvárajúcu proces interferencie viacerých lúčov. Prvýkrát ho opísali Charles Fabry a Albert Perot v roku 1899 [5].

Optické vláknové Fabryho-Perotove (FFP) filtre sa ľahko začlenia do optických vláknových komunikačných spojení. Pri tomto type sa nepoužívajú žiadne šošovky ani iné kolimujúce optické prvky. FP dutina je tvorená dvoma starostlivo zarovnanými koncovými plochami vlákien, na

ktorých je nanesená vysoko reflexná vrstva. V prípade väčších vzdialeností medzi zrkadlami sa do dutiny môže vložiť krátky kus jednovidového vlákna s antireflexnou vrstvou, ktorý sa postará o vhodné vlnovo obmedzené vedenie svetla a ktorého koncová strana má antireflexnú vrstvu, aby sa zabránilo vzniku sekundárnych dutín. Základná štruktúra takéhoto Fabryho-Perotovho vláknového filtra je znázornená na Obr. 3.



Obr. 3 Vláknový Fabry-Perot filter.

III. VÝSLEDKY A ZHODNOTENIE SIMULÁCIÍ

Rozhodli sme sa vytvoriť simuláciu pre 16-kanálový DWDM prenosový systém pracujúci pri prenosovej rýchlosti 10 Gb/s/kanál. Na tvorbu simulácie sme využili program OptiSystem od spoločnosti Optiwave. Vďaka komponentom, ktoré nám tento program ponúka sme boli schopní vytvoriť potrebné simulácie na skúmanie vlastností rôznych optických filtrov. Simulácia pozostávala z viacerých krokov: Prvým krokom bolo vytvorenie vysielacieho zariadenia. Druhým krokom bolo pridanie multiplexoru a optického prenosového kanálu. Multiplexor sme využili primárne na spojenie jednotlivých vysielaných dát do jedného optického kanálu. Rovnaký postup sme využili pri každej simulácii. Rozdielna bola len konečná tretia časť, kedy sme museli meniť komponenty, a to hlavne optické filtre. Výsledky sme nakoniec analyzovali a v závere porovnávali.

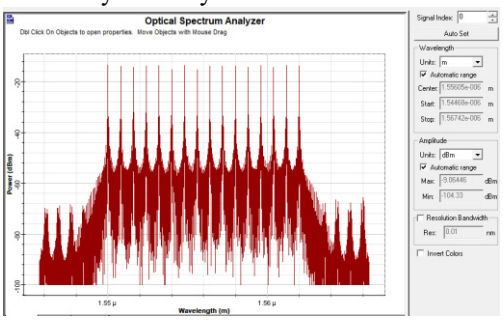
Hlavné kanály, ktoré sme si zvolili na pozorovanie boli prvý s vlnovou dĺžkou 1550 nm, šiesty s vlnovou dĺžkou 1554 nm, desiaty s vlnovou dĺžkou 1558 nm a predposledný teda pätnásty s vlnovou dĺžkou 1561,2 nm. Základné parametre, ktoré sme si zvolili a upravili podľa nich simulácie môžeme vidieť v Tabuľke 1.

Tabuľka 1
Vstupné parametre simulácie

Parametre	Hodnoty
Rýchlosť prenosu	10 Gb/s
Výkon vysielačov	2 mW
Vzdialenosť	50 km
Útlm na optickom vlákne	0,15 dB/km

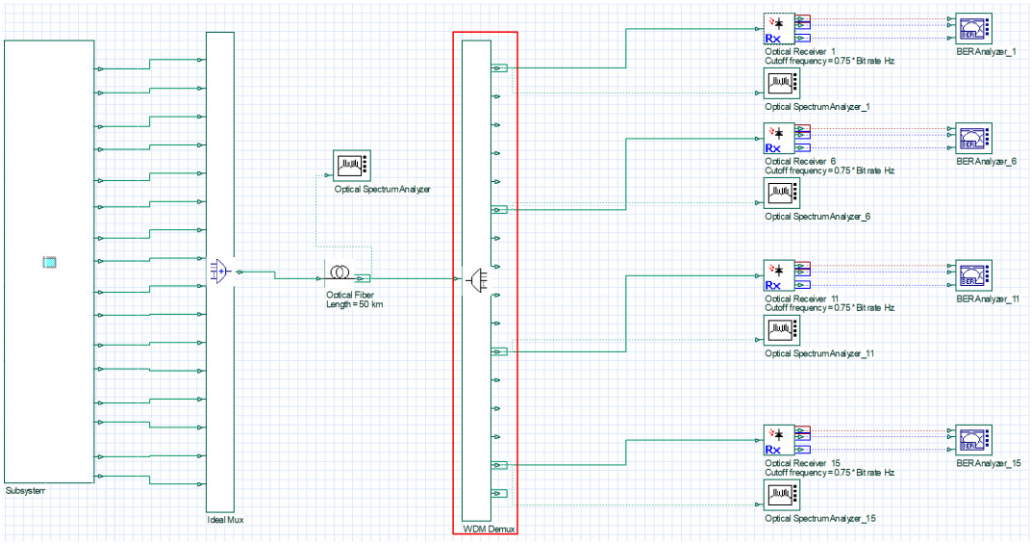
A. Simulácia s demultiplexorom

Na Obr. 5 môžeme vidieť simuláciu, v ktorej sa nachádza demultiplexor. Hlavná časť, v ktorej budeme robiť zmeny, bude posledná, v ktorej demultiplexor nahrádzame vybranými optickými filtrami. Na demultiplexore môžeme vidieť 16 možných výstupov, keďže sme navrhli 16 kanálový systém. Na pozorovanie sme však vybrali len 4 od seba vzdialené kanály, a to 1., 6., 11. a 15., z ktorých budeme zaznamenávať výsledky. Na konci linky vidíme optický prijímač, analyzátor optického spektra a analyzátor chybovosti BER.



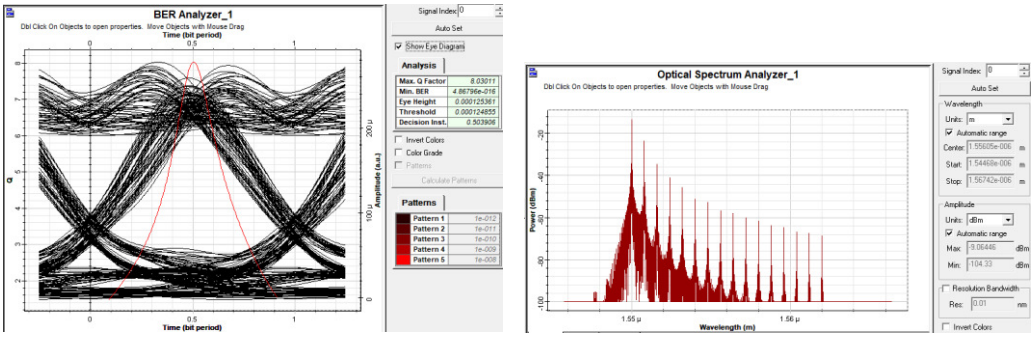
Obr. 4 Analyzátor optického spektra pre simuláciu s demultiplexorom.

Na základe analyzátorov budeme posudzovať efektívnosť navrhnutej simulácie a budeme porovnávať vybrané optické filtre. Optické spektrum simulácie s demultiplexorom je znázornené na Obr. 4. Na obrázku môžeme vidieť pekne rozčlenenie kanálov podľa vlnovej dĺžky.



Obr. 5 Simulačná schéma s demultiplexorom.

Pri vzdialenosti 50 km prvý kanál (vlnová dĺžka 1550 nm) v simulácii s demultiplexorom nemal problém s prenosom dát. Na Obr. 6a je výstup BER analyzátora prvého kanálu. Môžeme vidieť, že BER sa rovnalo približne 4,87e-16. Optické spektrum je znázornené na Obr. 6b. Ako vidíme, tak demultiplexor splnil svoju úlohu a dokázal odfiltrovať žiadaný kanál, teda prvý kanál.



Obr. 6 a) BER analyzátor pre 1. kanál b) Analyzátor optického spektra pre 1. kanál s demultiplexorom.

Pre simuláciu pri hraničných vzdialenostiach sme skúšali rôzne vzdialenosti a hľadali takú vzdialenosť, pri ktorej by bol prenos pre všetky simulácie možný. Zistili sme, že vzdialenosť 80 km vyhovovala pre všetky simulácie. Nižšie v Tabuľke 2a a 2b môžeme vidieť porovnanie výsledkov simulácie s demultiplexorom pri vzdialenosti 50 km a 80 km.

Tabuľka 2
Výsledky simulácie s demultiplexorom pri vzdialenosti a) 50 km a b) 80 km

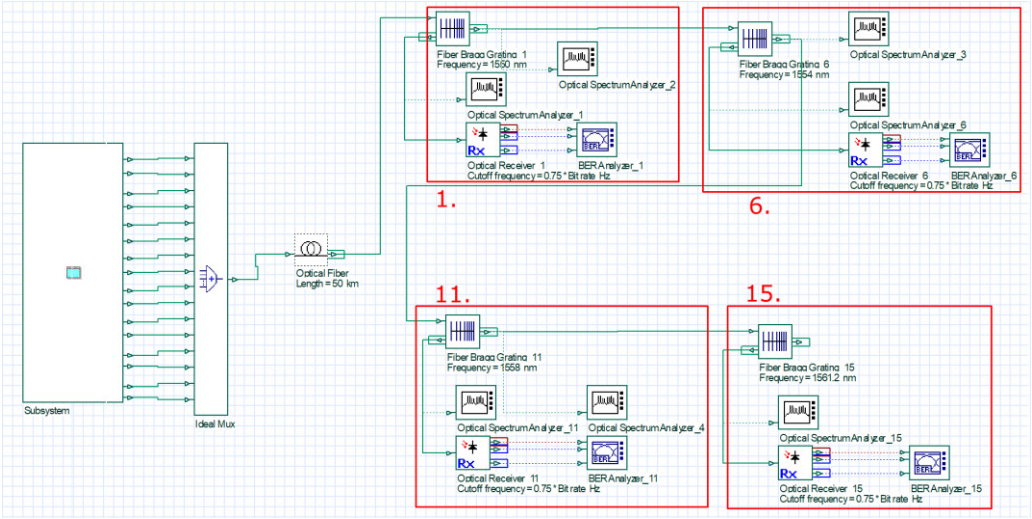
Kanál	Vlnová dĺžka	Max. Q-faktor	Min. BER
1.	1550,0 nm	8,03	4,87e-16
6.	1554,0 nm	5,86	2,31e-9
11.	1558,0 nm	5,61	1,00e-8
15.	1561,2 nm	5,55	1,43e-8

Kanál	Vlnová dĺžka	Max. Q-faktor	Min. BER
1.	1550,0 nm	2,82	2,09e-3
6.	1554,0 nm	2,64	3,34e-3
11.	1558,0 nm	2,71	2,95e-3
15.	1561,2 nm	2,77	2,47e-3

Z výsledkov simulácie je zrejmé, že zvýšenie vzdialenosti z 50 km na 80 km viedlo k zhoršeniu kvality prenosu na všetkých kanáloch, čo naznačujú ako hodnoty maximálneho Q-faktora, tak aj minimálneho BER. Tento úbytok kvality je očakávaný, pretože s väčšou vzdialenosťou sa zvyšuje úroveň šumu a útlmu signálu. Aj napriek tomu, že kvalita prenosu klesla, prenos dát bol stále možný, čo naznačuje, že systém bol schopný vyrovnať sa s väčšou vzdialenosťou. Je tiež pozoruhodné, že rozdiel v kvalite medzi kanálom číslo 1 a ostatnými kanálmi bol menej výrazný pri väčšej vzdialenosti, 80 km, ako pri 50 km. To môže byť spôsobené rozdielnymi úrovňami útlmu a interferencií medzi kanálmi pri rôznych vzdialenostiach.

B. Simulácia s FBG filtrom

Znázornenie navrhnutej simulácie s optickým mriežkovým Bragg filtrom vidíme na Obr. 7. Tak ako pri simulácii s demultiplexorom je aj v tomto prípade vzdialenosť 50 km. V tomto návrhu sú znázornené len 4 filtre na odrazenie požadovaných kanálov, a to 1., 6., 11. a 15. kanál.



Obr. 7 Simulačná schéma s FBG filtrom.

Hraničnú vzdialenosť pre simuláciu s FBG filtrom sme zvolili rovnakú, ako pri simulácii s demultiplexorom. Pre porovnanie sme výsledky pri vzdialenosti 50 km zapísali do Tabuľky 3a a pri vzdialenosti 80 km sú výsledky zapísané v Tabuľke 3b.

Tabuľka 3
Výsledky simulácie s FBG filtrom pri vzdialenosti a) 50 km a b) 80 km

Kanál	Vlnová dĺžka	Max. Q-faktor	Min. BER
1.	1550,0 nm	9,96	1,05e-23
6.	1554,0 nm	10,15	1,65e-24
11.	1558,0 nm	10,19	1,06e-24
15.	1561,2 nm	10,07	3,59e-24

Kanál	Vlnová dĺžka	Max. Q-faktor	Min. BER
1.	1550,0 nm	2,78	2,40e-3
6.	1554,0 nm	2,72	2,79e-3
11.	1558,0 nm	2,58	4,30e-3
15.	1561,2 nm	2,81	2,11e-3

Na základe týchto výsledkov vieme povedať, že väčšia vzdialenosť ovplyvnila simuláciu s FBG filtrom viac ako simuláciu s demultiplexorom.

Jednou z možností ako zlepšiť výkon FBG optického filtru je nastavenie rozostupu medzi Braggovými mriežkami. V navrhnutej simulácií bol rozostup medzi nimi 2 mm. V prípade zväčšenia, napríklad na 10 mm, sme dostali lepšie hodnoty Q-faktora a min. BER pri vzdialenosti 50 km. Výsledky sú zapísané v Tabuľke 4a a 4b. Pri vyššej vzdialenosti však rozdiel vo výsledkoch nebol až taký veľký oproti výsledkom s Braggovou mriežkou nastavenou na 2 mm. Môžeme teda povedať, že hlavnou limitáciou pri WDM systémoch je útlm, ktorý spôsobuje práve vyššia vzdialenosť v komunikačnom kanáli.

Tabuľka 4
Výsledky simulácie s Braggovou mriežkou 10 mm pri vzdialenosti a) 50 km a b) 80 km

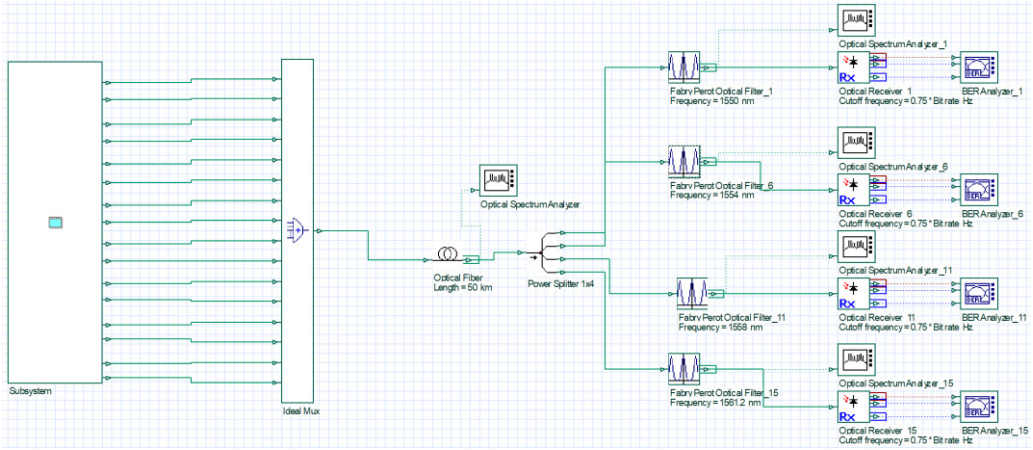
Kanál	Vlnová dĺžka	Max. Q-faktor	Min. BER
1.	1550,0 nm	14,74	1,72e-49
6.	1554,0 nm	14,79	7,62e-50
11.	1558,0 nm	14,48	8,05e-48
15.	1561,2 nm	13,78	1,59e-43

Kanál	Vlnová dĺžka	Max. Q-faktor	Min. BER
1.	1550,0 nm	3,51	2,12e-4
6.	1554,0 nm	3,35	3,77e-4
11.	1558,0 nm	3,44	2,71e-4
15.	1561,2 nm	3,52	2,07e-4

C. Simulácia s FP filtrom

Posledným sledovaným filtrom bol FP (Fabry-Perot) optický filter. V návrhu tejto simulácií sme oproti ostatným simuláciám urobili zmenu. Potrebovali sme použiť rozdeľovač výkonu 1x4 (z angl. power splitter), čo môžeme vidieť aj na Obr. 7. Tento krok sme museli urobiť z toho dôvodu, že FP filter nefunguje ako FBG filter alebo demultiplexor. FP nedokáže odfiltrovať a zároveň pokračovať v prenášaní dát z optického káblu tak, ako FBG filter z predošlej simulácie. Rovnako tak nedokáže fungovať ani ako demultiplexor, ktorému keď nastavíme na jednotlivé výstupy žiadané vlnové dĺžky pre 1., 6., 11., 15. kanál, tak ich podľa toho dokáže na žiadanom

výstupe odfiltrovať. Toto riešenie nám odstránilo problém viacnásobných simulácií. V prípade, ak by sme nepoužili rozdeľovač výkonu, pri každej simulácií museli by sme meniť nastavenia FP filtru na žiadanú vlnovú dĺžku a potom vykonať simuláciu.



Obr. 7 Simulačná schéma s FP filtrom.

Rovnako, ako pri predošlých simuláciách sme a aj pri poslednej simulácii nastavili ako hraničnú vzdialenosť pre komunikačný kanál 80 km. Všetky výsledky sú zapísané v nasledujúcich tabuľkách (Tabuľka 5a a 5b).

Tabuľka 5
Výsledky simulácie s FP filtrom pri vzdialenosti a) 50 km a b) 80 km

Kanál	Vlnová dĺžka	Max. Q-faktor	Min. BER
1.	1550,0 nm	13,30	1,12e-40
6.	1554,0 nm	12,77	1,15e-37
11.	1558,0 nm	12,66	4,36e-37
15.	1561,2 nm	11,82	1,39e-32

Kanál	Vlnová dĺžka	Max. Q-faktor	Min. BER
1.	1550,0 nm	2,98	1,31e-3
6.	1554,0 nm	2,96	1,33e-3
11.	1558,0 nm	3,03	1,09e-3
15.	1561,2 nm	3,27	4,84e-4

Z pomedzi všetkých simulácií pri vzdialenosti 50 km sme dosiahli najlepšie výsledky práve pri simulácií FP filtrom. Na druhú stranu pri vzdialenosti 80 km boli výsledky veľmi podobné. Z toho vieme usúdiť, že vyššia vzdialenosť vysoko ovplyvnila aj simuláciu s FP filtrom. Hodnoty Q-faktora a min. BER síce boli o niečo vyššie ako v simuláciách s FBG filtrom a demultiplexorom, ale rozdiel bol minimálny. Pri predošlých simuláciách bola vyššia vzdialenosť hlavným faktorom zníženia kvality prenosu, a to sa ukázalo ako limitujúce aj pri tejto simulácií.

IV. ZÁVER

Táto práca sa venovala vlastnostiam optických filtrov v DWDM optických prenosových systémoch. Hlavným cieľom práce bolo navrhnutie simulácie pre optické filtre vo WDM systéme a skúmanie vlastností jednotlivých filtrov. Na vyriešenie tejto problematiky sme použili simulácie pomocou programu OptiSystem. Výsledky a vyhodnotenie simulácií sú v kapitole III.

POĎAKOVANIE

Táto práca bola podporovaná nasledujúcimi výskumnými projektami: APVV-22-0400 a VEGA 1/0260/23.

LITERATÚRA

[1] HAMLITSCH, Ben. Basic components of a fiber optic cable. 2023. Dostupné na: https://www.truefiber.com/blogs/fiber-forum/basic-components-of-a-fiber-optic-cable.
[2] HYCGLOBAL. CWDM (CoarseWavelength Division Multiplexing) VS DWDM (Dense Wavelength Division Multiplexing). 2023. Dostupné na: https://medium . com / @hycfiberoptics / cwdm - coarse - wavelength - division -multiplexing- vs- dwdm- dense- wavelength- division- multiplexingc1af83b32087.
[3] WORLD WIDE TECHNOLOGY. CWDM or DWDM: Which Should You Use and When? 2023. Dostupné na: https://www.wwt.com/article/cwdmor-dwdm-which-should-you-use-and-when.
[4] ONOUFRIOU, Andreas; KALLI, Kyriacos; PUREUR, David; MUGNIER, Alain. Fibre Bragg Gratings. In: 2006, zv. 123, s. 189–269. Isbn 978-3-540-31769-2. Dostupné na: doi: 10.1007/3-540-31770-8_6.
[5] KOONEN, Ton. Fabry-Perot Interferometer Filters. In: VENGHAUS, Herbert (ed.). Wavelength Filters in Fibre Optics. Springer Berlin Heidelberg, 2006, s. 271–287. Springer Series in Optical Sciences. Isbn 978-3-540-31769-2. Dostupné na: doi: 10.1007/3-540-31770-8_7.

Využitie neurónových sietí na prepínanie liniek v hybridných FSO/RF systémoch

[1]*Zuzana LIŠČINSKÁ,* [2]*Ľuboš OVSENÍK,* [3]*Jakub ORAVEC*

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

[2,3] Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]zuzana.liscinska@student.tuke.sk, [2]lubos.ovsenik@tuke.sk, [3]jakub.oravec@tuke.sk

***Abstrakt* — Príspevok sa zaoberá súčasnými bezdrôtovými komunikačnými technológiami, najmä rádiofrekvenčnou komunikáciou (RF, Radio Frequency) a optickou komunikáciou voľným priestorom (FSO, Free Space Optics) a ich hybridnými formami. Sú analyzované vplyvy prenosového prostredia na prenos dát v hybridných FSO/RF systémoch. Je rozpracovaná problematika predikcie atmosférických vplyvov na parameter RSSI (Received Signal Strength Indicator), čo je sila prijatého signálu v hybridných FSO/RF systémoch a následne je navrhnutý a optimalizovaný model neurónových sietí na tvrdé prepínanie liniek v týchto systémoch. V závere sú porovnané vlastnosti navrhnutých riešení s existujúcimi prístupmi.**

***Kľúčové slová* —FSO systém, hybridný FSO/RF systém, LSTM, RF systém, RSSI**

I. Úvod

V súčasnej dobe digitálnej éry, kde je všetko prepojené s technológiou a dátami, je neustále narastajúci dopyt po spoľahlivej a rýchlej komunikácii kľúčovým faktorom v mnohých aspektoch nášho života, od komerčných telekomunikačných sietí až po kritické infraštruktúry, ako sú zdravotnícke zariadenia a priemyslové procesy. Jednou z najväčších výziev, s ktorými sa stretávajú bezdrôtové komunikačné systémy je ich citlivosť na atmosférické podmienky. Znečistenie (prach), sneh, dážď, hmla a iné faktory môžu dramaticky ovplyvniť kvalitu a spoľahlivosť prenosu signálu. Hybridné FSO/RF systémy, ktoré kombinujú optické a rádiové prenosové technológie, sa javia ako sľubná alternatíva pre riešenie niektorých týchto problémov, avšak ich účinnosť je často narušená práve v dôsledku atmosférických vplyvov. Táto práca sa preto zaoberá problematikou predikcie atmosférických vplyvov na parameter RSSI v hybridných FSO/RF systémoch. RSSI, ako dôležitý ukazovateľ kvality signálu, je kritický pre správne fungovanie bezdrôtových komunikačných sietí. Cieľom práce nie je len predikovať samotný parameter RSSI, ale najskôr predikovať atmosférické podmienky, ktoré najviac ovplyvňujú tento parameter. Následne sa na základe týchto predpovedaných atmosférických vplyvov vykonáva predikcia samotného parametra RSSI.

V rámci práce sa využíva metóda hlbokého učenia, konkrétne Long Short-Term Memory (LSTM) algoritmus. LSTM siete sú špecifický typ rekurentných neurónových sietí, ktoré sú navrhnuté na spracovanie a predikciu sekvenčných dát, ako sú časové rady. Jednou z kľúčových vlastností LSTM sietí je ich schopnosť zachytávať dlhodobé závislosti v týchto radoch. To znamená, že LSTM môžu efektívne uchovávať informácie o udalostiach, ktoré sa stali v minulosti, a využiť ich pri predikcii budúcich udalostí. Táto schopnosť je kritická pri predikcii v dynamických a komplexných prostrediach, ako je atmosféra, kde sa môžu vyskytovať rôzne faktory ovplyvňujúce správanie systému.

LSTM siete používajú špeciálnu štruktúru pamäte, ktorá sa skladá z jednotiek pamäte nazývaných "bunky pamäte". Tieto bunky pamäte majú schopnosť rozhodovať, ktoré informácie majú byť zachované a ktoré majú byť zabudnuté v závislosti od kontextu sekvencie vstupných dát. Tento mechanizmus umožňuje LSTM sieťam efektívne pracovať s dlhými časovými radmi a odhaľovať zložité vzory a závislosti v dátach.

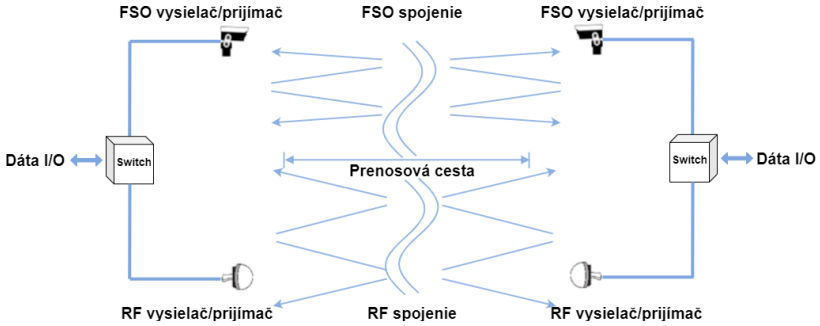
V kontexte predikcie atmosférických vplyvov na parameter RSSI v hybridných FSO/RF systémoch umožňuje použitie LSTM algoritmu modelom efektívne zachytiť vzťahy medzi

rôznymi atmosférickými premennými a ich vplyvom na parameter RSSI. Týmto spôsobom LSTM umožňuje vytvárať presné a robustné predikcie, ktoré môžu byť následne využité na lepšie riadenie a optimalizáciu prenosu signálu v týchto hybridných systémoch.

Výsledkom tejto práce sú modely, ktoré sú schopné predikovať atmosférické vplyvy s dostatočnou presnosťou. Tieto predikcie sú nástrojom na lepšie porozumenie a optimalizáciu prenosu signálu v týchto systémoch, čo vedie k zlepšeniu ich celkových výkonových vlastností a spoľahlivosti.

II. Hybridný FSO/RF prenosový systém

Predstavovaná aplikácia pre analýzu dát získaných prostredníctvom senzorov monitorujúcich poveternostné podmienky počasia slúži k správnemu nastaveniu vhodnej prenosovej cesty v hybridnom FSO/RF prenosovom systéme. FSO 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 SiO_2, 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 20 km), 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 vzniká hybridný FSO/RF systém (Obr. 1), vyzdvihujúci dobré vlastnosti jednotlivých systémov, čím sa stáva odolnejším voči vplyvom počasia [1, 3].



Obr. 1 Hybridný FSO/RF systém.

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 [2, 4]. 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 je odporúčaná v kontinentálnom podnebí, pretože sú obidve linky 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 RF liniek od 70 GHz do 80 GHz používané modernejšími zariadeniami dokážu prenášať dáta na vzdialenosti do 3 km rýchlosťou 1 Gb/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 [1, 5].

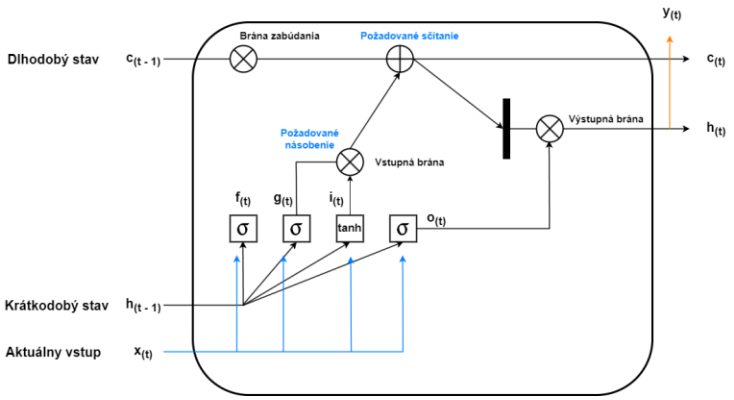
III. Neurónové siete a ich využitie v hybridných FSO/RF prenosových systémoch

Neurónovú sieť možno chápať ako spojenie dvoch atribútov a to umelej inteligencie a dizajnu inšpirovaného ľudským mozgom. Neurónová sieť s jej komplexnými vrstvami vzájomne

prepojených umelých neurónov, simuluje fungovanie ľudského mozgu čím umožňuje dosahovať pozoruhodné výsledky v strojovom učení. Existuje viacero typov neurónových sietí napríklad umelé neurónové siete, rekurentné alebo konvolučné neurónové siete. Každá zo spomenutých neurónových sietí je priamo prispôsobená na riešenie konkrétnych úloh. Táto práca sa zaoberá rekurentnými neurónovými sieťami a ich využitím pri predikcii samotného atmosférického vplyvu, ako aj predikcii parametra RSSI. Neurónové siete sú v dnešnej dobe čoraz obľúbenejšie pre ich efektívne a adaptívne učenie sa. Medzi ich hlavnú charakteristickú schopnosť patrí skutočnosť, že sa vedia rýchlo naučiť ako riešiť konkrétny problém. Pre učenie využívajú neurónové siete sadu tzv. trénovaniach dát.

A. Neurónová sieťová architektúra LSTM

LSTM neurónová sieťová architektúra založená na rekurentných neurónových sieťach (RNN, Recurrent Neural Network), je navrhnutá na analýzu a modelovanie dlhých časových sekvencií údajov. Aj keď existujú iné prístupy na prácu s sekvenciami, ako sú Markovove modely, podmienené náhodné polia a Kalmanove filtre, iba LSTM sa vyníma v schopnosti učiť sa dlhodobé vzťahy. Základným rozdielom medzi LSTM a RNN je štruktúra, ktorá využíva iný matematický vzorec na výpočet skrytého stavu. Namiesto tradičných neurónov používa LSTM špeciálne jednotky nazývané "pamäťové bunky“. Táto architektúra kombinuje rýchly tréning s efektívnym učením na úlohách, ktoré vyžadujú krátkodobé uchovávanie informácií počas mnohých časových krokov v experimente. LSTM je schopná naučiť sa preklenúť časové intervaly dlhšie ako 1000 diskrétnych časových krokov vďaka týmto špeciálnym pamäťovým bunkám. Tieto bunky rozhodujú, či je potrebné aktualizovať predchádzajúcu pamäť a vypočítajú hodnotu, ktorá sa má uložiť do aktuálnej pamäte na základe aktuálneho stavu a vstupnej hodnoty pamäťovej bunky. Táto štruktúra je veľmi účinná pri ukladaní a manipulácii s dlhými sekvenciami údajov.



Obr. 2 Architektúra bunky LSTM.

Na rozdiel od alternatívnych modelov, ako sú Markovove modely, podmienené náhodné polia a Kalmanove filtre, LSTM nevyžaduje expertné znalosti ani inžinierstvo príznakov a má väčšiu pravdepodobnosť objavovania nepredvídaných vzorov a štruktúr [6]. LSTM algoritmus je možné chápať aj ako RNN s pomyselným pamäťovým fondom, ktorý obsahuje dva základné vektory:

- **Krátkodobý stav:** udržuje výsledok v aktuálnom časovom okamihu.
- **Dlhodobý stav:** ukladá, číta a rozhoduje o prvkoch určených pre dlhodobé uchovávanie počas prechodu cez sieť.

Rozhodovanie o tom, či čítať, ukladať alebo zahodiť informácie, je založené na určitých aktivačných funkciách, ako je znázornené na obrázku Obr. 2. Výstup z týchto aktivačných funkcií sa nachádza v rozsahu (0, 1) [6].

B. Vstupné údaje a ich analýza

Na účelnú analýzu vstupných údajov je nevyhnutný správny formát. Vstupné dáta boli transformované z tabuľky .xlsx na súbor .csv, ktorý je nevyhnutný pre prácu s údajmi vo vývojových prostrediach. Údaje v súbore .csv obsahujú dátum Date (D) (13.10.2023) [D.M.R], čas Time (T) (8:55) [H:M], parameter RSSI (informácia o sile prijatého signálu RSSI (- 26[dB/m])), teplotu Temperature1 (TE1) (14.4 [°C]), tlak Pressure (P) (994.715 [hPa]), teplotu Temperature2 (TE2) (12.95 [°C]), vlhkosť vzduchu Dampness (H) (79.6500015[%]), teplotu Temperature3 (TE3) (10.437 [°C]), rýchlosť vetra WindSpeed1 (WS1) (0.3940265 [m/s]), rýchlosť vetra WindSpeed2 (WS2) (0.3965745 [m/s]), koncentráciu častíc Particles (CP) (11

[mg/m3]) a viditeľnosť Visibility (V) (4000 [m]).

Ukážka súboru .csv na obrázku nižšie (Obr. 3) zobrazuje dáta z dňa 13.10.2023 v časovom intervale od 8:55 do 9:14. Jednotlivé údaje v súbore .csv sú oddelené bodkočiarkou, ktorá ich rozdeľuje na jednotlivé stĺpce. Tieto dáta, ktoré boli získané z meteostanice Technickej univerzity v Košiciach, predstavujú časť celej databázy použitej v tejto práci, ktorá obsahuje 85 125 záznamov.

```
13.10.2023;8:55;-26;14.4;994.715;12.95;79.6500015;10.437;0.3940265;0.3965745;11;4000
13.10.2023;8:56;-26;13.9;994.63;12.9;77.950001;10.437;0.7807575;0.409075;14;4000
13.10.2023;8:57;-26;13.9;994.685;12.9;79.75;10.437;0.780757;0.409075;13.5;4000
13.10.2023;8:58;-26;13.95;994.725;12.8;79.800003;10.437;0.392093;0.396512;13;4000
13.10.2023;8:59;-26;13.7;994.775;12.85;79.199997;10.437;0.552586;0.4016995;13.5;4000
13.10.2023;9:00;-26;13.5;994.775;12.8;80.1500015;10.375;0.477174;0.399262;12.5;4000
13.10.2023;9:01;-26;13.65;994.74;12.85;79.900002;10.375;0.4133635;0.3971995;14;4000
13.10.2023;9:02;-26;13.8;994.73;12.7;79.7000005;10.375;0.7788235;0.4090125;12.5;4000
13.10.2023;9:03;-26;13.9;994.805;12.75;80.1500015;10.375;0.848435;0.4112625;12;4000
13.10.2023;9:04;-26;13.95;994.735;12.7;79.5;10.375;0.6299325;0.4042;12.5;4000
13.10.2023;9:05;-26;14.5;994.74;12.9;79.199997;10.375;0.4017615;0.3968245;12;4000
13.10.2023;9:06;-26;14.25;994.795;12.8;79.049999;10.375;0.55452;0.401762;13.5;4000
13.10.2023;9:07;-26;14.4;994.84;13;79.1499975;10.375;0.6299325;0.4042;12;4000
13.10.2023;9:08;-26;14.15;994.84;13;78.949997;10.375;1.0089285;0.4164505;11.5;4000
13.10.2023;9:09;-26;13.8;994.795;12.9;78.5;10.375;1.2525685;0.424325;14.5;4000
13.10.2023;9:10;-26;13.8;994.835;12.95;78.549999;10.375;0.693743;0.406262;11.5;4000
13.10.2023;9:11;-26;13.25;994.88;12.85;78.450001;10.375;1.467204;0.431263;13;4000
13.10.2023;9:12;-26;12.6;994.915;12.55;78.5499995;10.312;1.532948;0.433388;12;4000
13.10.2023;9:13;-26;12.55;994.9;12.4;80.299999;10.312;0.47524;0.3991995;14;4000
13.10.2023;9:14;-26;12.85;994.93;12.25;81.1000025;10.25;0.8581035;0.411575;12;4000
```

Obr. 3 Ukážka časti databázy určenej pre učenie neurónovej siete.

Pre lepšie pochopenie použitej databázy je možné interpretovať vstupné dáta aj prostredníctvom matice, kde sú jednotlivé dáta zapísané pomocou skratiek v riadku v nasledujúcom poradí: dátum, čas, parameter RSSI, teplota1 z prvého snímača, tlak, teplota 2 z druhého snímača, vlhkosť vzduchu, teplota3 z tretieho snímača, rýchlosť vetra nameraná z prvého snímača, rýchlosť vetra nameraná z druhého snímača, koncentrácia častíc a viditeľnosť.

$$M = \begin{bmatrix} D_1 & T_1 & RSSI_1 & TE_1 & P_1 & TE2_1 & H_1 & TE3_1 & WS_1 & WS2_1 & CP_1 & V_1 \\ D_2 & T_2 & RSSI_2 & TE_2 & P_2 & TE2_2 & H_2 & TE3_2 & WS_2 & WS2_2 & CP_2 & V_2 \\ D_3 & T_3 & RSSI_3 & TE_3 & P_3 & TE2_3 & H_3 & TE3_3 & WS_3 & WS2_3 & CP_3 & V_3 \\ D_4 & T_4 & RSSI_4 & TE_4 & P_4 & TE2_4 & H_4 & TE3_4 & WS_4 & WS2_4 & CP_4 & V_4 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ D_n & T_n & RSSI_n & TE_n & P_n & TE2_n & H_n & TE3_n & WS_n & WS2_n & CP_n & V_n \end{bmatrix}$$

C. Korelácia dát

Jedným z cieľov tejto práce je predikcia parametra RSSI. Z tohto dôvodu je nevyhnutné preskúmať vzťah medzi parametrom RSSI a ďalšími stĺpcami, teda jednotlivými údajmi z poskytnutého súboru .csv. V praxi sa k tomu využíva korelačný koeficient, ktorý slúži na identifikovanie vzťahu medzi dvomi premennými, konkrétne medzi parametrom RSSI a jednotlivými vstupnými údajmi zo súboru .csv. Výsledok korelačného koeficientu sa pohybuje v rozsahu od -1 do +1. Záporný korelačný koeficient -1 signalizuje existenciu záporného vzťahu medzi veličinami, zatiaľ čo kladný korelačný koeficient +1 naznačuje prítomnosť kladného vzťahu. Nulový korelačný koeficient ukazuje, že medzi veličinami neexistuje žiadny vzťah [7].

Na určenie korelačného koeficientu sme využili metódy, ktoré zahŕňajú Spearmanov korelačný koeficient, Kendall-tau korelačný koeficient a Pearsonov korelačný koeficient. Najrelevantnejším korelačným koeficientom v tomto prípade sa javí Pearsonov korelačný koeficient, nakoľko definuje a meria lineárny vzťah medzi dvoma spojitými premennými, pričom jeho hodnota sa pohybuje od -1 (úplný záporný lineárny vzťah) cez 0 (žiadny lineárny vzťah) po +1 (úplný kladný lineárny vzťah). Táto metóda je vhodná pre situácie, kde je predpoklad lineárneho vzťahu medzi premennými [7]. Naproti tomu Kendall-tau korelačný koeficient a Spearmanov korelačný koeficient sa často využívajú pre poradové alebo ordinálne dáta, kde sa neberie do úvahy presná veľkosť, ale iba poradie hodnôt. Vyššie spomenuté metódy boli použité pre skúmanie korelačného vzťahu medzi parametrom RSSI a ostatnými dátami. Korelačné koeficienty boli využité aj pre skúmanie vzťahu medzi parametrom RSSI a spriemerovanými hodnotami dát, ktorými boli teplota a rýchlosť vetra.

Tabuľky nižšie (Tabuľka 1 a 2) demonštrujú výsledky jednotlivých korelačných koeficientov. Z výsledných výpočtov korelačných koeficientov boli brané do úvahy najmä výsledky z Pearsonovho korelačného koeficientu. Na základe toho boli určené za najviac ovplyvňujúce faktory parametra RSSI teplota1, teplota2, teplota3 ako aj viditeľnosť a koncentrácia častíc. Identická situácia nastala v prípade, že sme spúšťali výpočet korelácie so spriemerovanými hodnotami teploty a rýchlosti vetra. V tomto prípade bol parameter RSSI ovplyvňovaný najmä priemernou hodnotou teplôt, viditeľnosťou a koncentráciou častíc.

Tabuľka 1
Výsledky jednotlivých korelačných koeficientov

Hodnota Y	Spearmanov kor. koef.	Kendall tau kor. koef.	Pearsonov kor. koef.
RSSI	1,000	1,000	1,000
Teplota1	-0,276	-0,234	0,145
Tlak	0,131	0,091	0,053
Teplota2	-0,286	-0,242	0,130
Vlhkosť	-0,130	-0,106	-0,007
Teplota3	-0,271	-0,230	0,167
Sila vetra1	0,158	0,120	-0,010
Sila vetra2	0,158	0,120	-0,010
Koncentrácia častíc	-0,041	-0,039	0,177
Viditeľnosť	-0,081	-0,072	0,264

Tabuľka 2
Výsledky jednotlivých korelačných koeficientov so spriemerovanými údajmi o teplote a sile vetra

Hodnota Y	Spearmanov kor. koef.	Kendall tau kor. koef.	Pearsonov kor. koef.
RSSI	1,000	1,000	1,000
Priemerná teplota	-0,289	-0,245	0,155
Tlak	0,131	0,091	0,053
Vlhkosť	-0,130	-0,106	-0,007
Priemerná sila vetra	0,158	0,120	-0,010
Koncentrácia častíc	-0,041	-0,039	0,177
Viditeľnosť	-0,081	-0,072	0,264

IV. ANALÝZA VÝSLEDKOV PRÁCE

Ako už bolo spomenuté vyššie práca sa venuje predikcii atmosférických vplyvov, ktoré na základe výpočtu Pearsonovho korelačného koeficientu ovplyvňujú parameter RSSI v najväčšej miere. Týmito parametrami sú teplota, ktorá bola pre zníženie výpočtovej náročnosti spriemerovaná, nakoľko vstupné dáta obsahovali nameranú teplotu z troch snímačov. Ďalšími parametrami, ktoré ovplyvňujú parameter RSSI sú koncentrácia častíc vo vzduchu a samotná viditeľnosť. Pre každý z týchto troch atmosférických vplyvov bol vytvorený samostatný model ktorý načítal vstupné dáta obsahujúce len daný predikovaný parameter a obsahoval záznamy od 12.10.2023 00:00 do 7.12.2023 do 23:59. Pôvodné dáta obsahovali aj deň 8.12.2023 od 00:00 do 23:59 avšak tie boli odstrihnuté aby tak mohla byt vykonaná samotná 24 hodinová predikcia pre daný dátum. Deň 8.12.2023 je posledným dňom z dostupného datasetu a bol zvolený na predikciu z hľadiska malej rozsiahlosti vstupných dát. Keďže sme chceli dosiahnuť čo najlepšie výsledky bolo potrebné poskytnúť modelu čo najviac dát, na ktorých sa mohol trénovať a testovať. Samotná predikcia atmosférických vplyvov bola spúšťaná pri takýchto rozdeleniach trénovacích dát:

- **90:10:** trénovacie dáta od 12.10. - 2.12.2023, testovacie dáta od 3.12. - 7.12.2023.
- **80:20:** trénovacie dáta od 12.10. - 26.11.2023, testovacie dáta od 27.11. - 7.12.2023.
- **70:30:** trénovacie dáta od 12.10. - 20.11.2023, testovacie dáta od 21.11. - 7.12.2023.
- **Manuálne rozdelenie vstupných dát:** trénovacie dáta od 12.10. - 6.12.2023, testovacie dáta od 7.12.2023.

Ďalšou dôležitou časťou bola 24 hodinová predikcia parametra RSSI pre časový interval od 8.12.2023 00:00 do 8.12.2023 23:59. Táto predikcia vychádzala už aj z predikovaných hodnôt atmosférických vplyvov, ktoré boli zahrnuté do dát pre predikciu daného parametra. Tak ako predikcie atmosférických vplyvov aj predikcia parametra RSSI sa vykonávala v rôznych pomeroch trénovacích a testovacích dát, konkrétne:

- **90:10:** trénovacie dáta od 12.10. - 2.12.2023, testovacie dáta od 3.12. - 7.12.2023.
- **80:20:** trénovacie dáta od 12.10. - 26.11.2023, testovacie dáta od 27.11. - 7.12.2023.
- **70:30:** trénovacie dáta od 12.10. - 20.11.2023, testovacie dáta od 21.11. - 7.12.2023.
- **Manuálne rozdelenie vstupných dát:** trénovacie dáta od 12.10. - 6.12.2023, testovacie dáta od 7.12.2023.

Výsledky predikcie atmosférických vplyvov je možné vidieť v tabuľkách nižšie (Tabuľka 3 až 6). Tabuľky demonštrujú výsledky predikcií kde boli dáta rozdelené na trénovaciu a testovaciu množinu manuálne, potom v pomeroch 90:10, 80:20, 70:30.

Tabuľka 3
Výsledky predikcií atmosférických vplyvov kde boli dáta rozdelené manuálne

Atmosférické vplyvy	Čas trénovania LSTM modelu	MSE	Percentuálna úspešnosť
Teplota	1413,69 sekúnd	0,24366	77,57%
Koncentrácia častíc	1413,98 sekúnd	0,49977	81,60%
Viditeľnosť	1418,28 sekúnd	6825,94009	99,65%

Tabuľka 4
Výsledky predikcií atmosférických vplyvov kde boli dáta rozdelené v pomere 90:10

Atmosférické vplyvy	Čas trénovania LSTM modelu	MSE	Percentuálna úspešnosť
Teplota	1297,77 sekúnd	0,01856	99,38%
Koncentrácia častíc	1291,39 sekúnd	1,14805	57,64%
Viditeľnosť	1277,99 sekúnd	7010,24859	99,86%

Tabuľka 5
Výsledky predikcií atmosférických vplyvov kde boli dáta rozdelené v pomere 80:20

Atmosférické vplyvy	Čas trénovania LSTM modelu	MSE	Percentuálna úspešnosť
Teplota	1236,66 sekúnd	0,10786	95,90%
Koncentrácia častíc	1240,19 sekúnd	1,07243	57,57%
Viditeľnosť	1236,79 sekúnd	6820,90747	99,86%

Tabuľka 6
Výsledky predikcií atmosférických vplyvov kde boli dáta rozdelené v pomere 70:30

Atmosférické vplyvy	Čas trénovania LSTM modelu	MSE	Percentuálna úspešnosť
Teplota	1142,74 sekúnd	0,24147	89,10%
Koncentrácia častíc	1136,53 sekúnd	0,79461	59,58%
Viditeľnosť	1144,09 sekúnd	5468,60477	99,86%

Na základe poskytnutých tabuliek je možné vydedukovať, že najlepšie predikcie podľa MSE a celkovej úspešnosti predikcie vykazuje teplota, nakoľko práve hodnoty namerané pre teplotu sa v čase radikálne nemenili. Naopak najhoršie výsledky predikcie vykazuje koncentrácia častíc aj keď je chybovosť MSE pomerne malá, úspešnosť predikcie nie je veľmi presná, nakoľko sa hodnoty koncentrácie častíc radikálne menili v malých časových intervaloch.

Tabuľka 7 demonštruje výsledky predikcie parametra RSSI, kde bol parameter predikovaný z atmosférických vplyvov teplota, koncentrácia častíc a viditeľnosť, kde boli do vstupných dát zahrnuté aj ich predikované hodnoty. Z tabuľky je zrejmé, že všetky výsledky predikcie parametra RSSI sú uspokojivé, avšak pri rozdelení trénovaniach a testovacích dát v pomere 70:30 bola percentuálna miera predikcie zanedbateľne lepšia.

Tabuľka 7
Výsledky predikcie parametra RSSI v rôznych pomeroch trénovacích a testovacích dát

Predikcia RSSI rozdelenie dát	Čas trénovania modelu	MSE	Percentuálna úspešnosť
Manuálne rozdelenie dát	116,085896 sekúnd	1,29583	96,39%
90:10	117,037107 sekúnd	1,27643	96,39%
80:20	94,418121 sekúnd	1,27628	96,39%
70:30	91,305367 sekúnd	1,47492	96,60%

Tabuľky 8 a 9 zobrazujú úspešnosť predikcie parametra RSSI pomocou algoritmov strojového učenia a to algoritmom DecisionTree a algoritmom DecisionTree obohateným o AdaBoost regresiu v pomere trénovacích a testovacích dát 90:10, 80:20 a 70:30. Tieto výsledky predikcií nadväzujú na výsledky predikcií parametra RSSI z našich predchádzajúcich výskumov, nakoľko súčasný výskum skúma vylepšenie predikcie parametra RSSI a rieši taktiež otázku predikcie samotných atmosférických vplyvov obmedzujúcich parameter RSSI. Tabuľky 8 a 9 sú teda demonštrované pre porovnanie rôznych prístupov predikcie parametra RSSI.

Tabuľka 8
Výsledky predikcie parametra RSSI pomocou algoritmu DecisionTree v pomere trénovacích a testovacích dát 90:10, 80:20 a 70:30

DecisionTree	Čas trénovania modelu	MSE	Percentuálna úspešnosť
90:10	115,0819 sekúnd	1,38459	82,83%
80:20	115,1798 sekúnd	1,49871	82,31%
70:30	114,9716 sekúnd	1,39485	81,55%

Z tabuliek je možné vidieť, že algoritmus DecisionTree obohatený o Adaboost regresiu má pomerne väčšiu úspešnosť ako samotný algoritmus DecisionTree.

Tabuľka 9
Výsledky predikcie parametra RSSI pomocou algoritmu DecisionTree obohateného o AdaBoost regresiu v pomere trénovacích a testovacích dát 90:10, 80:20 a 70:30

AdaBoost	Čas trénovania modelu	MSE	Percentuálna úspešnosť
90:10	120,0148 sekúnd	1,28451	82,83%
80:20	121,1784 sekúnd	1,74285	89,48%
70:30	121,2487 sekúnd	1,18504	88,98%

V. ZÁVER

Cieľom tejto práce bolo preskúmať a implementovať metódu využívajúcu neurónové siete na optimalizáciu prepínania liniek v hybridných systémoch FSO/RF. Neurónové siete sa osvedčili ako efektívny nástroj v oblasti predikcie a riadenia v rôznych odvetviach, a preto sa javia ako sľubná metóda aj pre tento konkrétny problém. Ich schopnosť adaptácie a učenia sa umožňuje dynamicky prispôsobovať sa meniacim sa podmienkam a prostrediu, čo môže viesť k zlepšeniu celkového výkonu systému. Integrácia neurónových sietí do hybridných systémov umožňuje efektívne riadiť prepínanie medzi FSO a RF linkami v závislosti na aktuálnych podmienkach a požiadavkách prenosu dát. Tento prístup otvára nové možnosti v oblasti optimalizácie a zlepšenia výkonu komunikačných sietí a môže prispieť k vytvoreniu spoľahlivejších a efektívnejších systémov pre prenos dát.

POĎAKOVANIE

Táto práca bola podporovaná nasledujúcimi výskumnými projektami: APVV-22-0400 – „Rozšírenie autonómnych aplikácií monitorovania, prostredníctvom viacpásmových UWB senzorov“ a VEGA 1/0260/23 – „Umelá inteligencia, robustné MANET a viacpreskokové D2D siete integrované do siete 6.generácie“.

LITERATÚRA

[1] M. N. Khan, et al., “Maximizing throughput of hybrid FSO-RF communication system: An algorithm”, in IEEE Access, Vol. 6, 2018, pp. 30039-30048.
[2] M. A. Amirabadi, et al., “A novel hybrid FSO/RF communication system with receive diversity”, in Optic, Vol. 184, 2019, pp. 293-298.
[3] A. Munandar, et al., “Design of real-time weather monitoring system based on mobile application using automaticweather station”, in *2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT 2017)*, IEEE, 2017, pp. 44-47.
[4] A. Sampathkumar, et al., “A novel scheme for an IoT-based weather monitoring system using a wireless sensor network”, in Integration of WSN and IoT for smart cities, 2020, pp. 181-191.
[5] S. Dewitte, et al., “Artificial intelligence revolutionises weather forecast, climate monitoring and decadal prediction”, in Remote Sensing, Vol. 13, no 16 (2021), pp. 3209.
[6] Mohammed Alhamid, 2021. LSTM and Bidirectional LSTM for Regression. In: Medium. Dostupné tiež z: https://towardsdatascience.com/lstm-andbidirectional-lstm-for-regression-4fddf910c655
[7] Rohit Sharma, 2023. Correlation Coefficient: Simple Definition, Formula, Easy Steps. In: Statistics How To. Dostupné tiež z: https://www.statisticshowto.com/probability-and-statistics/correlationcoefficient-formula/

Moderné metódy prevencie online obťažovania detí na Slovensku

Patrik GAJDOŠ, Eva KUPCOVÁ, Matúš PLEVA, Renát HALUŠKA

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Letná 9, 04001 Košice, Slovenská republika

patrik.gajdos@student.tuke.sk, eva.kupcova@tuke.sk, matus.pleva@tuke.sk, renat.haluska@tuke.sk

Abstrakt **– Tento príspevok sa zaoberá modernými metódami prevencie obťažovania detí na internete, pričom cieľom je zvýšiť povedomie o tejto problematike a preskúmať aktuálny stav v danej oblasti, s dôrazom na identifikáciu nových hrozieb. Príspevok analyzuje existujúce metódy a dostupné dáta využívajúce sa v modeloch prevencie obťažovania detí na internete. Na základe týchto analýz bola vytvorená webová aplikácia na infraštruktúre Technickej univerzity v Košiciach, zameraná na prevenciu online obťažovania detí na Slovensku. Aplikácia umožňuje hodnotiť bezpečnosť výrokov v konverzáciách na internete. V závere boli spracované a analyzované získané informácie z vytvorenej webovej aplikácie, ktoré pochádzajú od unikátnych respondentov.**

Kľúčové slová **– internetová bezpečnosť, konverzácie na internete, moderné detekčné metódy, prevencia obťažovania detí, webová aplikácia**

I. Úvod

Obťažovanie detí na internete predstavuje závažný problém s mnohými rizikami pre mladú generáciu. Rýchly rozvoj digitálnych technológií a rastúci počet detí používajúcich internet vedie k nárastu prípadov kyberšikany a kybernetického obťažovania. Tieto hrozby môžu mať dlhodobé psychické a fyzické následky pre obete, čo zdôrazňuje potrebu efektívnych preventívnych stratégií.

V rámci tohto príspevku sme sa zamerali na moderné metódy prevencie obťažovania detí na internete. Analyzovali sme súčasné hrozby a dostupné technológie na ich prevenciu. Vyvinuli sme webovú aplikáciu na hodnotenie bezpečnosti výrokov v online konverzáciách, ktorá bola testovaná respondentmi a analyzovaná na základe zhromaždených údajov.

Cieľom príspevku je analyzovať moderné metódy prevencie online obťažovania detí, identifikovať aktuálne hrozby a preskúmať technológie v tejto oblasti. Výsledkom je webová aplikácia na hodnotenie bezpečnosti slovenských výrokov v online konverzáciách, ktorá umožňuje jej ďalšie zlepšovanie.

II. Obťažovanie detí na internete

Obťažovanie detí na internete zahŕňa rôzne formy šikany, ktoré môžu mať vážne a dlhodobé následky na zdravie a psychickú pohodu dieťaťa. Tieto formy obťažovania sa môžu prejaviť ako fyzická šikana, psychická šikana, kyberšikana a kybernetické obťažovanie.

A. Šikana

Šikanovanie je forma obťažovania mierená na jednotlivca alebo skupinu. Ide o dlhodobé násilie, kde jednotlivec alebo skupina vyvíja fyzický alebo psychický tlak na osobu, ktorá sa nedokáže brániť [1]. Existujú dva typy násilia: priame a nepriame. Priame násilie je fyzické a zahŕňa bitky alebo hádzanie predmetov. Nepriame násilie je psychické a zahŕňa zastrašovanie, vyhrážky a psychické vydieranie. Často sa vyskytuje v online svete, napríklad na sociálnych sieťach alebo v počítačových hrách [2].

B. Fyzická a psychická šikana

Fyzická šikana zahŕňa fyzické násilie alebo zastrašovanie na získanie kontroly nad obeťou, čo môže spôsobiť zranenie obete alebo poškodenie osobných veci. Toto správanie je v školách časté a obete často odmietajú účasť na vyučovaní zo strachu z ďalšej šikany. To vedie k nízkemu sebavedomiu, izolácii, a zúfalstvu. Dieťa môže stratiť záujem o aktivity, ktoré ho predtým bavili, a uzatvárať sa do škodlivých vzorcov správania, vrátane sebapoškodzovania a vyhýbania sa komunikácii [3].

Psychická šikana zahŕňa verbálne útoky, zastrašovanie a manipuláciu, ktoré môžu viesť k vážnym psychickým problémom. Verbálna šikana sa často prejavuje prostredníctvom urážok, hanlivých prezývok a iných ponižujúcich činov, často aj v online prostredí. Psychická šikana môže mať dlhodobé účinky na psychické zdravie obete, vrátane úzkosti, depresie a nízkeho sebavedomia [4].

C. Kyberšikana

Kyberšikana je forma šikany, ktorá využíva digitálne technológie na obťažovanie, zastrašovanie alebo urážanie iných osôb. Táto forma šikany často vedie k sociálnemu vylúčeniu, šíreniu klebiet, zneužitiu účtu a kybernetickému prenasledovaniu. Kyberšikana môže byť anonymná, čo sťažuje identifikáciu páchateľov a zvyšuje jej dopad na obete [5].

Kyberšikana sa od tradičnej šikany líši tým, že agresor môže svoju identitu skrývať prostredníctvom anonymných účtov alebo falošných profilov, čo obeti sťažuje možnosť úniku a ochrany. Útočníci často využívajú sociálne siete, komunikačné aplikácie a iné online platformy na šírenie urážlivého obsahu. Táto nepretržitá forma obťažovania môže viesť k vážnym psychickým a emocionálnym problémom detí, vrátane úzkosti, depresie a v extrémnych prípadoch až k sebapoškodzovaniu [6].

D. Kybernetické obťažovanie

Kybernetické obťažovanie patrí medzi závažnú formu obťažovania páchaného na deťoch prostredníctvom digitálnych technológií. Páchateľ si vytvára falošný vzťah s dieťaťom, aby ho mohol zneužiť, čo je umožnené anonymitou online prostredia [7]. Tento proces prebieha v niekoľkých fázach, ktoré začínajú vytvorením falošnej identity na sociálnych sieťach alebo komunikačných platformách [8]. Páchateľ používa tieto falošné profily na získanie dôvery detí, predstierajúc podobné záujmy a vlastnosti, čím vytvára ilúziu blízkosti a dôvery [9].

E. Subjekty šikany

Subjekty šikany zahŕňajú *agresora*, *obeť* a *ďalšie osoby*, ktoré môžu byť nepriamo zapojené do šikany. Okrem dospelých ľudí, aj každé dieťa môže zastávať rôzne role šikany, či už ako agresor, obeť, alebo podporovateľ. Pochopenie týchto rolí je kľúčové pre efektívne riešenie problému šikanovania a vyžaduje individuálny prístup ku každému dieťaťu [10].

1) Agresor: Agresor šikanuje iných. Neustále ich obťažuje a ponižuje. Využíva svoju moc a silu na to, aby spôsoboval fyzickú alebo psychickú bolesť. V kontexte kybernetického obťažovania často ide o dospelú osobu, ktorá sa vydáva za niekoho iného, s cieľom získania dôvery dieťaťa. Po získaní dôvery ho zvyčajne následne zneužije alebo využije. Agresori používajú rôzne formy násilia, ako sú fyzické útoky, verbálna agresia a nepriame ubližovanie. Často pochádzajú z problematických rodinných prostredí, kde sami zažívajú násilie [11].

2) Obeť: Obeťami šikanovania môžu byť deti, ktoré sú buď pasívne alebo agresívne. Pasívne obete sú často plaché, málo sebavedomé a sociálne menej skúsené, čo ich robí zraniteľnými. Môžu trpieť psychosomatickými problémami a negatívnymi emóciami, ako sú smútok alebo beznádej. Naopak, agresívne obete sa bránia verbálne aj fyzicky, no dlhotrvajúce šikanovanie vedie k ich sociálnej izolácii [12].

3) Ďalšie osoby: Medzi ďalšie osoby patria svedkovia šikanovania, ktorí môžu zastávať rôzne úlohy: môžu byť pasívnymi pozorovateľmi, tzv. *neutráli*, ktorí nevstupujú do konfliktu, ďalej aktívnymi podporovateľmi agresora, tzv. *podporovatelia*, ktorí ho povzbudzujú a podnecujú tým, že reagujú na šikanovanie smiechom alebo prejavmi sympatie, alebo obhajcami obete, tzv. *obrancovia*, ktorí sa snažia zasiahnuť a poskytnúť jej pomoc [10].

III. Metódy prevencie obťažovania detí na internete

Prevencia obťažovania detí na internete zahŕňa rôzne stratégie a technológie, ktoré majú za cieľ znížiť riziko vystavenia detí nebezpečenstvám online prostredia. Medzi hlavné metódy prevencie patrí rodičovská kontrola, využívanie umelej inteligencie a vzdelávanie detí o bezpečnosti na internete [13].

A. Rodičovská kontrola

Rodičovská kontrola je jedným zo spôsobov, ako chrániť deti pred nebezpečenstvami na internete. Tento prístup zahŕňa využitie rôznych nástrojov na monitorovanie a obmedzenie online aktivity detí. Medzi najčastejšie používané metódy patria vzdelávacie a prispôsobené prehliadače, monitorovací softvér, časová kontrola a filtračný softvér. Vzdelávacie prehliadače sú špeciálne navrhnuté pre deti a blokujú prístup k nevhodným obsahom, zatiaľ čo inteligentné vyhľadávače vylučujú nevhodné webové stránky z výsledkov vyhľadávania [14].

B. Detekcia obťažovania pomocou umelej inteligencie

Umelá inteligencia je čoraz viac využívaná na detekciu a prevenciu kybernetického obťažovania. Algoritmy strojového učenia dokážu analyzovať veľké množstvo dát v reálnom čase a identifikovať rizikové správanie. Tieto systémy môžu analyzovať textové správy a hodnotiť ich obsah z hľadiska rizikovosti, čím pomáhajú identifikovať potenciálnych predátorov [15].

C. Vzdelávanie detí o bezpečnosti na internete

Otázka ochrany detí online je mimoriadne dôležitá a vyvoláva vážne obavy najmä preto, že deti sú na internete čoraz aktívnejšie. Trávia čoraz viac času online, či už na vzdelávacie alebo zábavné účely. Kvôli svojmu veku majú deti často problém rozpoznať riziká a môžu sa stať obeťami online zneužitia. Zvýšenie povedomia u detí je kľúčové pre prevenciu potencionálnych hrozieb. Cieľom je pomôcť deťom rozpoznať hrozby a aplikovať bezpečnostné opatrenia na ochranu svojich údajov a súkromia na internete [16], [17].

D. Linky pomoci

Linky pomoci sú určené na ochranu detí pred sexuálnym zneužívaním. Tieto linky umožňujú deťom bezpečne hovoriť o svojich problémoch a získať podporu od vyškolených profesionálov [18]. Ukázalo sa, že sú účinné a priemerne zaznamenávajú až 800 hovorov mesačne [19]. Rozšírenie týchto liniek by mohlo ešte viac posilniť dostupnosť pomoci a dosiahnuť lepšie výsledky v prevencii sexuálneho zneužívania detí [20].

IV. Návrh a implementácia webovej aplikácie

Prevencia obťažovania detí na internete je dnes mimoriadne dôležitá, preto sme sa zamerali na vývoj webovej aplikácie, ktorá by pomohla zvýšiť bezpečnosť detí v online prostredí. Naša aplikácia je navrhnutá tak, aby poskytovala interaktívne a vzdelávacie prostredie, ktoré pomáha deťom rozpoznať a reagovať na rizikové situácie na internete prostredníctvom rôznych konverzačných scenárov.

A. Návrh aplikácie

Aplikácia je navrhnutá tak, aby obsahovala rôzne scenáre a situácie simulujúce možné prípady obťažovania, nevhodného správania na internete alebo bežných konverzácií bez náznaku obťažovania. Deti sa tak učia identifikovať a správne reagovať na rizikové situácie online.

Webová stránka aplikácie obsahuje niekoľko hlavných častí:

1) Úvodná stránka: Poskytuje základné informácie, aktuálne novinky a kontaktný formulár.

2) Aplikácia: Hlavná časť aplikácie, kde sa používatelia zapájajú do interaktívnej hry, ktorá ich učí rozpoznávať a reagovať na potenciálne nebezpečné situácie na internete.

B. Implementácia databázy

Pre našu aplikáciu sme potrebovali vytvoriť dataset obsahujúci rôzne typy konverzácií. Tieto konverzácie boli prefiltrované, aby sme zabezpečili, že neobsahujú nevhodné slová alebo témy, ktoré by mohli byť pre deti nezrozumiteľné. Filtrovanie bolo nevyhnutné pre zachovanie vhodnosti a bezpečnosti obsahu pre deti.

Po vytvorení datasetu sme pristúpili k vytvoreniu a naplneniu tabuliek v databáze. Použili sme aplikáciu `Postico`, ktorá ponúka intuitívne rozhranie a skvelé možnosti práce s databázou. `Postico` nám umožnilo efektívne spravovať dátové štruktúry potrebné pre našu aplikáciu. Vytvorili sme tabuľky obsahujúce relevantné stĺpce na uchovávanie informácií o konverzáciách, ako napríklad ID konverzácie, obsah správy, časy a označenie, či sa jedná o konverzáciu s obťažovaním alebo nie.

Výsledný dataset obsahuje konkrétne 100 konverzácií, z ktorých 50 zahŕňa jednu z niektorých vyššie spomenutých foriem obťažovania, a zvyšné 50 sú bežné konverzácie bez náznaku akéhokoľvek obťažovania. Tento proces bol nevyhnutný na to, aby bola aplikácia nielen funkčná, ale aj bezpečná a vhodná pre mladšie publikum.

Po úspešnom vytvorení a naplnení tabuliek sme získali vyvážený a relevantný dataset, ktorý spĺňa všetky požiadavky na použitie v aplikácii pre deti.

V. Analýza výsledkov testovania respondentmi

Testovanie aplikácie bolo realizované medzi žiakmi základnej školy, pričom sa zameriavalo na rôzne vekové skupiny. Výsledky ukázali, že starší žiaci mali tendenciu dosahovať vyššie skóre a rýchlejšie dokončovať úlohy v porovnaní s mladšími žiakmi. Tento rozdiel môže byť spôsobený väčšími skúsenosťami a zručnosťami pri práci s počítačom. Mladší žiaci potrebovali viac času na dokončenie úloh a dosiahli nižšie skóre, čo naznačuje, že aplikácia môže byť pre túto vekovú skupinu náročnejšia.

Tabuľka 1 znázorňuje distribúciu respondentov podľa veku a pohlavia. Najväčšiu skupinu tvorili dievčatá vo veku $10-12$ rokov (13 respondentov), nasledované chlapcami vo veku $13-15$ rokov (7 respondentov). Dievčatá vo veku $13-15$ rokov tvorili 6 respondentov a najmenej početnou skupinou boli chlapci vo veku $10-12$ rokov (4 respondenti).

Tabuľka 1
Distribúcia respondentov podľa veku a pohlavia

Pohlavie/Vek	$10-12$ rokov	$13-15$ rokov
Dievča	13	6
Chlapec	4	7

Tabuľka 2 znázorňuje priemerný čas hrania podľa veku a pohlavia. Z výsledkov vyplýva, že dievčatá vo veku $10-12$ rokov dosahovali priemerný čas 5:31 minút, zatiaľ čo dievčatá vo veku $13-15$ rokov potrebovali na dokončenie hry priemerne 6:13 minút. Chlapci vo veku $10-12$ rokov dosahovali priemerný čas 7:37 minút, zatiaľ čo chlapci vo veku $13-15$ rokov potrebovali na dokončenie hry priemerne 5:31 minút.

Tabuľka 2
Priemerný čas hrania podľa veku a pohlavia

Pohlavie/Vek	$10-12$ rokov	$13-15$ rokov
Dievča	5:31	6:13
Chlapec	7:37	5:31

Tabuľka 3 znázorňuje priemerné skóre dosiahnuté respondentmi podľa veku a pohlavia. Dievčatá vo veku $10-12$ rokov dosahovali priemerné skóre 7 bodov, zatiaľ čo dievčatá vo veku $13-15$ rokov dosahovali priemerné skóre 8 bodov. Chlapci vo veku $10-12$ rokov dosahovali priemerné skóre 6 bodov a chlapci vo veku $13-15$ rokov dosahovali priemerné skóre 8 bodov.

Tabuľka 3
Priemerné skóre podľa veku a pohlavia

Pohlavie/Vek	$10-12$ rokov	$13-15$ rokov
Dievča	7	8
Chlapec	6	8

A. Návrhy na zlepšenie webovej aplikácie

Na základe získaných výsledkov sme navrhli niekoľko vylepšení aplikácie. Patria sem zavedenie úrovní obtiažnosti, ktoré by umožnili deťom pracovať so scenármi prispôsobenými ich schopnostiam, a postupné zvyšovanie obtiažnosti, čo by deťom umožnilo začať s jednoduchými konverzáciami a postupne prechádzať na zložitejšie. Implementácia funkcie sledovania času stráveného na jednotlivých rozhodnutiach by umožnila identifikovať konverzácie, ktoré sú pre deti najnáročnejšie, a prispôsobiť ich obtiažnosť.

VI. Vyhodnotenie

Aplikácia na prevenciu obťažovania detí na internete predstavuje významný krok v boji proti online hrozbám. Prostredníctvom interaktívnych scenárov a vzdelávacích nástrojov pomáha deťom rozpoznať a reagovať na rizikové situácie, čím prispieva k ich ochrane. Výsledky testovania a návrhy na zlepšenie poskytujú cennú spätnú väzbu, ktorá nám umožňuje neustále vylepšovať aplikáciu a prispievať k bezpečnejšiemu využívaniu internetu pre deti.

VII. Záver

Tento príspevok sa venoval moderným metódam prevencie obťažovania detí na internete. Rýchly rozvoj digitálnych technológií a rastúci počet detí používajúcich internet si vyžaduje vytváranie bezpečného online prostredia pre mladú generáciu.

V teoretickej časti sme objasnili problematiku obťažovania detí na internete, definovali kľúčové pojmy a analyzovali rôzne formy obťažovania. Skúmali sme existujúce metódy prevencie, ako sú rodičovská kontrola, detekcia obťažovania pomocou umelej inteligencie a vzdelávanie detí o bezpečnosti na internete.

Praktická časť sa zamerala na vytvorenie webovej aplikácie, ktorá hodnotí bezpečnosť výrokov v online konverzáciách. Testovanie aplikácie ukázalo, že staršie deti dosahovali lepšie výsledky, čo naznačuje jej intuitívnosť. Identifikovali sme oblasti na ďalšie vylepšenie, ako zavedenie rôznych úrovní obtiažnosti a poskytovanie okamžitej spätnej väzby.

Príspevok zdôraznil, že prevencia obťažovania detí na internete vyžaduje aktívnu spoluprácu všetkých zainteresovaných strán a zvýšenie povedomia o rizikách na internete. Edukačné programy a technológie zamerané na ochranu detí musia byť pravidelne aktualizované, aby reagovali na nové hrozby a trendy v online prostredí, čím môžeme efektívnejšie chrániť deti a podporovať ich bezpečný rozvoj.

Poďakovanie

Táto práca bola čiastočne podporená Vedeckou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky a Slovenskou akadémiou vied v rámci projektu VEGA 2/0165/21, Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky cez projekt KEGA 049TUKE-4/2024, čiastočne z projektu FakeDetect-Early Fake Profile Detection FBR-PDI-019 financovaného z Nórskych fondov a EHP 2014-2021, a čiastočne v rámci projektov podporených Agentúrou na podporu výskumu a vývoja APVV-22-0261 a APVV-22-0414.

Literatúra

[1] C. R. Center, "What is Bullying? - Cyberbullying Research Center — cyberbullying.org," https://cyberbullying.org/what-is-bullying, 2023.
[2] U. D. of Health and H. Services, "What Is Cyberbullying — stopbullying.gov," https://www.stopbullying.gov/cyberbullying/what-is-it, 2021.
[3] P. A. Staff, "Bullying Prevention and Intervention Strategies & Nebraska Department of Education — education.ne.gov," https://www.education.ne.gov/safety/bullying-prevention-and-intervention-strategies/, 2023.
[4] N. Kurniati, I. Purnamasari, and I. Rahmawati, "Analysis of the impact of verbal bullying on elementary school children," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 20, pp. 383–393, 2023.
[5] P. R. Borj, K. Raja, and P. Bours, "Online grooming detection: A comprehensive survey of child exploitation in chat logs," *Knowledge-Based Systems*, vol. 259, p. 110039, 2023.
[6] C. Dodds, "How is cyberbullying different from 'traditional' bullying? – Cybersmile — cybersmile.org," https://www.cybersmile.org/blog/how-is-cyberbullying-different-from-traditional-bullying, 2014.
[7] P. Jirásek, L. Novák, and J. Požár, *Výkladový slovník kybernetické bezpečnosti*. Policejní akademie ČR v Praze, 2015.
[8] J. Suler, "The online disinhibition effect," *Cyberpsychology & behavior*, vol. 7, no. 3, pp. 321–326, 2004.
[9] S. Wachs, G. K. Jiskrova, A. T. Vazsonyi, K. D. Wolf, and M. Junger, "A cross-national study of direct and indirect effects of cyberbullying on cybergrooming victimization via self-esteem," *Psicologia educativa*, vol. 22, no. 1, pp. 61–70, 2016.
[10] U. D. of Health and H. Services, "The Roles Kids Play in Bullying — stopbullying.gov," https://www.stopbullying.gov/bullying/roles-kids-play, 2024.
[11] J. Smith, S. W. Twemlow, and D. W. Hoover, "Bullies, victims and bystanders: A method of in-school intervention and possible parental contributions," *Child psychiatry and human development*, vol. 30, pp. 29–37, 1999.
[12] X. Garcia-Continente, A. Pérez-Giménez, A. Espelt, and M. Nebot Adell, "Bullying among schoolchildren: differences between victims and aggressors," *Gaceta Sanitaria*, vol. 27, pp. 350–354, 2013.
[13] E. Buçaj and F. Haziri, "Cyberbullying and cyberstalking: Parents' role to protect their children," *Multidisciplinary Reviews*, vol. 7, no. 4, pp. 2 024 079–2 024 079, 2024.
[14] H. H. Altarturi, M. Saadoon, and N. B. Anuar, "Cyber parental control: A bibliometric study," *Children and Youth Services Review*, vol. 116, p. 105134, 2020.
[15] A. Faraz, J. Mounsef, A. Raza, and S. Willis, "Child safety and protection in the online gaming ecosystem," *IEEE Access*, vol. 10, pp. 115 895–115 913, 2022.
[16] N. E. Willard, *Cyber-safe kids, cyber-savvy teens: Helping young people learn to use the Internet safely and responsibly*. John Wiley & Sons, 2007.
[17] E. Kupcová, R. Haluška, M. Popovič, M. Pleva, M. S. Heng, and P. Bours, "Early recognition of the speaker's age," in *2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, 2023, pp. 341–347.
[18] L. Faithfull, "Who We Are | Stop It Now — stopitnow.org," https://www.stopitnow.org/about-us/who-we-are, 2024.
[19] S. I. Now, "Call examples - Stop It Now — stopitnow.org.uk," https://www.stopitnow.org.uk/how-we-prevent-child-sexual-abuse/call-examples/, 2024.
[20] S. Craven, S. Brown, and E. Gilchrist, "Current responses to sexual grooming: Implication for prevention," *The Howard Journal of Criminal Justice*, vol. 46, no. 1, pp. 60–71, 2007.

HMI rozhranie a komunikácia s Master zariadením pre ovládanie žalúzií

[1]*Milan LACKO,* [2]*Andrej KLEIN*

[1] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika
[2] SEZ Krompachy, a.s., Krompachy, Slovenská republika

[1]milan.lacko@tuke.sk, [2]andrejklein123@gmail.com

***Abstrakt* — Článok opisuje návrh a realizáciu systému na ovládanie vonkajších žalúzii v rodinnom dome. Stručne je opísané zloženie celého systému po hardvérovej stránke a tiež vytvorenie HMI rozhrania pre ovládanie a nastavovanie systému. Opísaný je aj spôsob komunikácie a vytvorený komunikačný protokol s konkrétnymi pokynmi medzi displejom a Master zariadením.**

***Keywords* — HMI, Nextion, displej, mikrokontrolér, komunikácia**

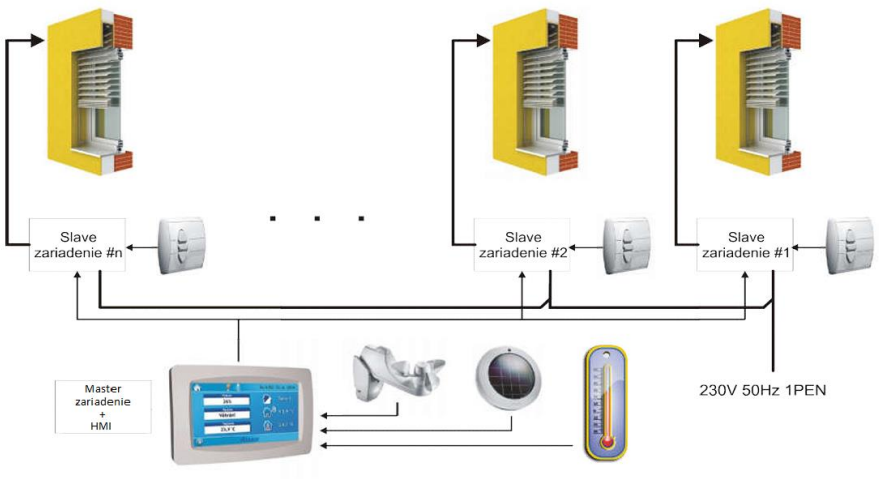
I. Úvod

V dnešnej dobe je automatizácia inteligentných domácností stále viac prítomná. A to najmä v oblastí smart technológií a riadenia domácností pomocou užívateľských rozhraní alebo embedded systémov. Do tejto kategórie inteligentných domácností spadá aj riadenie vonkajších žalúzií. Efektívne a pohodlné ovládanie týchto zariadení môže prispieť k zvýšeniu komfortu užívateľov a optimalizácii energetickej náročnosti domu. Takéto systémy sú často riešené ako automatizované, avšak užívateľ si praje mať možnosť zasahovať do systému podľa vlastných požiadaviek, a to prostredníctvom vhodného rozhrania. V dnešnej dobe často pomocou smartfónu prostredníctvom aplikácie alebo pomocou dotykového displeja daného systému. Práve systém pomocou dotykového displeja je použitý pre tento systém ovládania tienenia.

II. Opis celého systému ovládania vonkajších žalúzii

Celkový systém ovládania žalúzii na konkrétnom rodinnom dome (Obr. 1) je možné rozdeliť na štyri dôležité časti, a to:

- Samotné vonkajšie žalúzie, ktoré sú motoricky ovládané (230V AC)
- Slave zariadenie, ktoré riadi konkrétnu žalúziu s prepojením na lokálny ovládač
- Master zariadenie pre komplexné ovládanie a nastavovanie systému
- HMI rozhranie v podobe dotykového displeja Nextion, ktorý komunikuje s Master zariadením po sériovej linke.



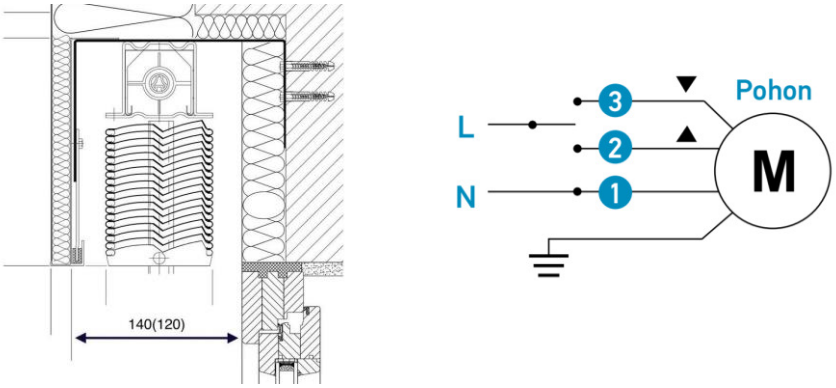
Obr. 1. Grafické znázornenie celého systému ovládania žalúzii [1]

Systém pozostáva z ôsmich vonkajších žalúzii, z toho 3 na prízemí a 5 na poschodí a v rámci orientácie svetových strán sú 4 žalúzie orientované na juh a 4 na západ. Ku každej žalúzii prislúcha Slave zariadenie osadené v inštalačnej krabičke priamo za lokálnym ovládačom.

Informácie o vonkajšom prostredí (teplota, vlhkosť, rýchlosť vetra, intenzita osvetlenia juh/západ) je spracovaná iným systémom a údaje sú posielané do Master zariadenia, ktoré tieto informácie pravidelne posiela všetkým Slave zariadenia, ktoré v automatickom režime nastavujú polohu a natočenie jednotlivých žalúzii podľa vopred naprogramovanej logiky. Užívateľ á možnosť zasiahnuť do tohto automatického režimu buď lokálnym ovládačom alebo Master zariadením prostredníctvom dotykového displeja.

A. Vonkajšie žalúzie

V konkrétnej aplikácii tienenia rodinného domu je použitým typom žalúzia Z-90 (Obr. 2 vľavo). Ide o lamely v tvare písmena „Z“ o šírke 90 cm. Vonkajšie žalúzie Z-90 sú technologicky najvyspelejšie žalúzie v oblasti exteriérovej tieniacej techniky, ktorá je vhodná k zatieneniu rodinných domov i administratívnych budov. V uzatvorenom stave vytvárajú lamely ucelenú plochu, ktorá zaručuje dokonalé zatienenie. Guma, vlisovaná po celej dĺžke lamely zvyšuje termoregulačný efekt žalúzie a postará sa aj o odhlučnenie, v prípade narážania či trenia lamiel o seba pri vetre. Odolnosť žalúzie voči vetru je udávaná do 50 km/h.



Obr. 2. Nákres žalúzie Z-90 a schéma zapojenie pohonu vonkajšej žalúzie [2]

B. Slave zariadenie

Jedná sa o zariadenie na báze mikrokontroléra Atmega32, ktoré priamo ovláda pohon vonkajšej žalúzie, t.j. pripája 230 V AC na motor pre pohon žalúzie buď smerom hore alebo dole (Obr. 2 vpravo). Na toto prepínanie slúžia 2 relé. Celé zariadenie je umiestnené v krabičke za samotným ovládačom Inos Uno VB od fy Somfy (Obr. 3 vľavo). Hlavnou úlohou zariadenia je autonómne ovládanie žalúzie na základe informácii od Master zariadenia príp. nastavovanie žalúzie podľa pokynov z lokálneho ovládača alebo z dotykového displeja, čo z pohľadu systému znamená zadanie pokynu z Master zariadenia (priame nastavenie, budík, umývanie okien).

Samotné zariadenie ešte obsahuje svorkovnicu pre prívod a dva vývody fázového vodiča na motor, svorkovnicu na napájanie a komunikáciu s Master zariadením (12V, GND, TxD, RxD), konektor pre pripojenie lokálneho ovládača, konektor pre programovanie (ISP), konektor pre pripojenie fotorezistora a prepínač servisného režimu. Pohľad na osadenú DPS Slave zariadenia je na Obr. 3 vpravo.

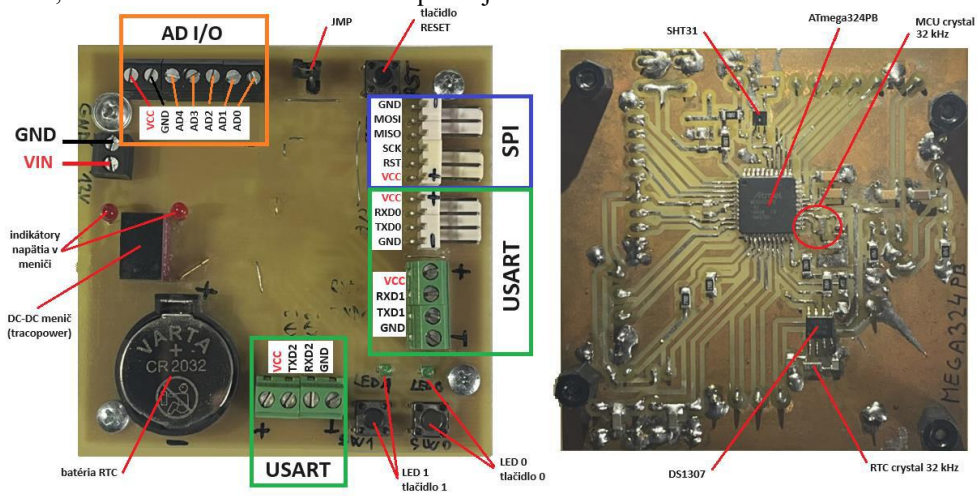


Obr. 3. Ovládač Inis Uno VB [2] a osadená DPS Slave zariadenia [9]

C. *Master zariadenie*

Pre projekt ovládania exteriérových žalúzií bola navrhnutá a zhotovená vývojová doska MASTER, na ktorej sa nachádza čip Atmega324PB. Pohľad na zhotovenú Master DPS aj s legendou je na Obr. 4. Táto vývojová doska má zabudovaný RTC modul (DS1307) so zabudovanou batériou a vlastným kryštálom 32,768 kHz, snímač relatívnej vlhkosti a teploty (SHT31), 2 tlačidlá a 2 LED indikátory, ktoré budú využité len počas procesu ladenia programu. Okrem toho má predpripravenú analógovo-digitálnu vstupno-výstupnú zbernicu pre budúce využitie priamo na doske. V budúcnosti sa uvažuje s rozšírením systému o signály zo snímačov vetra, intenzity vonkajšieho osvetlenia a vonkajšej teploty. Niektoré z týchto údajov môžu byť pripojené práve cez tieto I/O. V prípade potreby bude možné využiť aj UART2 zbernicu, keďže ATmega324PB disponuje troma UART zbernicami. Jedna je vyhradená na komunikáciu s displejom Nextion (UART0) a druhá (UART1) je určená na komunikáciu so Slave zariadeniami. Master zariadenie je možné programovať pomocou sériovej SPI zbernice, čo je štandard pre AVR zariadenia známy ako ISP – In-System Programming.

Komunikácia so Slave zariadeniami je typu Master – multi slave, kde na jednej zbernici UART sú súčasne pripojené 8 zariadenia Slave. Adresovanie jednotlivých Slave zariadení je podľa ID žalúzie, keďže konkrétnu žalúziu ovláda práve jedno Slave zariadenie.



Obr. 4. Zhotovené MASTER zariadenie pre ovládanie žalúzii a komunikáciu s displejom

D. *Nextion Displej*

Samotné HMI rozhranie je realizované pomocou dotykového displeja fy. Nextion. Jedná sa o 5″ dotykový displej Nextion s rozlíšením 800x480 pixelov základnej série typu NX8048T050 (Obr. 5). Displej je vybavený 48 MHz mikroprocesorom, ktorý umožňuje samostatné fungovanie bez potreby externého MCU. Displej obsahuje 3584 bajtov pamäte RAM a 16 MB internej Flash pamäte pre ukladanie grafických prvkov, textov, čísiel a iných dát na zobrazenie GUI pre náš HMI projekt. Má zabudovaný SD slot pre nahrávanie programu a uloženie dát. Dotyková obrazovka má 65536 RGB farieb a podsvietenie je typu LED.[5][9]



Obr. 5. Nextion 5” dotykový displej NX8048T050

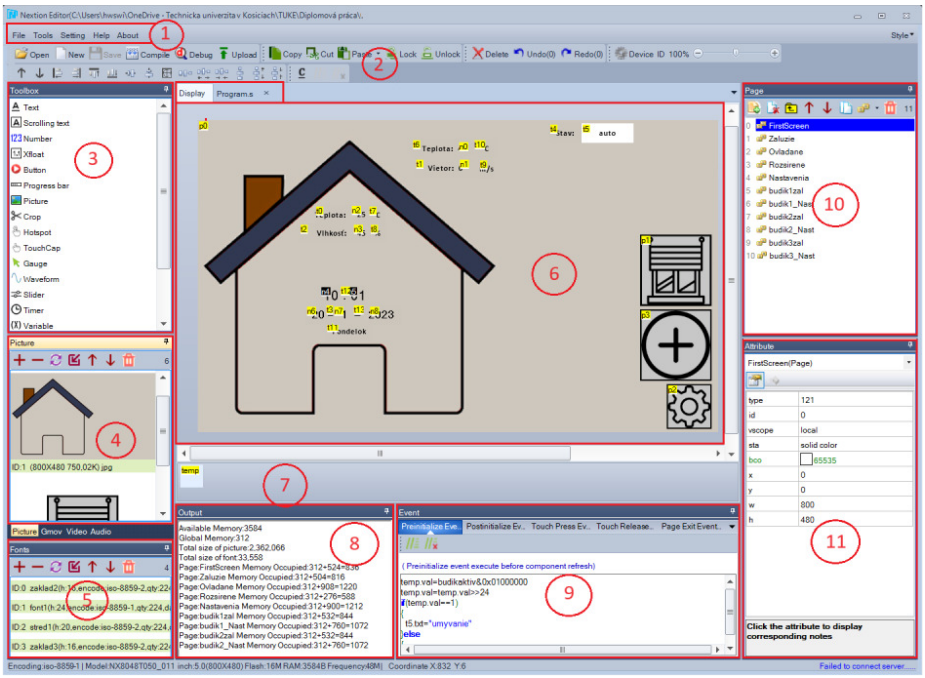
Odporúčané napájacie napätie je 5 V a prevádzkový prúd pri jase 100 % dosahuje 410 mA. Nextion displej pracuje aj v režime spánku, kedy prúd klesne na hodnotu 15 mA. Displej nemá USB rozhranie, ale zabudovaný slot na SD kartu (formát FAT32), ktorý podporuje maximálne 32G micro SD kartu. Táto karta sa používa výhradne na aktualizáciu firmvéru a nami vytvorenej aplikácie GUI. Pre komunikáciu s displejom je využívaný sériového portu typu TTL, ktorý

pozostáva so 4 pinov (VCC, RX, TX a GND). Štandardná prenosová rýchlosť portu môže byť 2400 bps, 9600 bps alebo 115200 bps. Nextion displej využíva na ukladanie textov, písma a obrázkov 16 MB Flash pamäť. Na ukladanie premenných využíva pamäť RAM. [9]

III. GRAFICKÝ EDITOR NEXTION

Najväčšou výhodou dotykových displejov Nextion je grafický editor, ktorý výrobca poskytuje, a v ktorom sa dá ľahko a bez problémov nadefinovať grafické a dotykové prostredie a vykonať simulácie.

Nextion editor je bezplatný vývojový softvér, ktorý sa používa na rýchle vytváranie používateľských rozhraní GUI pre dotykové displeje Nextion [10]. Tento editor obsahuje nástroje na grafické navrhovanie rozhraní, pridávanie interaktívnych prvkov a programovanie funkcionality pomocou jednoduchého programovacieho jazyka EDL (Event-Driven Language). Rýchle vytvorenie *HMI* pre Nextion displej nevyžaduje extrémne zručnosti, no očakávajú sa aspoň základné zručnosti v programovaní. Záťaž spojená s grafikou je priamo prenesená do ARM procesora, ktorý sa stará o samotné vykresľovanie na displej. Pohľad na grafické rozhranie editora je uvedený na Obr. 6.



Obr. 6. Pohľad na grafický editor Nextion

Pre lepšie pochopenie pracovného prostredia editora sú jednotlivé časti editora označené legendou. Význam legendy je nasledovný:

1 – hlavné menu, 2 – panely nástrojov, 3 – panel s nástrojmi, 4 – panel zdrojov obrázkov, 5 – panel zdrojov písma, 6 – dizajnové plátno aktuálnej obrazovky (vizuálne komponenty), 7 – nevizuálne komponenty – lokálne premenné, 8 – výstup, 9 – karty udalostí, 10 – panel všetkých obrazoviek (stránok), 11 – panel atribútov komponentu.

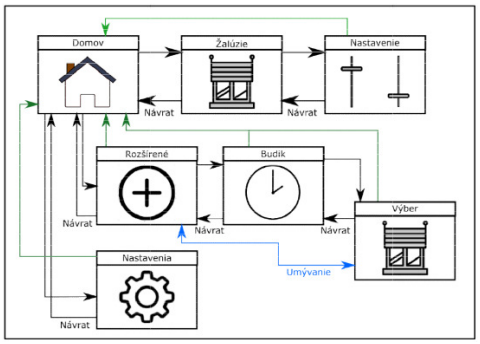
Tu možno spomenúť, že asi najdôležitejšia časť je karta udalostí (9), kde si podľa svojich potrieb vieme jednoduchými príkazmi naprogramovať činnosť, ktorá sa ma udiať pri stlačení/uvoľnení ľubovoľného elementu na konkrétnej obrazovke alebo môžeme vykonať nejaké akcie pred a po zobrazením danej obrazovky alebo pri opustení danej obrazovky. Tento systém práce s displejom sme využili aj pri návrhu nášho systému, čím sme preniesli veľkú časť zaťaženia na procesor displeja a mikrokontrolér nášho Master zariadenia už spracovával len konkrétne povely (správy) z displeja. Nemusel sa zaoberať každým ťuknutím či uvoľnením rôznych tlačidiel na displeji a nemusel spracovávať veľké množstvo premenných. Tým sa nám samotný program Master zariadenia značne zjednoduší.

IV. NÁVRH HMI POMOCOU NEXTION DISPLEJA

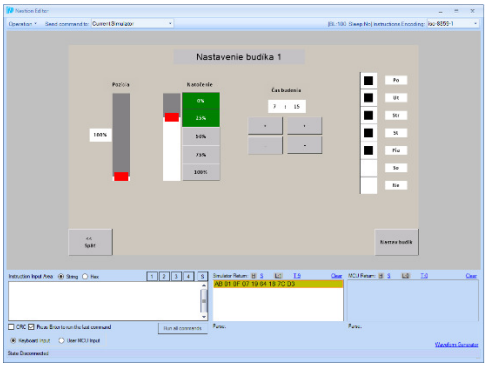
A. *Dizajn obrazoviek*

Bol navrhnutý dizajn jednotlivých obrazoviek (stránok) displeja podľa potrieb užívateľa pre konkrétnu aplikáciu, čiže konkrétny rodinný dom. V dome je osadených 8 vonkajších žalúzii, ktoré sú orientované na juh a západ a tiež na prízemí a poschodí. Bol urobený návrh toho, čo užívateľ bude potrebovať nastavovať a ovládať. Pomocou grafického editora Nextion sme vytvorili jednotlivé obrazovky, ktoré na seba nadväzujú v závislosti od stlačenia príslušných ikon. V rámci jednotlivých vetiev (nastavení) sa užívateľ prekliká až ku koncu vetvy, kde potvrdením je odoslaná celá správa (paket dát) do Master zariadenia, ktorá obsahuje všetky potrebné údaje. Na základe tejto správy Master rozpošle vybraným Slave zariadeniam správu, ktorá jasne definuje, čo má Slave zariadenie urobiť, či už zmeniť nejaké nastavenia pre konkrétnu žalúziu alebo danú žalúziu priamo nastaviť (zmeniť polohu alebo natočenie). Navrhnutý pohyb po jednotlivých obrazovkách displeja je zjednodušene naznačený na Obr. 7.

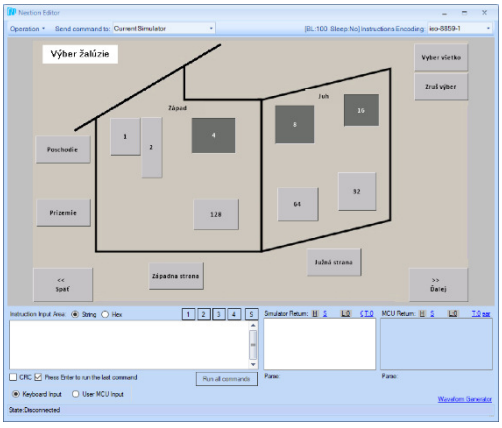
Základnou obrazovkou je *Domovská* obrazovka, kde je stále zobrazený aktuálny čas a dátum, teplota a vlhkosť v miestnosti. Z tejto obrazovky si užívateľ vyberá pomocou 3 ikon, či chce priamo nastavovať žalúzie, alebo vojsť do rozšírených možností na nastavenie budíka alebo režimu umývania okien alebo, do tretice, má možnosť zvoliť si nastavenia, kde nastavuje systém (čas, dátum, jas displeja). Konkrétny pohľad na vybrané vytvorené obrazovky (stránky) v simulačnom prostredí je zobrazený na Obr. 8 až Obr. 10.



Obr. 7. Schematické znázornenie pohybu po jednotlivých obrazovkách displeja [10]



Obr. 8. Obrazovka pre nastavenie budíka č. 1



Obr. 9. Obrazovka na výber žalúzii



Obr. 10. Obrazovka na nastavenie vybraných žalúzií

A. *Formát odosielaných správ z displeja Nextion do Master zariadenia*

Aby bolo možné efektívne komunikovať medzi displejom a Master zariadením, bolo potrebné navrhnúť komunikačný protokol. Pre komunikáciu smerom do Master zariadenia bola navrhnutá správa (paket), ktorá sa odosiela z displeja, keď užívateľ prejde jednotlivými obrazovkami, navolí jednotlivé parametre a na konci vetvy stlačí príslušnú ikonu, ktorá má konkrétny význam, pokyn, čo sa má vykonať. Správa bola stanovaná na pevnú dĺžku 9 bajtov, kde 1. bajt je hodnota, ktorá znamená konkrétny pokyn. Každý pokyn má jedinečnú hodnotu. Za tým nasleduje 7 bajtov s parametrami, kde je presne určené, aké parametre sa posielajú s daným pokynom. Ak niektorý pokyn nepotrebuje toľko parametrov, správa je doplnená nulami. Posledný 9. bajt správy je

„checksum“, čo je kontrolný súčet predošlých bajtov orezaný na 8 najmenej významových bitov. Slúži pre kontrolu komunikácie. Ak je správa doručená v poriadku, na obrazovke sa vysvieti nápis „*Čakajte.....prebieha komunikácia*“ a po krátkom zobrazení tohto oznamu sa na displeji zobrazí *Domovská obrazovka*. Takto užívateľ vie, že pokyn bol Master zariadením prijatý a spracovaný. V Tabuľke 1 sú znázornené všetky potrebné pokyny s jednotlivými parametrami pre navrhnutý systém.

Tabuľka 1
Zoznam všetkých pokynov pre Master zariadenie so znázornenými parametrami

Pokyn	1.bajt	2.bajt	3. bajt	4. bajt	5.bajt	6. bajt	7. bajt	8.bajt	9.bajt
Priame zadanie	0xA1	ID žal	Pozícia	Natočenie	0	0	0	0	CHCKS
Obmedzenie pozície	0xA3	ID žal	Pozícia	0	0	0	0	0	CHCKS
Zadanie doby (lokálna a master)	0xA5	ID žal	Minúty lokálnej doby	Minúty master doby	0	0	0	0	CHCKS
Umývanie okien	0xA7	ID žal	Stav 0/1	0	0	0	0	0	CHCKS
Zrušenie man. režimu	0xA8	ID žal	0	0	0	0	0	0	CHCKS
Nastavenie dátumu a času	0xA9	Rok	Mesiac	Deň	Hodina	Minúta	Deň	0	CHCKS
Nastavenie budíka 1	0xAB	Stav	Minúta	Hodina	Natočenie	Pozícia	ID žal	Deň	CHCKS
Nastavenie budíka 2	0xAD	Stav	Minúta	Hodina	Natočenie	Pozícia	ID žal	Deň	CHCKS
Nastavenie budíka 3	0xAF	Stav	Minúta	Hodina	Natočenie	Pozícia	ID žal	Deň	CHCKS

Je potrebné podotknúť, že generovanie týchto správ spolu s parametrami je výsledkom vhodného naprogramovania Nextion displeja za pomoci grafických elementov používaných na jednotlivých obrazovkách, využitím lokálnych a globálnych premenných a jednoduchých programov napísaných v kartách udalostí grafického editora (Obr. 6 - legenda 9) pre jednotlivé obrazovky a jednotlivé grafické prvky.

Práve takto je napr. vytvorený systém identifikovania zvolených žalúzii (parameter ID žal). Keďže je žalúzii 8, zvolili sme si, že každá pozícia bitu v bajte *ID žal* predstavuje práve jednu konkrétnu žalúziu. Ak je zvolená daná žalúzia, potom na príslušnej pozícii je „log. 1“, inak je tam „log. 0“. Tento systém adresovania je následne využitý aj v komunikácii Master – multi slave, kde práve adresa Slave zariadenia bude daná parametrom *ID žal*, a všetky Slave zariadenia, ktoré na svojej pozícii príjmu „log. 1“ vedia, že daná správa sa ich týka.

B. Formát odosielaných správ z Master zariadenia do displeja Nextion

Pre odosielanie správ na displej je potrebné dodržať formát správy stanovený výrobcom. Každá správa musí byť ukončená 3 bajtami 0xFF. Zároveň formát správ je znakový, tzn. Musíme posielať konkrétne príkazy aj s parametrami po jednotlivých znakoch podľa ASCII tabuľky. Čo sa týka príkazov, displej Nextion disponuje množstvom príkazov pre nastavovanie hodnôt konkrétnych elementov alebo pre zobrazenie stránky, či konkrétneho objektu na danej stránke alebo príkazy pre samotné systémové nastavenie displeja. Dôležité je spomenúť, že ak posielame príkaz, ktorý ovplyvňuje konkrétny prvok, tak to platí len pre práve zobrazenú obrazovku (stránku). Kompletný zoznam príkazov aj s parametrami je k dispozícii v inštrukčnej sade displeja [7].

V našom systéme sme využívali príkazy na nastavovanie hodnôt jednotlivých elementov na úvodnej obrazovke, ako je dátum, čas, teplota či vlhkosť a taktiež po odoslaní pokynu z displeja do Master zariadenie Master odpovedal zviditeľnením oznamu a prepnutím obrazovky.

Napr. príkaz pre zobrazenie Domovskej obrazovky je „page 0“, pretože *Domovská* obrazovka (stránka) má označenie 0. Preto je potrebné zaslať správu s týmito bajtami: 0x70 0x61 0x67 0x65 0x20 0x30 0xFF 0xFF 0xFF. V ASCII formáte sme v podstate do displeja vyslali tieto znaky: „p“, „a“, „g“, „e“, „ “, „0“ a 3 ukončovacie bajty 0xFF.

V. ZÁVER

Článok opisuje návrh a realizáciu systému riadenia vonkajších žalúzii pre konkrétny rodinný dom. Boli stručne opísané a ukázané základné komponenty systému, ktoré sa použili, ich vlastnosti a tiež načrtnutá predstava o fungovaní systému riadenia vonkajších žalúzii. Bližšie bol opísaný systém vytvorenia obrazoviek pre HMI rozhranie pomocou grafického editora a tiež spôsob komunikácie medzi dotykovým displejom Nextion a Master zariadením v oboch smeroch.

Boli vytvorené jednotlivé obrazovky podľa požiadaviek užívateľa tak, aby bolo užívateľské rozhranie schopné pokryť všetky potreby, čo sa týka nastavovania a ovládania systému. Funkčnosť systému HMI a Master zariadenia bola aj reálne otestovaná. Ďalej je možné zaoberať

sa vylepšením estetického vzhľadu HMI rozhrania a doprogramovaním Slave zariadení, aby bolo možné systém uviesť do ostrej prevádzky.

POĎAKOVANIE

Tento článok vznikol s podporou Kultúrnej a edukačnej grantovej agentúra MŠVVaM SR v rámci projektu KEGA 059TUKE-4/2024 a v rámci projektu 032TUKE-4/2024

LITERATÚRA

[1] LACKO, Milan – BIROŠ, Milan. 2017. Inteligentné riadenie vonkajších žalúzii v rodinnom dome pomocou mikrokontrolérov. In: Elektrotechnické Listy, [online]. Roč. 2, č. 1 [cit. 2020-3-14]. ISSN: 2453-898. Dostupné na internete: http://elektrotechnickelisty.eu/casopis/rocnik_II/clanky/EL_07_2017.pdf

[2] SOMFY, spol. s.r.o., J4_WT_CZ_Úplný návod k použití_A_2017 (PDF). In: Somfy, 2017 [online]. Dostupné na internete: <https://www.kasko-vs.cz/uploads/navody/St%C3%ADn%C3%ADc%C3%AD%20technika/Venkovn%C3%AD%20%C5%BEaluzie/MOTORY/WT/J4_WT_CZ_%C3%9Apln%C3%BD%20n%C3%A1vod%20k%20pou%C5%BEit%C3%AD_A_2017.pdf>.

[3] Datasheet ATmega324PB, [online PDF] [Dátum: 2024] Dostupné na webovej stránke: https://www.microchip.com/ en-us/ product/ ATMEGA324PB

[4] Datasheet ATmega32, [online PDF] Dostupné na webovej stránke: https://ww1.microchip.com/ downloads/en/DeviceDoc/doc2503.pdf

[5] Nextion, Basic Series Introduction, [online] [Dátum: 2024] Dostupné na webovej stránke: <https://nextion.tech/basic-series-introduction>.

[6] Nextion, NX8048T050 datasheet, [online] [Dátum: 2024] Dostupné na webovej stránke: https://nextion.tech/datasheets/nx8048t050/#7

[7] The Nextion Instruction Set, [online], Dostupné na webovej stránke: https://nextion.tech/instruction-set/

[8] A., Klein: Master zariadenie s dotykovým displejom, Diplomová práca, 2024, Fakulta elektrotechniky a informatiky, TU Košice

[9] D., Marcin: Slave zariadenie pre autonómne ovládanie žalúzii, Bakalárska práca, 2020, Fakulta elektrotechniky a informatiky, TU Košice

[10] D., Sekeľ: Master zariadenie pre ovládanie žalúzii, Diplomová práca, 2020, Fakulta elektrotechniky a informatiky, TU Košice

Vyhodnocovanie nameraných údajov a ich vizualizácia pomocou MATLAB aplikácie

[1]*Simona KIREŠOVÁ,* [2]*Milan GUZAN*

[1,2] Katedra teoretickej a priemyselnej elektrotechniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]simona.kiresova@tuke.sk, [2]milan.guzan@tuke.sk

***Abstrakt* — Tento článok sa zaoberá vytvorením MATLAB aplikácie, ktorá umožňuje importovať, analyzovať a vizualizovať údaje zo senzorickej siete, ktorá slúži na monitorovanie poveternostných podmienok a kvality ovzdušia. Pomocou modulu MATLAB App Designer bola vytvorená aplikácia s grafickým rozhraním, ktorá slúži na vyhodnocovanie, sledovanie a porovnávanie nameraných hodnôt, výpočet priemerných hodnôt, korelácií a indexu kvality ovzdušia.**

***Kľúčové slová* — aplikácia, MATLAB, vizualizácia, vyhodnocovanie údajov**

I. Úvod

V súčasnosti existuje mnoho zdrojov, ktoré generujú veľké množstvo údajov potrebných na analýzu. Senzorické siete tvorené častokrát desiatkami až stovkami uzlov, zhromažďujú údaje z meracích staníc rozmiestnených po rôznych lokalitách [1]. Zariadenia IoT (internet vecí) využívame v každodennom živote v domácnosti (napr. inteligentné termostaty regulujúce teplotu v domácnosti, fitness náramky či inteligentné osvetlenie), ale aj v priemysle, ako sú napríklad zariadenia a systémy na monitorovanie výrobného procesu [2]. Sociálne siete taktiež produkujú obrovské objemy dát [3]. Veľké online obchody využívajú systémy pre zaznamenávanie transakcií a správanie zákazníkov [4]. V zdravotníctve sú zhromažďované údaje z elektronických zdravotných záznamov a medicínskych zariadení [5]. Tieto rozsiahle datasety predstavujú výzvu aj príležitosť pre ich analýzu a vizualizáciu. Je teda kľúčové ich efektívne spracovať.

S veľkým množstvom údajov sme sa stretli aj pri vytvorení senzorickej siete pre monitorovanie kvality ovzdušia a poveternostných podmienok. Vytvorená senzorická sieť [6] je umiestnená prevažne na východnom Slovensku a produkuje obrovské množstvo údajov o meteorologických a environmentálnych parametroch. V čase písania tohto článku ju tvorí 23 senzorických uzlov v 16 lokáciách, pre merania v interiéri, exteriéri a priemysle. V budúcom vývoji tejto senzorickej siete bude samozrejme prioritou túto sieť rozšíriť o ďalšie lokality aj mimo východné Slovensko. Tieto údaje je potrebné nielen zhromažďovať, ale aj analyzovať a vizualizovať, preto bolo potrebné vytvoriť aplikáciu umožňujúcu efektívne spracovanie nameraných údajov – od ich načítania, cez ich spracovanie a analýzu, až po ich vizualizáciu. Namerané údaje sú odosielané a ukladané do databázy InfluxDB [7]. Pre vizualizáciu bolo použitých viacero platforiem, každá z nich slúžiaca špecifickým potrebám používateľov. Zatiaľ čo databáza InfluxDB ponúka určité možnosti vizualizácie pomocou ovládacích panelov, viac možností vizualizácie ponúka nástroj Grafana, ktorej implementácia bola popísaná v článku [8]. Avšak pre prezeranie vytvorených grafických vizualizácií na mobilných zariadeniach, ktoré majú vertikálnu obrazovku, je vhodnejšie vytvoriť mobilnú aplikáciu, ktorá by komunikovala s databázou InfluxDB a zobrazovala by aktuálne namerané údaje, čo bolo implementované v článku [9]. Stále však ostala nepokrytá potreba vyhodnocovania údajov zapísaných do *.csv súboru (vykonávanie rôznych výpočtov, napr. hodinové a denné priemery, AQI a korelácie). Pre tieto vyhodnocovanie účely bude slúžiť MATLAB aplikácia (spustiteľná na PC). V nasledujúcich podkapitolách budú podrobne predstavené jednotlivé možnosti vizualizácie údajov.

Ako platforma pre vývoj aplikácie bol zvolený MATLAB (produkt spoločnosti MathWorks). Ide o výkonný nástroj pre vykonávanie numerických výpočtov a vizualizáciu dát. Svoje využitie má v rôznych technických a vedeckých disciplínach vďaka bohatej knižnici nástrojov a funkcií, ktoré umožňujú efektívne spracovanie veľkého množstva údajov a ich prezentáciu v grafickej

podobe [10]. Pre vyhodnotenie údajov bez potreby grafického prostredia si používateľ vystačí s jednoduchým skriptom. Ak je však potrebné vytvoriť aplikáciu s grafickým rozhraním, potom je možné použiť nástroj MATLAB App Designer. Aplikáciu je možné vytvoriť pomocou drag-and-drop komponentov a naprogramovať ich správanie priamo v prostredí MATLABu. Okrem toho je možné vytvoriť rôzne pomocné funkcie, ktoré sú používané napr. pri opakovaných výpočtoch [11].

Účelom vytvorenej aplikácie je zjednodušenie vyhodnocovacích procesov, ako je načítanie údajov zo senzorov, výpočet priemerov, korelácií a indexu kvality ovzdušia (AQI). Vďaka tomu umožňuje používateľom efektívne analyzovať a interpretovať zhromaždené údaje, čo vedie k lepšiemu pochopeniu meraných veličín ako aj vzťahov, ktoré medzi nimi existujú.

II. VYHODNOCOVANIE ÚDAJOV

Vytvorenú senzorickú sieť tvorí 17 meracích staníc určených na merania v exteriéri, pričom takmer všetky meracie stanice sa nachádzajú v mestách a obciach na východnom Slovensku, okrem dvoch staníc v malej dedinke na severovýchode Maďarska (Füzérkomlós), neďaleko od hraníc Maďarska a Slovenska. Zo 17 staníc sú 2 stanice neaktívne – už ukončili svoje merania. Meracie zariadenia sú umiestnené na rôznych miestach ako sú parapety, balkóny a záhrady rodinných domov alebo univerzitná budova.

Na meranie hmotnostnej a početnej koncentrácie pevných častíc, ich typickej veľkosti, VOC a NOx indexu, teploty a vlhkosti používame senzor SEN5x [12]. Konkrétne model SEN55 meria všetky zo spomenutých veličín, pričom SEN54 meria všetky veličiny okrem NOx indexu. (Väčšina staníc obsahuje model SEN54.) Na meranie atmosférického tlaku je použitý senzor MS5611 [13]. Niektoré stanice obsahujú aj senzor SCD41 na meranie koncentrácie CO_2 v ovzduší [14] a anemometer na meranie rýchlosti vetra [15].

Uzly v senzorickej sieti merajú mnoho parametrov, čím sa vytvára veľké množstvo údajov a vzniká potreba ich efektívneho vyhodnocovania. Okrem odosielania nameraných hodnôt do databázy InfluxDB, ich taktiež zapisujú na SD kartu do *.csv súboru v nasledujúcom formáte: **Datum,PM1,PM2p5,PM4,PM10,NC0p5,NC1,NC2p5,NC4,NC10,typSize,VOC,Temp,Hum, Pres**(**,NOx,CO2,WS**). Na vyhodnocovanie údajov zapísaných do *.csv súboru bola vytvorená MATLAB aplikácia použitím nástroja MATLAB App designer [11].

Z nameraných hodnôt sú pomocou funkcie *mean()* vypočítané hodinové a denné priemery. Hodinové priemery (ktoré minimalizujú šum a vyhladzujú krátkodobé fluktuácie a extrémne hodnoty) sú použité na výpočet Pearsonovho korelačného koeficientu (r) pre časový interval jedného dňa medzi PM a ostatnými veličinami (teplota, vlhkosť, tlak, rýchlosť vetra, VOC, NOx, CO_2). Na výpočet je použitá funkcia *corrcoef()* [16].

Na určenie štatistickej významnosti vypočítaných korelačných koeficientov bol použitý post-hoc test. Nulová hypotéza znie, že medzi PM a ostatnými meranými veličinami neexistuje významná korelácia, teda H_0: $r = 0$. P-hodnota bola testovaná pri úrovni významnosti $\alpha = 0{,}05$ (t. j. ak platí, že $p < \alpha$, tak môžeme zamietnuť nulovú hypotézu a r je štatisticky významný). P-hodnota je vypočítaná spolu s r pomocou funkcie v nasledujúcom formáte: $[r\ p]$ = *corrcoef*(X, Y), pričom X a Y sú merané veličiny, medzi ktorými sa počíta korelácia [16].

Denné priemery boli zase použité na výpočet AQI, ktorý udáva mieru znečistenia ovzdušia z hľadiska verejného zdravia. Do úvahy sa berie 6 znečisťujúcich látok, a to: PM2.5, PM10, SO_2, NO_2, CO a O_3. Najprv je potrebné vypočítať čiastkové indexy z 1-hodinových (SO_2 a NO_2), 8-hodinových (CO a O_3) alebo 24-hodinových priemerov (PM2.5 a PM10) [17]. Zameriavame sa na PM2.5 a PM10, keďže boli merané len tieto častice. AQI má pre každú znečisťujúcu látku definovaných 6 rôznych kategórií kvality ovzdušia (Tab. 1).

Tab. 1
Kategórie AQI

Kvalita ovzdušia	AQI	PM2.5 (µg/m^3)	PM10 (µg/m^3)
Dobrá	0 – 50	0.0 – 12.0	0 – 54
Mierna	51 – 100	12.1 – 35.4	55 – 154
Nezdravá pre citlivé skupiny*	101 – 150	35.5 – 55.4	155 – 254
Nezdravá	151 – 200	55.5 – 150.4	255 – 354
Veľmi nezdravá	201 – 300	150.5 – 250.4	355 – 424
Nebezpečná	301 – 500	250.5 – 500.4	425 – 604

*Ľudia s ochorením srdca alebo pľúc, seniori, deti a ľudia s nižším sociálno-ekonomickým statusom.

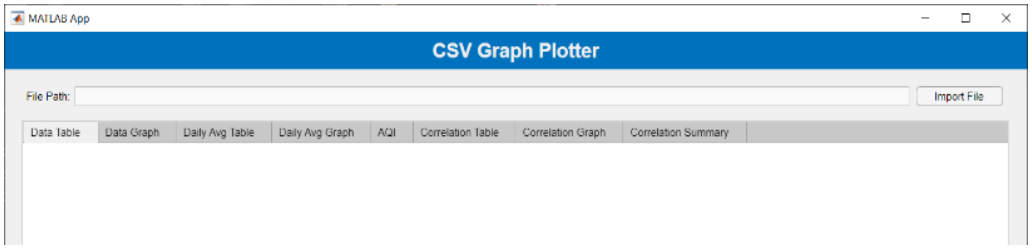
Pre čiastkový index I_p pre znečisťujúcu látku p platí:

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}} \left(C_p - BP_{Lo}\right) + I_{Lo}, \quad (1)$$

pričom C_p je koncentrácia znečisťujúcej látky p, BP_{Hi} je horná hraničná hodnota koncentrácie, BP_{Lo} je dolná hraničná hodnota koncentrácie, I_{Hi} je horná hraničná hodnota AQI a I_{Lo} je dolná hraničná hodnota AQI. Výsledný AQI je určený ako maximum čiastkových indexov pre každú znečisťujúcu látku, v našom prípade maximum medzi PM2.5 a PM10. V MATLABe bola na výpočet AQI vytvorená pomocná funkcia *aqicalc(C_p, p)*. Argumenty funkcie sú koncentrácia (C_p) a typ (p) znečisťujúcej látky, na základe ktorých je vypočítaný index AQI.

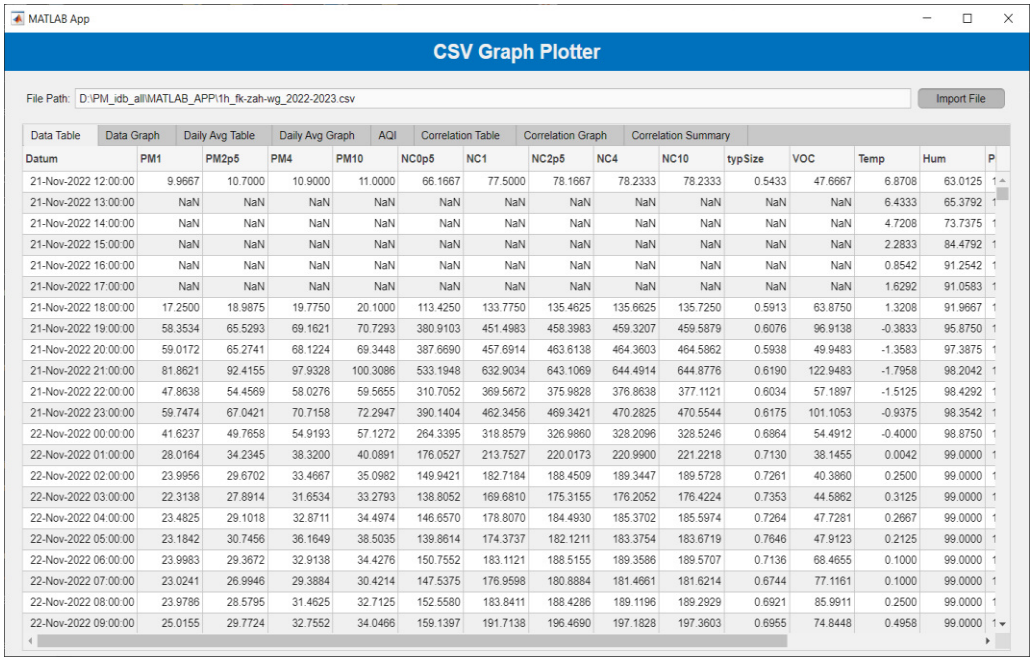
III. GUI ROZHRANIE APLIKÁCIE

Samotné používateľské prostredie aplikácie obsahuje tlačidlo na načítanie nameraných údajov z *.csv súboru (príklad vyhodnocovania bude znázornený na údajoch nameraných stanicou v dedine Füzérkomlós), needitovateľné textové pole slúžiace na výpis cesty k importovanému súboru a niekoľko záložiek, v ktorých budú zobrazené tabelárne hodnoty alebo grafické vizualizácie vyhodnotených údajov. Prvá záložka sa nazýva *Data Table*, ktorá pred načítaním údajov obsahuje prázdnu tabuľku.



Obr. 1 Úvodná obrazovka aplikácie.

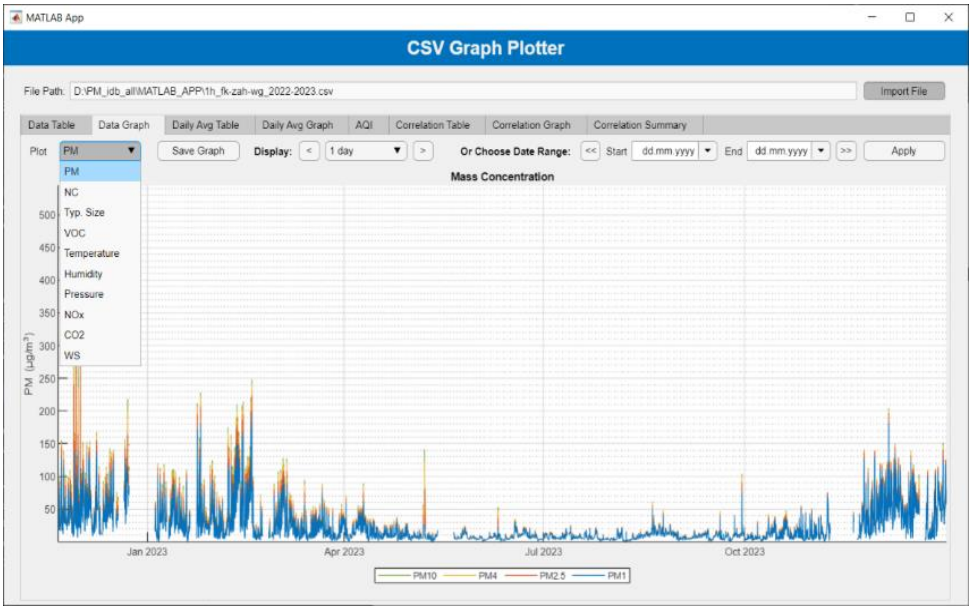
Aplikácia je ošetrená pred prepínaním medzi záložkami pred importovaním údajov z *.csv súboru, ako aj pred importovaním údajov v nesprávnom formáte. Po importovaní a vyhodnotení údajov sa v aplikácii zobrazí tabuľka s údajmi (Obr. 2) a používateľovi je umožnené prepínanie medzi záložkami. Podobne vyzerajú aj tabuľky v záložkách *Daily Avg Table* a *Correlation Table,* ktoré obsahujú vypočítané denné priemery a korelácie medzi meranými veličinami.



Datum	PM1	PM2p5	PM4	PM10	NC0p5	NC1	NC2p5	NC4	NC10	typSize	VOC	Temp	Hum	P
21-Nov-2022 12:00:00	9.9667	10.7000	10.9000	11.0000	66.1667	77.5000	78.1667	78.2333	78.2333	0.5433	47.6667	6.8708	63.0125	1
21-Nov-2022 13:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6.4333	65.3792	1
21-Nov-2022 14:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.7208	73.7375	1
21-Nov-2022 15:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.2833	84.4792	1
21-Nov-2022 16:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.8542	91.2542	1
21-Nov-2022 17:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.6292	91.0583	1
21-Nov-2022 18:00:00	17.2500	18.9875	19.7750	20.1000	113.4250	133.7750	135.4625	135.6625	135.7250	0.5913	63.8750	1.3208	91.9667	1
21-Nov-2022 19:00:00	58.3534	65.5293	69.1621	70.7293	380.9103	451.4983	458.3983	459.3207	459.5879	0.6076	96.9138	-0.3833	95.8750	1
21-Nov-2022 20:00:00	59.0172	65.2741	68.1224	69.3448	387.6690	457.6914	463.6138	464.3603	464.5862	0.5938	49.9483	-1.3583	97.3875	1
21-Nov-2022 21:00:00	81.8621	92.4155	97.9328	100.3086	533.1948	632.9034	643.1069	644.4914	644.8776	0.6190	122.9483	-1.7958	98.2042	1
21-Nov-2022 22:00:00	47.8638	54.4569	58.0276	59.5655	310.7052	369.5672	375.9828	376.8638	377.1121	0.6034	57.1897	-1.5125	98.4292	1
21-Nov-2022 23:00:00	59.7474	67.0421	70.7158	72.2947	390.1404	462.3456	469.3421	470.2825	470.5544	0.6175	101.1053	-0.9375	98.3542	1
22-Nov-2022 00:00:00	41.6237	49.7658	54.9193	57.1272	264.3395	318.8579	326.9860	328.2096	328.5246	0.6864	54.4912	-0.4000	98.8750	1
22-Nov-2022 01:00:00	28.0164	34.2345	38.3200	40.0891	176.0527	213.7527	220.0173	220.9900	221.2218	0.7130	38.1455	0.0042	99.0000	1
22-Nov-2022 02:00:00	23.9956	29.6702	33.4667	35.0982	149.9421	182.7184	188.4509	189.3447	189.5728	0.7261	40.3860	0.2500	99.0000	1
22-Nov-2022 03:00:00	22.3138	27.8914	31.6534	33.2793	138.8052	169.6810	175.3155	176.2052	176.4224	0.7353	44.5862	0.3125	99.0000	1
22-Nov-2022 04:00:00	23.4825	29.1018	32.8711	34.4974	146.6570	178.8070	184.4930	185.3702	185.5974	0.7264	47.7281	0.2667	99.0000	1
22-Nov-2022 05:00:00	23.1842	30.7456	36.1649	38.5035	139.8614	174.3737	182.1211	183.3754	183.6719	0.7646	47.9123	0.2125	99.0000	1
22-Nov-2022 06:00:00	23.9983	29.3672	32.9138	34.4276	150.7552	183.1121	188.5155	189.3586	189.5707	0.7136	68.4655	0.1000	99.0000	1
22-Nov-2022 07:00:00	23.0241	26.9946	29.3884	30.4214	147.5375	176.9598	180.8884	181.4661	181.6214	0.6744	77.1161	0.1000	99.0000	1
22-Nov-2022 08:00:00	23.9786	28.5795	31.4625	32.7125	152.5580	183.8411	188.4286	189.1196	189.2929	0.6921	85.9911	0.2500	99.0000	1
22-Nov-2022 09:00:00	25.0155	29.7724	32.7552	34.0466	159.1397	191.7138	196.4690	197.1828	197.3603	0.6955	74.8448	0.4958	99.0000	1

Obr. 2 Tabuľka údajov importovaných z *.csv súboru.

V záložke *Data Graph* sú vizualizované importované údaje vo forme čiarových grafov (Obr. 3). V záložke je možnosť výberu rôznych meraných veličín, a to PM (hmotnostné koncentrácie častíc PM1, PM2.5, PM4 a PM10), NC (početné koncentrácie častíc PM0.5, PM2.5, PM4 a PM10), typická veľkosť častíc, VOC index, teplota, vlhkosť, tlak a ak to importovaný *.csv súbor obsahuje, tak aj NOx index, CO_2 a rýchlosť vetra.



Obr. 3 Vizualizácia importovaných údajov pomocou čiarového grafu s možnosťou výberu veličiny.

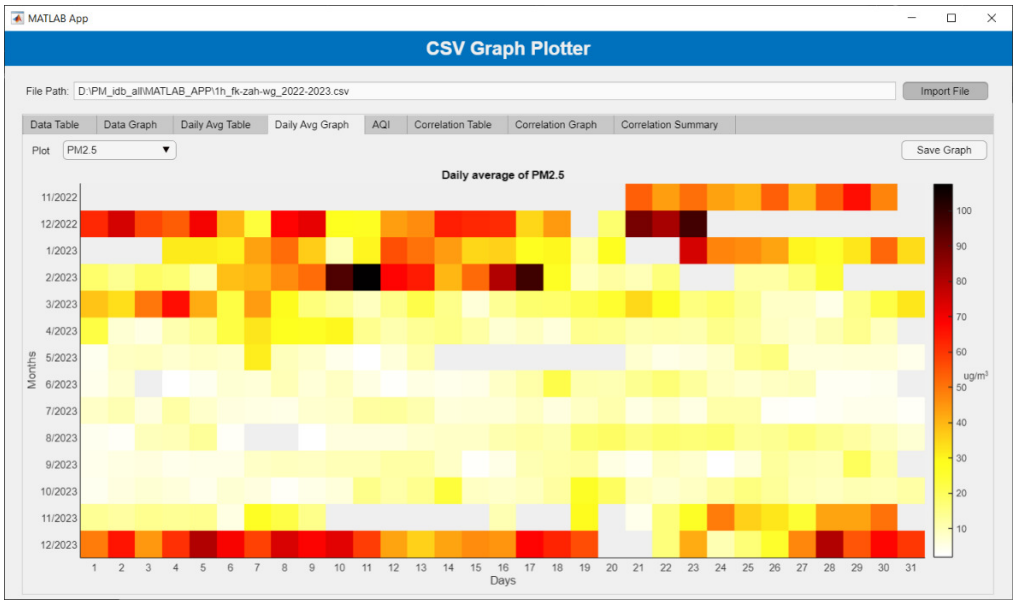
Je možné aj priblížiť zobrazenie v kratšom časovom intervale, a to buď výberom z prednastavených intervalov (Obr. 4a), alebo vlastným výberom začiatočného a konečného dátumu (Obr. 4b). Šípkami vpravo a vľavo taktiež nastáva posun v čase.



a) b)

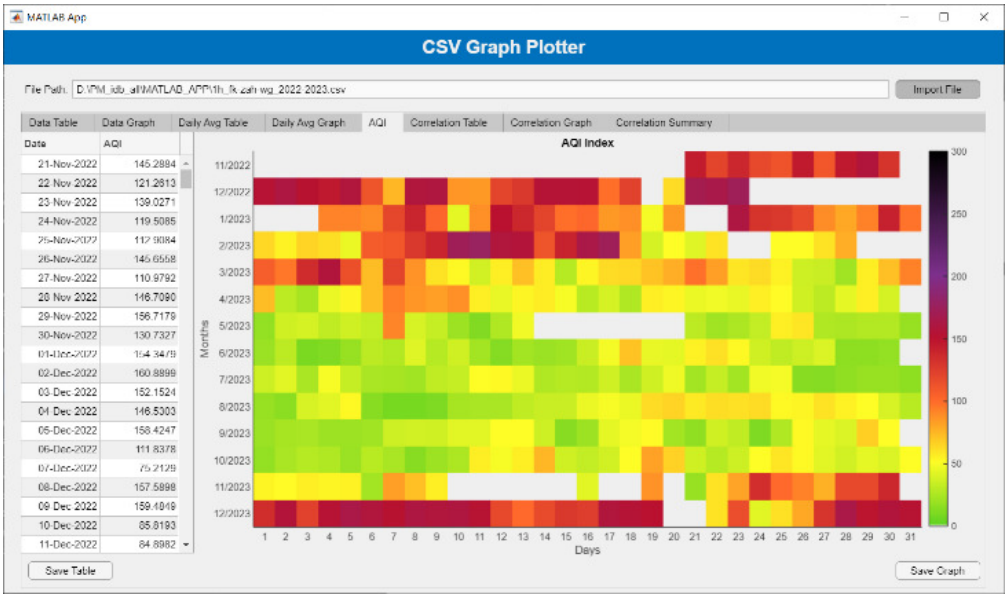
Obr. 4 Výber intervalu zobrazenia grafu: a) prednastavené intervaly, b) výber začiatočného a konečného dátumu.

V záložke *Daily Avg Graph* (Obr. 5), je zobrazená vizualizácia denných priemerov pomocou kalendárnej tepelnej mapy. Na osi *y* sa nachádzajú mesiace a na osi x sa nachádzajú dni. Každý deň je reprezentovaný jedným štvorcom, ktorého farba odpovedá vypočítanému dennému priemeru – od nižších hodnôt v odtieňoch bielej a žltej farby, k vyšším hodnotám odpovedajúcim červenej až tmavej bordovej farbe. Vďaka menu v ľavej hornej časti záložky je možné vybrať si zobrazovanú veličinu pomocou kalendárnej tepelnej mapy. Konkrétne na Obr. 5 je zvolená hmotnostná koncentrácia častíc PM2.5, kde je možné vidieť bielu až žltú farbu pre väčšinu dní merania, čo odpovedá koncentráciám nižším ako 30 µg/m^3. Zvýšená koncentrácia PM je hlavne počas jesenných a zimných mesiacov, prípadne na začiatku jari (od novembra 2022 do marca 2023 a potom opäť v novembri a decembri 2023).



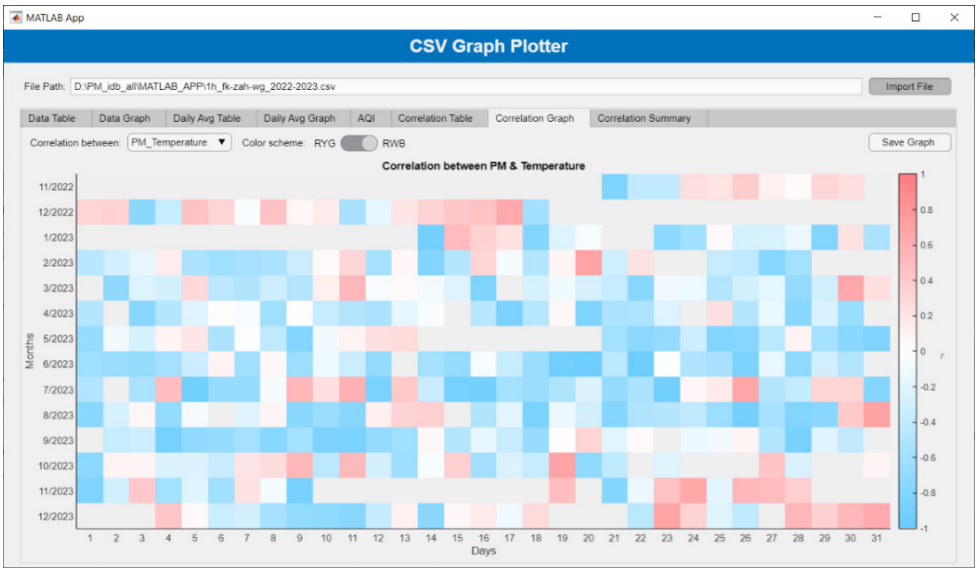
Obr. 5 Vizualizácia denných priemerných hodnôt pomocou tepelnej kalendárnej mapy (výber PM2.5).

Záložka *AQI* (Obr. 6) obsahuje tabuľku s vypočítanými hodnotami indexu AQI, ako aj ich vizualizáciu pomocou kalendárnej tepelnej mapy, pričom (podľa kategórií AQI) odpovedí zelená farba dobrej kvalite ovzdušia, žltá farba miernej kvalite ovzdušia, oranžová farba nezdravej kvalite ovzdušia pre citlivé skupiny, červená farba nezdravej kvalite ovzdušia, fialová veľmi nezdravej kvalite ovzdušia a čierna farba nebezpečnej kvalite ovzdušia. Nakoľko najväčší vplyv majú na kvalitu ovzdušia častice PM2.5, kalendárna tepelná mapa v záložke AQI sa podobá na mapu priemernej dennej hmotnostnej koncentrácie častíc PM2.5 v predchádzajúcej záložke. Dobrá (zelená farba) a mierna (žltá farba) kvalita ovzdušia prevládajú v jari a lete 2023, kým v jesennom a zimnom období prevláda nezdravá kvalita ovzdušia.



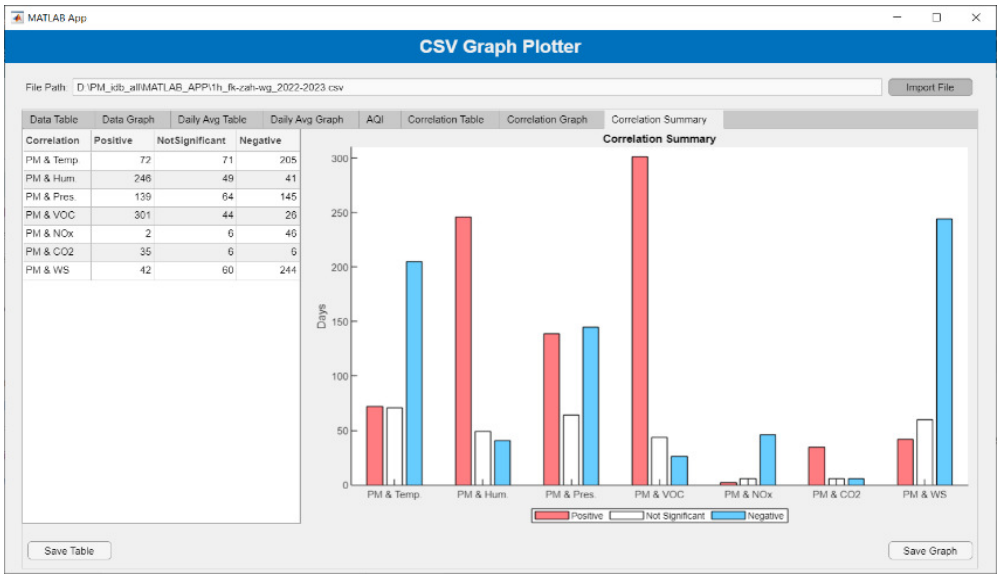
Obr. 6 Záložka obsahujúca tabuľku s hodnotami AQI a ich vizualizáciu pomocou tepelnej kalendárnej mapy.

Vizualizácia korelácie medzi PM a ostatnými veličinami (teplota, vlhkosť, tlak, VOC a ak boli merané, tak aj NOx, CO_2 a rýchlosť vetra) sú zobrazené v záložke *Correlation Graph*, kde bola pre vizualizáciu opäť vytvorená kalendárna tepelná mapa s farebnou škálou, vďaka ktorej je používateľovi jasné, či prevláda pozitívna (odtiene červenej/ružovej farby), negatívna (odtiene modrej farby) alebo žiadna korelácia (biela farba), prípadne, či sa v kalendárnej tepelnej mape nachádza rovnomerne rozložený mix všetkých farieb. Na Obr. 7 je zobrazená korelácia medzi PM a teplotou, pričom prevláda negatívna korelácia. Samozrejme, používateľ si môže zobraziť aj ostatné korelácie (PM/teplota, PM/vlhkosť, PM/tlak, PM/VOC, PM/NOx a PM/CO_2) pomocou menu v ľavej hornej časti.



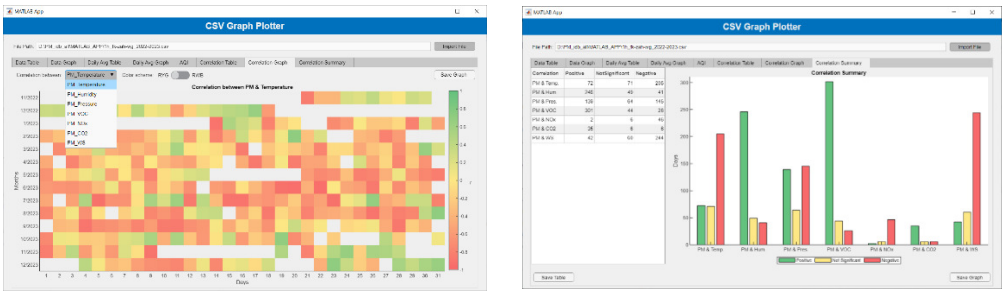
Obr. 7 Vizualizácia korelácie pomocou kalendárnej tepelnej mapy (výber PM/teplota).

Aj keď si používateľ môže vybrať, ktoré korelácie zobrazí v aplikácii, tak stále je v záložke *Correlation Graph* (Obr. 7) naraz zobrazená len jedna kalendárna tepelná mapa. Porovnať všetky korelácie však umožní záložka *Correlation Summary* (Obr. 8), kde sa nachádza tabuľka so zhrnutím o korelácii medzi PM a ostatnými veličinami aj vizualizácia tejto sumarizačnej tabuľky v podobe stĺpcového grafu. V grafe aj tabuľke sa zobrazuje počet dní so štatisticky významnou pozitívnou a negatívnou koreláciou, ako aj koreláciou, ktorá nebola štatisticky významná – teda nebola vyvrátená nulová hypotéza *p*-testu. Hodnota korelácie, pri ktorej $p = 0{,}05$ sa pohybuje okolo $|r| = 0{,}4$, čo odpovedá spodnej hranici pre miernu koreláciu [18]. Z takejto sumarizačnej tabuľky a grafu môže používateľ vidieť, že pre vyhodnocovanú meraciu stanicu silno prevládala pozitívna korelácia medzi PM/teplotou a PM/VOC, zatiaľ čo pre PM/teplotu a PM/rýchlosť vetra prevládala negatívna korelácia. Korelácia medzi PM/tlakom obsahovala pomerne vyrovnanú distribúciu medzi pozitívnou a negatívnou koreláciou (s rozdielom iba 6 dní). Senzory na snímanie NOx a CO_2 boli pridané neskôr, pričom kratšiemu meraniu odpovedá aj výška stĺpcov týchto korelácií. Prevláda negatívna korelácia medzi PM/NOx a pozitívna korelácia medzi PM/CO_2, ale samozrejme je stále potreba dlhších meraní pre tieto veličiny.



Obr. 8 Sumarizácia korelácie v tabelárnej a grafickej podobe (stĺpcový graf).

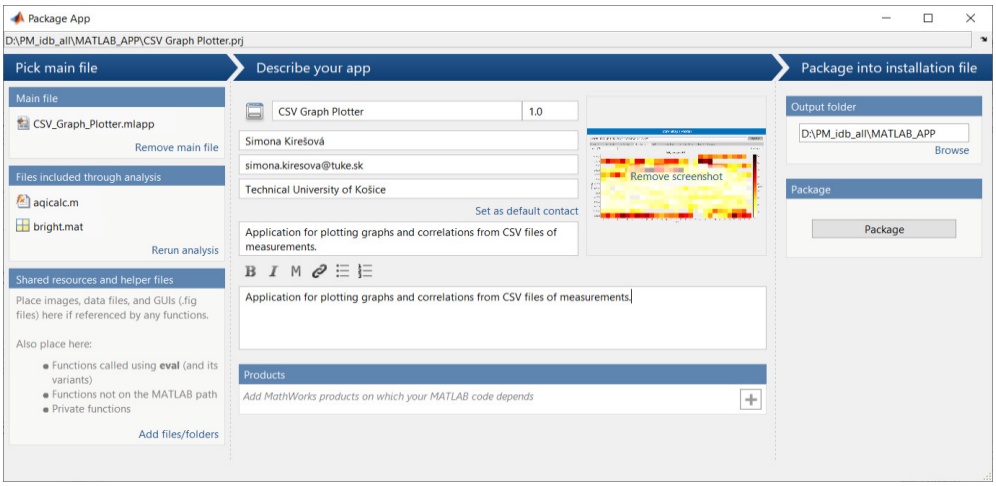
V tabuľke *Correlation Graph* je možné prepínať medzi dvoma farebnými škálami (Obr. 9), a to červeno-bielo-modrou a červeno-žlto-zelenou farebnou škálou. Voľba farebnej škály neovplyvňuje len záložku *Correlation Graph* (Obr. 7 a Obr. 9a), ale aj záložku *Correlation Summary* (Obr. 8 a Obr. 9b). Napokon aplikácia umožňuje aj uloženie všetkých vygenerovaných grafov a tabuliek v aplikácii do PC vo formáte *.fig, resp. *.xlsx.



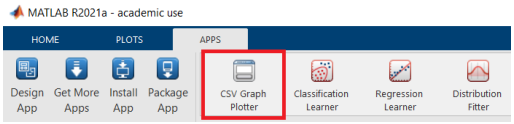
a) b)

Obr. 9 Zmena farebnej škály: a) záložka *Correlation Graph*, b) záložka *Correlation Summary.*.

MATLAB umožňuje vytvorenie balíka *(package)* pre aplikáciu (Obr. 10), pomocou ktorého je možné aplikáciu nainštalovať do záložky s aplikáciami (Obr. 11). Balíku je potrebné definovať potrebné polia obsahujúce informácie o aplikácii a jej autorovi a zahrnúť všetky externé súbory s funkciami alebo premennými, ktoré aplikácia využíva.



Obr. 10 Vytvorenie balíka aplikácie



Obr. 11 Nainštalovaná aplikácia

IV. ZÁVER

Keďže vznikla potreba efektívneho spracovania a vizualizácie veľkého množstva údajov, ktoré generujú stanice v senzorickej sieti, bolo potrebné vytvoriť MATLAB aplikáciu s intuitívnym používateľským rozhraním, ktoré umožňuje importovať, analyzovať a vizualizovať namerané údaje. Vďaka tejto aplikácii vytvorenej za pomoci nástroja MATLAB App Designer, sa výrazne zjednodušil proces vyhodnocovania nameraných údajov. Aplikácia umožňuje používateľovi sledovať a porovnávať hodnoty nameraných veličín, vypočítať priemery, korelácie a AQI, čo prispieva k lepšiemu porozumeniu vzťahov a korelácie medzi jednotlivými meranými veličinami. V neposlednom rade, ukladanie grafických a tabelárnych výstupov, ktoré môžu byť použité na ďalšie spracovanie, výskum či archiváciu, zvyšuje hodnotu aplikácie pre odborníkov, ktorí pracujú s veľkým objemom (mrakom) údajov.

Napriek tomu, že MATLAB poskytuje vynikajúce nástroje na analýzu a vizualizáciu údajov, je potrebné uznať, že ide o komerčnú aplikáciu s platenou licenciou, čo môže obmedzovať prístupnosť pre širšiu komunitu používateľov (najmä ak nemajú prístup k univerzitnej licencii). Doteraz sme implementovali riešenia pomocou InfluxDB a Grafana na vizualizáciu údajov a vyvinuli sme mobilnú aplikáciu pre zobrazenie aktuálne nameraných údajov na mobilných zariadeniach. Pre budúci vývoj by mohlo byť výhodné vyvinúť open-source aplikáciu, napr. v jazyku C#, ktorá by ponúkala podobné funkcie vyhodnocovania nameraných údajov z *.csv súboru, čím by bola zabezpečená širšia dostupnosť aplikácie pre ďalších používateľov.

REFERENCES

[1] J. A. Stankovic, "Wireless Sensor Networks," in Computer, vol. 41, no. 10, pp. 92-95, Oct. 2008, doi: 10.1109/MC.2008.441.

[2] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645-1660. doi:10.1016/j.future.2013.01.010

[3] Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241-251. doi:10.1016/j.bushor.2011.01.005

[4] Davenport, T. H. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review*, 76(4).

[5] Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., ... & Blumenthal, D. (2009). Use of electronic health records in US hospitals. *New England Journal of Medicine*, 360(16), 1628-1638. doi:10.1056/NEJMsa0900592

[6] S. Kirešová: "The Use of Wireless Sensor Network in the Long-Term Measurement and Correlation Analysis of PM and Meteorological Factors," In SCYR 2024: 24th Scientific Conference of Young Researchers, Proceedings from Conference, pp 35-37, Apr 2024.

[7] S. Kirešová, M. Guzan, B. Sobota, V. Fedák, R. Bača and D. Bakši, "The Use of Time Series Database in Measurements," 2023 International Conference on Electrical Drives and Power Electronics (EDPE), The High Tatras, Slovakia, 2023, pp. 1-8, doi: 10.1109/EDPE58625.2023.10274054.

[8] S. Kirešová, M. Guzan, B. Fecko, O. Somka, V. Rusyn and R. Yatsiuk, "Grafana as a Visualization Tool for Measurements," 2023 IEEE 5th International Conference on Modern Electrical and Energy System (MEES), Kremenchuk, Ukraine, 2023, pp. 1-5, doi: 10.1109/MEES61502.2023.10402486.

[9] S. Kirešová, M. Guzan, B. Sobota, V. Melnykov, R. Yatsiuk and B. Peryvertailo, "Mobile Application for Visualizing the Measurements," 2023 IEEE 5th International Conference on Modern Electrical and Energy System (MEES), Kremenchuk, Ukraine, 2023, pp. 1-5, doi: 10.1109/MEES61502.2023.10402445.

[10] MathWorks. "MATLAB Documentation." [Online]. Dostupné na: <https://www.mathworks.com/help/matlab/>. [cit. 2024-05-30].

[11] MathWorks. "MATLAB App Designer." [Online]. Dostupné na: <https://www.mathworks.com/products/matlab/app-designer.html>. [cit. 2024-05-30].

[12] Sensirion. "Datasheet SEN5x: Environmental Sensor Node for HVAC and Air Quality Applications." [Online]. Dostupné na: <https://sensirion.com/media/documents/6791EFA0/62A1F68F/Sensirion_Datasheet_Environmental_Node_SEN5x.pdf>. [cit. 2024-05-20].

[13] TE connectivity. "MS5611-01BA03: Barometric Pressure Sensor with stainless steel cap." [Online]. Dostupné na: <https://www.te.com/commerce/DocumentDelivery/DDEController?Action=showdoc&DocId=Data+Sheet%7FMS5611-01BA03%7FB3%7Fpdf%7FEnglish%7FENG_DS_MS5611-01BA03_B3.pdf%7FCAT-BLPS0036>. [cit. 2024-05-20].

[14] Sensirion. "SCD4x: Breaking the size barrier in CO2 sensing." [Online]. Dostupné na: <https://sensirion.com/media/documents/48C4B7FB/64C134E7/Sensirion_SCD4x_Datasheet.pdf>. [cit. 2024-05-20].

[15] Argent Data Systems. "Weather Sensor Assembly p/n 80422." [Online]. Dostupné na: <https://www.argentdata.com/files/80422_datasheet.pdf>. [cit. 2024-05-20].

[16] MathWorks. "Correlation coefficient - MATLAB." [Online]. Dostupné na: <https://www.mathworks.com/help/matlab/ref/corrcoef.html>. [cit. 2024-05-20].

[17] U.S. Environmental Protection Agency. "Technical Assistance Document for the Reporting of Daily Air Quality—The Air Quality Index (AQI)." [Online]. Dostupné na: <https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf>. [cit. 2024-05-17].

[18] Hanák, R. "Korelácia – Štatistika v PSPP." [Online]. Dostupné na: <https://statistikapspp.sk/korelacia/>. [cit. 2024-05-17].

Merateľné indikátory kvality vzdelávania získané z voľne dostupných dát

[1]*Miroslava MATEJOVÁ,* [2]*Ján PARALIČ,* [3]*Viktória SANISLOVÁ*

[1,2,3] Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]miroslava.matejova@tuke.sk, [2]jan.paralic@tuke.sk, [3]viktoria.sanislova@student.tuke.sk

***Abstrakt* — Predložený článok sa zameriava na hodnotenie kvality vzdelávania na Technickej univerzite v Košiciach (TUKE) pomocou dashboardovej PowerBI aplikácie. Ciele výskumu zahŕňajú navrhnutie a implementáciu systémového riešenia na zhromažďovanie údajov o kvalite vzdelávania a vytvorenie interaktívnej aplikácie na ich analýzu a vizualizáciu. Prvá časť článku sa zaoberá definíciou a identifikáciou faktorov kvality vzdelávania. Druhá časť zahŕňa popis implementácie navrhnutého riešenia v PowerBI, umožňujúc vytvorenie interaktívnej aplikácie na sledovanie a hodnotenie kvality vzdelávania na základe voľne dostupných údajov. Výsledky práce poskytujú prehľad o indikátoroch kvality prispôsobených stratégii Technickej univerzity v Košiciach. Analýza a výsledky môžu slúžiť ako inšpirácia pre ostatné slovenské univerzity a vzdelávacie inštitúcie pri využívaní podobných analytických nástrojov na hodnotenie a zlepšovanie kvality vzdelávania.**

***Kľúčové slová* — analýza dát, indikátory vzdelávania, hodnotenie kvality, vzdelávací proces**

I. Úvod

Indikátory kvality vzdelávania predstavujú metrické a hodnotiace ukazovatele, ktoré slúžia najmä na kvantitatívnu, ale aj kvalitatívnu analýzu rôznych aspektov a efektívnosti vzdelávacieho procesu a prostredia. V dnešnej dobe je nevyhnutné, aby vzdelávacie inštitúcie neustále monitorovali a zlepšovali kvalitu poskytovaného vzdelávania. Tieto indikátory poskytujú vzdelávacím inštitúciám významné informácie na zlepšenie ich procesov, aby mohli efektívne reagovať na potreby svojich študentov, tým prispieť k úspechu a rozvoju univerzity. Tento článok reaguje na uvedenú potrebu prostredníctvom zamerania sa na získanie a spracovanie väčšieho množstva údajov z voľne dostupných zdrojov. Následná vizualizácia týchto údajov pomocou dashboardovej aplikácie v PowerBI prispieva k lepšiemu porozumeniu a monitorovaniu kvality vzdelávania na univerzite.

II. Kvalita vzdelávania

V šesťdesiatych rokoch 20. storočia sa sledovanie kvality vzdelávania zameriavalo predovšetkým na množstvo naučených informácií a kvalitu učebných pomôcok a metód [1]. V posledných desaťročiach minulého storočia však tieto názory boli spochybnené v kontraste so schopnosťami študenta použiť nadobudnuté vedomosti v ďalšom vzdelávaní, resp. v živote.

Rastúci dopyt po zručnostiach si vyžaduje, aby vysokoškolské inštitúcie reagovali na situáciu novými spôsobmi. Aj vďaka tomu, ministri zodpovední za vysokoškolské vzdelávanie v Európskom priestore vysokoškolského vzdelávania (ESG) prijali v roku 2015 Štandardy a usmernenia na zabezpečovanie kvality [2]. Hlavným cieľom ESG je prispieť k spoločnému chápaniu zabezpečovania kvality naprieč geografickými hranicami. Jednotlivé štandardy a smernice sú zakotvené na systéme vnútornej kvality a naň nadväzujúcom systéme externého hodnotenia, ktoré zahŕňajú odporúčania pre zabezpečenie kvality v Európskom vysokoškolskom priestore. Štandardy definované Slovenskou akreditačnou agentúrou pre vysoké školstvo (SAAVŠ)[1], ktoré v súčasnosti platia, vychádzajú priamo z ESG, a slúžia ako referenčný dokument pre interné a externé systémy zabezpečovania kvality vysokých škôl.

[1] https://saavs.sk/

SAAVŠ hrá dôležitú úlohu v zabezpečovaní kvality vo vysokoškolskom vzdelávaní na území Slovenskej republiky. Ako verejná inštitúcia nesie zodpovednosť za monitorovanie a zabezpečovanie integrity a kvality vysokoškolských inštitúcií, čím zaručuje spoľahlivosť slovenského vysokoškolského systému vo svetovom kontexte. Každá vysoká škola na Slovensku má za povinnosť implementovať politiky na zabezpečenie kvality vo svojom riadení a preberá zodpovednosť za kvalitu poskytovaného vzdelávania.

A. Definovanie kvality vzdelávania

Rada Európy vo svojich vyhláseniach [3] zdôrazňuje, že kvalita vzdelávania by mala byť jedným z rozhodujúcich cieľov všetkých druhov a typov škôl a zabezpečenie kvality vzdelávania by malo byť prioritou na všetkých úrovniach a vo všetkých oblastiach vzdelávania.

Niektoré z hlavných bodov, ktoré kvalita vzdelávania zahŕňa sú:

- celková kvalita školy;
- kvalita vyučovacieho procesu;
- kvalita vzdelávacieho systému v štáte (regióne);
- kvalita pedagogického tímu;
- kvalita vyučovacej jednotky;
- podmienky v triedach a v učebniach;
- proces učenia sa študentov a žiakov;
- kvalita študijných materiálov.

B. Meranie kvality vzdelávania na slovenských vysokých školách

Univerzita Komenského (UK) v Bratislave využíva Microsoft PowerBI platformu na systematické porovnávanie širokého spektra faktorov, vrátane počtu študentov na jednotlivých fakultách, ich pomeru podľa štátnej príslušnosti a rozloženia pedagógov podľa pohlavia a akademických titulov. Tento súbor analýz zahŕňa aj vyhodnotenie študijných programov a ich príslušnosti k jednotlivým fakultám. Projekt DISK a aktivita HUB kvality UK [4] zefektívnili vnútorný systém monitorovania, vyhodnocovania a distribúcie výsledkov merania kvality. Systém HUB kvality UK nielen spracováva a vyhodnocuje dáta, ale ich aj efektívne sprístupňuje cieľovej skupine používateľov, pričom prispieva k debyrokratizácii hodnotenia kvality vzdelávania, vedy a výskumu.

Paneurópska univerzita sa podobne ako aj ostatné univerzity na Slovensku systematicky zaoberá hodnotením kvality svojich študijných programov, učiteľov, podporných služieb a prostredia [5]. Toto hodnotenie je realizované pravidelne a zohľadňuje vnútorné i vonkajšie faktory ovplyvňujúce fungovanie univerzity. Týmto spôsobom sa univerzita neustále snaží identifikovať silné a slabé stránky vlastnej činnosti, ako aj možnosti a hrozby vyplývajúce z vonkajšieho prostredia. Ústredným nástrojom v tomto procese je SWOT analýza, ktorá je vypracovaná na základe kvantitatívnych hodnotení a dát z rôznych zdrojov.

Žilinská univerzita vytýčila ambiciózne ciele a indikátory kvality v oblasti vzdelávania na základe dlhodobého zámeru do roku 2027 [6]. Tieto ciele zahŕňajú poskytovanie atraktívnej ponuky študijných programov, zvyšovanie kvality štúdia a tvorbu motivačného a nediskriminačného prostredia pre študentov. Žilinská univerzita vytvára široký rámec cieľov a indikátorov, ktoré reflektujú jej snahu o kontinuálne zlepšovanie a dosiahnutie excelentnosti vo vzdelávacom procese.

III. MERATEĽNÉ INDIKÁTORY KVALITY VZDELÁVANIA NA TUKE

TUKE systematicky sleduje a vyhodnocuje súbor ukazovateľov, ktoré sa delia podľa smernice OS/TUKE/H1/01 na ukazovatele vstupu do vzdelávania, ukazovatele vzdelávania a ukazovatele výstupu zo vzdelávania. Tieto ukazovatele slúžia ako nástroje na hodnotenie vývoja v čase v súlade s poslaním a cieľmi TUKE. Ukazovatele sú následne vyhodnocované v rámci periodických správ o činnosti Technickej univerzity v Košiciach, hodnotenia úrovne vzdelávacej činnosti, správ o hodnotení výkonnosti manažmentu kvality a ďalších vnútorných hodnotiacich správ. Jednotlivé študijné programy na univerzite využívajú tieto ukazovatele pri spracovaní vlastných správ o priebežnom monitorovaní a periodickom hodnotení (raz ročne). Tento proces zahŕňa analýzu ukazovateľov v primeranom rozsahu za obdobie spravidla posledných desiatich rokov pre jednotlivé študijné programy, fakulty a celú univerzitu.

TUKE v nedávno zverejnenom dokumente "Metodický pokyn o procesoch zabezpečenia kvality študijných programov" stanovila hlavné indikátory a procesy, na ktoré sa zameriava. Základom tejto práce sú primárne indikátory definované univerzitou v prílohe č.30 smernice OS/TUKE/H1/01 Vzdelávanie. Najpočetnejšie sú indikátory zamerané na proces vzdelávania, ktoré sú ďalej členené nasledovne:

A. Učitelia

Ukazovatele slúžia na sledovanie zloženia pedagogického zboru z hľadiska kvalifikácie, veku a mobilite učiteľov. Ide najmä o nasledovné ukazovatele:

- Podiel učiteľov s praxou dlhšou ako jeden rok na zahraničnej vysokej škole alebo na výskumnej inštitúcii v zahraničí.
- Počet prijatých učiteľov na mobility zo zahraničia v príslušnom akademickom roku.
- Podiel vyslaných učiteľov na mobility do zahraničia v príslušnom akademickom roku.
- Počet učiteľov na funkčnom mieste profesor, docent, odborný asistent, asistent, lektor, ostatní.

B. Učenie sa, vyučovanie a hodnotenie orientované na študenta

Ukazovatele slúžia na zhodnotenie prístupu k vzdelávaniu z hľadiska študenta, hodnotenia kvality poskytovaného vzdelávania zo strany študentov a posúdenia podpory študentov. TUKE vyhodnocuje raz ročne nasledovné ukazovatele:

- Pomer počtu študentov a učiteľov.
- Počet záverečných prác vedených vedúcim záverečnej práce.
- Podiel vyslaných študentov na mobility do zahraničia z celkového počtu študentov.
- Počet prijatých študentov na mobility zo zahraničia v príslušnom akademickom roku.

C. Prijímacie konanie, priebeh a ukončenie štúdia

Výsledky a vývoj ukazovateľov slúžia TUKE na monitorovanie výberu a posudzovania študentov, ich pokroku a mieru predčasného ukončenia štúdia. Škola vyhodnocuje raz ročne nasledovné ukazovatele:

- Počet študentov prekračujúcich štandardnú dĺžku štúdia.
- Počet študentov TUKE/ŠP (študijný program) v jednotlivých rokoch štúdia.
- Podiel študentov prvého roka štúdia, ktorí predčasne ukončili štúdium v štruktúre podľa dôvodu (vylúčenie pre neprospech, zanechanie štúdia, zmena ŠP).
- Miera predčasného ukončenia štúdia v ďalších rokoch štúdia.

D. Ukazovatele vstupu do vzdelávania

Výsledky a vývoj ukazovateľov vstupu do vzdelávania naznačujú zhodu medzi ponukou študijných programov a záujmom o štúdium na TUKE. Tieto ukazovatele zahŕňajú:

- Počet uchádzačov o štúdium v príslušnom akademickom roku.
- Počet uchádzačov o štúdium v príslušnom akademickom roku s iným ako slovenským občianstvom.
- Podiel zapísaných študentov zo všetkých prihlásených záujemcov o štúdium v príslušnom akademickom roku.

IV. DOSTUPNÉ DÁTA, SLÚŽIACE NA HODNOTENIE KVALITY VZDELÁVANIA NA TUKE

V tejto kapitole analyzujeme voľne dostupné dáta a zdroje informácií, ktoré môžu slúžiť na vyhodnotenie indikátorov kvality vzdelávania na TUKE.

A. Štatistická ročenka

Štatistická ročenka [7] je jedným z hlavných zdrojov štatistických informácií o slovenských vysokých školách a ich školských zariadeniach, ktorý je každoročne zrealizovaný vďaka Centru vedecko-technických informácií Slovenskej republiky (CVTI SR). CVTI SR je priamo riadená organizácia Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky (MŠVVaM SR) zabezpečuje prevádzku interdisciplinárnych výskumno-vývojových centier a národných infraštruktúr pre výskum, vývoj, inovácie a vzdelávanie.

Tento publikovaný dokument obsahuje rozsiahle údaje a tabuľky, ktoré poskytujú prehľad o počte škôl, fakúlt a študentov na rôznych stupňoch štúdia. Taktiež obsahuje informácie o ďalších študijných formách a rozdelenie študentov podľa rôznych kritérií, ako sú napríklad fakulty, odbory a ročníky. Okrem toho, ročenka tiež zahŕňa štatistiky o novoprijatých študentoch slovenského a iného štátneho občianstva, ako aj údaje o pedagogických zamestnancoch a absolventoch vysokých škôl.

B. Výročná správa TUKE

Správa o činnosti Technickej univerzity v Košiciach [8] za jednotlivé roky uvádza aktuálne a presnejšie informácie o vývoji univerzity a jej jednotlivých fakúlt. Tieto správy detailne analyzujú

štatistiky týkajúce sa počtu študentov, pedagogických zamestnancov a výskumných aktivít na každej fakulte. Zameriavajú sa na údaje o prijatých a absolvovaných študentoch, počte diplomových prác, publikačnej činnosti pedagógov a získaných výskumných grantoch. Okrem toho správy poskytujú informácie o infraštruktúre univerzity, investíciách do modernizácie výučby a výskumných zariadení, ako aj o spolupráci s priemyslom a zahraničnými inštitúciami.

C. Akademická štvrťhodinka

Cieľom študentského prieskumu Akademickej štvrťhodinky [9] je systematicky zhodnotiť úroveň spokojnosti študentov vysokých škôl na Slovensku. Tento prieskum, realizovaný SAAVŠ v roku 2021 je zameraný na hodnotenie kvality vzdelávania na slovenských vysokých školách, pričom TUKE je jednou z týchto inštitúcií. Prostredníctvom tohto prieskumu sú získané názory študentov, ktoré sú následne využité na identifikáciu oblastí, v ktorých je potrebné zlepšiť kvalitu vzdelávania v celom sektore, vrátane TUKE. Detailnejšie prehľady výsledkov za jednotlivé inštitúcie sú sprístupnené vysokým školám, čo im umožní získať spätnú väzbu o ich činnosti a úrovni spokojnosti študentov v celoslovenskom kontexte.

D. Výročná správa o stave vysokého školstva

MŠVVaM SR každoročne vypracúva výročnú správu o stave vysokého školstva [10], ktorá predkladá informácie o poskytovanom vysokoškolskom vzdelávaní, podpore výskumu a ďalších relevantných oblastiach. Zo správy môžeme získať dáta o kvalite vzdelávania, ako je poskytovanie vysokoškolského vzdelávania (počty študentov a absolventov, informácie o prijímacom konaní), podpora výskumu, vývoja a ďalšej tvorivej činnosti na vysokých školách, sociálnej podpory študentov, podpora rozvoja vysokých škôl, ale aj financovaní a hospodárení verejných vysokých škôl.

V. NÁVRH RIEŠENIA

V tejto kapitole sa zameriavame na návrh dashboardovej aplikácie v PowerBI na zbieranie, analýzu dostupných údajov a jednotlivých vybraných indikátorov.

Výber vhodného modelu životného cyklu je zásadným prvkom pri realizácii softvérového vývoja. V kontexte vyvíjania dashboardovej aplikácie pre hodnotenie kvality vzdelávania na TUKE je tento výber zdôvodnený z niekoľkých hľadísk. Jedným z týchto hľadísk je flexibilita a adaptabilita modelu, ktoré sú dôležitými faktormi v prostredí neustále sa meniacej oblasti vzdelávania a hodnotenia kvality. Zároveň je nutné zohľadniť potrebu neustáleho zlepšovania a prispôsobovania sa požiadavkám používateľov, čo je typické pre projekt tejto povahy. V tomto ohľade evolučný model životného cyklu softvérového vývoja sa javí ako ideálna voľba.

Na Obr. *1* sú znázornené jednotlivé fázy evolučného modelu. Prvým krokom pri vývoji je zadefinovanie hlavných cieľov návrhu riešenia, identifikácia cieľových používateľov a definovanie účelu poskytnutých informácií a analýz (spolu nazvané koncept softvéru). Nasleduje fáza zberu a analýzy dostupných dát o kvalite vzdelávania na TUKE. Tento proces zahrňuje zbieranie relevantných údajov v súlade so stanovenými ukazovateľmi, ich konsolidáciu do jedného dokumentu a následnú transformáciu s cieľom dosiahnuť jednotný formát a štruktúru údajov. Nasleduje návrh a vytvorenie vizualizácií, ako sú grafy, tabuľky a interaktívne filtre. Ďalším krokom je implementácia a testovanie vizualizácií v PowerBI, aby sme zistili, či fungujú správne a poskytujú relevantné informácie. Po dokončení implementácie získame spätnú väzbu od používateľov a zainteresovaných strán, ktorá nám pomôže identifikovať prípadné nedostatky a oblasti na vylepšenie. Na základe tejto spätnej väzby vykonáme finálne úpravy a aplikácia bude pripravená na nasadenie do produkčného prostredia.

A. Koncept softvéru

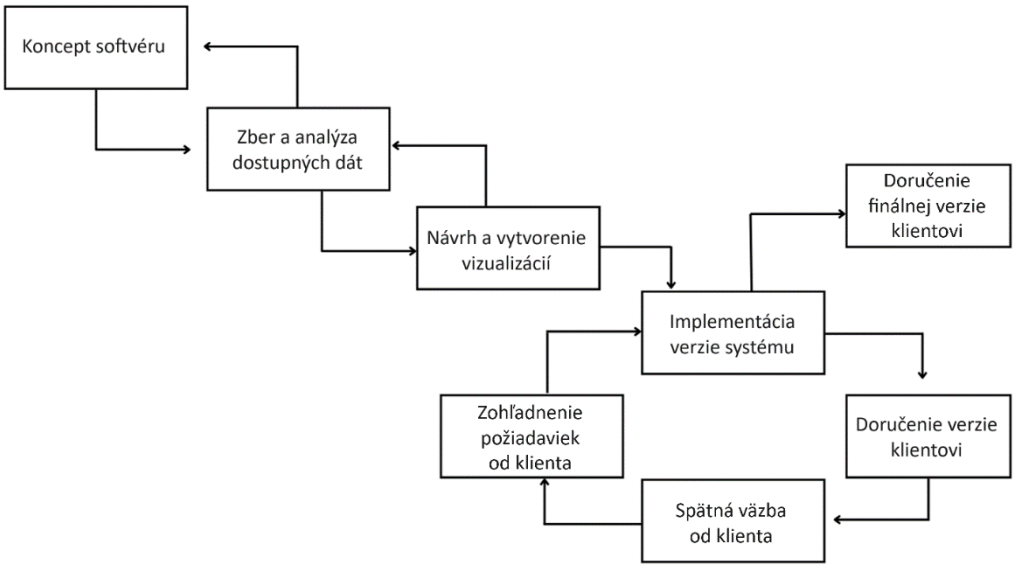
Cieľová skupina koncových používateľov zahŕňa vedenie univerzity, pedagogický zbor, študentov a administratívny personál, ktorí budú aktívne využívať aplikáciu PowerBI na vizualizáciu a analýzu indikátorov kvality vzdelávania. Používateľské požiadavky boli kategorizované podľa metódy MoSCoW [11] s cieľom ich prioritizácie.

Stanovenie hlavných cieľov aplikácie je esenciálne pre úspešný vývoj a nasadenie tejto aplikácie. Medzi tieto ciele patria:

- **Zobrazenie trendov:** Sledovanie dlhodobých trendov v kvalite vzdelávania na TUKE, ktoré zahŕňajú dynamiku študentov, zamestnancov, mobility, prijímacích procesov a ďalších kľúčových ukazovateľov. Takéto monitorovanie umožňuje identifikovať dlhodobé trendy, ktoré môžu mať významný vplyv na vzdelávacie politiky a procesy.
- **Identifikácia slabých miest:** Identifikácia slabých miest zahŕňa prehľad vybraných

oblastí, čo umožňuje podrobnejšie pozorovanie potenciálnych oblastí pre zlepšenie a faktorov, ktoré môžu ovplyvniť výsledky indikátorov kvality vzdelávania.

- **Monitorovanie vývoja:** Sledovanie a hodnotenie účinnosti implementovaných zlepšení a opatrení na základe identifikovaných slabých miest a porovnávanie trendov vývoja indikátorov kvality vzdelávania súvisiacich s inými vysokými školami.
- **Filtrovanie podľa rokov:** Možnosť filtrovania a porovnávania údajov o kvalite vzdelávania v rôznych rokoch umožní identifikovať trendové vzory a dlhodobé zlepšenia alebo oblasti na zlepšenie.
- **Filtrovanie podľa fakúlt:** Schopnosť špecifikovať a analyzovať údaje podľa jednotlivých fakúlt umožní detailné zhodnotenie kvality vzdelávania v jednotlivých akademických jednotkách univerzity a identifikáciu špecifických potrieb a výziev.



Obr. 1 Evolučný model životného cyklu vývoja aplikácie

Pri výbere softvérového riešenia je dôležité mať na pamäti rôzne faktory a kritériá, ktoré ovplyvňujú rozhodnutie. Hodnotová analýza poskytuje štruktúrovaný prístup k porovnaniu rôznych možností a pomáha identifikovať najvhodnejšie riešenie na základe stanovených parametrov. V kontexte našej štúdie sme sa rozhodli použiť hodnotovú analýzu na porovnanie softvérových produktov PowerBI, Tableau a SQL z dôvodu ich významu a rozšírenosti v oblasti podnikovej analytiky a analýzy dát. Prvým krokom bolo stanovenie rozhodovacích kritérií, ktoré budeme používať na porovnanie jednotlivých riešení. Tieto kritériá zahŕňajú faktory ako používateľská prívetivosť, flexibilita vizualizácií, cenová dostupnosť, integrácia s existujúcimi systémami, špeciálne vedomosti pri použití a rozsah vizualizácií. Po stanovení kritérií je nevyhnutné priradiť im váhy na základe ich dôležitosti. Tento krok umožňuje vytvoriť hierarchiu priorít a jasne určiť, ktoré kritériá sú pre nás kľúčové pri výbere riešenia. Druhý krok hodnotovej analýzy zahŕňa stanovenie škály hodnotenia pre miery splnenia jednotlivých kritérií, ktorý umožní objektívne porovnanie a hodnotenie softvérových produktov vzhľadom na stanovené kritériá. V treťom kroku sme detailne zanalyzovali každú alternatívu a ohodnotili, do akej miery spĺňajú stanovené kritériá. PowerBI dosiahol najvyšší vážený súčet hodnotení, čo naznačuje, že je najvhodnejšou alternatívou pre našu dashboardovú aplikáciu na prezentovanie dostupných dát ohľadom kvality vzdelávania na TUKE. Tento výber je založený na analýze a porovnaní výkonu a schopností jednotlivých softvérových riešení v súvislosti s našimi požiadavkami a cieľmi.

B. Zber, analýza a úprava dostupných dát

Fáza zberu a analýzy dostupných údajov o kvalite vzdelávania na TUKE zahŕňa technické postupy na vytvorenie spoľahlivej a rozsiahlej databázy v nástroji PowerBI. Tento proces bol iniciovaný detailným prieskumom dostupných údajov, následne boli všetky relevantné informácie systematicky zhromaždené do jedného súboru formátu Excel.

Nasledovala fáza zoskupovania a zjednotenia údajov. Táto fáza sa venovala čisteniu a úpravám údajov, čo zahŕňalo vymazanie, nahradenie alebo doplnenie chýbajúcich dát. Tento proces zabezpečil, že všetky dáta boli spoľahlivé a kompletné, čo bolo nevyhnutné pred samotným nahraním údajov do PowerBI aplikácie.

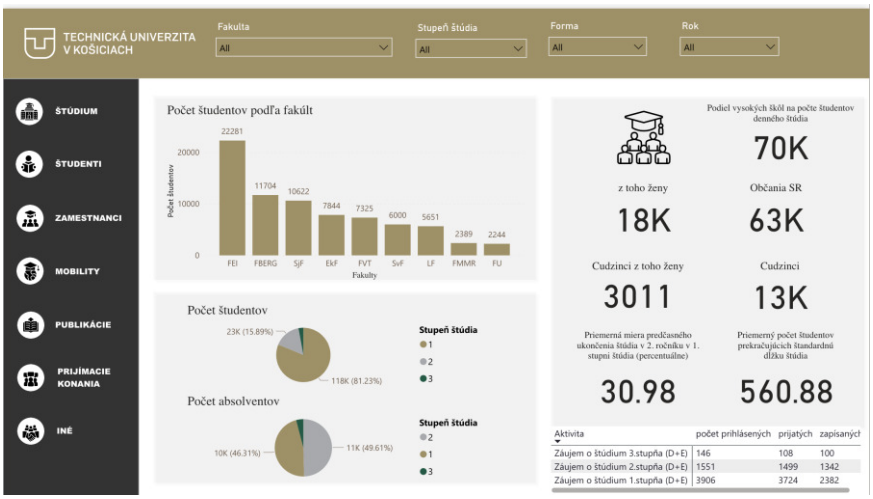
Po úspešnom zjednotení a úprave údajov sa pokračovalo s implementáciu databázy v nástroji PowerBI. Naimportovali sme už upravený Excel dokument a v prípadoch, kde bolo potrebné, bolo aplikované manuálne dopĺňanie údajov priamo do datasetu v PowerBI. Tento dataset bol

následne podrobený rôznym úpravám a transformáciám priamo v prostredí PowerBI, vďaka ktorým bolo možné presné a flexibilné spracovanie údajov podľa potrieb aplikácie.

Rozpracovaný proces vyžadoval definíciu vzťahov pre jednotlivé kategórie indikátorov a medzi jednotlivými databázami v PowerBI, aby sa zabezpečilo, že všetky filtre a údaje budú fungovať správne a konzistentne pre všetky grafy a tabuľky na jednotlivých stránkach.

C. Návrh a vytvorenie vizualizácií

Jednou z hlavných úloh dashboardovej aplikácie je prezentovať a analyzovať jednotlivé indikátory, ktoré sú relevantné pre hodnotenie kvality vzdelávania. Pri návrhu vizualizácií a výberu vhodných grafov sa prihliadalo na špecifické vlastnosti dát, aby boli vizualizácie zrozumiteľné a informatívne. Po vypracovaní filtračných možností a vizualizácií dát nasledovala fáza úpravy dizajnu celej aplikácie. Cieľom bolo vytvoriť používateľsky prívetivé prostredie, ktoré umožňovalo používateľom intuitívne interakcie s dátami. Celkový vizuálny výsledok tejto práce približuje Obr. 2, na ktorom vidieť, že informačný systém poskytuje používateľom jasný a prehľadný pohľad na dáta týkajúce sa hodnotenia kvality vzdelávania na univerzite.



Obr. 2 Dashboardová aplikácia zobrazujúca ukazovatele kvality vzdelávania TUKE

Po otvorení aplikácie sa na ľavej strane zobrazuje panel menu, kde sú jednotlivé kategórie indikátorov. Tieto kategórie zahŕňajú informácie o študentoch, zamestnancoch, štúdiu, mobilite, prijímacích konaniach a ďalšie relevantné aspekty. Vpravo od panela menu používateľ uvidí jednotlivé indikátory a ich vizualizácie. Tieto vizualizácie môžu zahŕňať koláčové grafy, stĺpcové diagramy, iné typy grafov alebo zobrazenie štatistických údajov. Každý indikátor je reprezentovaný príslušnou vizualizáciou, ktorá zobrazuje užitočné informácie a umožňuje používateľovi rýchlo a ľahko pochopiť trendy, vzťahy alebo rozloženie dát v danom indikátore. Celkovým cieľom tohto usporiadaného prístupu je umožniť užívateľom efektívne využívať všetky dostupné dáta a výsledky, aby mohli čo najlepšie využiť potenciál aplikácie. PowerBI aplikáciu sme vylepšili pridaním širokej škály filtrovacích možností, čo ponúka používateľom ešte podrobnejšie vyhľadávanie a analyzovanie indikátorov. Filtre umožňujú používateľom špecifikovať a vybrať iba tie dáta, ktoré sú pre nich relevantné, a to na základe rôznych kritérií.

Nasadenie[2] navrhovaného riešenia pre dashboardovú aplikáciu v PowerBI prebehlo podľa zvoleného modelu životného cyklu a preddefinovanými požiadavkami používateľov.

VI. TESTOVANIE IMPLEMENTOVANEJ DASHBOARDOVEJ APLIKÁCIE

Testovanie dashboardovej aplikácie sa uskutočnilo so skupinou respondentov, ktorá pozostávala z 5 študentov a 5 zamestnancov Technickej univerzity v Košiciach. Samotné testovanie prebiehalo formou dotazníka, ktorý pozostával z pätnástich otázok. Desať otázok z nich bolo navrhnutých pomocou štandardného dotazníka System usability scale (SUS) [12] a ďalších päť otázok bolo otvorených, pričom sa viac zamerali na konkrétne postrehy používateľov a prípadné návrhy na zlepšenie aplikácie.

SUS je meradlo použiteľnosti informačných systémov, ktoré sa predstavuje ako sada otázok , kde respondent hodnotí stupeň súhlasu alebo nesúhlasu s tvrdením na základe Likertovej škály.

[2] Aplikácie je dostupná pre TUKE používateľov na https://app.powerbi.com/links/kyFUV4fWLz?ctid=1c9f27ef-fee6-45f4-9a64-255a8c8e25a5&pbi_source=linkShare&bookmarkGuid=e624c91b-0493-4849-a1af-c1f0476ccf03

Hlavným zámerom bolo získať štandardizované, spoľahlivé a jednoducho zrozumiteľné hodnotenie používateľskej skúsenosti dashboardovej aplikácie .

Na základe výsledkov dotazníka môžeme uviesť, že zamestnanci TUKE považovali aplikáciu za ľahko použiteľnú a intuitívnu. Celkové skóre podľa SUS bolo vypočítané individuálne pre každého respondenta z kruhu zamestnancov, a následne bol získaný priemer, ktorý činil 84,5 boda. Toto skóre indikuje úroveň A, čo znamená, že respondenti prejavili vysokú spokojnosť a pozitívne hodnotenie aplikácie. Respondenti vyjadrili veľkú spokojnosť s dizajnom, opísali ho ako príjemný a veľmi pekný. Identifikovali niekoľko nedokonalostí v implementácii filtrov a interakcií s údajmi. Konkrétne zistili, že niektoré filtre nefungovali tak, ako by mali, čo spôsobilo problémy pri spracovaní všetkých dostupných údajov.

Celkové skóre u študentov bolo vypočítané individuálne pre každého respondenta a následne bol získaný priemer, ktorý činil 90,5 boda. Toto skóre tiež indikuje úroveň A, čo znamená, že respondenti pozitívne hodnotili aplikáciu. Ich všeobecné pocity z dizajnu boli veľmi pozitívne a vyjadrili celkovú spokojnosť.

VII. Záver

Hlavným cieľom tohto výskumu bolo navrhnúť a implementovať systémové riešenie na zhromažďovanie dostupných údajov o kvalite vzdelávania na TUKE a vytvoriť interaktívnu dashboardovú aplikáciu na ich analýzu a vizualizáciu. Pre dashboardovú aplikáciu sme využili údaje z viacerých zdrojov, pričom sme sa primárne zamerali na Výročné správy o činnosti TUKE, ktoré sme kombinovali s informáciami z MŠVVaM SR a Štatistickej ročenky pre vysoké školy od CVTI SR. Ako vhodné riešenie na vzniknutú situáciu sme zvolili vytvorenie dashboardovej aplikácie pomocou PowerBI, ktorú sme navrhli tak, aby umožňovala efektívnu analýzu a vizualizáciu týchto dát. Poslednou fázou bolo otestovanie vytvorenej dashboardovej aplikácie. Za týmto účelom sme vytvorili dotazník, ktorý bol zameraný na použiteľnosť a vyjadrenie názoru respondentov na dizajn, manipuláciu s údajmi a filtrovanie, či celkový dojem z aplikácie, s cieľom zhromaždiť komplexné pohľady na jej funkčnosť a spokojnosť zo strany zamestnancov a študentov TUKE. V závere možno konštatovať, že stanovené ciele boli úspešne dosiahnuté, čo potvrdzujú pozitívne výsledky a hodnotenia získané v priebehu testovania. V budúcnosti môžu byť do aplikácie doplnené ďalšie merateľné indikátory.

Poďakovanie

Tento článok vznikol vďaka spolupráci s Ústavom výpočtovej techniky TUKE a naše poďakovanie patrí Ing. Marte Dupkalovej a Ing. Helene Tóthovej za ich konzultáciu a poskytnutie časti dát v štruktúrovanej podobe.

Použitá literatúra

[1] Z. Juščáková, "Indikátory kvality vzdelávania v hodnotení škôl," Národný ústav certifikovaných meraní vzdelávania, 2013.
[2] "Štandardy a usmernenia na zabezpečovanie kvality v Európskom priestore vysokoškolského vzdelávania (ESG)," Brusel, 2015.
[3] Európska komisia, "Zabezpečovanie kvality," 2024. Accessed: Jun. 01, 2024. [Online]. Available: https://eurydice.eacea.ec.europa.eu/sk/national-education-systems/slovakia/zabezpecovanie-kvality
[4] "DISK: Aktualizácia a digitalizácia vnútorného systému kvality." Accessed: Jun. 01, 2024. [Online]. Available: https://uniba.sk/o-univerzite/vnutorny-system-kvality/digitalny-system-kvality/
[5] Paneurópska vysoká škola, "Systém kvality." Accessed: Jun. 03, 2024. [Online]. Available: https://www.paneurouni.com/pevs/uradne-informacie/system-kvalita/
[6] Žilinská univerzita v Žiline, "Ciele a indikátory kvality UNIZA." Accessed: Jun. 03, 2024. [Online]. Available: https://www.uniza.sk/index.php/ciele-a-indikatory-kvality-uniza
[7] Centrum vedecko-technických informácií SR, "Štatistická ročenka - vysoké školy." Accessed: Jun. 01, 2024. [Online]. Available: https://www.cvtisr.sk/cvti-sr-vedecka-kniznica/informacie-o-skolstve/statistiky/statisticka-rocenka-publikacia/statisticka-rocenka-vysoke-skoly.html?page_id=9596
[8] Technická univerzita v Košiciach, "Výročné správy a dokumenty." Accessed: Jun. 01, 2024. [Online]. Available: https://www.tuke.sk/wps/portal/tuke/university/vyrocne-spravy-a-dokumenty/
[9] Slovenská akreditačná agentúra pre vysoké školstvo, "O prieskume Akademická štvrťhodinka." Accessed: Jun. 01, 2024. [Online]. Available: https://prieskum.saavs.sk/o-prieskume/
[10] MŠVVaM SR, "Výročné správy o stave vysokého školstva | Ministerstvo školstva, výskumu, vývoja a mládeže Slovenskej republiky." Accessed: Jun. 01, 2024. [Online]. Available: https://www.minedu.sk/vyrocne-spravy-o-stave-vysokeho-skolstva/
[11] D. Clegg and R. Barker, Case Method Fast-Track: A Rad Approach. USA: Addison-Wesley Longman Publishing Co., Inc., 1994.
[12] J. Brooke, "SUS: A quick and dirty'usability scale," in *Usability evaluation in industry*, USA: CRC Press: Boca Raton, 1996.

Implementácia algoritmu PSO do MANET-IoT siete v simulátore NS-3

[1]Tomáš PETRILÁK, [1]Ján PAPAJ

[1]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská Republika

[1]jan.papaj@tuke.sk

***Abstract* – V dynamickom prostredí MANET-IoT, kde tradičná stabilná infraštruktúra chýba, integrácia umelej inteligencie ponúka významné výhody pre zlepšenie procesov smerovania a čiastočné eliminovanie bežných problémov ako sú nestabilita, energetická náročnosť a bezpečnostné riziká. Táto práca sa zameriava na dôkladnú analýzu existujúcich smerovacích protokolov a na výber techniky umelej inteligencie, ktorá by mohla prispieť k optimalizácii smerovacieho procesu. Na základe uskutočnenej analýzy bol vybraný najvhodnejší algoritmus optimalizácia roja častíc (PSO), ktorý bol následne uplatnený na optimalizáciu parametrov smerovacieho protokolu, s cieľom dosiahnuť maximálne efektívne riadenie a výkonnosť smerovania. Praktická časť práce zahŕňa implementáciu PSO algoritmu, ktorý bol aplikovaný na vylepšenie smerovacích rozhodnutí v simulačnom prostredí NS-3. Hodnotenie efektívnosti navrhnutého riešenia bolo realizované porovnávaním výsledkov získaných pred a po aplikácii PSO algoritmu.**

***Keywords* – MANET, IoT, MANET-IoT, PSO**

I. Úvod

Mobilné uzly v mobilných ad-hoc sieťach (MANET) sú mobilné a ich spojenia sú oproti tradičným sieťovým infraštruktúram veľmi dynamické, čo umožňuje zariadeniam efektívne spájať sa a zdieľať užitočné informácie aj v zložitých situáciách a na vzdialených miestach, kde nie sú dostupné tradičné sieťové služby. Vďaka týmto vlastnostiam sa MANET využíva vo veľkom množstve aplikácií. Integrácia MANET a internet vecí (IoT) do prostredí 5G a následne 6G sietí predstavuje kľúčový krok k realizácii spoľahlivej, dostupnej a vysoko efektívnej komunikácie, ktorá je nevyhnutná pre široké spektrum aplikácií [1].

MANET sú druhom bezdrôtových sietí, ktoré nepotrebujú žiadnu infraštruktúru [1]. To znamená, že v MANET neexistuje žiadny centrálny orgán a všetky uzly fungujú ako vysielače, aj prijímače pre účely komunikácie. Komunikácia v MANET je dosiahnutá tak, že keď je cieľ ľahko dosiahnuteľný zo zdroja, teda cieľ leží v dosahu vysielania zdroja, tak komunikujú priamo. Ak sú však od seba vzdialené, využívajú pomoc susedných uzlov. To znamená, že každý uzol sa správa ako smerovač v MANET.

Jednou z najväčších výhod bezdrôtových sietí je ich schopnosť umožniť rôznym uzlom komunikovať a zároveň si zachovať mobilitu. Vzhľadom k tomu, že mobilné ad-hoc siete sa nespoliehajú na žiadnu infraštruktúru, všetky uzly sú nezávislé a môžu sa voľne pohybovať. Prenosový rozsah MANET je však obmedzený, čo znamená, že priama komunikácia medzi zdrojom a cieľom nie je možná, keď sa nachádzajú mimo svojej zóny prenosových uzlov, na čom sa podieľajú sprostredkujúce uzly. Komunikácia v MANET sa delí na dva typy: jednoskoková a viacpreskoková komnunikácia. V prvom prípade existujú uzly, ktoré sú v dosahu rádiového signálu a môžu komunikovať priamo. V prípade komunikácie cez viacero skokov, keď cieľový uzol nie je v dosahu zdrojových uzlov, medziľahlé uzly pomáhajú pri prenose správ do ich cieľových uzlov. Tieto siete môžu fungovať nezávisle alebo sa môžu pripojiť k väčšej sieti, ako je napríklad internet. Tieto siete využívajú bezdrôtové spojenia, pri ktorých mobilní hostitelia komunikujú priamo, bez akéhokoľvek ďalšieho podporného zariadenia. Tieto siete sa nazývajú bezdrôtové siete s viacerými skokmi alebo mobilné rádiové siete [2].

Internet vecí (IoT) je technológia, ktorá sa rýchlo rozvíja a sľubuje revolúciu v globálnom svete prostredníctvom vzájomne prepojených fyzických objektov. IoT sa zameriava na zariadenia s nízkou spotrebou energie, ktoré spolu komunikujú cez internet. Koncept IoT pritiahol pozornosť výskumnej komunity s cieľom prepojiť nositeľné technológie, senzory, inteligentné spotrebiče, práčky, tablety, smartfóny, inteligentné dopravné systémy a iné entity do spoločného rozhrania,

ktoré im umožňuje navzájom komunikovať. IoT umožňuje komunikáciu medzi strojmi (Machine-to-Machine M2M), čo je forma komunikácie dát medzi heterogénnymi zariadeniami bez ľudskej intervencie prostredníctvom bezproblémového komunikačného média [3]. Internet vecí vytvorí priaznivé prostredie, ktoré ovplyvní mnohé aspekty každodenného života, podnikových aplikácií a prispeje k rastu svetovej ekonomiky prostredníctvom masívnych a kritických IoT aplikácií. Aplikácie IoT si vyžadujú, aby obrovské množstvo inteligentných zariadení bolo prepojených, čo môže byť nasadené v prostrediach ako lodná doprava, inteligentné domy (budovy) a mestá, inteligentné energetické systémy a monitorovanie poľnohospodárskych prostredí. Tie vyžadujú časté aktualizácie do cloudu pri nízkych nákladoch na koncové spojenie [4].

Uzly v sieťach MANET sú mobilné a ich spojenia sú oproti tradičným sieťovým infraštruktúram veľmi dynamické, čo umožňuje zariadeniam efektívne sa spájať a zdieľať užitočné informácie aj v zložitých situáciách a na vzdialených miestach, kde nie sú dostupné tradičné sieťové služby. Vďaka týmto vlastnostiam sa MANET využíva vo veľkom množstve aplikácií. Integrácia MANET a IoT do prostredí 5G a následne 6G sietí predstavuje kľúčový krok k realizácii spoľahlivej, dostupnej a vysoko efektívnej komunikácie, ktorá je nevyhnutná pre široké spektrum aplikácií [5].

V súčasnosti sú siete MANET považované za jednu z najlepších nových technológií pre mobilnú výpočtovú techniku. MANET sú najrýchlejšie rastúcou sieťou vďaka zariadeniam, ktoré sú prenosné, výkonné a cenovo dostupné. Oproti káblovým sieťam majú MANET niektoré jedinečné vlastnosti, ktoré prinášajú výzvy v oblasti bezpečnosti a smerovania. Komunikačné médium je zdieľané, čo znamená, že prístup k nemu môžu získať aj nelegitímni používatelia, čo narušuje bezpečnosť. V MANET sieťach neexistuje jasná hranica obrany ako v káblových sieťach, kde sú pevné trasy. V MANET sieťach každé zariadenie môže byť považované za smerovač [2]. Zabezpečenie takejto siete je výzvou, pretože rozsah možných útokov je obrovský. Ďalšie výzvy sú viazané na zdroje servera a dynamiku topológie siete. Siete MANET ponúkajú rôzne aplikácie, ale pred ich nasadením v komerčnom rozsahu je potrebné riešiť niektoré typické výzvy a problémy, ktoré vyžadujú zváženie. Medzi ne patria [5]:

- **zmena topológií** – Uzly v MANET sú vysoko dynamické, pohybujú sa nezávisle, často sa pripájajú alebo odpojujú zo siete a pohybujú sa v nej podľa vlastného uváženia. Zabezpečenie takéhoto dynamického prostredia predstavuje výzvu, pretože uzly sa v sieti ľahko pohybujú a vyžadujú bezpečnosť kdekoľvek a kedykoľvek,
- **smerovanie** – Dynamika topológie spôsobuje problémy aj pri smerovaní. Vzhľadom na neustále pohyby uzlov nie sú použiteľné proaktívne protokoly a je nutné využiť reaktívne smerovacie protokoly. Multicastové smerovanie sa stáva výzvou, pretože uzly sa voľne pohybujú a multicastový strom už nie je statický. Okrem toho zdroj a cieľ nemusia byť v rádiovom dosahu jeden druhého, čo vyžaduje viacskokovú komunikáciu, ktorá je zložitejšia než jednoskoková,
- **objavovanie zariadení** – Dynamické aktualizácie sú potrebné na identifikáciu dôveryhodných uzlov vstupujúcich do siete a informovanie o ich príchodoch, aby sa zabezpečil výber optimálnej trasy. Autentifikácia uzlov je nevyhnutná pre tento proces,
- **obmedzené zdroje** – Bezdrôtové spojenia majú obmedzenú kapacitu a menia sa, čo predstavuje množstvo výziev. Okrem toho, takmer všetky uzly MANET fungujú na zdrojoch energie, ktoré sa dajú vyčerpať. Preto je dôležité, aby dizajn mobilných uzlov vážne zohľadňoval úsporu energie. Spotreba energie by mala byť čo najnižšia, čo si vyžaduje použitie energeticky úsporných smerovacích protokolov, ktoré sú zároveň bezpečné,
- **spoľahlivosť** – V MANET sa vyskytuje množstvo problémov so spoľahlivosťou, ako sú konštantný dosah bezdrôtového prenosu, problém skrytých uzlov, straty paketov spôsobené mobilitou a chyby pri prenose dát,
- **kvalita služby (QoS)** – Zabezpečenie rôznych úrovní QoS pre zariadenia v dynamickom prostredí, ako je MANET, je náročnou úlohou. Vzhľadom k nepredvídateľnej povahe komunikácie v MANET nie je možné zaručiť stabilnú QoS. Preto je nevyhnutné implementovať prispôsobivú QoS,
- **medzisieťovanie** – Občas je nevyhnutné, aby MANET mohli komunikovať aj so sieťami s infraštruktúrou.

V takom prípade by mobilné zariadenia mali mať súbežne fungujúce smerovacie protokoly. Táto súbežná existencia predstavuje výzvu pre riadenie mobility [6].

A. Umelá inteligencia a algoritmus PSO v MANET-IoT

Pojem Inteligencia roja (Swarm Intelligence SI) sa týka strojového učenia, ktoré je inšpirované biologickým správaním sa spoločenského hmyzu, ako sú mravce, včely atď. Hlavným konceptom je distribuovaná povaha algoritmov, kde jednotliví agenti disponujú len veľmi obmedzenými

pamäťovými a výpočtovými zdrojmi. SI je vhodná pre distribuované siete, kde mobilita a zmeny topológie sú kľúčové, pričom energia nie je obmedzená, ako napríklad siete MANET [7].

Optimalizácia roja častíc (Particle swarm optimization PSO) je výpočtová metóda, ktorá rieši problémy iteratívne tým, že sa snaží zlepšiť kandidátske riešenia zadané danou mierou kvality. Táto metóda pracuje s populáciou kandidátskych riešení, nazývaných častice, a pohybuje týmito časticami v priestore hľadania podľa jednoduchých matematických vzorcov, ktoré zahŕňajú polohu a rýchlosť častíc. Pohyb každej častice je ovplyvnený jej najlepšou známou polohou v lokálnom okolí, ale pohybuje sa aj smerom k najlepším známym polohám v priestore hľadania, ktoré sa aktualizujú, keď iné častice nájdu lepšie polohy. Cieľom tohto procesu je posunúť roj smerom k najlepším možným riešeniam.

Princíp fungovania PSO algoritmu [8]:

1) *Inicializácia konštánt algoritmu* – Nastavenie konštánt, ako sú koeficienty akcelerácie a inercie, ktoré ovplyvňujú pohyb častíc.
2) *Inicializácia riešenia z priestoru riešenia* – Počiatočné umiestnenie častíc v priestore riešenia s náhodnými rýchlosťami, predstavujúce potenciálne riešenia.
3) *Vyhodnotenie fitness funkcie pre všetky častice* – Posúdenie, ako dobre aktuálne pozície častíc riešia daný problém.
4) *Aktualizácia individuálne a globálne najlepšie hodnot*y – Identifikácia a zapamätanie si najlepších riešení pre jednotlivé častice (pbest) a pre celý roj (gbest).
5) *Aktualizácia rýchlosti a polohy každej častice* – Úprava rýchlostí a polôh častíc na základe pbest a gbest, čo ich vedie k lepším riešeniam.
6) *Iterácia* – Opakovanie krokov 3 až 5, až kým nie je splnená podmienka ukončenia, ako je maximálny počet iterácií alebo dosiahnutie dostatočnej konvergencie k optimálnemu riešeniu [9].

II. Implementácia algoritmu umelej inteligencie do smerovacieho protokolu v MANET-IoT

Predtým, ako sme začali s optimalizáciou parametrov AODV protokolu pomocou PSO algoritmu, bolo potrebné urobiť niekoľko dôležitých krokov. Prvým krokom bolo vyhodnotenie existujúcich knižničných implementácií PSO algoritmu, aby sme pochopili ich možnosti a obmedzenia. Nakoľko sme zistili, že štandardné implementácie, napríklad PySwarm alebo Torch-PSO, nie sú vhodné pre špecifické požiadavky našej sieťovej simulácie, rozhodli sme sa pre vlastnú implementáciu PSO algoritmu. Táto implementácia musela obsahovať flexibilnú fitness funkciu, ktorá by efektívne hodnotila rôzne konfigurácie AODV protokolu. Každá častica v "Particle"triede reprezentovala možný súbor parametrov AODV a ich polohy v riešenom priestore boli kľúčové pre hľadanie optimálneho riešenia.

A. Implementácia PSO algoritmu na optimalizovanie vstupných parametrov AODV protokolu do MANET-IoT

Pri implementácii optimalizačného procesu bol použitý simulačný nástroj NS-3 s jeho integrovaným modulom AODV protokolu [10]. S rozšírením NS3-AI pre NS3, bolo možné do simulácie integrovať PSO algoritmus pomocou Python skriptu, ktorý bol zameraný na spúšťanie a optimalizáciu parametrov AODV protokolu. PSO algoritmus kontinuálne spracovával údaje zaznamenávané do súboru vo formáte JSON, ktorý obsahoval výsledky predchádzajúcich simulácií. Na základe analýzy týchto údajov, PSO algoritmus vyhodnotil aktuálny stav konfigurácie a prispôsobil parametre AODV. Tieto získané poznatky sa iteratívne využívali na ďalšie zlepšenia, čo umožnilo algoritmu neustále sa zlepšovať pri hľadaní optimálnej sady parametrov pre zvýšenie výkonnosti siete. Tento prístup k úpravám parametrov zabezpečil, že každá nová iterácia simulácie bola efektívnejšia, ako tá predchádzajúca. Výsledky a údaje, ktoré boli kľúčové pre hodnotenie úspešnosti a smerovanie ďalšieho zlepšenia optimalizácie, boli dôležité pre iteratívne cykly zlepšovania, ktoré sú súčasťou PSO algoritmu.

Simulačný nástroj NS-3 je široko uznávaný v oblasti sieťového výskumu, najmä pre svoju podporu rôznych protokolov a technológií potrebných pre ad-hoc siete. Jednou z hlavných výhod NS-3 je priamo implementovaný modul pre AODV protokol. Simulácie v NS-3 sú napísané v programovacom jazyku C++. NS-3 poskytuje realistickú modeláciu fyzickej vrstvy pre bezdrôtové kanály, vrátane štatistických modelov pre simuláciu rôznych podmienok prenosu signálu. Tieto schopnosti sú nevyhnutné pre analyzovanie vplyvov prostredia a mobility na sieťovú výkonnosť. NS-3 taktiež obsahuje sofistikované propagačné a mobilné modely, ako aj metódy pre alokáciu zdrojov a generovanie premávky, čo umožňuje presné a komplexné simulácie. NS-3 neustále rozširuje svoje schopnosti s novými protokolmi a modelmi, a to nielen v rámci tradičných sieťových protokolov.

```
def update_velocity(self, global_best_particle_position):
    for i in range(nv):
        r1 = random.random()
        r2 = random.random()

        cognitive_velocity = c1 * r1 * (self.local_best_particle_position[i] - self.particle_position[i])
        social_velocity = c2 * r2 * (global_best_particle_position[i] - self.particle_position[i])
        self.particle_velocity[i] = w * self.particle_velocity[i] + cognitive_velocity + social_velocity
```

Obr. 1: Ukážka funkcie $upadate\_velocity$ pre MANET-IoT v simulátore NS-3.

Väčšina algoritmov umelej inteligencie sa opiera o open-source frameworky ako TensorFlow a PyTorch, ktoré sú vyvíjané nezávisle od NS-3, čo ich integráciu sťažuje. Rozšírenie NS3-AI pre NS-3 ponúka efektívne riešenie na umožnenie dátovej interakcie medzi NS-3 a ďalšími Python založenými AI frameworkmi. V NS3-AI, interakcia dát medzi simuláciou NS-3 a Python skriptom prebieha cez definované štruktúry Env (prostredie) a Act (akcie), ktoré sú navzájom prepojené. Tieto štruktúry reprezentujú stav simulovaného prostredia a rozhodnutia, ktoré majú byť vykonané. Trieda $Experiment$ potom tieto dáta posiela a prijíma medzi NS-3 a Python scriptom prostredníctvom zdieľanej pamäťovej oblasti, čo umožňuje Python skriptu analyzovať stav prostredia a podľa toho určovať najvhodnejšie akcie. Rozhranie $py_interface$, ktoré je súčasťou NS3-AI modulu, poskytuje podrobné pokyny na implementáciu interakcie medzi NS-3 a AI modelmi. Toto rozhranie obsahuje príklady a návody, ktoré ukazujú, ako sa dajú definovať a použiť štruktúry Env a Act na prenos dát a ako s pomocou triedy Experiment efektívne integrovať AI algoritmy do NS-3 simulácií. Trieda Experiment v NS3-AI slúži ako kontrolný bod pre NS-3 simulácie. Atribút $mempool_{key}$ je identifikátor používaný pre prístup k zdieľanej pamäti medzi NS-3 a AI procesom, mem_{size} definuje veľkosť tejto pamäťovej oblasti, a názov súboru simulácie je cesta k skriptu NS-3, ktorý sa má spustiť. Táto trieda zjednodušuje nastavenie a spustenie simulácií a umožňuje efektívnu výmenu údajov nevyhnutných pre AI spracovanie.

B. Proces integrovania PSO do AODV protokolu pre MANET-IoT

Pre optimalizáciu parametrov AODV protokolu pomocou PSO algoritmu, bol použitý integrovaný prístup, kde jedna iterácia PSO zahŕňa spustenie desiatich simulácií v simulačnom prostredí NS-3. Celkový proces zahŕňa 30 iterácií, čo predstavuje 300 simulácií na dosiahnutie optimálneho nastavenia parametrov. Algoritmus bol navrhnutý tak, aby efektívne spravoval a vykonával simulácie počas každej iterácie:

1) ***Inicializácia*** – Na začiatku každej iterácie PSO algoritmus inicializuje roj častíc, kde každá častica predstavuje jednu potenciálnu konfiguráciu parametrov protokolu AODV. Každá častica má svoju unikátnu pozíciu v priestore parametrov a rýchlosť, ktorá určuje smer a rýchlosť jej pohybu.
2) ***Spúšťanie simulácií*** – Každá častica spustí desať nezávislých simulácií s jej súčasným nastavením parametrov v NS-3. Výsledky týchto simulácií sa zhromažďujú a agregujú, čím sa získava spoľahlivý odhad fitness hodnoty pre danú pozíciu častice.
3) ***Aktualizácia stavu častíc*** – Na základe vypočítanej fitness hodnoty každá častica aktualizuje svoju pozíciu a rýchlosť v závislosti od svojich osobných najlepších riešení a najlepšieho riešenia nájdeného celým rojom. Tento krok zahŕňa kognitívne a sociálne komponenty, ktoré napomáhajú zdieľaniu informácií a učeniu sa v rámci roja.

Funkcia $objective\_function()$ je základná fitness funkcia v implementácii PSO algoritmu pre optimalizáciu parametrov AODV protokolu v simulačnom prostredí NS-3. Táto funkcia prijíma vektor parametrov X, ktorý obsahuje nastavenia rôznych parametrov AODV, ako sú $TtlStart$, $TtlIncrement$, $TtlThreshold$, a ďalšie, ktoré sú kľúčové pre správanie a výkon sieťovej simulácie. Tieto výstupy sú potom normalizované a zahrnuté do fitness funkcie.

Trieda Particle, ktorá reprezentuje jednu časticu v rámci PSO, používa $objective\_function(X)$ na vyhodnotenie a porovnanie svojej aktuálnej pozície v priestore riešení. Každá častica udržuje záznam o svojej najlepšej doteraz nájdenej pozícii ($local\_best\_particle\_position$) a jej fitness hodnotu. Keď je nová pozícia častice vyhodnotená ako lepšia, častica aktualizuje svoje lokálne najlepšie hodnoty.

Funkcia $upadate\_velocity$ riadi pohyb častíc v priestore riešení (Obr.1). Táto metóda aktualizuje rýchlosť každej častice založenú na jej predchádzajúcej rýchlosti a pozícii, pozícii jej osobného najlepšieho riešenia, a pozícii najlepšieho riešenia nájdeného v celom roji. Je základným prvkom pre dynamické prispôsobovanie sa častíc pri hľadaní optimálneho riešenia. Parameter $global\_best\_particle\_position$ predstavuje najlepšie globálne riešenie (najlepšiu pozíciu), ktoré bolo nájdené celým rojom častíc. Slúži ako referenčný bod pre sociálnu zložku vplyvu na rýchlosť častice. Správne nastavenie koeficientov inercie, kognitívnej a sociálnej zložky je kritické pre dosiahnutie dobrej konvergencie algoritmu a zabránenie zasekávaniu sa v lokálnych minimách alebo nadmernému blúdeniu bez zmysluplného smerovania.

Tabuľka 1: Nastavie simulačných parametrov simulátore NS-3

Parameter	Hodnoty
Rozmer simulačného prostredia	300 x 300 m
Veľkosť paketov	64 bajtov
Komunikačný protokol	$IEEE802\_11b$
Počet uzlov v sieti	15
Rýchlosť uzlov	0 – 20 m/s
Interval posielania paketov (za sekundu)	1 paket
Dĺžka simulácie	100s

Tabuľka 2: Porovnanie výkonnostných metrík pred a po implementácii PSO algoritmu do protokolu AODV v MANET-IoT

Parameter	AODV	AODV - PSO	Zmena
PDR [%]	76,44	88,69	16,02
E2EDelay$[s]$	0,1958	0,0435	77,79
Throughput$[Mb/s]$	0,0107	0,0124	15,89

III. Analyzované sieťové parametre

Na vykonanie výkonovej analýzy AODV protokolu pred a po aplikácii PSO algoritmu boli využité tieto parametre:

- ***priemerný počet doručených paketov PDR***- pomer medzi počtom paketov prijatých cieľovým uzlom (P_r) a celkovým počtom paketov odoslaných zo zdrojového uzla (P_s). Tento pomer poskytuje informácie o efektívnosti smerovacích protokolov tým, že zobrazuje konkrétne údaje o strate paketov počas prenosu.

$$PDR_{avr} = \frac{P_r}{P_s} 100\%$$

- ***end-to-end oneskorenie E2E***- definovaný ako čas, ktorý je potrebný na prenos paketu od zdrojového uzla k cieľovému uzlu. Jednotka času je sekunda. Tento časový údaj je zásadný pre posúdenie reaktívnosti a efektivity smerovania v rámci siete, a je určený na základe meraní doby, ktorú paket strávi cestou medzi dvoma koncovými bodmi.

$$Oneskorenie_{avr} = \frac{\sum_{i=1}^{n} (t_r - t_s)}{P_r}$$

- ***Priepustnosť***- definovaná ako súčin počtu dátových paketov a veľkosti paketu v určenej časovej jednotke, pričom jednotkou je bit za sekundu. Priepustnosť sa určuje na základe celkového množstva dát prenesených v sieti (KT) počas stanoveného časového intervalu (T), čo umožňuje posúdiť účinnosť a kapacitu sieťovej prenosovej rýchlosti.

$$Priepustnosť_{avr} = \frac{P_r KT}{T}$$

Tieto normalizované hodnoty sú potom kombinované do váženého súčtu (costu), kde váhy odrážajú prioritu každej metriky vzhľadom na celkový sieťový výkon. Vyššia váha 0.4 pre PDR odráža jeho vysokú dôležitosť pre celkovú spoľahlivosť a efektívnosť siete. Váha 0.3 pre priepustnosť znamená, že táto metrika je dôležitá, ale mierne menej prioritná v porovnaní s PDR. Negatívna váha -0.3 pre oneskorenie naznačuje, že pri optimalizácii sa snažíme dosiahnuť čo najnižšie oneskorenie.

A. Analýza dosiahnutých výsledkov

Cieľom bolo identifikovať a implementovať optimálne nastavenia pre AODV, aby sme zlepšili jeho efektivitu a spoľahlivosť v simulačnom prostredí NS-3 (Tab. 1). Porovnanie výkonnostných metrík pred a po implementácii PSO algoritmu jasne ukazuje, že aplikácia tohto optimalizačného prístupu na sieťový protokol AODV viedla k celkovému zlepšeniu všetkých sledovaných metrík. PSO algoritmus bol efektívny pri hľadaní optimálnych nastavení pre AODV.

Výsledky ukázali, že PSO algoritmus nielenže zlepšil jednotlivé metriky, ale aj celkovú stabilitu a spoľahlivosť siete. To znamená, že sieť je schopná efektívnejšie spracovávať a distribuovať dátové pakety, minimalizovať straty, znižovať oneskorenia pri prenose a zvyšovať celkovú priepustnosť. Tieto zlepšenia sú kritické pre aplikácie vyžadujúce vysokú mieru spoľahlivosti a nízke oneskorenia. V tabuľke č. 2 sú zobrazené priemerné hodnoty sledovaných metrík pred a po implementovaní

PSO algoritmu do sieťového protokolu AODV. Spustených bolo celkom 150 simulácií protokolu AODV, z ktorých sme vybrali priemer pre porovnanie s 15 spusteniami PSO algoritmu

Pri optimalizácii sieťového protokolu AODV pomocou PSO algoritmu bol identifikovaný najefektívnejší súbor parametrov, ktorý výrazne zlepšil výkonnosť siete. Táto optimálna konfigurácia, ktorá bola dosiahnutá po 30 iteráciách, kde každá iterácia zahŕňala 10 simulácií, dosiahla nákladovú hodnotu 8,12, čo predstavuje globálne najlepšiu pozíciu častice, a to naznačuje výrazné zlepšenie v rámci všetkých kľúčových metrík sieťového výkonu. Pomocou najnižšej nákladovej hodnoty 8,12 daného spustenia PSO algoritmu, bolo možné identifikovať optimálne hodnoty parametrov AODV protokolu.

IV. Záver

V článku sme sa zamerali na možnosti využívania PSO algoritmu v smerovacom protokolole AODV pre MANET - IoT sieť. Porovnanie výkonnostných metrík pred a po implementácii PSO algoritmu jasne ukazuje, že aplikácia tohto optimalizačného prístupu na sieťový protokol AODV viedla k celkovému zlepšeniu všetkých sledovaných metrík. PSO algoritmus bol efektívny pri hľadaní optimálnych nastavení pre AODV.

Výsledky ukázali, že PSO algoritmus nielenže zlepšil jednotlivé metriky, ale aj celkovú stabilitu a spoľahlivosť siete. To znamená, že sieť je schopná efektívnejšie spracovávať a distribuovať dátové pakety, minimalizovať straty, znižovať oneskorenia pri prenose a zvyšovať celkovú priepustnosť. Tieto zlepšenia sú kritické pre aplikácie vyžadujúce vysokú mieru spoľahlivosti a nízke oneskorenia. V tabuľke č.2 sú zobrazené priemerné hodnoty sledovaných metrík pred a po implementovaní PSO algoritmu do sieťového protokolu AODV.

Celkové zistenia tejto štúdie potvrdzujú, že PSO je významným nástrojom pre inžinierov a výskumníkov v oblasti sietí, ktorý umožňuje lepšiu kontrolu a správu sietí v náročných prostrediach, ako je MANET-IoT. Výsledky tejto práce poskytujú podklady pre ďalší vývoj a aplikácie PSO v praxi, kde môže byť tento algoritmus využitý na zlepšenie rôznych aspektov sieťovej infraštruktúry a aplikácií založených na internete vecí

Poďakovanie

Tento výskum bol financovaný agentúrou na podporu výskumu a vývoja SR, výskumný grant č. APVV-17-0208 a VEGA 1/0260/23.

Literatúra

[1] M. N. Siddiqui, K. R. Malik, and T. S. Malik, "Performance analysis of blackhole and wormhole attack in manet based iot," in *2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, 2021, pp. 1–8.

[2] V. K. Quy, V. H. Nam, D. M. Linh, and L. A. Ngoc, "Routing algorithms for manet-iot networks: A comprehensive survey," *Wirel. Pers. Commun.*, vol. 125, no. 4, p. 3501–3525, aug 2022. [Online]. Available: https://doi.org/10.1007/s11277-022-09722-x

[3] R. E. Ahmed, "A novel multi-hop routing protocol for d2d communications in 5g," in *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, 2021, pp. 0627–0630.

[4] G. Dinesh, Y. S. Pawar, P. Dinshi, S. S, S. S. Husain, and C. Venkata Krishna Reddy, "Machine learning based secure data transmission and improvement in manet through internet of things (iot)," in *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, 2023, pp. 1006–1012.

[5] Z. Niu, Q. Li, C. Ma, H. Li, H. Shan, and F. Yang, "Identification of critical nodes for enhanced network defense in manet-iot networks," *IEEE Access*, vol. 8, pp. 183 571–183 582, 2020.

[6] A. Gupta and T. Sasikala, "Secure routing protocols for manet-enabled iot," in *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, 2021, pp. 1–4.

[7] M. Achankunju, R. Pushpalakshmi, and A. V. A. Kumar, "Particle swarm optimization based secure qos clustering for mobile ad hoc network," in *2013 International Conference on Communication and Signal Processing*, 2013, pp. 315–320.

[8] J. J. Liang and B. Y. Qu, "Large-scale portfolio optimization using multiobjective dynamic mutli-swarm particle swarm optimizer," in *2013 IEEE Symposium on Swarm Intelligence (SIS)*, 2013, pp. 1–6.

[9] W. Chuanjun, W. Ling, and R. Xuejing, "General particle swarm optimization algorithm," in *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, 2023, pp. 1204–1208.

[10] T. W. Ching, A. H. M. Aman, W. M. H. Azamuddin, and Z. S. Attarbashi, "Performance evaluation of aodv routing protocol in manet using ns-3 simulator," in *2021 3rd International Cyber Resilience Conference (CRC)*, 2021, pp. 1–4.

Analytic system for solving the magnetic field distribution in anisotropic magnetic materials

[1]*Šimon GANS,* [2]*Ján MOLNÁR*

[1,2] Department of Theoretical and Industrial Electrical Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

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

***Abstract* — This paper addresses the problem of solving 2D magnetic field equations in anisotropic magnetic media. Usually in engineering textbooks that address electromagnetic problems only isotropic problems are considered for the sake of simplicity. When the need arises for describing fields in anisotropic media numerical simulation software is used to compute the distribution usually via the finite element method (FEM). In this work, we developed an analytical approach that yields the same results as the numerical one while some simplifying assumptions hold.**

***Keywords* — analytic, anisotropy, COMSOL Multiphysics, magnetic fields**

I. Introduction

When solving magnetic field problems analytically usually linearity and isotropy are assumed. There are many analytical tools to obtain the solution [1]. However, for obtaining the solution for anisotropic media, not many tools are presented in the literature and FEM simulations are usually used for this task [2].

The magnetic characteristics of most ferromagnetic materials are usually at least a bit anisotropic [3]. If the anisotropy is weak (as in hot-rolled electrical steel) it can be neglected [4]. If it is strong (as in cold-rolled steel) it has to be taken into account [5].

The FEM is a useful tool in the design processes in mechanical [6] and electrical [7] engineering. The FEM solves for a linearized version of nonlinear systems. With a finer discretization, higher solution accuracy is obtained.

If the model is complex then employing the FEM comes with long computation times. Model simplifications can be employed like reduction of the model due to symmetry [8] or neglecting some physical behaviors of the model due to the operating conditions (ignoring eddy currents due to low magnetizing frequency) [9].

Analytical solutions of systems are preferred because they offer an exact solution whereas the FEM offers just an approximation (but a sufficient one if done correctly). Analytical solutions usually require advanced mathematical tools to obtain[1]. Many partial differential equations (PDEs) cannot be solved at all.

This work was inspired by the work of Nowicki [10], in which a winding placement in anisotropic media must be found so that the magnetic flux through it will be zero. This paper offers a method to describe magnetic field distribution in anisotropic media around a single conductor. In linear media, this solution can be expanded to an infinite amount of conductors by summing their fields.

II. Magnetic vector potential in isotropic media

The magnetic field (the $\boldsymbol{B}$ field) around a current-carrying wire can be solved by multiple approaches, like using Ampere's law. In this paper, we modified the solution using the magnetic vector potential (the $\boldsymbol{A}$ field) [11].

The $\boldsymbol{A}$ field distribution around a wire with a current I flowing through it can be seen below (Fig. 1). The field lines form concentric circles centered around the wire obeying equation (1)[12],

$$A_z(r) = -\frac{I\mu}{2\pi}\ln\left(\frac{r}{r_0}\right) \tag{1}$$

where r is the distance from the center of the wire, μ is the permeability of the medium and r_0 is the distance at which $A_z = 0$ (which can be chosen arbitrarily).

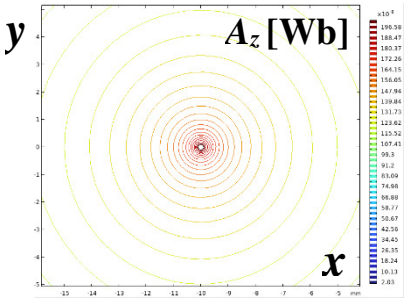


Fig. 1: Simulated $\boldsymbol{A}$ field distribution around a current-carrying wire in the xy-plane.

The $\boldsymbol{B}$ field is defined by the $\boldsymbol{A}$ field by eq. (2) by applying the rotation operator [12].

$$\boldsymbol{B} = \nabla \times \mathbf{A} \tag{2}$$

If only the z-component of the $\boldsymbol{A}$ field is non-zero (which is true in 2D systems), then the $\boldsymbol{B}$ field is defined by eq. (3)[12],

$$\boldsymbol{B} = \frac{\partial A_z}{\partial y}\vec{\imath} + \frac{\partial A_z}{\partial x}\vec{\jmath} \tag{3}$$

where $\boldsymbol{i}$ and $\boldsymbol{j}$ are the unit vectors of the x and y axes respectively.

III. Effect of Anisotropy on the $\boldsymbol{A}$ Field Distribution

If the permeability in the x direction is μ_x and the permeability in the y direction is μ_y, a permeability tensor $\boldsymbol{\mu}$ can be defined using eq. (4),

$$\boldsymbol{\mu} = \begin{vmatrix} \mu_x & 0 \\ 0 & \mu_y \end{vmatrix} = \begin{vmatrix} \mu_x & 0 \\ 0 & s \cdot \mu_x \end{vmatrix} \tag{4}$$

where s was defined as the "anisotropy factor". The solutions for a case where $s = 0.5$ ($\mu_x = 400$, $\mu_x = 200$) for the $\boldsymbol{B}$ and the $\boldsymbol{A}$ fields are shown in Fig. 2 and Fig. 3. The $\boldsymbol{B}$ field lines are no longer circular and a simple analytical expression was not found to describe them directly.

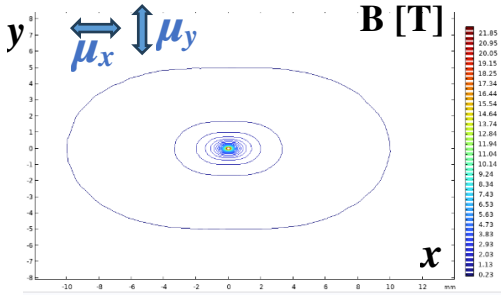


Fig. 2: The $\boldsymbol{B}$ field distribution in anisotropic media ($\mu_x = 400$ and $\mu_y = 200$).

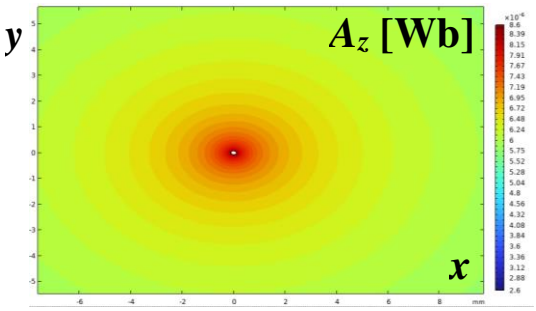


Fig. 3: The $\boldsymbol{A}$ field distribution in anisotropic media ($\mu_x = 400$ and $\mu_y = 200$).

The ***A*** field lines are elliptic, which was found out by the following empirical test:

- In multiple simulations, an ellipse was drawn around the wire. Various axes length ratios of the ellipse were tested (Fig. 4).
- The A_z potential was analyzed along the circumference of the ellipse. If the potential is constant, then the given ellipse is an equipotent line (Fig. 4).

The general equation of an ellipse is given by eq. (5), where *a* and *b* are the lengths of its semi-major and semi-minor axes.

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \tag{5}$$

Empirically it was found that if the ratio between *b* and *a* obeys eq. (6), then the given ellipse (if centered around the wire) has the shape of equipotential lines (Fig. 4).

$$\frac{b}{a} = \sqrt{s} \tag{6}$$

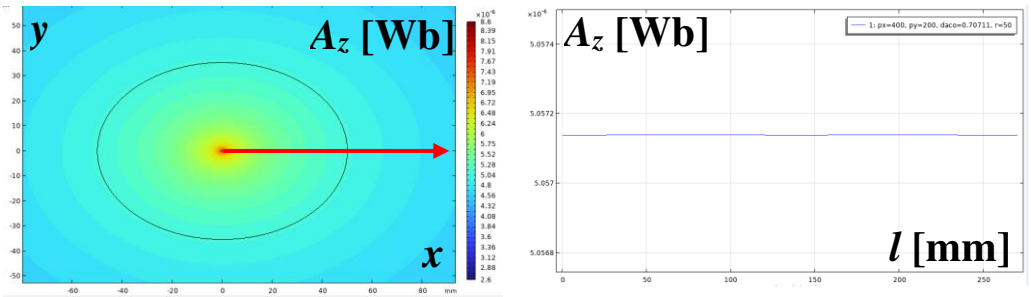


Fig. 4: Left - the **A** field of an anisotropic medium, where $s = 0.5$. An ellipse was drawn according to eq. (6). Right – the A_z potential along the circumference of the ellipse.

IV. Analytic Expression for the *A* Field

If we plot the simulated A_z distribution on the positive half of the *x*-axis (the red arrow in Fig. *5)* we obtain the graph in Fig. 5. The distribution shows a clear logarithmic trend.

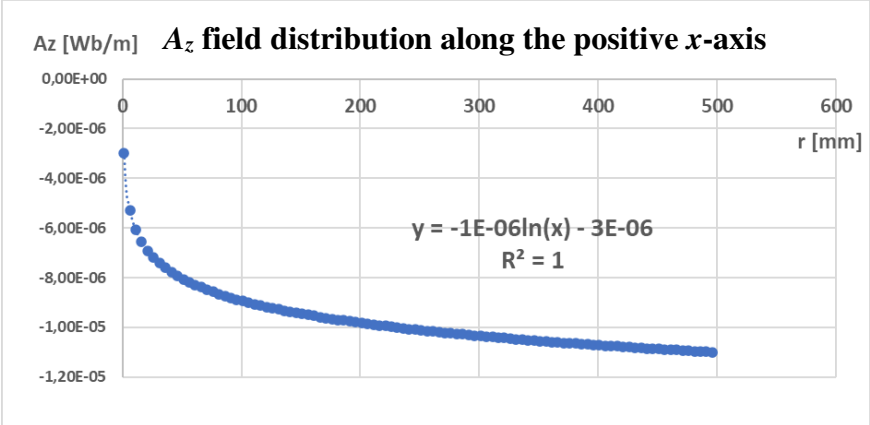


Fig. 5: The A_z field distribution along the x-axis (in the case of $s = 0.5$).

The trend is expressed by eq. (7), which is similar to eq. (1), where *C* is a constant.

$$A_z(x, 0) = -C\ln(x) + A_{z0} \tag{7}$$

From empirical testing, it was found that the *C* is defined well by eq. (8). The value for A_{z0} is not needed since the ***B*** field is defined from the ***A*** field by differentiation.

$$C = \frac{I\mu_x\sqrt{s}}{2\pi} \tag{8}$$

The expression for the A_z values of points along the x-axis is known. Every elliptic field line crosses the x-axis. Then the expression for the ***A*** field in an arbitrary point is obtained by finding the *a* semi-axis length of the ellipse on which the given point lies.

After combining eq. (5) and eq. (6) and substituting the arbitrary point coordinates x_0 and y_0 eq. (9) is obtained.

$$\frac{x_0^2 s + y_0^2 - a^2 s}{a^2 s} = 0 \tag{9}$$

It follows that $a \neq 0$, which means that the $\boldsymbol{A}$ field cannot be defined at the center of the conductor, which is a drawback of the model and it will be analyzed further. From eq. (9) the length a is expressed by eq. (10).

$$a = \sqrt{x_0^2 + \frac{y_0^2}{s}} \tag{10}$$

The expression for a can be substituted into eq. (7) for x. The final expression for the $\boldsymbol{A}$ field is given by eq. (11).

$$A_z(x,y) = -\frac{I\mu_0\mu_x\sqrt{s}}{2\pi} \ln\left(\sqrt{x^2 + \frac{y^2}{s}}\right) + A_{z0} \tag{11}$$

The $\boldsymbol{B}$ field is given by eq. (12) by using eq. (3) on eq. (11).

$$\vec{B} = \left(-\frac{I\mu_0\mu_x}{2\pi\sqrt{s}} \cdot \frac{y}{x^2 + \frac{y^2}{s}}\right)\vec{\imath} - \left(\frac{I\mu_0\mu_x\sqrt{s}}{2\pi} \cdot \frac{x}{x^2 + \frac{y^2}{s}}\right)\vec{\jmath} \tag{12}$$

The region close to the conductor may show some differences that get smaller as the distance from the conductor increases which is analyzed in the next section.

The solution assumes the absence of demagnetizing fields and end effects, which is also assumed in most analytic solutions to electromagnetic problems.

V. Comparison of the Mathematical Model with Simulation Results

To compare the simulated results with the mathematical model a FEM model was created with a circular conductor placed in the center with a radius of 1 mm surrounded by an infinite ferromagnetic plane as shown in Fig. 6. At the test points TP1 to TP4 the $\boldsymbol{B}$ field magnitude was enumerated and compared with the mathematical model. The same was done along the test line.

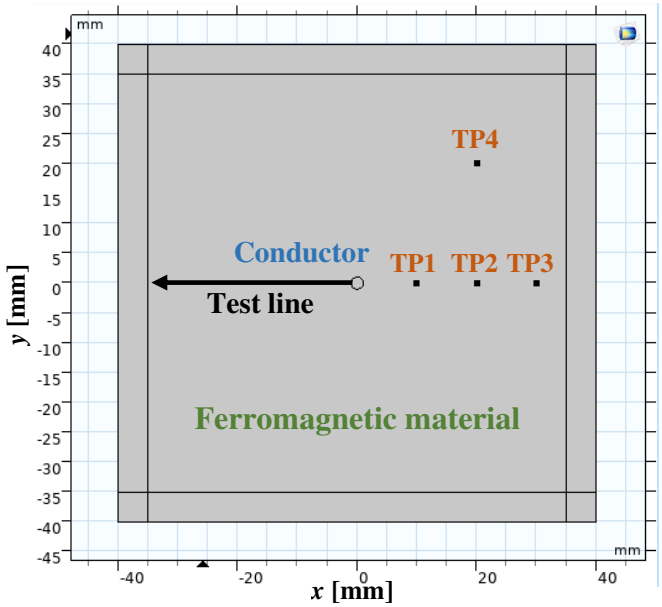


Fig. 6: The simulation model of the system with probe points (TP1 to TP4) and a test line. The ferromagnetic material was described with μx = 5000 and μy = 20000.

The simulated and the analytical distribution of the $\boldsymbol{B}$ field is shown in Fig. 7. The magnetic field shapes in the analyses are similar which suggests the correctness of the model.

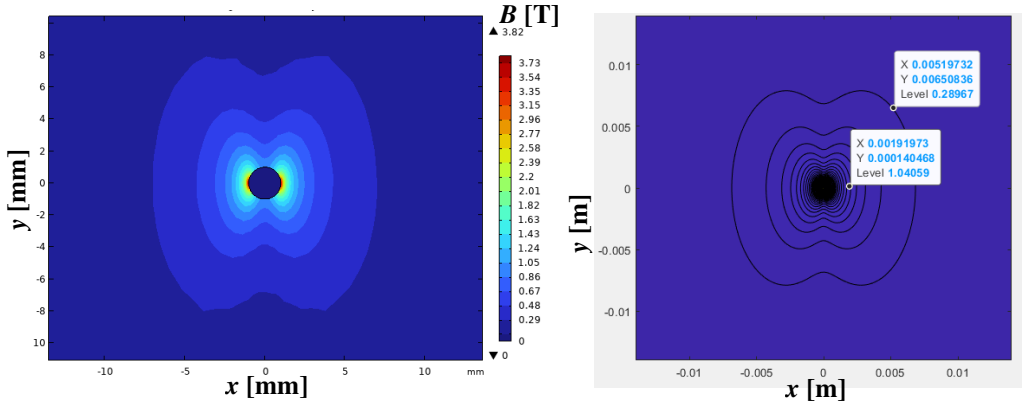


Fig. 7: Visual comparison between the simulated results (left) and the mathematical model (right).

The positions of the testing points are shown in the table below (TABLE 1).

TABLE 1: Positioning of the testing points.

Testing point	*x*-coordinate	*y*-coordinate
TP1	10 mm	0 mm
TP2	20 mm	0 mm
TP3	30 mm	0 mm
TP4	20 mm	20 mm

The comparison of the testing point ***B*** field between the mathematical model and the simulation is shown in TABLE 2. As the distance from the conductor increases the relative error between the mathematical model and the simulations decreases.

TABLE 2: Comparison of the ***B*** field norm in the given testing points between the simulation and the mathematical model.

Testpoint	B_{model} [T]	B_{sim} [T]	δ_{model} [%]
TP1	0,2	0,200274	0,136
TP2	0,1	0,099995	0,00472
TP3	$6{,}666 \cdot 10^{-2}$	$6{,}66469 \cdot 10^{-2}$	0,02958
TP4	$8{,}24621 \cdot 10^{-2}$	$8{,}2441 \cdot 10^{-2}$	0,02549

The comparison between the two approaches along the testing line from Fig. 6 with the expression of the relative error is shown in Fig. 8. The largest error of approximately 90% is found directly at the surface of the conductor. At approx. 1 mm from the surface the relative error decreases to 10% and at 3 mm (and larger) the two approaches are identical.

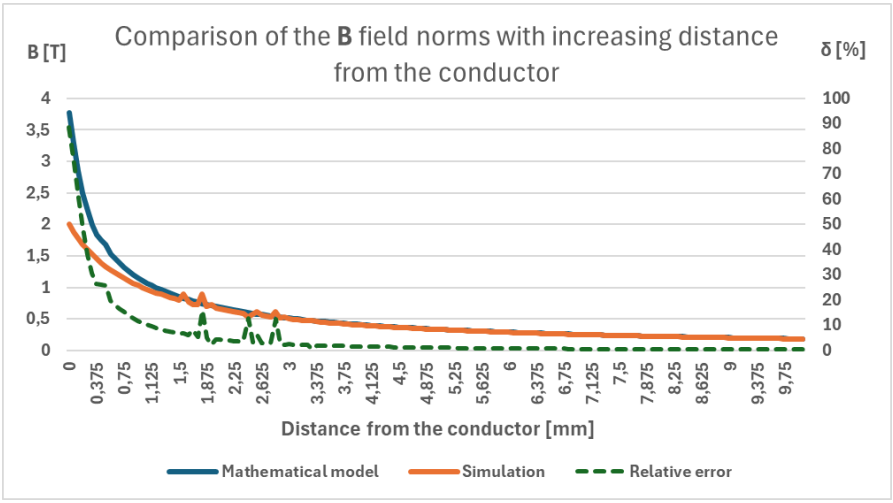


Fig. 8: Comparison of the error between the mathematical model and the simulation results along the test line. The mathematical model was the ground truth in the relative error computation.

VI. Conclusion

The derived expression can be used effectively to analyze linear anisotropic 2D systems with various current carrying conductors inside them. The given model arose from a desire to analytically solve field distribution in analytical media which occur in various physical systems.

From the comparison of the model with the simulation results it was found that the mathematical model is trustworthy at a distance of 3 mm or more from the conductor surface (or 3 times the conductor radius). This approach can be used in problems where an accurate description of the field in such regions is not needed (for example the field at the center of an solenoid with an inner diameter of 6 times or more than the conductor diameter).

The following simple solution was not seen before in scientific literature by the authors. In the future work an accurate solution of the field around the conductor is being developed.

References

[1] Lubin T., Mezani S., Rezzoug A., Analytical Computation of the Magnetic Field Distribution in a Magnetic Gear, IEEE Transactions on Magnetics, 2010, doi: 10.1109/TMAG.2010.2044187. Available on the internet at: https://ieeexplore.ieee.org/document/5427037

[2] Enokizono M., Matsumura K., Magnetic Field Analysis of Anisotropic Permanent Magnet Problems by Finite Element Method, IEEE Transactions on magnetics, 1997, doi: 10.1109/20.582576. Available on the internet at: https://ieeexplore.ieee.org/document/582576

[3] Rheinboth H., Vlastnosti a použití magnetických materiálů, Praha, Státní nakladatelství technické literatury, 1975, 1. Vydanie, ISBN (viaz.)

[4] Elgamli E., Anayi F., Advancements in Electrical Steels: A Comprehensive Review of Microstructure, Loss Analysis, Magnetic Properties, Alloying Elements, and the Influence of Coatings, MDPI Applied sciences, 2023, 13(18), doi: 10.3390/app131810283. Available on the internet at: https://www.mdpi.com/2076-3417/13/18/10283?type=check_update&version=2

[5] Price K., GoodeB., Power D., Grain-oriented electrical steels for power and distribution transformers, Ironmaking & Steelmaking, 2016, 43:9, doi: 10.1080/03019233.2016.121122. Available on the internet at: https://doi.org/10.1080/03019233.2016.1211122

[6] Valero-Páez L. B., Chacón-Velasco J. L., González-Estrada O. A., Optimization of the connecting rod of a two-stroke engine using finite element analysis, Journal of Physics: Conference Series, Volume 1386, 5th Internation Meeting for Researchers in Materials and Plasma, 2019, doi: 10.1088/1742-6596/1386/1/012114. Available on the internet at: https://iopscience.iop.org/article/10.1088/1742-6596/1386/1/012114

[7] Tümbek M., Oner Y., Kesler S., Optimal Design of Induction Motor with Multi-Parameter by FEM Method, International Conference on Electrical and Electronics Engineering, ELECO, 2015, doi: 10.13140/RG.2.1.3149.5762. Available on the internet at: https://www.researchgate.net/publication/287760082_Optimal_Design_of_Induction_Motor_with_Multi-Parameter_by_FEM_Method?channel=doi&linkId=567924d708ae6041cb49ec06&showFulltext=true

[8] Gans Š., Molnár J., Using model symmetry to shorten computation time in COMSOL Multiphysics, JIEE Časopis priemyselnej elektrotechniky, 2021, vol. 5, issue 2, ISSN 2454-0900. Available on the internet at: https://www.jiee.eu/using-model-symmetry-to-shorten-computation-time-in-comsol-multiphysics/

[9] Gans Š., Molnár J., Kováč D., Kováčová I., Fecko B., Bereš M., Jacko P., Dziak J., Vince T., Driving Signal and Geometry Analysis of a Magnetoelastic Bending Model Pressductor Type Sensor, MDPI Sensors, 2023, 23(20), 8393, doi: 10.3390/s23208393. Available on the internet at: https://www.mdpi.com/1424-8220/23/20/8393

[10] Nowicki M., Tensductor – Amorphous Alloy Based Magnetoelastic Tensile Force Sensor, MDPI Sensors, 2018, 18(12), 4420, doi: 10.3390/s18124420. Available on the internet at: https://www.mdpi.com/1424-8220/18/12/4420

[11] Kvasnica J., Teorie elektromagnetického pole, Academia Praha 1985, 1. vydanie, 03/5-1466 21-081-85, strany 99-105.

[12] Mayer D., Teorie elektromagnetického pole 1. díl, Západočeská univerzita, 2004. ISBN 80-7082-826-9.

Meranie koncentrácie CO_2 a možný jeho vplyv v pedagogickom procese

[1]*Milan GUZAN,* [2]*Simona KIREŠOVÁ*

[1,2] Katedra teoretickej a priemyselnej elektrotechniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1] milan.guzan@tuke.sk, [2] simona.kiresova@tuke.sk

***Abstrakt* — Článok sa zaoberá meraním koncentrácie CO_2 v miestnostiach posluchární a laboratórií Technickej univerzity v Košiciach počas prebiehajúcich semestrov s následným porovnaním nameraných hodnôt s už odsledovanými hranicami vplyvu na študentov z predchádzajúcich prác zahraničných autorov. Sleduje sa vývoj CO_2 počas vyučovania i po jeho skončení na základe začiatočných hodnôt pred ďalším blokom výučby a ponúka riešenie na minimalizáciu CO_2 v učebniach. Takto je možné pozitívne ovplyvniť odozvu študentov na aktivity počas prebiehajúceho pedagogického procesu.**

***Kľúčové slová* — aplikácia, senzor, CO_2, vizualizácia, vyhodnocovanie údajov**

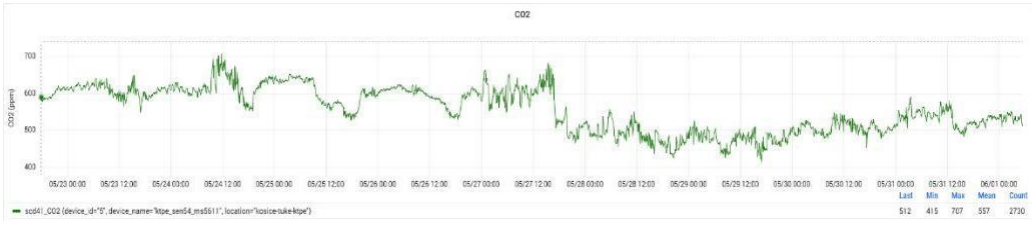
I. Úvod

Súčasná doba je charakteristická meraním všetkého vôkol nás. Meranie teploty ovzdušia, vody, teploty povrchu ciest, vlhkosti či tlaku sú samozrejmosťou. Začíname si všímať aj iné parametre ako napr. pevné častice (PM) na niektorých križovatkách, počet cyklistov ktorí prešli cez križovatku, organické chemické zlúčeniny (VOC), parametre vplývajúce na index kvality ovzdušia (PM2.5, PM10, O_3, NO_2 a SO_2) a v neposlednom rade aj CO_2. Meranie CO_2 je zaujímavé nielen v exteriéri, ale pokiaľ ide aj o pedagogický proces tak aj v interiéri. Podnetom merania CO_2 v interiéroch boli práce [1], [2], ktoré sledovali úroveň CO_2 v školách a zisťovali vplyv na študenta či absencie študentov počas vyučovania. Negatívne vplyvy CO_2 [3] na osoby vyskytujúce sa v danom prostredí je možné zhrnúť v Tab. 1 [4].

Tab. 1
Koncentrácia CO_2 a vplyv na človeka

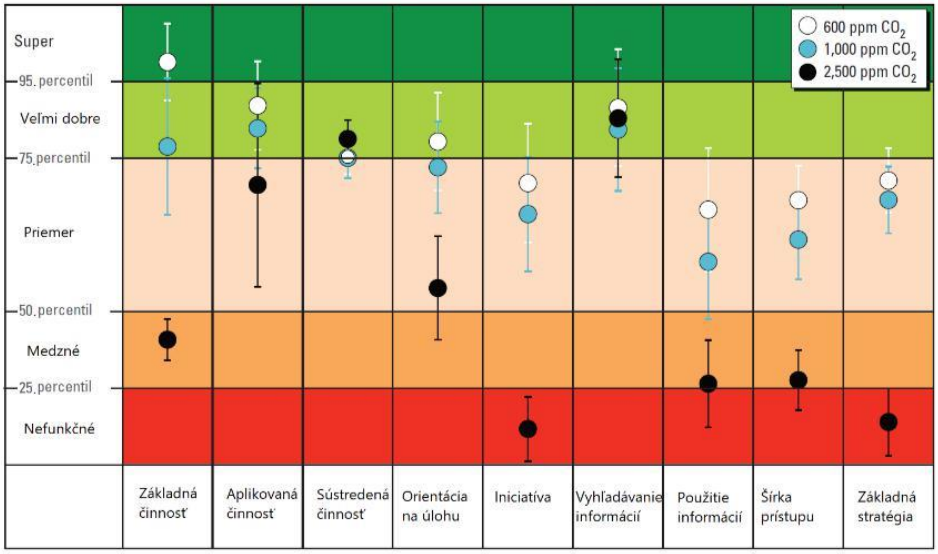
CO_2 koncentrácia (ppm)	Vplyv CO_2 na ľudské zdravie
400 – 1 000	Nie sú známe žiadne významné účinky na zdravie väčšiny osôb. Ide o hodnoty bežne merané vo vnútri budov.
1000 – 2000	Úrovne, pri ktorých sa ľudia môžu cítiť unavení a kvalita ovzdušia sa považuje za zlú.
2000 – 5000	Koncentrácia spojená s bolesťami hlavy, ospalosťou a pocitom starého a vydýchaného vzduchu. Môžu sa vyskytnúť problémy so sústredením, zvýšená srdcová frekvencia a mierna nevoľnosť.
40 000	Bezprostredné nebezpečenstvo v dôsledku nedostatku kyslíka, škodlivé pri expozícii.

Bežne o ovzduší je koncentrácia CO_2 v rozmedzí 400 – 700 ppm. Takého hodnoty boli namerané v areáli TUKE, konkrétne na budove PK3, ktorá je sídlom pracoviska autorov. Zmeny CO_2 za 9 dní od 23. 5. 2024 do 1. 6. 2024 ilustruje Obr. 1.



Obr. 1 Zmeny koncentrácie CO_2 v areáli TUKE, budova PK3 od 23.5.2024 do 1.6.2024.

Práca [1] tiež sledovala pomocou dotazníkov aj vplyv CO_2 o koncentrácii 600, 1000 a 2500 ppm na jednotlivé činnosti prebiehajúce počas vyučovania. Sumárny prehľad je na Obr. 2. Z neho je zrejmé, že niektoré činnosti je možné úspešne realizovať takmer bez ohľadu na koncentráciu CO_2 v miestnosti ako napr. pri sústredenej pozornosti alebo vyhľadávaní informácií. Zhoršené výsledky až znemožnenie aktívneho zapojenia sa študentov do vyučovacieho procesu nastane pri koncentrácii 2500 ppm CO_2 v miestnosti v takmer polovici z testovaných činností, čo pri 90-minútových prednáškach resp. cvičeniach môže zohrať dôležitú úlohu pri vnímaní, aktivite, uvažovaní či rozhodovaní sa ako riešiť danú úlohu. Otázkou je, aké koncentrácie budú namerané v laboratóriách či prednáškových miestnostiach TUKE, kde prebieha pedagogický proces.



Obr. 2 Prehľadová tabuľka vplyvu troch testovaných koncentrácií na aktivity študentov.

II. Možnosti merania CO_2

V súčasnosti sa ponúka niekoľko možností merania CO_2 senzorom SCD41 od firmy Sensirion. Prvou je možnosť pripojenia senzora (Obr. 3a) CO_2 na USB port počítača – ide o senzor SCD4x CO_2 Gadget [5]. Nainštalovaním programu je možné získať prehľad o zmenách CO_2 v miestnosti, komunikácia je cez Bluetooth. Iné riešenie je pripojenie na power banku s nasledujúcou inštaláciou aplikácie do mobilného telefónu. Druhá možnosť je meranie senzorom SCD 41_BOB
[6] (Obr. 3b). Ide o senzor pripojiteľný cez zbernicu I2C k mikrokontroléru, pričom existuje aj jeho SMD verzia [7]. Týmito senzormi boli namerané výsledky opísané v nasledujúcej kapitole.



Obr. 3 Senzor SDC41 na meranie CO_2 pripojiteľný na USB port (vľavo) a vyžadujúci mikrokontrolér (vpravo).

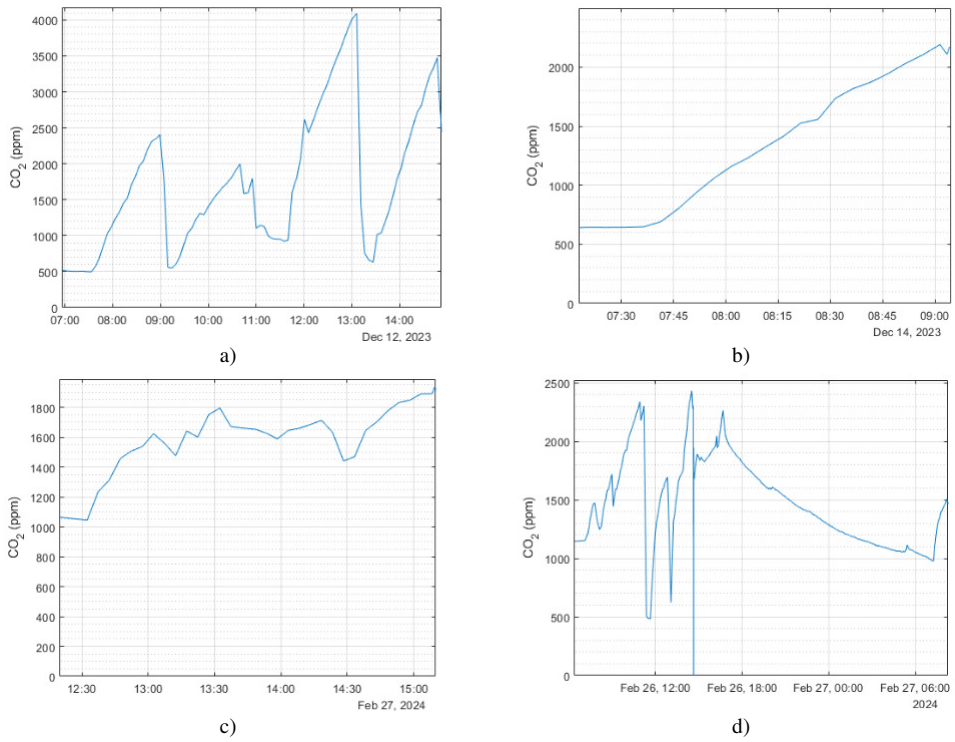
III. Výsledky merania CO_2 počas vyučovania

Merania koncentrácie CO_2 zobrazené na Obr. 4 prebiehali počas akademického roku 2023/2024 v zimnom i letnom semestri. Vysvetlenie jednotlivých obrázkov bude širšie obsiahnuté v komentári k Obr. 4a) a ďalší komentár k zvyšným obrázkom bude podobný a stručnejší. Na vizualizáciu priebehov bol použitý MATLAB.

Na Obr. 4a je ukázané meranie CO_2 počas vyučovania od 7:30 hod. do 15:00 hod., na pracovisku autorov, budova PK3, v laboratóriu L1. Jeho rozmer je 5×11 m. Pred začiatkom vyučovania bolo nameraných 500 ppm CO_2. Vyučovanie jedného predmetu prebiehalo v čase 7:30 - 9:00 hod. (prednáška) a 9:10 – 10:40 hod. (cvičenie) s 8 študentmi. Počas prednášky sa nevetralo a koncentrácia CO_2 dosiahla úroveň takmer 2500 ppm. Prudký pokles CO_2 počas prestávky do 9:10 hod. naznačuje vetranie miestnosti a od 9:10 hod. zo začiatkom cvičenia, koncentrácia CO_2 opäť stúpa avšak nedosiahne takú vysokú hodnotu, kvôli jednému pootvorenému oknu. Maximum na konci cvičenia je 2000 ppm. Po skončení cvičenia aj vďaka

možnému prievanu a dlho otvoreným dverám laboratória L1 nastal pokles koncentrácie CO_2. V prípade len otvorených dverí nastáva „vyliatie“ CO_2 do susedného priestoru – v tomto prípade ide o chodbu s rozmerom 3×11 m. Takéto „vyliatie“ CO_2 bolo pozorované aj v domácnosti, keď napr. ráno boli otvorené dvere (a neotvorené okno), tak v susednej miestnosti bola zaznamenaná zvýšená koncentrácia CO_2. Faktom je, že do začiatku ďalšieho bloku prednáška/cvičenie iného predmetu so začiatkom o 11:35 hod. úroveň CO_2 nedosiahla úroveň pred 7:30 hod., ale sa ustálila na hodnote okolo 900 ppm. Na ďalšej prednáške od 11:35 do 13:05 hod. presiahla koncentrácia CO_2 pri zatvorených oknách hodnotu 4000 ppm, ale počas prestávky a intenzívnom vetraní bol zaznamenaný pokles na prijateľných 600 ppm. Na nasledujúcom cvičení pri zatvorených oknách koncentrácia CO_2 dosiahla 3500 ppm.

Obr. 4b) ilustruje nárast CO_2 počas cvičenia predmetu ZEI v posluchárni budovy PK7, kde boli spojené 2 študijné skupiny – spolu 55 študentov. Zo začiatočných 640 ppm o 7:30 hod., za 90 minút cvičenia dosiahla koncentrácia CO_2 hodnotu 2200 ppm pri otvorenom 1 okne. V miestnosti je požiadavka pri odchode pozatvárať okná, takže podobne veľká skupina ktorá príde o 9:10 hod. začína s koncentráciou CO_2 okolo 2000 ppm, čo aj pri otvorení troch okien nemusí byť vyhovujúce pre ďalší priebeh cvičenia.



Obr. 4 Namerané koncentrácie CO_2 počas zimného semestra v rôznych miestnostiach a dňoch naznačených vľavo dole a) laboratórium L1 na PK3, b) v posluchárni budovy PK7, c) v posluchárni P27, d) v kancelárii počas 24 hodín.

Zmeny CO_2 počas trojhodinového cvičenia s 59 študentmi v miestnosti P27 sú na Obr. 4c). Ako je z obrázka zrejmé, miestnosť nebola dostatočne vyvetraná po ukončení predchádzajúceho bloku výučby, pretože začiatočná hodnota CO_2 je 1100 ppm. Nasledujúco však boli vyklopené obe okná v posluchárni a koncentrácia CO_2 nepresiahla 1900 ppm. V miestnosti však bolo cítiť chlad, čo pre pedagóga z hľadiska trojhodinového rozprávania môže znamenať problém s hornými dýchacími cestami v nasledujúcich dňoch. Po ukončení výučby po 15:10 hod. museli byť obe okná zatvorené a koncentrácia CO_2 klesala prirodzeným únikom cez netesnosti okien či dverí.

Prirodzený pokles CO_2 v inej miestnosti ako na Obr. 4c) ilustruje Obr. 4d). Ide o 24 hodinové meranie v kancelárii pedagóga. Po príchode do práce o 7:00 hod. začala koncentrácia CO_2 stúpať zo začiatočnej hodnoty 1150 ppm z predchádzajúceho dňa. Postupným vetraním počas neprítomnosti a opätovnou prítomnosťou pedagóga v kancelárii dosiahne CO_2 takmer 2500 ppm a to až do odchodu z kancelárie o 17:00 hod. Po tejto hodine, za počas nesledujúcich 12 hodín je vidieť exponenciálny pokles koncentrácie CO_2 na podobnú hodnotu ako v predchádzajúcom dni avšak 2× vyššiu ako v ovzduší, čo je dôsledok nedostatočného vetrania.

Bolo by možné uviesť ešte niekoľko ďalších ilustratívnych obrázkov zmien CO_2 v ďalších väčších, či menších miestnostiach TUKE kde prebieha výučba s menším počtom študentov až po pospájané, v prvom ročníku, 2 či 3 študijné skupiny - princíp vzrastu a poklesu CO_2 by bol stále

rovnaký. Ak si uvedomíme, že podľa [1], [8], [9] existuje vzťah medzi CO_2 a absenciami študentov v dôsledku ochorení, bolo by vhodné venovať dostatok pozornosti nadmernej koncentrácii CO_2 v posluchárňach a laboratóriách počas vyučovacieho procesu.

IV. ZÁVER

Problém vysokých, naakumulovaných koncentrácií CO_2 v budovách je nielen možné, ale priam nutné riešiť vetraním. Priebeh vetrania sa bude líšiť podľa aktuálneho ročného obdobia nielen v posluchárňach ale aj v domácnostiach. Súčasný tlak na zníženie spotreby energie má za následok nízku mieru vetrania, čo vedie zasa k vyšším koncentráciám CO_2 v miestnosti. Vyššie koncentrácie CO_2 majú napokon škodlivý vplyv na rozhodovaciu schopnosť študenta/pedagóga čím sa narušuje cieľ pedagogického procesu.

Výhodu majú také miestnosti, kde sa môže vetrať aj nasledujúcich napr. 45 min. pred začatím ďalšieho bloku výučby. Také sú umiestnené napr. na 1.poschodí budovy, kde nehrozí prienik tretej, neželanej strany s následnou krádežou. Tiež je vhodné, keď po väčšom bloku výučby sa môže v miestnosti upratovať a zároveň aj vetrať. To sú však prípady po skončení výučby medzi jednotlivými blokmi.

Ak by malo ísť o vetranie počas vyučovania, je vhodné na konci väčšej miestnosti ponechať otvorené aspoň 1 okno avšak tak, aby nebola narušená tepelná pohoda sediacich, teda študentov a v neposlednom rade aj pedagóga samotného, ktorý problematiky vysvetľuje. Vhodné by bolo aj pred samotným začiatkom vyučovania, aby bola miestnosť vyvetraná (ako napr. na Obr. 4a,b)), čím sa predíde vysokým začiatočným koncentráciám CO_2 s následným rýchlym zvýšením nad hranicu 2000 ppm podobne, ako to ilustruje Obr. 4a) po 11:00 hod., alebo Obr. 4d). Miera vetrania bude tak kontrolovaná koncentráciou CO_2 v miestnosti a preto by bolo vhodné, aby každá učebňa bola vybavená napr. informačným panelom merajúcim, vyhodnocujúcim a ilustrujúcim aktuálny stav koncentrácie CO_2. Okrem toho by bolo v budúcom výskume CO_2 vhodné zahrnúť objektívne metódy merania kognitívnych schopností, ako sú logické a postrehové úlohy, a tak kvantifikovať vplyv CO_2 na študentov a pedagógov, čo by následne umožnilo vytvoriť konkrétne a efektívne odporúčania pre zlepšenie vnútornej kvality ovzdušia na univerzitách.

REFERENCES

[1] Satish, U. et al. "Is CO_2 an Indoor Pollutant? Direct Effects of Low-to-Moderate CO_2 Concentrations on Human Decision-Making Performance Attendance in Washington and Idaho." Environmental Health Perspectives • volume 120, number 12, December 2012, pp. 1671 – 1677.

[2] Shendell, D.G. et al. "Associations Between Classroom CO_2 Concentrations and Student. " Indoor Air (LBNL-54413), 2004, p.14.

[3] Wisconsin Department of Health Services: Carbon Dioxide [Online]. [cit. 2024-02-03]. Dostupné na internete: <https://www.dhs.wisconsin.gov/chemical/carbondioxide.htm#:~:text=1%2C000%E2%80%932%2C000%20ppm%3A%20level%20associated,nausea%20may%20also%20be%20present.>

[4] Šoffa, S. "Automatizácia merania vybraných veličín v interiéri, ich vizualizácia a vyhodnocovanie. " Technická univerzita v Kočiciach, FEI, Diplomová práca 2024, p. 100, vedúci práce: doc. Guzan Milan.

[5] SOS Electronic. "SCD4x CO_2 Gadget." [Online]. Dostupné na: <https://soselectronic.com/sk/products/sensirion/scd4x-co2-gadget-358536?query=scd41>. [cit. 2024-5-31].

[6] SOS Electronic. "SCD4x_BOB." [Online]. Dostupné na: <https://soselectronic.com/sk/products/various/scd-41-bob-398492?query=scd41>. [cit. 2024-5-31].

[7] SOS Electronic. "SCD4x-D-R1." [Online]. Dostupné na: <https://soselectronic.com/sk/products/sensirion/scd41-d-r1-347339?query=scd41>. [cit. 2024-5-31].

[8] Erdmann, C.A., Steiner, K.C,, Apte, M.G., "Indoor carbon dioxide concentrations and SBS symptoms in office buildings revis- ited: Analyses of the 100 building BASE Study dataset." *Proceedings of Indoor Air 2002* Conference, Monterey, CA, Vol. 3, pp. 443-448. Indoor Air 2002, Santa Cruz, CA.

[9] Milton D.K., Glencross P.M., Walters M.D. "Risk of sick leave associated with outdoor ventilation level , humidification, and building-related complaints. " *Indoor Air 2000*, 10 (4): 212-221.

Vplyv bypass diód na zatienenie fotovoltických panelov

[1]*Tomáš MANIAK,* [2]*Marek PAVLÍK*

[1] Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]tomas.maniak@student.tuke.sk, marek.pavlik@tuke.sk

***Abstrakt* — Príspevok skúma vplyv bypass diód na výkon fotovoltických (FV) panelov pri zatienení. So zvyšujúcim sa povedomím o globálnom otepľovaní je potrebné prejsť na obnoviteľné zdroje energie, ako sú fotovoltické (FV) systémy. Nárast cien elektriny v minulom roku so sebou priniesol obavu laickej verejnosti, ktorá čím ďalej tým viac siaha po možnosti zníženia závislosti na distribučnej sieti. Ako najlepšia alternatíva sa ukazuje využívanie slnečnej energie kvôli jednoduchosti inštalácie. Avšak prevádzka fotovoltickej elektrárne na streche domu vyžaduje aj potrebné znalosti, ktoré môžu ovplyvniť množstvo vyrobenej elektrickej energie. Napríklad zatienenie jedného fotovoltického panela zapojeného v stringu ovplyvňuje výrobu elektrickej energie z celého stringu. Tento príspevok sa zameriava na vplyv bypass diód na prevádzku fotovoltických panelov. Poukazuje na množstvo energie, ktoré je možné získať použitím bypass diód. Napríklad, pokles výkonu P_{MPP} je oproti nezatienenému FVP o 58%. Použitie bypass diód teda zabránilo poklesu P_{MPP} o 76 % - prípad zatieneného FVP, na čo poukazuje tento príspevok.**

***Kľúčové slová* — fotovoltický panel, bypass dióda, zatienenie fotovoltického panela**

I. Úvod

S rastúcim povedomím o globálnom otepľovaní a jeho vplyve na životné prostredie je stále zrejmejšie, že musíme zmeniť spôsob, akým vyrábame a spotrebúvame energiu. Fosílne palivá, ktoré tvoria základ tradičných energetických systémov, emitujú veľké množstvo skleníkových plynov, ktoré prispievajú k zvýšeniu emisií a klimatickým zmenám. Obnoviteľné zdroje energie, akými sú slnečná, veterná a vodná energia, pred-stavujú čistejšiu alternatívu, ktorá má menší negatívny dopad na životné prostredie.

Obnoviteľné zdroje energie sú kľúčové pre udržateľný energetický systém, avšak sú spojené s niekoľkými výzvami, ktoré je potrebné vyriešiť. Hlavnou nevýhodou je ich premenlivá výroba energie spôsobená výkyvmi počasia a zmenou ročných období a teda aj menšia spoľahlivosť oproti konvenčným zdrojom elektrickej energie. Túto nevýhodu vieme eliminovať záložnými fosílnymi zdrojmi alebo skladovaním energie, avšak v sú-časnosti je skladovanie technologický náročne a drahé a okrem toho sú potrebné rozsiahle investície do prenosových sústav. Ďalšími negatívami sú vysoké investičné náklady a potreby veľkých plôch. Napriek tomu, prechod na obnoviteľné zdroje je nevyhnutný pre udržateľnú energetickú budúcnosť, a technologické inovácie sa snažia minimalizovať tieto nevýhody.

Do skupiny obnoviteľných zdrojov energie patrí aj fotovoltika. Svojím potenciálom poskytuje riešenia pre znižovanie emisií skleníkových plynov, diverzifikáciu energetických zdrojov a vytváranie udržateľnejších energetických systémov a stáva sa kľúčovým prvkom pre dosahovanie energetických a environmentálnych cieľov. Fotovoltické panely sa neustále vyvíjajú, čoho výsledkom je zvyšovanie ich účinnosti a životnosti [1].

Fotovoltické články sa zvyčajne delia na tri generácie. Kategorizácia vychádza z materiálov a techník používaných pri ich výrobe. Prvá generácia je založená na kryštalických kremíkových doštičkách. Výrobné technológie prvej generácie si vyžadujú vysoké energetické a pracovné vstupy, ktoré bránia významnému pokroku pri znižovaní výrobných nákladov. Jednotlivé kremíkové zariadenia sa blížia k teoretickej účinnosti 33 %. V súčasnosti bola dosiahnutá laboratórna účinnosť pri monokryštalickom článku 27,6 % a pri polykryštalickom 23,3 %. Materiáli pre FV články druhej generácie boli vyvinuté s cieľom riešiť nižšie energetické požiadavky na výrobu a nižšie výrobné náklady. Najviac používanými materiálmi druhej generácie sú amorfný kremík, mikromorfný kremík, telurid kadmia (CdTe) a materiál CIGS (

Copper indium gallium selenide). Tieto materiály sa nanášajú v tenkej vrstve na nosný materiál ako je sklo alebo keramika, čím sa znižuje hmotnosť materiálu, a tým aj náklady na výrobu. Tieto technológie sú prísľubom vyššej účinnosti a ponúkajú výrazne nižšie výrobné náklady. Cieľom technológií tretej generácie je zlepšiť horšie elektrické parametre druhej generácie tenkovrstvových technológií pri zachovaní veľmi nízkych výrobných nákladov. Súčasný výskum sa zameriava na účinnosť 30 - 60 % pri zachovaní nízko nákladových materiálov a výrobných techník [1].

Súčasne sa viacerí výskumníci pokúšali o nové prístupy, ako je využitie fotovoltických článkov citlivých na farbivo (DSSC), perovskitových článkov, organických fotovoltických článkov, fotochemických článkov, kvantových bodiek (QD) a nanoštruktúrovaných článkov na riadenie svetla. Štvrtá generácia fotovoltických článkov patrí do triedy kompozitov [2].

II. V-A CHARAKTERISTIKA FOTOVOLTICKÉHO ČLÁNKU/PANELA

Parametre fotovoltického článku, resp. panela definujú jeho vlastnosti a správanie. Niektoré parametre odzrkadľujú kvalitu fotovoltického článku, niektoré parametre sú potrebné pri návrhu fotovoltických systémov. Viaceré parametre sa určujú pri takzvaných STC podmienkach. Štandardné testovacie podmienky "Standard Test Condition - STC" sú podmienky, za ktorých sú merané charakteristiky fotovoltických panelov, resp. článkov, tj. Intenzita žiarenia 1000 W/m^2, spektrum AM1,5 Global a teplota panelu 25 °C .

Napätie naprázdno U_{OC} je definované ako napätie na výstupných svorkách FV panela pri nezaťaženom stave a teda bez pripojenej záťaže. Toto napätie je možné považovať za maximálne napätie na FV paneli i keď, ako bolo vyššie spomenuté, parametre FV panela sa môžu meniť. Hodnota napätia naprázdno sa pri súčasných FV paneloch pohybuje radovo v desiatkach voltov.

Skratový prúd I_{SC} FV panela je maximálna hodnota prúdu pri danom osvetlení. Pri hodnote skratového prúdu sa uvažuje, že hodnota sériového odporu je rovná nule. Veľkosť skratového prúdu je taktiež závislá od intenzity osvetlenia a pohybuje sa v horných jednotkách ampérov až vyše 10A.

Jedným z najdôležitejších parametrov FV panela je výkon PMPP. Tento výkon je definovaný vzťahom: [3] [4]

$$P_{MPP} = U_{MPP}.I_{MPP} \tag{1}$$

Pričom U_{MPP} a I_{MPP} je napätie a prúd v bode MPP - Obr. 1. Bod MPP je bod na volt-ampérovej charakteristike FV panela, ktorý je charakterizovaný maximálnym výkonom FV panela - $U_{MPP}.I_{MPP} = P_{MPP}$. Volt-ampérová charakteristika je definovaná postupným zvyšovaním napätia s následným postupným poklesom prúdu. Na tejto charakteristike sa však nachádza oblasť (takzvané „koleno“), v ktorom sa MPP bod nachádza. V prípade, že fotovoltický panel pracuje v tejto oblasti a striedač dokáže tento bol identifikovať, vtedy je účinnosť premeny slnečnej energie na elektrickú najvyššia.

Pomer výkonu PMPP k výkonu dopadajúceho slnečného žiarenia je definovaný ako účinnosť FV panela – viď (2). V súčasnosti dosahuje účinnosť hodnoty 15-22% [5].

$$\eta_{ef} = \frac{P_{MPP}}{P_{rad}} = \frac{P_{MPP}}{E.A_c} \tag{2}$$

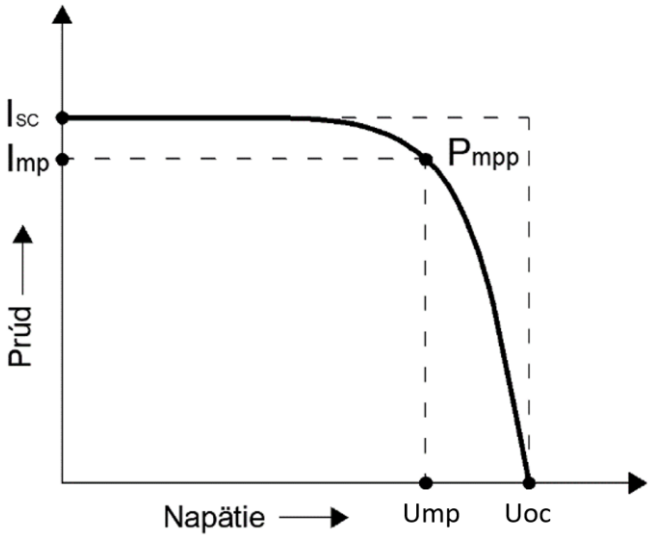
pričom P_{rad} je výkon dopadajúceho slnečného žiarenia, E je intenzita osvetlenia pri štandardizovaných skúšobných podmienkach, A_C je plocha fotovoltického článku. Z vyššie uvedených parametrov teda vyplýva, že výkon FV panela nie je dostatočná informácia nakoľko samotný výkon FV panela je možné zmeniť rozmermi FV panela. Dôležitejším parametrom je účinnosť FV panela, poprípade výkon FV panela na m^2.

Ako bolo skôr spomenuté, výkon PMPP je výkon v bode MPP, v ktorom však FV panel nemusí byť v praxi prevadzkovaný. V práci bude táto problematika podrobne popísaná neskôr. Avšak výkon v bode MPP a maximálny výkon sa však líšia. Maximálny výkon je definovaný najvyššou hodnotou napätia (U_{OC}) a najvyššou hodnotou prúdu (I_{SC}). Pomer týchto dvoch parametrov určuje faktor plnenia (Fill factor, resp. *FF*). *FF* faktor je definovaný vzťahom: [6-7]

$$FF = \frac{P_{MPP}}{I_{SC}.U_{OC}} = \frac{U_{MPP}.I_{MPP}}{I_{SC}.U_{OC}} \tag{3}$$

Fill factor (*FF*) v kontexte fotovoltiky je dôležitým parametrom fotovoltického článku a

charakterizuje jeho kvalitu. Je to miera, akou ideálny fotovoltický článok (so 100 percentnou účinnosťou a bez sériových a paralelných strát) sa blíži k skutočnému výkonu článku. Hodnota fill factoru sa zvyčajne pohybuje medzi 0,7 a 0,85 pre väčšinu komerčne dostupných fotovoltických článkov. Vysoká hodnota *FF* znamená, že fotovoltický článok je schopný efektívne prevádzať absorbovanú slnečnú energiu na elektrickú energiu. Ak je hodnota FF nízka, môže to znamenať problémy s kvalitou materiálu, výrobnými chybami alebo inými faktormi, ktoré znižujú účinnosť fotovoltického článku [3-5].



Obr. 1 Volt-ampérová charakteristika FV panela

III. VÝSKUM POUŽITIA BYPASS DIÓD PRI ZATIENENÍ FOTOVOLTICKÝCH PANELOV

Na Obr. 2 je zobrazený simulačný obvod pre skúmanie vplyvu použitia bypass diód, kde je možné vidieť tri bypass diódy (D4, D5 a D6).

Skúmaný bol vplyv bypass diódy na V-A charakteristiku a výkonovú charakteristiku a *FF* faktor. Z výsledkov výskumu je možné vyvodiť niekoľko záverov:

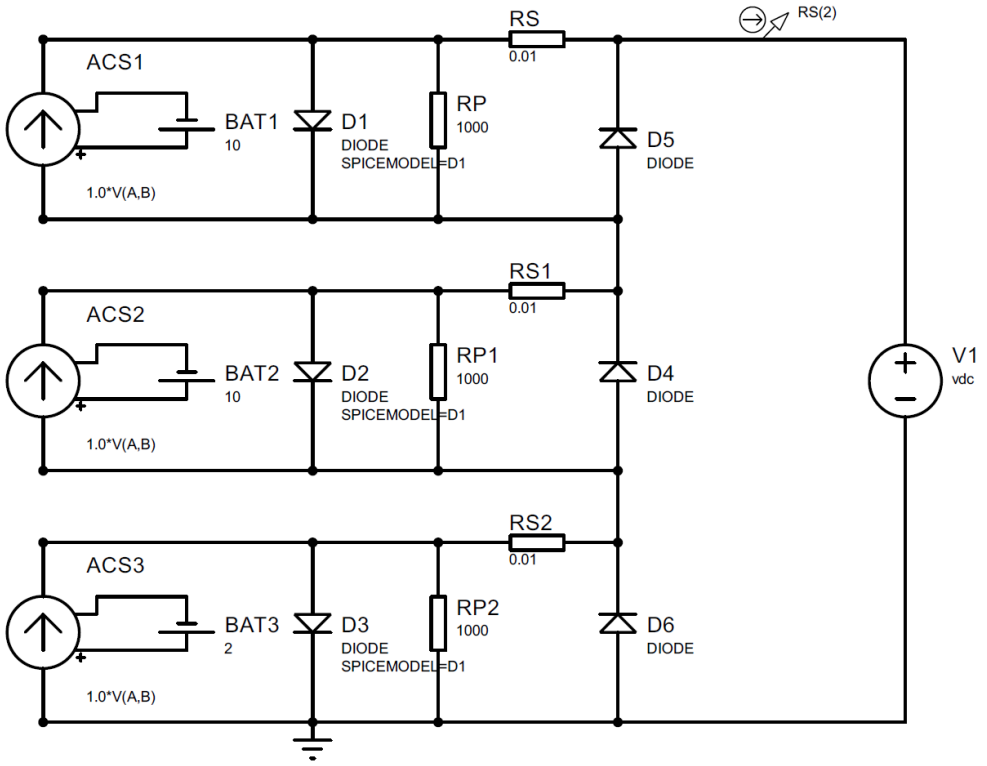
a) Na Obr. 2 sú zobrazené výsledky výskumu použitia bypass diód paralelne k FVP. Zelená V-A charakteristika reprezentuje nezatienený FVP a teda počas slnečného počasia bez zatienenia. Červená charakteristika reprezentuje V-A charakteristiku v prípade, že bude jeden FVP zatienený na úrovni 4/5 z celej plochy FVP. V tomto prípade neboli použité bypass diódy. Z výsledkov je možné vidieť, že V-A charakteristika zmenila tvar a podstatne klesol prúd z 10 A na 2 A, čo spôsobí pokles vo výrobe z FVP. Značne klesol prúd v bode P_{MPP}, napätie sa zmenilo minimálne, navyše hodnota napätia klesla, avšak minimálne. Výkon P_{MPP} značne klesol kvôli poklesu I_{MPP}. V Tab. I je možné vidieť výsledky, kde je pozorovaný pokles P_{MPP} z hodnoty 493 W na hodnotu 118 W, čo je pokles o 78 %. Toto je pokles výkonu P_{MPP} bez použitia bypass diódy.

b) Modrá V-A charakteristika na Obr. 2 reprezentuje použitie bypass diód pri zatienení FVP. Z výsledkov vyplýva, že tvar V-A charakteristiky sa po zatienení jedného FVP zmenil oproti V-A charakteristike pri nezaťaženej situácii. Avšak zmena je iná ako v prípade zatieneného FVP bez použitia bypass diódy. V prípade použitia bypass diódy je možné pozorovať pokles napätia U_{MPP} pričom prúd I_{MPP} sa takmer nezmenil. Pokles výkonu P_{MPP} je oproti nezatienenému FVP o 58%. Použitie bypass diód teda zabránilo poklesu P_{MPP} o 76 % - prípad zatieneného FVP.

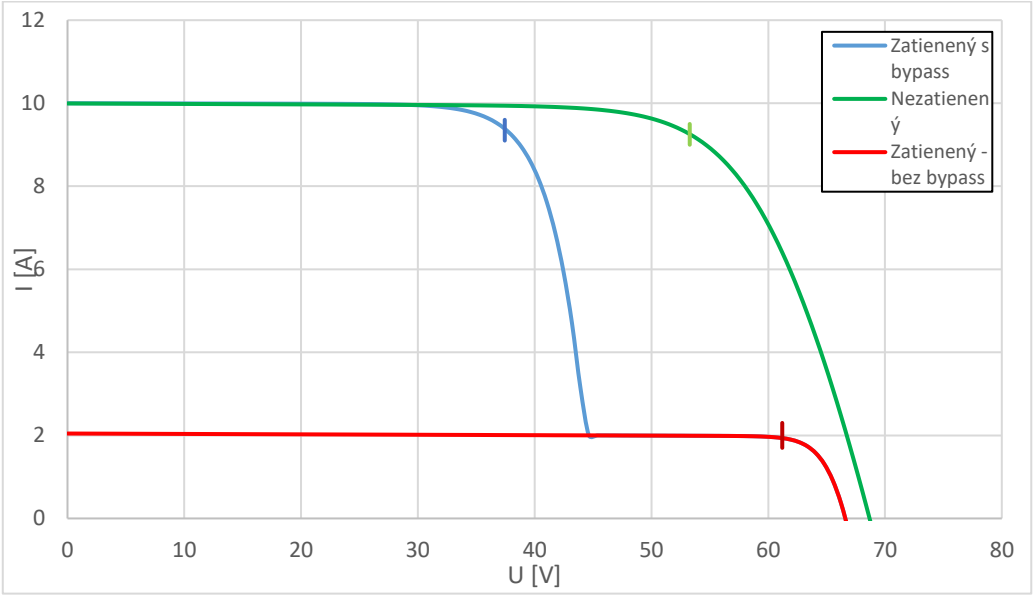
c) Napätie U_{MPP} sa s použitím bypass diód zmenilo z hodnoty 53,28 V (prípad nezatieneného FVP) na hodnotu 37,44 V, čo je pokles o približne 30 %.

d) Pri porovnaní *FF* faktora je možné pozorovať v Tab. I pokles *FF* faktora z hodnoty 0,69 (nezatienený stav) na hodnotu 0,49 (zatienený s použitím bypass diód), resp. na hodnotu 0,8 (zatienený bez použitia bypass diód). Z výsledkov výskumu je možné pozorovať, že bez použitia bypass diód je hodnota *FF* najvyššia. Ako bolo skôr spomenuté „Fill factor v kontexte fotovoltiky je dôležitým parametrom fotovoltického článku a charakterizuje jeho kvalitu. Je to miera, akou ideálny fotovoltický článok (s 100 percentnou účinnosťou a bez sériových a paralelných strát) sa blíži k skutočnému výkonu článku. Z výsledkov sa teda môže zdať, že s použitím bypass diód klesla hodnota *FF* faktora a teda aj „hodnota kvality“ fotovoltického panela, resp. a aj jeho účinnosť. Avšak je potrebné si uvedomiť, že s použitím bypass diód sa zvýšil výkon FV systému z 118,4 W (zatienený FVP bez použitia bypass diód) na hodnotu 351,3 W (zatienený s použitím bypass diód), čo je zvýšenie generovaného výkonu o 297 %. Z výskumu teda vyplýva, že *FF*

faktor nie v každej situácii hovorí o kvalite FVP, resp. jeho účinnosti, hlavne pri použití bypass diód. Na Obr. 3 sú zobrazené V-A charakteristiky s vyznačení oblasti *FF* faktora pre prípad zatienenia s a bez bypass diódy. Principiálne je *FF* pomer medzi plochami oblastí I_{SC}*U_{OC} (Na Obr. 3 zobrazené plnou čiarou) a I_{MPP}*U_{MPP} (na Obr. 3 zobrazené čiarkovanou čiarou). Ako bolo spomenuté, podiel medzi týmito plochami je hodnota *FF* faktora a ako je zrejmé z Obr. 3 preto dosahuje *FF* faktor pri použití bypass diód vyššie hodnoty.



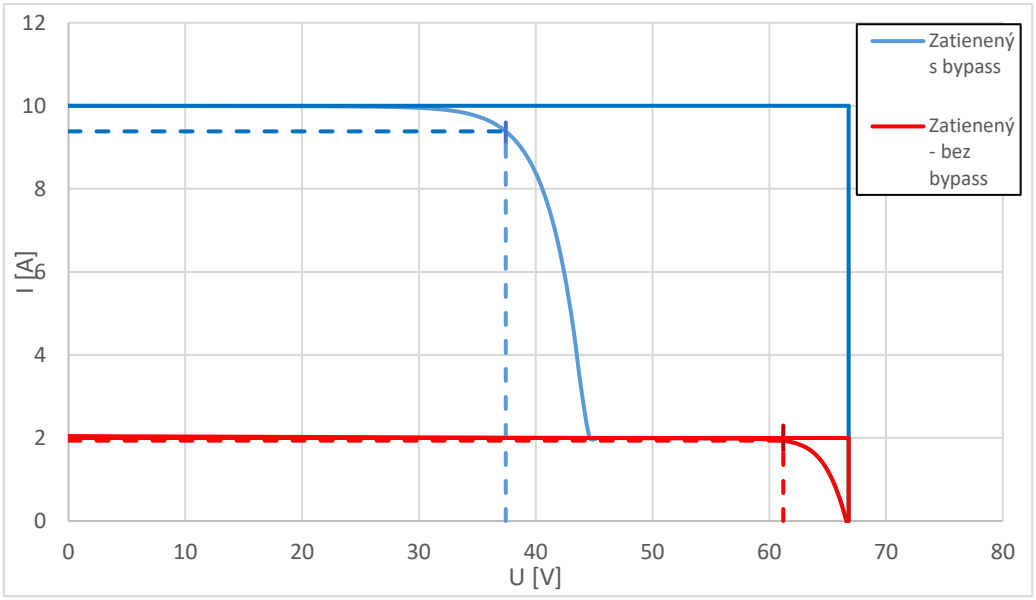
Obr. 2 Volt-ampérová charakteristika FV panela



Obr. 2 V-A charakteristiky FVP po pripojení bypass diódy

Tabuľka 1
Výsledky výskumu po pripojení bypass diódy

Stav	I_{SC} [A]	U_{OC} [V]	P_{MPP} [W]	I_{MPP} [A]	U_{MPP} [V]	FF [-]
Nezatienený	9.994	72	493.1176	9.25521	53.28	0.685297
Zatienený s bypass	9.99946	72	351.3437	9.38418	37.44	0.488004
Zatienený bez bypass	2.04514	72	118.4379	1.93526	61.2	0.804332



Obr. 4 FF faktor pri použití bypass diód

IV. Záver

Vplyv bypass diód je možné vidieť z výsledkov simulácii. Z výsledkov je možné vidieť, že V-A charakteristika zmenila tvar a podstatne klesol prúd z 10 A na 2 A, čo spôsobí pokles vo výrobe z FVP nakoľko výkon z fotovoltického panela je daný súčinom napätia a prúdu. Značne klesol prúd v bode P_{MPP}, napätie sa zmenilo minimálne, navyše hodnota napätia klesla, avšak minimálne. Výkon P_{MPP} značne klesol kvôli poklesu I_{MPP}. V Tab. I je možné vidieť výsledky, kde je pozorovaný pokles P_{MPP} z hodnoty 493 W na hodnotu 118 W, čo je pokles o 78 %. Toto je pokles výkonu P_{MPP} bez použitia bypass diódy. V prípade použitia bypass diódy je možné pozorovať pokles napätia U_{MPP} pričom prúd I_{MPP} sa takmer nezmenil. Pokles výkonu P_{MPP} je oproti nezatienenému FVP o 58%. Použitie bypass diód teda zabránilo poklesu P_{MPP} o 76 % - prípad zatieneného FVP.

Poďakovanie

Tento príspevok bol podporený Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-19-0576 „Sebestačnosť elektroenergetiky v podmienkach liberalizovaného trhu s elektrinou“ a Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-21-0312 „Dynamické prideľovanie kapacít elektrickej energie“.

References

[1] M. Pavlík, L. Beňa, D. Medved', Z. Čonka and M. Kolcun,“Analysis and Evaluation of Photovoltaic Cell Defects and Their Impact on Electricity Generation,“ Energies 2023, 16, 2576, https://doi.org/10.3390/en16062576

[2] A. S. Mahmood, M. Teke, R. K. Ibrahim, A. Hussein Ali, A. A. Abdulrazzaq and M. M. Kareem, "Tracking the MPP of a PV system using an advanced fuzzy logic control technique," 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2022, pp. 1-7, doi: 10.1109/ICAECT54875.2022.9808009.

[3] R. B. Bollipo, S. Mikkili and P. K. Bonthagorla, "Hybrid, optimal, intelligent and classical PV MPPT techniques: A review," in CSEE Journal of Power and Energy Systems, vol. 7, no. 1, pp. 9-33, Jan. 2021, doi: 10.17775/CSEEJPES.2019.02720.

[4] M. Pavlík, S. Bucko, J. Zbojovský a Ľ. Beňa, “Monitoring of Photovoltaic Cell Parameters,“ 2022 IEEE 5th International conference and worshop Obuda on Electrical and power Engineering, Cando-EPE, 2022, pp.27-30, DOI10.1109/CANDO-EPE57516.2022.10046365.

[5] M. Pavlík, M. Kolcun a S. Bucko,“Analysis and evaluation of photovoltaic cell defects and their impact on electricity generation,“ Proceedings of the 11th International Scientific Symposium on Electrical Power Engineering, ELEKTROENERGETIKA 2022, Pages 57 – 62, Stara Lesna,12-14 September 2022, ISBN: 978-805534104-0.

[6] I. I. R. Gomes, R. Melicio, V. M. F. Mendes, P. Gordo and T. C. D. Pardal, "Aerostat Powered by PV Cells: Hot-Spot Effect," 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA), Brasov, Romania, 2019, pp. 477-482, doi: 10.1109/ICRERA47325.2019.8996803.

[7] L. Chen, B. Liu, K. Sun and J. Zhao, "A Hot Spot Detection Method of Photovoltaic Module Based on Image Semantic Segmentation," 2023 IEEE 6th International Electrical and Energy Conference (CIEEC), Hefei, China, 2023, pp. 837-841, doi: 10.1109/CIEEC58067.2023.10167018.

Analýza vplyvu vlastností elektrického vozidla na dojazd

[1]*Peter GIROVSKÝ,* [2]*Jaroslava ŽILKOVÁ,* [3]*Ján Kaňuch,*

[1, 2, 3] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

[1]peter.girovsky@tuke.sk, [2]jaroslava.zilkova@tuke.sk, [3]jan.kanuch@tuke.sk

***Abstrakt* — V tomto príspevku sa zaoberáme analýzou vplyvu elektromobilu na kapacitu batérie a teda na samotný dojazd elektromobilu. Model elektromobilu bol zostrojený v programe MATLAB a rovnako aj v tomto programe prebehlo simulačné testovanie. V závere príspevku sú uvedené výsledky analýzy dojazdu elektromobilu, ktoré boli robené pre jazdný cyklus FTP-75.**

***Kľúčové slová* — elektromobil, dojazd, batéria, simulácia**

I. Úvod

Elektromobilita zaznamenáva v súčasnej dobe rastúci trend pričom významným spôsobom mení spôsob dopravy a energetickej závislosti. S narastajúcim trendom, ktorý čoraz viac prihliada na životné prostredie, sa elektrické vozidlá stali atraktívnou alternatívou k tradičným vozidlám so spaľovacími motormi. Rovnako s pokrokom v batériových technológiách [3], poklesom nákladov na ich výrobu a rozsiahlym rozvojom infraštruktúry nabíjacích staníc sa elektromobilita stáva reálnou možnosťou pre mnoho ľudí a spoločností. Významnou výhodou elektromobilov je, že ponúkajú nižšie emisie skleníkových plynov, tichý chod, nižšie prevádzkové náklady a v neposlednom rade príležitosti pre inteligentné riadenie v oblasti energetiky. Aj napriek nesporným výhodám, čelia elektromobily aj niektorým problémom, pričom najpálčivejším problémom je kapacita batérie elektromobilu a tým aj samotný dojazd elektrických vozidiel [1], [2], [4], [6].

II. Simulácia dojazdu elektrického vozidla

Pre testovanie dojazdu elektromobilu [20], [21], [22], [23], [24] bol vytvorený simulačný model elektromobilu v programe MATLAB [28]. Simulačný model elektromobilu principiálne pozostával z niekoľkých častí:

- Vstupný signál – zdroj jazdného cyklu FTP-75
- Elektrická časť vozidla – batéria, menič, motor
- Mechanická časť – karoséria, kolesá, prevodovka
- Riadenie

A. Jazdný cyklus

Ako vstupný signál pre simuláciu je potrebné zvoliť niektorý zo štandardných jazdných cyklov určených pre testovanie spotreby elektromobilov. Jazdné cykly [26] vytvárajú rôzne krajiny a organizácie na posúdenie výkonnosti vozidiel rôznymi spôsobmi, ako je napríklad spotreba paliva, emisie, využitie batérie pri elektrických vozidlách atď.

- FTP-75 - Mestský jazdný cyklus definovaný Americkou agentúrou na ochranu životného prostredia (EPA) na meranie emisií a spotreby paliva osobných automobilov.
- NEDC – Európsky jazdný cyklus, pozostáva zo 4 opakujúcich sa mestských jazdných cyklov ECE-15 (UDC) a jedného mimo mestského jazdného cyklu (EUDC). Meria spotrebu paliva alebo spotreby elektrickej energie spolu s meraním emisií.

Pre potreby simulácie bol použitý jazdný cyklus FTP-75

B. Elektrická časť vozidla

Elektrická časť simulovaného vozidla pozostávala principiálne z troch základných častí –

z batérie, výkonového meniča a elektrického motora.

- Ako elektrický motor bol pre pohon použitý jednoduchý DC motor s permanentnými magnetmi, kvôli jednoduchému spôsobu riadenia.
- Výkonovú časť predstavoval H-mostík, ktorý bol riadený pomocou PWM modulácie.
- Batéria elektromobilu bola simulovaná pomocou všeobecného dynamického modelu batérie, pomocou ktorého je možné simulovať najpoužívanejšie typy nabíjateľných batérií. Tento všeobecný model batérie môžeme voliť tak, aby reprezentoval konkrétny typ batérie a jej charakteristiky vybíjania, ktorá pozostáva z 3 častí:
    1. 1. časť – predstavuje exponenciálny pokles napätia ak je batéria nabitá. Veľkosť tohto poklesu závisí od typu batérie.
    2. 2. časť – predstavuje veľkosť nabitia batérie, ktoré je možné odobrať z batérie, kým napätie neklesne pod menovité napätie batérie
    3. 3. časť – predstavuje úplné vybitie batérie pri rapídnom poklese napätia

C. Mechanická časť vozidla

Mechanická časť vozidla pozostávala z troch základných častí, karosérie, kolies a prevodovky. Parametre mechanickej časti vozidla patria rovnako k dôležitým faktorom, ktoré ovplyvňujú dojazd elektromobilu, či už je to aerodynamika vozidla, valivý odpor kolies, samotná hmotnosť vozidla alebo prevodový pomer.

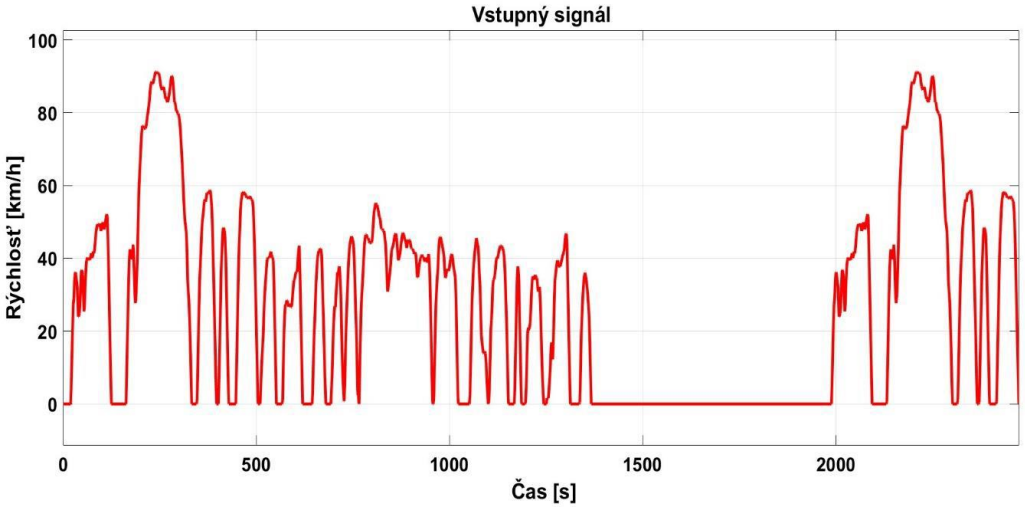
- Pre simuláciu bola použitá dvojnápravová karoséria vozidla s pohonom zadných kolies, pričom pri simulácií sme uvažovali zmenu hmotnosti karosérie a skúmali vplyv tejto zmeny na dojazd elektrického vozidla.
- Prevodovka predstavuje mechanické spojenie medzi motorom vozidla a kolesami vozidla. Zmenou prevodového pomeru vieme ovplyvniť správanie sa vozidla, ako akcelerácia, krútiaci moment, či maximálnu rýchlosť vozidla, pričom zmena týchto parametrov rovnako ovplyvní dojazd elektromobilu.

D. Riadenie

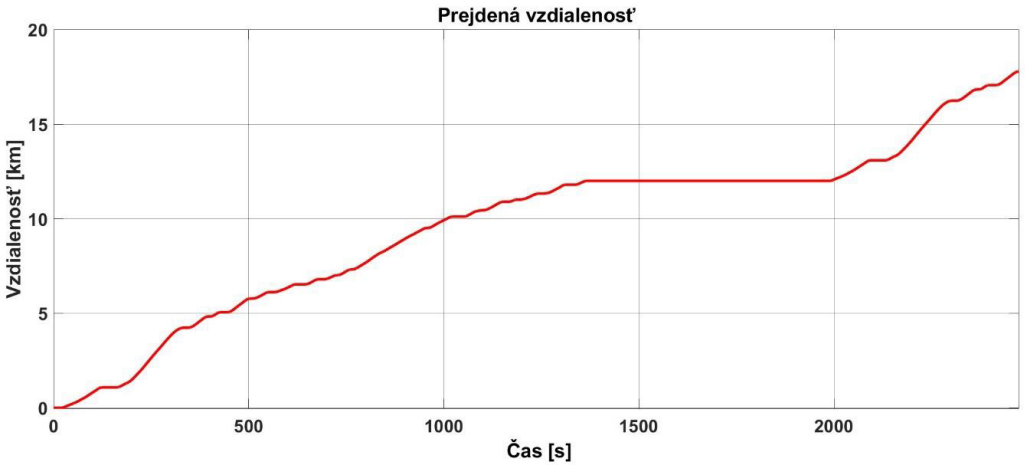
Riadenie principiálne predstavoval blok vodiča, ktorý riadi elektromobil, pričom toto riadenie zohľadňovalo žiadané hodnoty na základe zvoleného vstupného signálu – jazdného cyklu. Pomocou riadenia sme menili hodnotu referenčnej rýchlosti vozidla a tak skúmali vplyv zmeny rýchlosti na celkový dojazd.

III. SIMULAČNÉ VÝSLEDKY VPLYVU PARAMETROV VOZIDLA NA JEHO DOJAZD

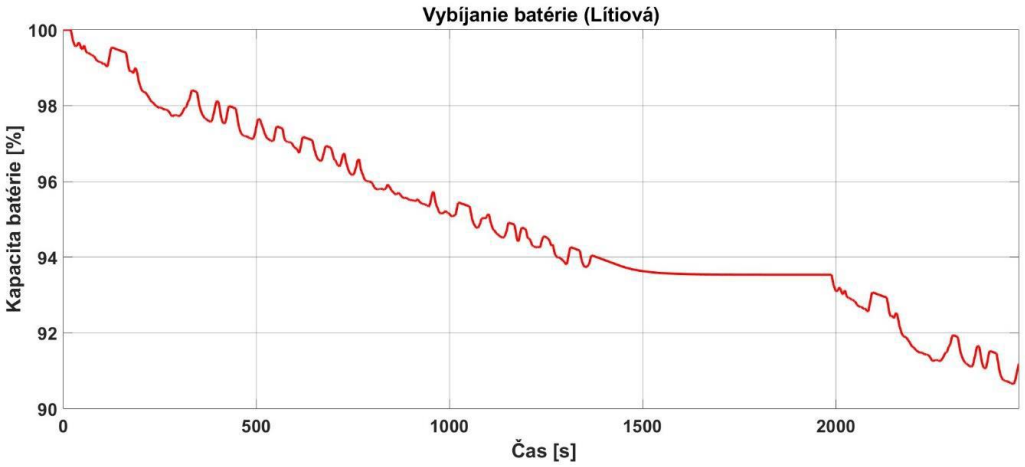
Simulácia bola robená na modeli elektrického vozidla v programe MATLAB, pričom pre vstup simulácie bol zvolený jazdný cyklus FTP-75 a simulácia trvala približne 2500s. Výsledky simulácie pri počiatočných parametroch sú zobrazené na obr. 1-5.



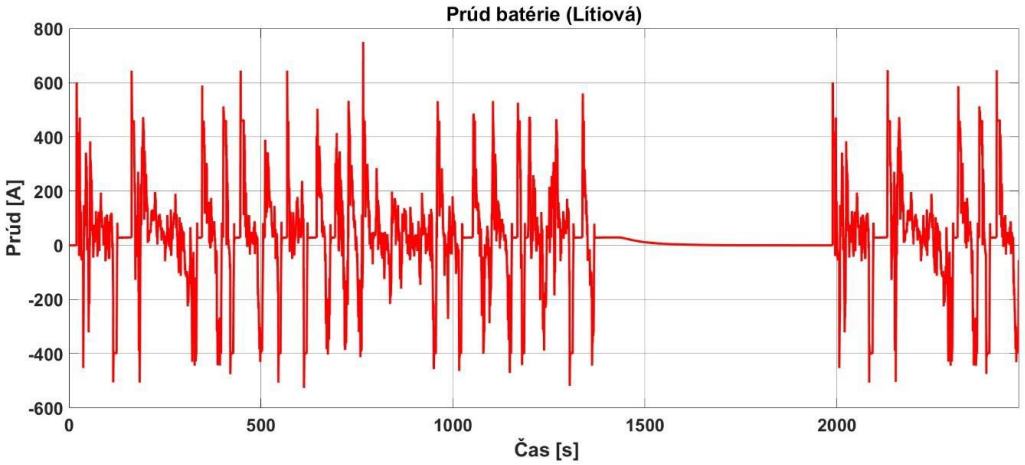
Obr. 1 Priebeh rýchlosti podľa jazdného cyklu FTP-75.



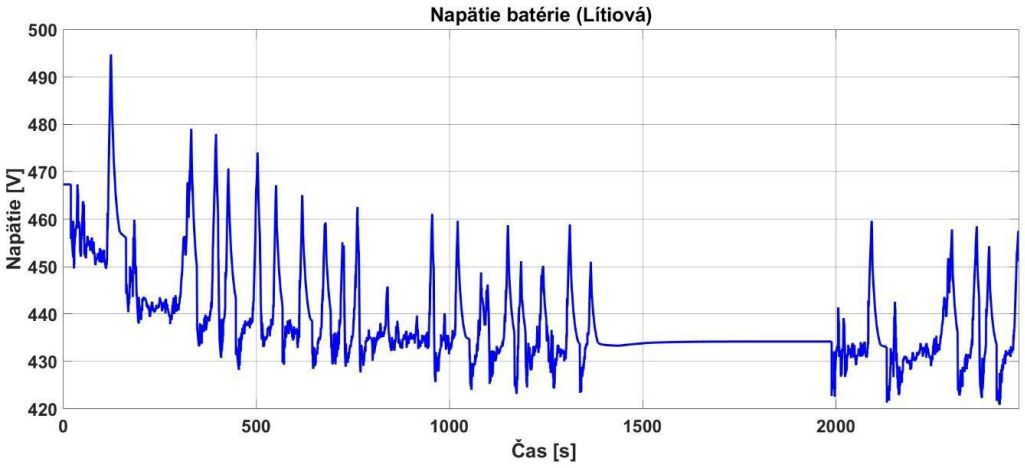
Obr. 2 Priebeh prejdenej vzdialenosti.



Obr. 3 Priebeh kapacity batérie.



Obr. 4 Priebeh prúdu batérie.

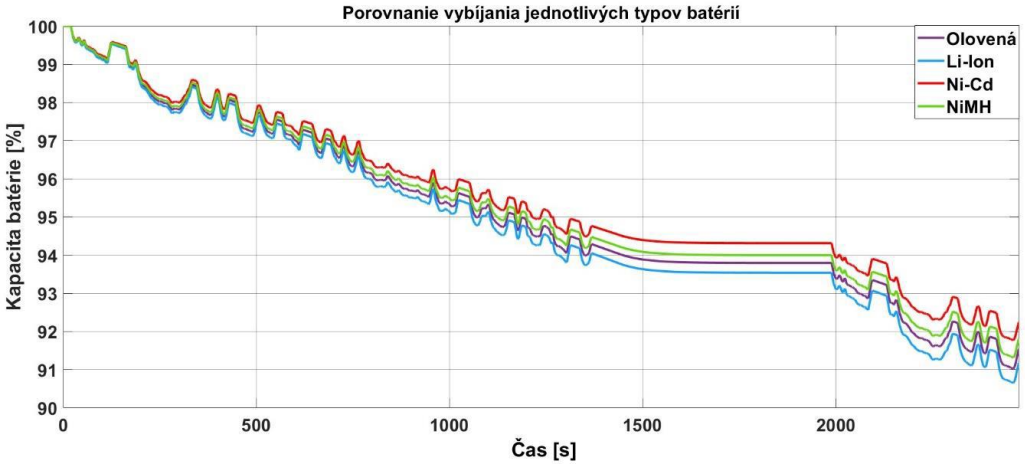


Obr. 5 Priebeh napätia batérie.

Ako vidno z uvedených priebehov, vozidlo prešlo za približne 2500s dráhu 17,79 km, pričom na konci jazdného cyklu bola batéria nabitá na 91,18%.

Simulácia zmeny typu batérie

Pri tejto simulácii bolo testované vybitie, resp. kapacita batérie na konci jazdného cyklu pre rôzne typy batérie (Lítium-iónová, Olovená, Nikel-kadmiová, Nikel-metal-hydridová). Výsledky sú zobrazené na obr. 6 a v tab.1.



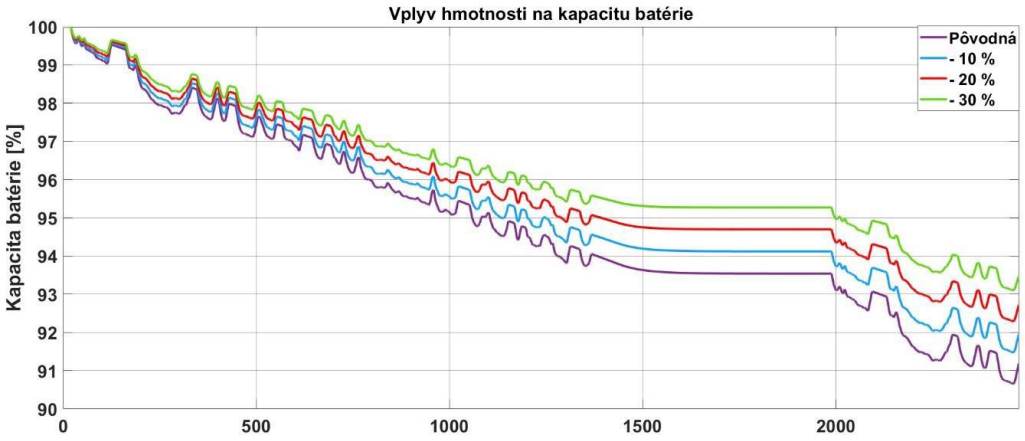
Obr. 6 Porovnanie kapacity batérie pre rôzne typy batérie.

Table 1
Porovnanie kapacity na konci jazdného cyklu a prepočítaná zmena dojazdu

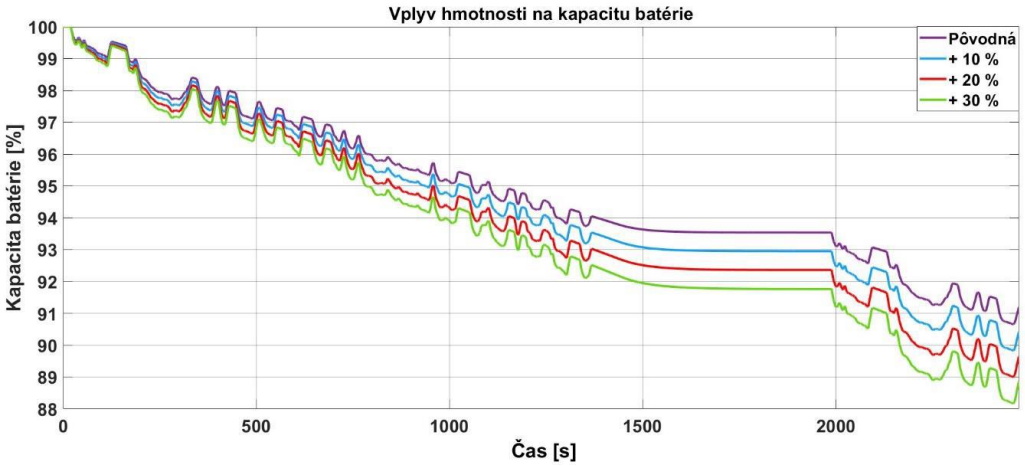
Batérie	Kapacita batérie [%]	Dojazd [km]
Lítium-iónová	91,18	201,701
Olovená	91,53	210,035
Nikel-kadmiová	92,24	229,253
Nikel-metal hybrid	91,8	216,951

Simulácia zmeny hmotnosti vozidla

Pri tejto simulácii bolo testované vybitie, resp. kapacita batérie na konci jazdného cyklu pre zmenu hmotnosti vozidla v rozsahu ±30% z celkovej hmotnosti. Výsledky simulácie sú zobrazené na obr. 7 a obr. 8 a v tab.2.



Obr. 7 Porovnanie vplyvy zníženia hmotnosti vozidla na kapacitu batérie a dojazd.



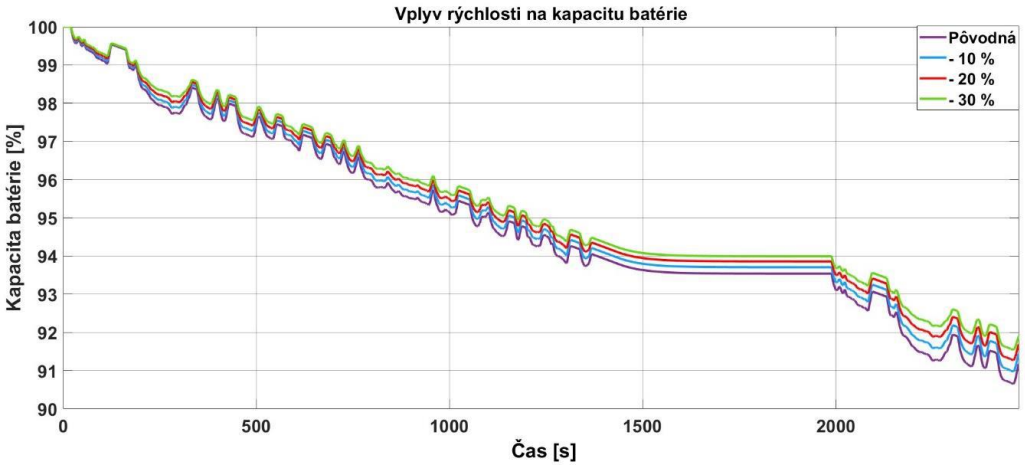
Obr. 8 Porovnanie vplyvy zvýšenia hmotnosti vozidla na kapacitu batérie a dojazd.

Table 2
Porovnanie kapacity na konci jazdného cyklu a prepočítaná zmena dojazdu v závislosti na zmene hmotnosti vozidla

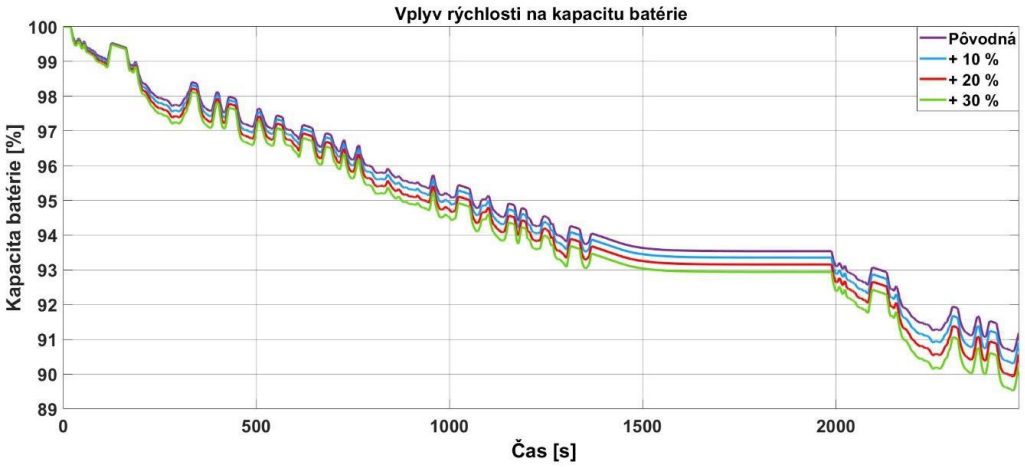
Hmotnosť [kg]	Kapacita batérie [%]	Dojazd [km]
1050 (-30%)	93,47	272,435
1575 (-20%)	92,71	244,033
1890 (-10%)	91,95	220,994
2100	91,18	201,701
2310 (+10%)	90,41	185,506
2520 (+20%)	89,63	171,553
2730 (+30%)	88,85	159,552

Simulácia zmeny rýchlosti vozidla

Pri tejto simulácii bolo testované vybitie, resp. kapacita batérie na konci jazdného cyklu pre zmenu rýchlosti vozidla v rozsahu ±30% z priemernej hodnoty rýchlosti daného jazdného cyklu. Výsledky simulácie sú zobrazené na obr. 9 a obr. 10 a v tab.3.



Obr. 9 Porovnanie vplyvy zníženia rýchlosti vozidla na kapacitu batérie a dojazd.



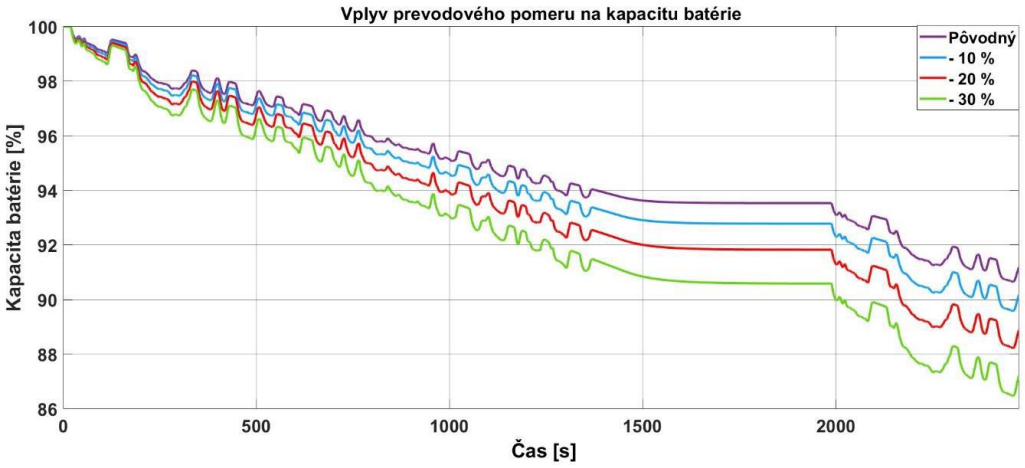
Obr. 10 Porovnanie vplyvy zvýšenia rýchlosti vozidla na kapacitu batérie a dojazd.

Table 3
Porovnanie kapacity na konci jazdného cyklu a prepočítaná zmena dojazdu v závislosti na zmene priemernej rýchlosti vozidla

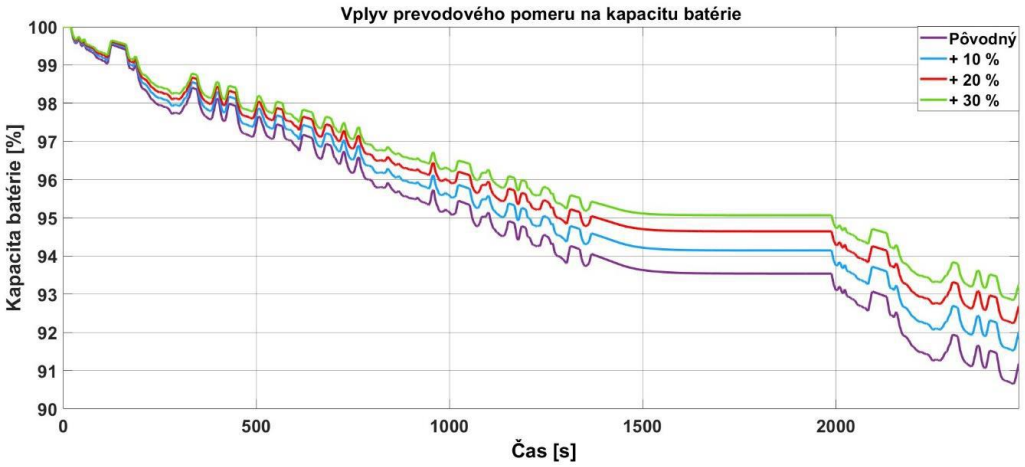
Rýchlosť [m/s]	Kapacita batérie [%]	Dojazd [km]
17,5 (-30%)	91,90	219,630
20,0 (-20%)	91,68	213,822
22,5 (-10%)	91,45	208,070
25,0	91,18	201,701
27,5 (+10%)	90,89	195,280
30,0 (+20%)	90,56	188,453
32,5 (+30%)	90,23	182,088

Simulácia zmeny prevodového pomeru vozidla

Pri tejto simulácii bolo testované vybitie, resp. kapacita batérie na konci jazdného cyklu pre zmenu prevodového pomeru vozidla v rozsahu ±30% hodnoty počiatočného prevodového pomeru. Výsledky simulácie sú zobrazené na obr. 11 a obr. 12 a v tab.4.



Obr. 9 Porovnanie vplyvy zníženia prevodového pomeru na kapacitu batérie a dojazd.



Obr. 10 Porovnanie vplyvy zvýšenia prevodového pomeru na kapacitu batérie a dojazd.

Table 4
Porovnanie kapacity na konci jazdného cyklu a prepočítaná zmena dojazdu v závislosti na zmene prevodového pomeru

Prevodový pomer	Kapacita batérie [%]	Dojazd [km]
6,328 (-30%)	87,20	138,984
7,232 (-20%)	88,88	159,982
8,136 (-10%)	90,16	180,793
9,040	91,18	201,701
9,944 (+10%)	92,01	222,653
10,848 (+20%)	92,68	243,033
11,752 (+30%)	93,25	263,556

IV. ZÁVER

V tomto príspevku bola predstavená analýza dojazdu elektromobilu v závislosti od zmeny parametrov elektrického vozidla. Z výsledkov simulácie možno vidieť, že vzhľadom na celkový dojazd elektromobilu je najlepším typom batérie nikel-kadmiová batéria, ktorá v súčasnosti nie je v elektromobiloch využívaná ako napr. lítium-iónová batéria. Ďalším významným faktorom, ktorý ovplyvňuje dojazd elektromobilu, sú hmotnosť a rýchlosť vozidla, ako vyplýva z výsledkov simulácie. Znížením hmotnosti a znížením rýchlosti vieme zvýšiť dojazd elektromobilu podobne ako aj zmenou prevodového pomeru. Pri vyšších prevodových pomeroch vozidla je na kolesách väčší krútiaci moment, teda lepšia akcelerácia, naopak pri nižších prevodových pomeroch dosahuje vozidlo vyššie maximálne rýchlosti a lepšiu spotrebu paliva.

POĎAKOVANIE

Tento príspevok vznikol s podporou Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied v rámci projektu VEGA 1/0363/23.

Táto práca bola podporená projektom KEGA 032TUKE-4/2024.

LITERATÚRA

[1] LARMINIE, James, LAWRY, John. 2012. Eletric Vehicle Technology Explained. 2. vyd. John Wiley a Sons Ltd, 344 s. ISBN 9781119942733.
[2] Abd El Halim, A. A. E.; Bayoumi, E. H. E.; El-Khattam, W.; Ibrahim, A. M. Electric vehicles: a review of their components and technologies. International Journal of Power Electronics and Drive Systems, 2022, Vol. 13, s. 2041–2061. DOI: http://doi.org/10.11591/ijpeds.v13.i4.pp2041-2061
[3] ZHAO, Weijie. A forum on batteries: from lithium-ion to the next generation. National Science Review. 2020, roč. 7, č. 7, s. 1263–1268. ISSN 2095-5138. DOI: https://doi.org/10.1093/nsr/nwaa068.
[4] MARK, Warner. The electric vehicle conversion handbook: How to Convert Cars, Trucks, Motorcycles, and Bicycles Includes EV Components, Kits, and Project Vehicles. Hudson Street, New York: by the Penguin Group, 2011, s. 1263–1268. eISBN : 978-1-101-51579-2
[5] SLOVAKIA, Kia. Aký dojazd ma elektromobil? 2024. Dostupné tiež z: https://www.kia.com/sk/predaj/elektromobilita-elektrifikacia/aky-dojazd-ma-elektomobil/
[6] SPELLMAN, Frank R. The Science of Electric Vehicle. Broken Sound Parkway NW: by CRC Press, 2023, 226 s. ISBN: 978-1-032-36628-9
[7] René, Hubinský; Peter Badík, Ivan Hyžák. ELEKTROMOBILITA. NEXTECH. 2020, s. 172. ISBN 978-80-973581-9-8. Dostupné tiež: https://www.nextech.sk/library/PDF%20ARCHIV/Elektromobilita/Elektromobilita%202022.pdf
[8] Roshandel, E.; Mahmoudi, A.; Kahourzade, S.; Yazdani, A.; Shafiullah, G. Losses in Efficiency Maps of Electric Vehicles: An Overview. Energies 2021, 14, 7805. https://doi.org/10.3390/en14227805
[9] Rickard Wretlind.; William Wärn. Adaptive Rolling Radius Estimation, 2023, Dostupné tiež z: https://liu.diva-portal.org/smash/get/diva2:1762057/FULLTEXT01.pdf
[10] MATHWORKS. Simscape Bus. Dostupné tiež z https://www.mathworks.com/help/simscape/ref/simscapebus.html
[11] MATHWORKS. Inertia. Dostupné tiež z: https://www.mathworks.com/help/simscape/ref/inertia.html
[12] MATHWORKS. Torsional Spring-Damper. Dostupné tiež z:
[13] https://www.mathworks.com/help/sdl/ref/torsionalspringdamper.html
[14] MATHWORKS. Motor and Drive System Level. Dostupné tiež z: https://www.mathworks.com/help/sps/ref/motordrivesystemlevel.html
[15] MATHWORKS. Simple Gear. Dostupné tiež z: https://www.mathworks.com/help/sdl/ref/simplegear.html
[16] MATHWORKS. Longitudinal Vehicle. Dostupné tiež z: https://www.mathworks.com/help/sdl/ref/longitudinalvehicle.html
[17] MATHWORKS. Drive Cycle Source. Dostupné tiež z: https://www.mathworks.com/help/autoblks/ref/drivecyclesource.html
[18] SLOVAKIA, Volkswagen. Čo je to WLTP. 2024. Dostupné tiež z: https://www.vwuzitkove.sk/wltp
[19] MARTIN MRUZEKA, Igor Gajdáča, Ľuboš Kučera. The possibilities of increasing the electric vehicle range. ScienceDirect. 2017, roč. 192, s. 621–625. Dostupné z : https://doi.org/10.1016/j.proeng.2017.06.107
[20] A SKUZA1, R S Jurecki. Analysis of factors affecting the energy consumption of an EV vehicle. International conference on Automotive Safety. 2022, roč. 13, s. 1–8. Dostupné z: DOI 10.1088/1757-899X/1247/1/012001
[21] DAVID SEBASTIAN PUMA BENAVIDES, Javier Izquierdo-Reyes. A Syste- matic Review of Technologies, Control Methods, and Optimization for Extended- Range Electric Vehicles. Applied Sciences. 2021, roč. 11, s. 1–25. Dostupné z https://doi.org/10.3390/app11157095
[22] NIKITA V. MARTYUSHEV Boris V. Malozyomov, Ilham H. Khalikov. Re- view of Methods for Improving the Energy Efficiencyof Electrified Ground Transport by OptimizingBattery Consumption. Energies. 2023, s. 1–39. Dostupné z : https://doi.org/10.3390/en16020729
[23] PARTH KADAV, Zachary D. Asher. Improving the Range of Electric Vehic- les. Electric Vehicles International Conference. 2019, s. 1–6. Dostupné z DOI: 10.1109/EV.2019.8892929.
[24] El Hadraoui, H.; Zegrari, M.; Chebak, A.; Laayati, O.; Guennouni, N. A Multi-Criteria Analysis and Trends of Electric Motors for Electric Vehicles. World Electr. Veh. J. 2022, 13, 65. https://doi.org/10.3390/wevj13040065
[25] SLOVAKIA, Kia. How to extend EV battery life? 2022. Dostupné tiež z: https://www.kia.com/dm/discover-kia/ask/how-to-extend-ev-battery-life.html
[26] Grüner, J., and Marker, S., "A Tool for Generating Individual Driving Cycles - IDCB," Commercial Vehicles 9(2):417-428, 2016, https://doi.org/10.4271/2016-01-9019
[27] Citation: Wellings, J.; Greenwood, D.; Coles, S.R. Understanding the Future Impacts of Electric Vehicles—An Analysis of Multiple Factors That Influence the Market. Vehicles 2021, 3, 851–871. https:// doi.org/10.3390/vehicles3040051
[28] MathWorks. Battery Electric Vehicle Model in Simscape. Dostupné tiež z: https://www.mathworks.com/matlabcentral/fileexchange/82250-battery-electric-vehicle-model-in-simscape?s_tid=FX_rc3_behav

Návrh elektroinštalácie pre medicínske centrum

[1]*Diana HOZOVÁ,* [2]*Ján ZBOJOVSKÝ*

[1] Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic
[2] Department of Electric Power Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]diana.hozova@student.tuke.sk, [2]jan.zbojovsky@tuke.sk

***Abstract* — Cieľom práce je vypracovať realizačný projekt elektroinštalácie pre medicínske centrum v súlade s príslušnými normami. Súčasťou práce je rovnako aj sumarizácia základných poznatkov, analýza legislatívneho rámca pri tvorbe projektovej dokumentácie, zdôvodnenie dôležitosti jednotlivých jej častí, vymedzenie principiálnych požiadaviek pre realizáciu. Článok pojednáva aj o realizácii projektu elektrickej prípojky a ochrany pred bleskom pre predmetný objekt. V závere sa nachádzajú všeobecné odporúčania pre prax a rovnako bude poukázané na prípadné vylepšenia, ktoré by mohli byť zvážené pri samotnej realizácii.**

***Keywords* — elektroinštalácia, elektrická prípojka, bleskozvod, systém ochrany pred bleskom**

I. Úvod

V súčasnosti sa zvyšuje počet obyvateľov sveta, a preto sa zvyšuje aj počet spotrebiteľov elektrickej energie. Zároveň medziročne narastá aj ich samotná spotreba elektrickej energie, pretože život dnešného človeka závisí od rôznych moderných technológií. Každý novopostavený ale aj rekonštruovaný objekt je pripojený do elektrizačnej sústavy. Z toho dôvodu je potrebné sa zaoberať návrhmi elektroinštalácií, ktoré budú v čo najväčšej možnej miere spĺňať požiadavky príslušných noriem a spoľahlivo tak slúžiť svojim užívateľom.

Účelom tejto práce je vypracovať projektovú dokumentáciu elektroinštalácie, elektrickej prípojky a bleskozvodu pre novopostavené medicínske centrum.

II. STN 33 2000-7-710 : Požiadavky na osobitné inštalácie alebo priestory – Zdravotnícke priestory

Požiadavky pre realizáciu elektroinštalácií v zdravotníckych priestoroch rieši STN 33 2000-7-710, ktorá bola vydaná v roku 2013, ako náhrada za STN 33 2140 z roku 1986. Platí však, že zdravotnícke zariadenia, v ktorých boli elektroinštalácie prevedené v minulosti podľa momentálne už neaktuálnej normy, sa nachádzajú v tzv. prechodnom období. To znamená, že do nasledujúcej rekonštrukcie môže prevedenie elektrickej inštalácie ostať v prevádzke, ak v čase jej uvedenia do prevádzky vyhovovala požiadavkám. V opačnom prípade je potrebné, aby boli vykonané bezprostredné zmeny v súlade so súčasnou normou.

Cieľom vypracovania tohto dokumentu bolo zaistiť čo najvyššiu spoľahlivosť a bezpečnosť elektrického napájania v budovách slúžiacich na poskytovanie zdravotnej starostlivosti na rôznych úrovniach, či už ide o bežnú ambulantnú starostlivosť alebo o intenzívnu starostlivosť, počas ktorej nemôže dôjsť k výpadku elektrickej energie. Každá z týchto úrovní má preddefinované jednoznačné požiadavky, ktoré musia byť čo v najvyššej možnej miere splnené, za účelom dosiahnutia adekvátnej bezpečnosti pacientov, ale aj zdravotníckeho personálu. [1-7]

A. Rozsah platnosti uvedenej normy

Požiadavky STN 33 2000-7-710 sa týkajú elektroinštalácií prevedených v akomkoľvek medicínskom prostredí. Vo všeobecnosti môže ísť o : nemocnicu, medicínske centrum, súkromnú kliniku, ambulantnú ošetrovňu praktického, resp. špecializovaného lekára, zariadenie určené pre zdravotnícky výskum, sanatórium, ošetrovňa športového lekára, kontajnerovú ošetrovňu a pod. V niektorých prípadoch je možné túto normu použiť aj ako záväznú pri projektovaní a zhotovení elektrických inštalácií veterinárnej nemocnice. [8] [9]

B. Rozdelenie zdravotníckych priestorov

Klasifikácia jednotlivých zdravotníckych priestorov je kľúčová pre návrh novej elektrickej inštalácie, ale rovnako aj pre rekonštrukciu starej elektroinštalácie v takomto objekte. Vzhľadom na to, že na základe tohto rozdelenia sa určujú jednotlivé požiadavky, ktoré sú v niektorých prípadoch v rámci porovnania skupín od seba diametrálne odlišné, je potrebné ho vykonať dôsledne. Správny postup návrhu zahŕňa aj dohodu projektanta s investorom, resp. osobou, ktorá je zodpovedná za bezpečnosť v medicínskom objekte a ktorá s určitosťou vie povedať, aké prístroje sa budú v priestoroch využívať. Potrebné je však určiť, do akej miery budú aplikačné časti elektrického zdravotníckeho zariadenia prichádzať do kontaktu s telom pacienta a rovnako aj to, do akej miery môže byť život pacienta ohrozený, v prípade výpadku elektrického napájania tohto prístroja. [8] [9]

Na základe týchto kritérií sa zdravotnícke priestory delia na nasledujúce 3 skupiny:

a) skupina 0 – ide o zdravotnícky priestor, kde aplikačná časť elektrického zdravotníckeho zariadenia nedochádza do styku s telom pacienta, resp. funkčnosť tohto zariadenia neovplyvní bezpečnosť pacienta a nijakým spôsobom ho neohrozí na živote,

b) skupina 1 – ide o zdravotnícky priestor, v ktorom prerušenie dodávky elektrickej energie nespôsobí ohrozenie života pacienta, avšak aplikačná časť elektrického zdravotníckeho zariadenia prichádza do kontaktu s telom pacienta buď externe (vonkajší dotyk s telom) alebo invazívne (vnútorný dotyk s telom) v prípadoch, ktoré nie sú určené pre zaradenie do skupiny 2,

c) skupina 2 – ide o zdravotnícky priestor, v ktorom prerušenie dodávky elektrickej energie môže spôsobiť ohrozenie života pacienta a zároveň aplikačná časť elektrického zdravotníckeho zariadenia prichádza do kontaktu s telom pacienta napr. pri procedúrach súvisiacich s primárnymi životnými funkciami človeka, chirurgických zákrokoch alebo pri intrakardiálnych procedúrach, počas ktorých môže elektrický vodič prísť do priameho styku so srdcom pacienta. [9]

Samotná STN 33 2000-7-710 uvádza vo svojej prílohe zoznam príkladov zdravotníckych priestorov a ich zadelenia do príslušných skupín ako návod.

C. Záložné zdroje elektrickej energie a bezpečnostné technické prostriedky budov

Vo všeobecnosti je pri tomto type objektov potrebné voliť rozvodnú sieť, ktorá umožní čo najspoľahlivejší prechod z normálnej siete na záložný napájací zdroj. V prípade, ak nastane prerušenie dodávky elektrickej energie pre zdravotnícke priestory, ktoré sú zaradené do skupín 1 a 2, je potrebné zabezpečiť automatický prechod na záložný napájací zdroj. Pripojenie na záložnú sieť sa musí vykonať vtedy, ak na čas ≥0,5 s napätie na jednom alebo na viacerých vstupných pracovných vodičoch klesne pod hodnotu 90 % z hodnoty menovitého napätia. [9]V Tab. 1 sa nachádza rozdelenie bezpečnostných technických prostriedkov do tried podľa dĺžky prerušenia hlavného napájania. [8] [9]

Tabuľka 1
Klasifikácia bezpečnostných technických prostriedkov zdravotníckej budovy [8]

Trieda :	**Automaticky pripájané záložné napájanie dostupné bez prerušenia dodávky počas:**
0 – bez prerušenia	bez prerušenia dodávky el. energie
0,15 – veľmi krátke prerušenie	≤ 0,15 s
0,5 – krátke prerušenie	≤ 0,5 s
5 – štandardné prerušenie	≤ 5 s
15 – stredné prerušenie	≤ 15 s
>15 – dlhé prerušenie	> 15 s
- Nie je však povinné pre všetky elektrické zdravotnícke zariadenia realizovať aj záložné zdroje elektrickej energie, keďže zdravotnícke prostredia zatriedené do skupiny 0 to nevyžadujú. - V prípade, ak sú jednotlivé miestnosti objektu zaradené do rôznych tried podľa tejto tabuľky, za východiskovú triedu pre daný zdravotnícky objekt sa berie tá, ktorá spĺňa prísnejšie kritériá.	

III. Realizácia Projektu

Projektová dokumentácia elektroinštalácie a ochrany pred bleskom bude zhotovená pre novovybudované medicínske centrum v katastrálnom území mesta Moldava nad Bodvou na

parcele s číslom 659/19. V interiéri sa bude nachádzať 5 ambulancií, v ktorých budú svoju činnosť vykonávať všeobecní doktori, pediater, ortopéd a endokrinológ. Rozmery predmetného objektu sú:

- šírka – 32,67 m,
- dĺžka – 13,23 m,
- výška – 8,9 m.

Zastavaná plocha je 330,92 m2.

A. Elektrická prípojka pre predmetné medicínske centrum

Pre návrh projektovej dokumentácie elektrickej prípojky zdravotníckeho centra je kľúčové zaradiť daný objekt do skupiny na ktoré sa delia zdravotnícke priestory podľa STN 33 2000-7-710: 2013. Toto medicínske centrum bolo zaradené do skupiny 0 – to znamená, že sa v ňom nenachádzajú žiadne prístroje s aplikačnými časťami, ktoré by dochádzali do styku s telom pacienta a výpadkom napájania týchto prístrojov nemôže dôjsť k ohrozeniu ľudského života.

Z tohto dôvodu je pre predmetný objekt definovaný 3. stupeň dôležitosti dodávky elektrickej energie v zmysle STN 34 1610. Elektrická NN prípojka bude realizovaná napojením z existujúcej rozpojovacej a istiacej skrine, ktorá je pripojená na už existujúce podzemné nn vedenie. Z tohto miesta bude vo výkope pokračovať kábel CYKY-J 4x25 mm2 v dĺžke 10 m na hranicu pozemku s predmetným objektom. Na tomto mieste sa bude nachádzať elektromerový rozvádzač, ktorý bude slúžiť na meranie spotreby elektrickej energie, ale aj na istenie hlavného rozvádzača a prívodného kábla CYKY-J 5x16 mm2 v dĺžke 20 m. V elektromerovom rozvádzači dôjde k rozdeleniu sústavy TN-C na TN-S. Tento bod rozdelenia je potrebné uzemniť.

V rovnakom výkope sa bude nachádzať aj optický kábel uložený v samostatnej chráničke a v čo najväčšej možnej vzdialenosti od nn kábla vrámci výkopu. Súbežne s káblom sa do výkopu uloží aj uzemňovacia pásovina FeZn 30x4, ktorá prepojí uzemnenie rozpojovacej a istiacej skrine, elektromerového rozvádzača a hlavného rozvádzača. Na svojom konci bude táto pásovina pripojená aj k obvodovému uzemňovaču vonkajšej ochrany pred bleskom predmetného objektu. Svojou dĺžkou tak zaručí dostatočne nízky zemný odpor. [10][11] [12]

B. Hlavný rozvádzač

Hlavný rozvádzač sa bude nachádzať pri vstupe do budovy na ľavej strane v priestoroch recepcie. Pôjde o 42 modulový zapustený rozvádzač, ktorý bude osadený vo výške 1,5 m. Z elektromerového rozvádzača bude napojený káblom CYKY-J 5x16 mm2. Nachádzať sa v ňom bude hlavný istič LTE B32A/3, pomocou ktorého sa vypne elektrická energia v celej budove. Rovnako sa v ňom bude nachádzať prepäťová ochrana, ističe a prúdové chrániče s integrovanou nadprúdovou ochranou, ktoré budú chrániť jednotlivé obvody. Schéma navrhovaného rozvádzača sa nachádza na Obr. 10. Na Obr. 11 sa nachádza vizualizácia uloženia jednotlivých prvkov v ňom.

C. Svetelné obvody

Svetelné obvody sú vo výkresovej dokumentácii označované ako EL1-EL4. Ku všetkým svietidlám sú priradené dané typy spínačov osvetlenia, ktoré budú inštalované vo výške 1,3 m. V objekte budú tieto obvody prevedené káblami CYKY 3x1,5 mm2 . Chrániť ich budú kombinované prúdové chrániče s ističom 2-pólové, In=10A, char.B, InΔ=30mA.

Vo výkresovej dokumentácii sa nachádza pôdorys stavby so zakresleným rozmiestnením svietidiel, ich presný typ je však potrebné určiť na základe svetelnotechnického prepočtu, ktorý nebol súčasťou zadania tejto bakalárskej práce. STN EN 12464-1 (Svetlo a osvetlenie pracovísk, Časť Vnútorné pracoviská) vymedzuje podmienky pre osvetlenie vnútorných pracovných priestorov na základe ktorých sa dosiahne adekvátna zraková pohoda pre daný druh činnosti.

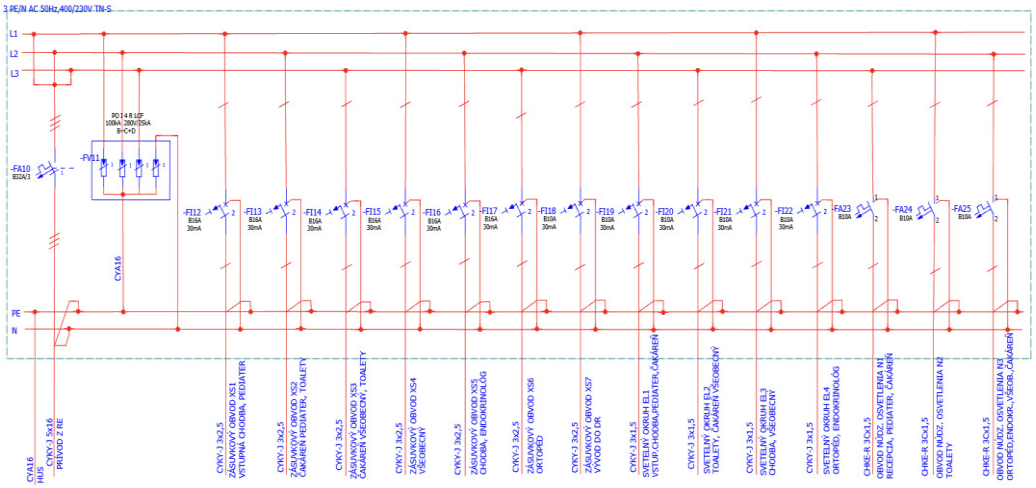
D. Zásuvkové obvody

Zásuvkové obvody sú vo výkresovej dokumentácii označované ako XS1-XS7. Prevedené budú káblami CYKY-J 3x2,5 mm2 a chrániť ich budú kombinované prúdové chrániče s ističom 2-pólové, In=16A, char. B, InΔ=30mA.

Vo výkresovej dokumentácii sa nachádza pôdorys stavby so zakresleným rozmiestnením jednotlivých typov zásuviek. Ich rozmiestnenie je navrhované na základe predpokladaného situovania elektrických zariadení v miestnostiach. V ambulanciách budú inštalované do výšky 0,6 m nad podlahou, v ostatných miestnostiach do výšky 1,2 m.

E. Obvody núdzového osvetlenia

Obvody núdzového osvetlenia boli vo výkresovej dokumentácii označované ako N1-N3. Realizované by mali byť káblami so zvýšenou odolnosťou voči požiaru, v tomto prípade CHKE-R 3Cx1,5. V hlavnom rozvádzači budú istené jednopólovými ističmi s char. B a menovitým prúdom 10 A. Výkresová dokumentácia obsahuje aj pôdorys s rozmiestnením jednotlivých núdzových svietidiel, použité boli dva druhy. Prvým druhom sú svietidlá s akumulátorom (1 hod.) pre núdzovú prevádzku – ide o tzv. antipanikové osvetlenie núdzových ciest. Úlohou týchto svietidiel je len osvetliť únikovú cestu. Druhým druhom sú svietidlá s akumulátorom (1 hod.) pre núdzovú prevádzku s piktogramom označujúcim smer úniku, ktoré majú za úlohu naviesť človeka k najbližšiemu únikovému východu. Rozmiestnenie týchto svietidiel je navrhnuté tak, aby z každej pozície v interiéri objektu bol výhľad na minimálne jedno takéto svietidlo s piktogramom. Dôležité je však poznamenať, že rovnako ako v prípade návrhu pracovného osvetlenia je aj v tomto prípade potrebné zrealizovať svetelnotechnický prepočet v zmysle platných noriem, predovšetkým STN EN 1838: 2001 -Požiadavky na osvetlenie - núdzové osvetlenie.



Obr. 1 Schéma hlavného rozvádzača

F. Dátové rozvody

Pre medicínske centrum bude realizovaná aj podzemná optická prípojka. Optický kábel bude uložený v samostatnej chráničke v maximálnej možnej vzdialenosti od nn kábla v rámci výkopu. Optický kábel bude vyvedený až do nástenného dátového rozvádzača, ktorý sa bude nachádzať v mieste hlavného vchodu do budovy a recepcie vo výške 1,5 m. Konvertor s Wi-Fi routerom v dátovom rozvádzači zabezpečí premenu optického signálu na elektrický. Ten bude pomocou dátových káblov Cat.6 F/UTP uložených v trubkách rozvedený do ambulancií. Dátové rozvody budú ukončené dátovými dvojzásuvkami Cat.6 STP.

G.Doplnkové pospájanie

Vodičom CYA 10 mm2 zž v trubke FXP-TURBO bude realizované pospájanie v zmysle STN 33 2000-4-41, čl. 411.3.1.2. K ochrannému pospájaniu budú pripojené vodivé časti konštrukcie budovy, vodovodného potrubia, plynového potrubia, odpadovej kanalizácie a ústredného kúrenia. Rovnakým vodičom bude realizovaný aj vývod do dátového rozvádzača a vodičom CYA 16 mm2 zž bude prevedený vývod do hlavného rozvádzača vo vzdialenosti < 0,5 m. Hlavná uzemňovacia svorkovnica bude použitím vodiča FeZn10 cez skúšobnú svorku pripojená k obvodovému uzemňovaču medicínskeho centra.

IV. NÁVRH OCHRANY PRED BLESKOM PRE DANÝ OBJEKT

Primárnym krokom pre správny návrh ochrany pred bleskom je manažérstvo rizika, ktoré je výstupom programu OEZ Prozik. Ide o softvér, pomocou ktorého sa dá pomerne jednoduchým spôsobom vypočítať riziko škôd a strát na stavbe v dôsledku úderu blesku. Na základe zadaných informácií, tento program vykoná výpočet v zmysle STN EN 62305-2 a určí triedu LPS potrebnú pre návrh ochrany pred bleskom.

Na základe analýzy rizika bola budova medicínskeho centra zaradená do triedy ochrany LPS II. Predmetné medicínske centrum má rovnú strechu, a preto najlepšou voľbou bude preň mrežová zachytávacia sústava. [13] Pre objekty zaradené do triedy LPS II s takýmto typom zachytávacej sústavy sú odporúčania nasledovné:

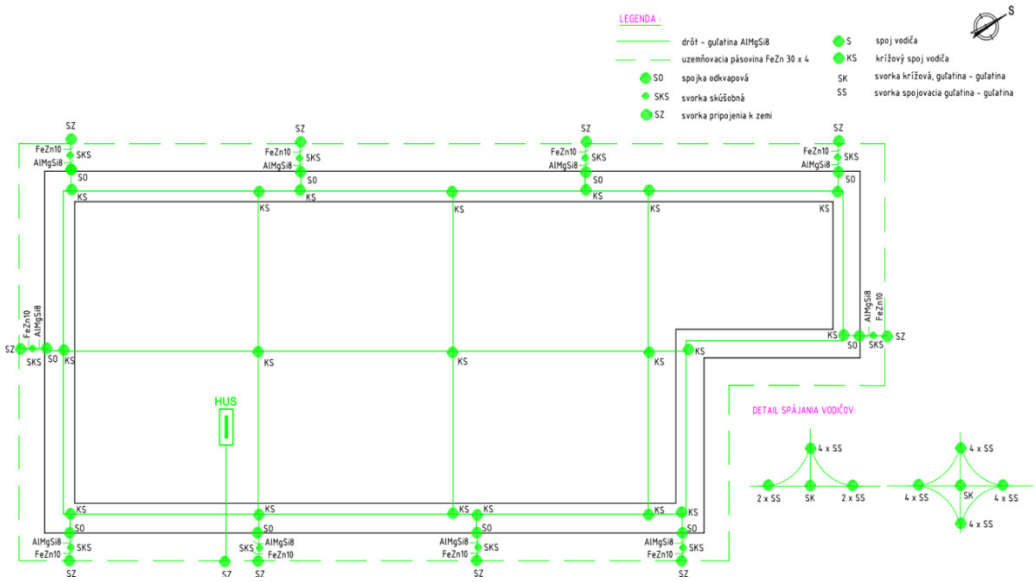
1. doporučená vzdialenosť medzi zvodmi 10 m,
2. maximálna veľkosť ôk 10m x 10m.

V tomto konkrétnom prípade bude zachytávacia sústava prevedená vodičmi AlMgSi 8 uloženými na podperách každý 1 m. Veľkosť ôk bude 7,8 m x 5,89 m. Presné rozloženie zbernej sústavy je známe z výkresovej dokumentácie, ktorej ukážka sa nachádza na Obr. 2.

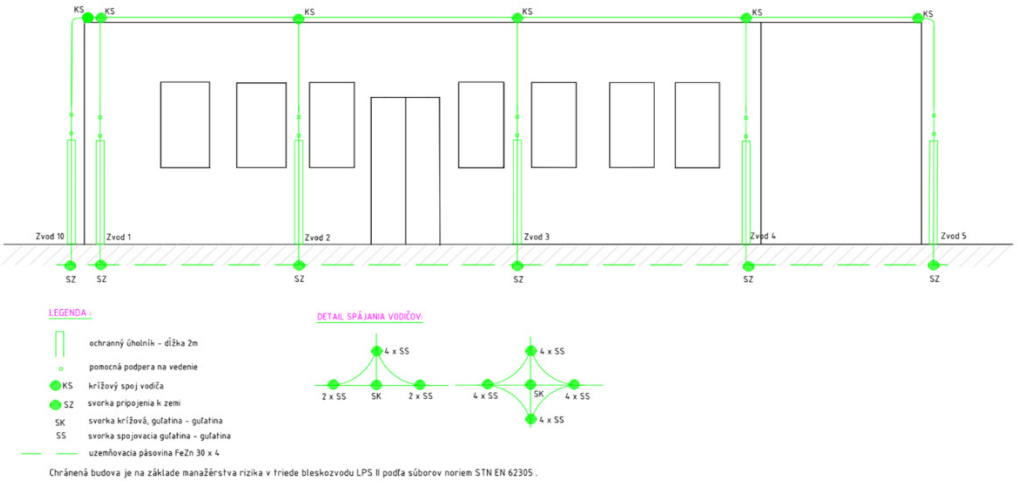
Počet zvodov pre túto stavbu je 10. Medzi nimi je približne rovnaká vzdialenosť. Guľatinou FeZn10 sú zvody a rovnako aj hlavná uzemňovacia svorkovnica pripojené na obvodový uzemňovač s celkovou dĺžkou 98,68 m. Pásovina FeZn 30x4 bude uložená vo vzdialenosti 1 m od budovy v hĺbke 0,8 m. Pre kontrolu je potrebný výpočet pre odpor uzemnenia:

$$R = 2 * \frac{\rho}{l_p} = 2 * \frac{100}{90} = 2{,}03\ \Omega \quad (1)$$

V miestach s prechodom rôznymi prostrediami a v zemi je potrebné chrániť vedenie voči korózii asfaltovým náterom. Na Obr. 3 sa nachádza ukážka z výkresovej dokumentácie, konkrétne výkres pre návrh bleskozvodu - pohľad na budovu spredu.



Obr. 2 Návrh bleskozvodu – pohľad na budovu zhora



Obr. 3 Návrh bleskozvodu – pohľad na budovu spredu

Pre elimináciu vzniku nechcených nebezpečných výbojov medzi zvodmi a okolitými uzemnenými vodivými časťami je potrebné vypočítať tzv. bezpečnú vzdialenosť s. V tomto prípade bol výpočet realizovaný pre dva prípady úderu blesku:

B1 – úder blesku do stredu mrežovej sústavy (do stredu strechy objektu).

B2 – úder blesku do okraju mrežovej sústavy.

Výpočet dostatočnej vzdialenosti je realizovaný použitím vzorca:

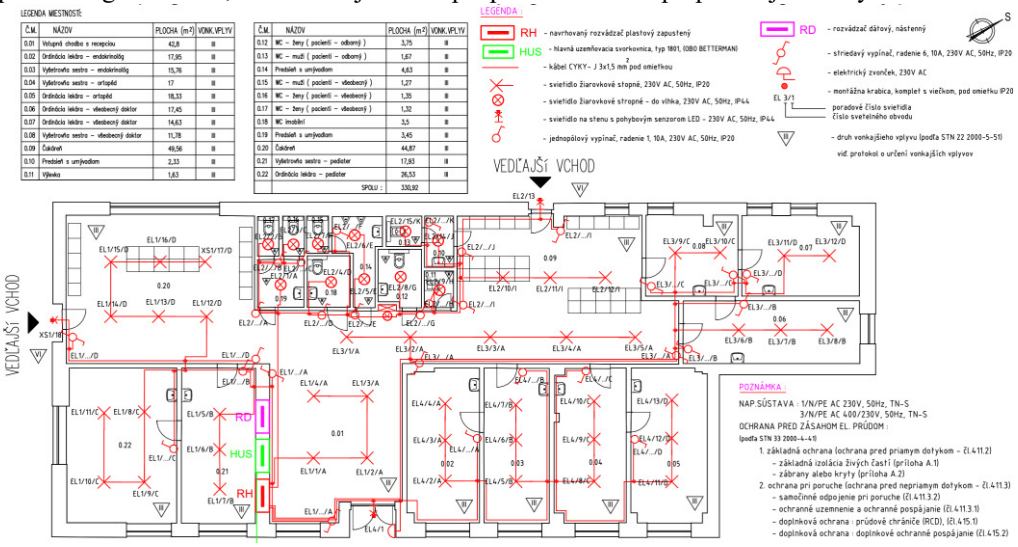
$$s = ki \cdot \frac{\text{kc}}{\text{km}} \cdot l \qquad (2)$$

kde:
- koeficient kc pre > 4 zvody pri mrežovej sústave: 0,5,
- koeficient ki pre LPS 2: 0,06,
- koeficient km izolácia vonkajšieho LPS – pevná látka „0,5“, pre vzduch je „1“, volím 0,5, čo je horší prípad. V praxi ide zväčša o kombináciu oboch základných materiálov,
- l - dĺžka pozdĺž zachytávacej sústavy.

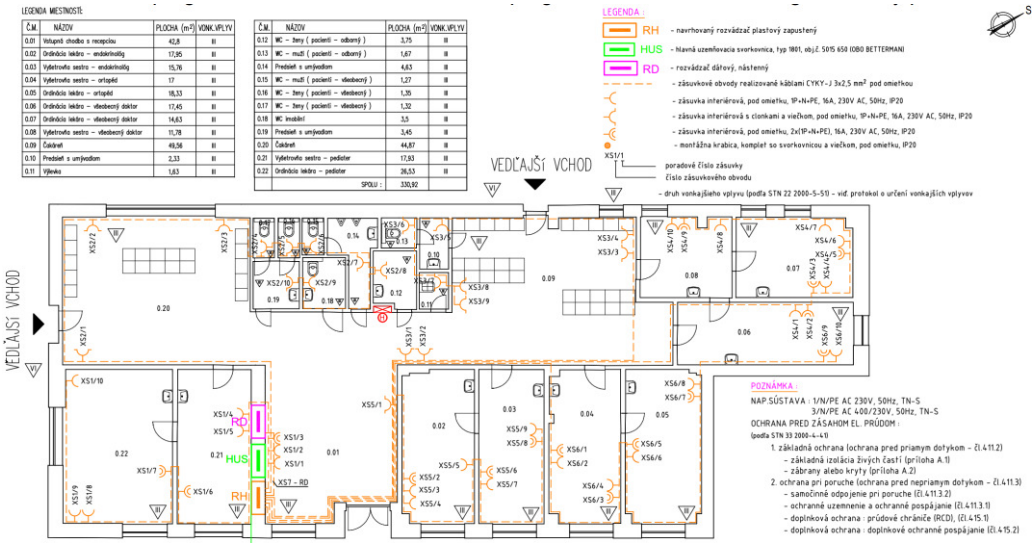
Tabuľka 1
Vypočítané dostatočné vzdialenosti „s“

Označenie	k_c	k_i	k_m	Dĺžka zvodu „l“ [m]	Dostatočná vzdialenosť „s“ [m]
B1	0,5	0,06	0,5	14,4	0,86
B2	0,5	0,06	0,5	8,9	0,53

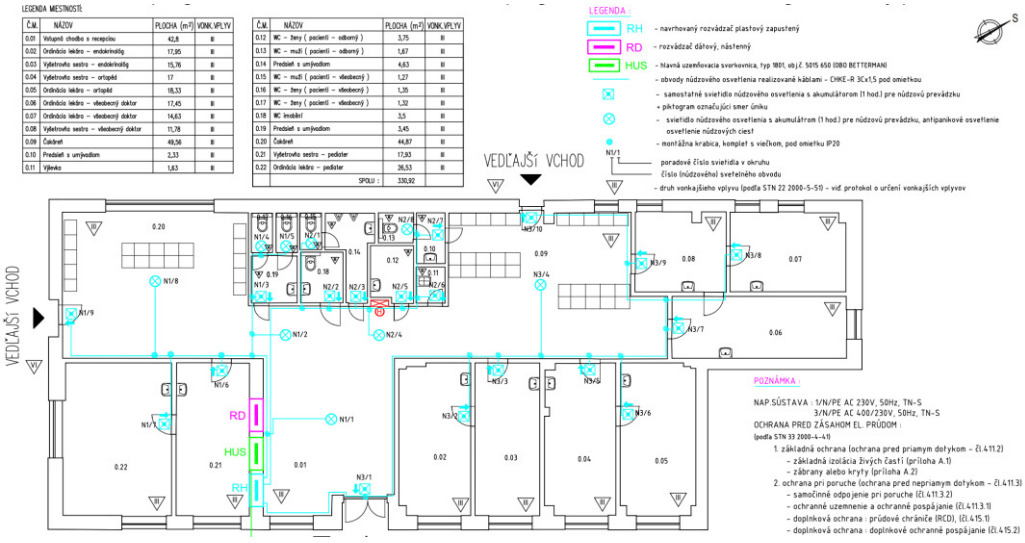
V hlavnom rozvádzači bude za hlavným ističom osadený kombinovaný (B+C+D) zvodič prepätia značky KIWA, konkrétne typ PO I 4 R LCF 100kA 280V/25kA . Tento typ prepäťovej ochrany je vhodný pre triedu LPL II a sústavu TN-S. Je to štvorpólová prepäťová ochrana s celkovým spoločným impulzným prúdom Iimp = 100 kA. Zabezpečí ochranu pre prístroje v hlavnom rozvádzači v rozsahu T1, T2, (T3) do 5 m. Zariadenia vo všetkých ambulanciách, ktoré budú pripojené do zásuviek budú chránené zásuvkovými prepäťovými ochranami ZPO D ATA1 (resp. 2) iS-3kV inštalovanými v inštalačných krabiciach. Tieto zásuvkové ochrany sú vybavené optickou signalizáciou, čo umožňuje mať lepší prehľad o stave prepäťovej ochrany.



Obr. 4 Svetelné obvody



Obr. 4 Zásuvkové obvody



Obr. 4 Núdzové osvetlenie

V. Záver a odporúčania pre prax

Projektová dokumentácia je základom pre montážnych technikov a aj revízneho technika, preto je podstatné ju vypracovať čo najvhodnejšie a najefektívnejšie. V prípade návrhu elektroinštalácie pre medicínske centrum, ktoré má slúžiť práve na poskytovanie zdravotnej starostlivosti pacientov, je o to dôležitejšie dbať na bezpečnosť a spoľahlivosť elektroinštalácie. Navrhnutá projektová dokumentácie je v prvom rade na základe obsahu slovenských technických noriem a iných špecifických požiadaviek. Pre návrh elektroinštalácie bol primárne využitý softvér ProgeCAD. Pre návrh usporiadania hlavného rozvádzača a jeho schému bol využitý softvér EPLAN, ktorý disponuje s databázou elektrických prvkov, a preto je vhodnou voľbou najmä pre tvorbu zapojení, 2D a 3D rozvádzačov. Je vhodné poukázať na niektoré odporúčania, ktoré boli mimo rozsahu práce, avšak boli by jej náležitým vylepšením. Prvým odporúčaním je realizácia svetelnotechnického prepočtu v programe Dialux. Prepočet by bolo vhodné zrealizovať pre pracovné osvetlenie. Vhodnými svietidlami by sa zaručila adekvátna zraková pohoda pre ošetrujúce sestry, lekárov a iných administratívnych pracovníkov objektu. Rovnako by bolo potrebné vykonať aj svetelnotechnický prepočet pre núdzové osvetlenie. Vhodným typom svietidiel by sa tak zabezpečila minimálna osvetlenosť interiérových priestorov pre evakuáciu osôb. Druhým odporúčaním pre prax je doplnenie interiérového a exteriérového kamerového systému, ktorý by zvýšil bezpečnosť osôb v rámci objektu.

References

[1] ILENIN, Stanislav: Elektrické inštalácie a stanice, Technická univerzita v Košiciach, s. 66, ISBN: 978-80-553-1176-0
[2] ASB: Technické zariadenia budov [online]. [cit. 2024-2-12]. Dostupné na internete: <https://www.asb.sk/stavebnictvo/technicke-zariadenia-budov/nove-trendy-vnudzovom-osvetleni>
[3] MERAVÝ, Ján: Elektrotechnická spôsobilosť pre elektrikárov, s. 411 , ISBN: 978-80-89576-09-8
[4] SEZ: Katalóg produktov [online]. Dostupné na internete: <https://www.sez-krompachy.sk/files/2018-05-11-123916-sez_modularne_pristroje_slovenska_verzia.pdf>
[5] ELECTREND: Elektroinštalácie [online]. [cit. 2024-3-23]. Dostupné na internete: <https://legrand.sk/electrend/elektroinstalacie/b-c-d-co-znamenaju-tieto-pismena-na-isticoch/>
[6] BLESKER: Prúdový chránič [online]. [cit. 2024-3-23]. Dostupné na internete: <https://blesker.6f.sk/index.php?stranka=2prudovychranic>
[7] ELECTREND: Prepäťová ochrana [online]. [cit. 2024-3-23]. Dostupné na internete: <[https://legrand.sk/electrend/elektroinstalacie/prepatova-ochrana-vam-zachrani-tisice-eur/>
[8] STN 33 2000-7-710: Požiadavky na osobitné inštalácie alebo priestory – Zdravotnícke priestory
[9] ELDRO: Elektroinštalácia a zdravotnícke priestory [online]. [cit. 2024-4-1]. Dostupné na internete: <https://www.eldro.sk/elektroinstalacia-a-zdravotnicke-priestory-a7-23>
[10] DŽMURA J. a kol., Projektovanie v elektroenergetike, Technická univerzita v Košiciach, s.273,ISBN: 978-80-553-3417-2
[11] KREMNIČAN, Kazimír: Odborná spôsobilosť elektrotechnikov a podnikateľov, s. 347, ISBN: 80-7094-307-6
[12] SCADDAN, Brian: Electrical Installation Work, s.322, ISBN 978-1032-3489-19
[13] COORAY, Vernon: Lightning protection, s.1072, ISBN-13: 978-0-86341-744-3

Meranie indukčnosti fáz spínaného reluktančného motora

[1]*Peter BOBER,* [2]*Želmíra FERKOVÁ*

[1, 2] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická Univerzita v Košiciach, Slovensko

[1]peter.bober@tuke.sk, [2]zelmira.ferkova@tuke.sk

***Abstrakt* — V posledných rokoch stúpol záujem akadémie a priemyslu o výskum v oblasti konštrukcie a riadenia spínaných reluktančných motorov (SRM). Modelovanie a simulácia spínaného reluktančného motora vyžaduje dostatočne presné merania jeho parametrov. Jedným zo základných podkladov pre modelovanie je priebeh indukčnosti jednotlivých fáz v závislosti od uhla otočenia rotora a veľkosti prúdu. Tento článok ponúka prehľad niekoľkých metód merania indukčnosti, opisuje realizáciu metódy priameho merania odozvy prúdu na pripojené napätie a obsahuje výsledky experimentálneho merania.**

***Keywords* — indukčnosť, magnetický tok, model, meranie, spínaný reluktančný motor**

I. INTRODUCTION

Spínaný reluktančný motor (SRM) sa skladá z rotora a statora s vyjadrenými pólmi, pričom na statorových póloch sú fázové vinutia. Po pripojení jednej fázy na zdroj sa rotor začne otáčať v smere znižovania magnetickej reluktancie. Po dosiahnutí polohy s najnižšou reluktanciou sa napätie prepne na ďalšie fázové vinutie a pohyb rotora pokračuje. Vynález SRM sa pripisuje Róbertovi Davidsonovi, ktorý v roku 1839 zostrojil prvý funkčný motor [1]. Hoci sa s nástupom výkonovej elektroniky postavenej na moderných spínacích súčiastkach oživil záujem o SRM na začiatku 80-ty rokoch dvadsiateho storočia, jeho rozšírenie do praxe nastáva až o ďalších 30 rokov. Okrem praktických dôvodov, akými sú jednoduchá konštrukcia, vysoká mechanická a tepelná odolnosť a neprítomnosť magnetov, je dôležitým faktorom potenciálne použitie motorov v elektrických vozidlách [2],[3] a možnosť postaviť motory s vysokou rýchlosťou otáčania približujúcou sa k jednému miliónu otáčok za minútu ([4]–[6]). Návrh konštrukcie a riadenia súčasných SRM sa opiera o modely, ktorých základ formuloval Lawrenson v roku 1980 [7].

II. MODELY SPÍNANÉHO RELUKTANČNÉHO MOTORA

Modely SRM sú analytické, numerické založené na analytických rovniciach a modely vytvárané pomocou metódy konečných prvkov. Na vytvorenie týchto modelov sú potrebné údaje. Modely, založené na analytických rovniciach, vyžadujú meranie alebo identifikáciu elektrického odporu, indukčnosti, momentu, zotrvačnosti, trecieho momentu a strát. Modely vytvárané pomocou metódy konečných prvkov sú postavené na znalostiach mechanických rozmerov a materiálových vlastnostiach a vedia vypočítať magnetický tok, indukčnosť a moment pre konkrétne pracovné podmienky motora. Napriek tomu, že metóda konečných prvkov vie lepšie zohľadniť elektromagnetické deje v stroji, sa analytické modely naďalej používajú, pretože sú podstatne menej náročné na výpočtový čas a to aj o dva rády.

V tomto článku sa budeme zaoberať identifikáciou elektrického odporu a indukčnosti fáz SRM pre analytické modely, ktoré sú numericky riešené v prostredí Matlab-Simulink.

Všeobecná rovnica, ktorá určuje vzťah medzi napätím jednej fázy u a prúdom i vyzerá nasledovne [8]:

$$u = Ri + \frac{d\psi}{dt}, \tag{1}$$

pričom ψ je magnetický tok prechádzajúci vinutím. Indukčnosť L ako vzťah medzi magnetickým tokom a prúdom je určená rovnicou:

$$L(\theta, i) = \frac{\psi(\theta, i)}{i}, \tag{2}$$

V rovnici (2) je už vyjadrená závislosť magnetického toku $\psi(\theta, i)$ a indukčnosti $L(\theta, i)$ na uhle otočenia rotora θ a prúde i. Rovnicu (1) môžeme za pomoci (2) prepísať na napäťovú rovnicu analytického modelu založeného na indukčnosti nasledovne [8]:

$$u = Ri + L(\theta, i)\frac{di}{dt} + i\frac{\partial L(\theta, i)}{\partial \theta}\omega, \tag{3}$$

kde ω je mechanická uhlová rýchlosť rotora. Hodnoty napätia, prúdov a uhlovej rýchlosti je možné relatívne jednoducho merať. Preto sú na týchto veličinách postavené viaceré metódy určenia indukčnosti fáz motora.

III. Meranie indukčnosti fáz motora

V elektrických motoroch je kvôli nasýteniu magnetického obvodu indukčnosť závislá od prúdu. Ďalej sa indukčnosť mení aj so zmenou polohy rotora voči statoru. Preto je pri meraní potrebné zohľadniť hodnotu prúdu a fixovať polohu rotora. Meranie bežným LCR metrom sa z uvedených dôvodnou väčšinou nepoužíva.

Indukčnosť je možné merať viacerými metódami. Na meranie indukčnosti fáz motora sú použiteľné metódy založené na harmonických napätiach a prúdoch a na prúdových impulzoch.

A. Metódy využívajúce harmonické napätia a prúdy

Pri tejto metóde sa fáza motora napája harmonickým napätím a merajú sa napätia na svorkách vinutia a prúd pretekajúci vinutím. Z nameraných hodnôt amplitúdy napätia, prúdu a fázového posunu medzi nimi je možné určiť elektrický odpor a indukčnosť. Medzi tieto metódy patrí volt-ampérová metóda s meraním fázového posunu osciloskopom ([9], [10]) a watmetrická metóda, kde je fázový posun určený z merania činného výkonu watmetrom [11]. Zohľadnenie nasýtenia sa pri volt-ampérovej metóde dosiahne moduláciou harmonického napätia na jednosmerný prúd [10] a pri watmetrickej metóde sa amplitúda prúdu volí dostatočne veľká.

B. Metódy využívajúce impulzné napätia a prúdy

Fázu motora je možné napájať aj krátkym napäťovým impulzom a zaznamenávať nárast prúdu [12], [13]. Veľkosť napätia a šírka impulzu sa musí voliť tak, aby prúd dosiahol hornú hodnotu meraného rozsahu. Príklad zaznamenaného priebehu napätia a prúdu pre jednu hodnotu uhla rotora θ je na obr. 1 b). Výpočet indukčnosti vychádza z rovnice (3). Pri stojacom motore je hodnota $\omega = 0$ a indukčnosť L pre každú hodnotu uhla θ a prúdu $i(\tau)$ v čase τ je možné vyjadriť vzťahom:

$$L[\theta, i(\tau)] = \frac{u(\tau) - Ri(\tau)}{\frac{di(\tau)}{dt}}. \tag{4}$$

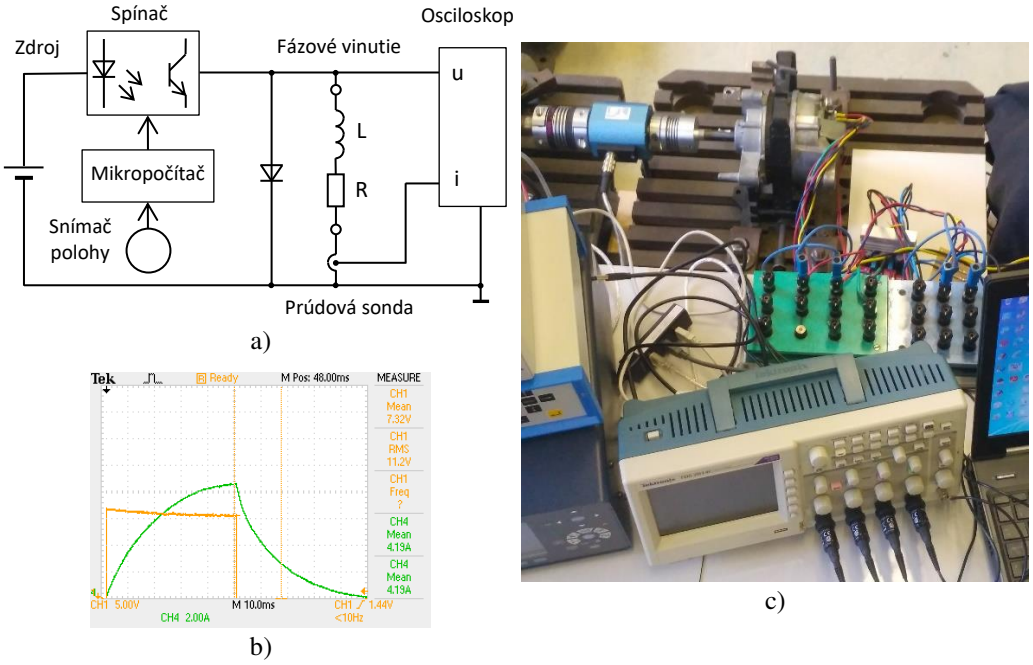
Ak by bolo určenie derivácie prúdu nepresné, potom existuje aj integrálny tvar na vyjadrenie magnetického toku a následný prepočet na indukčnosť:

$$\psi[\theta, i(\tau)] = \int_0^{\tau} [u(t) - Ri(t)]dt, \tag{5}$$

$$L[\theta, i(\tau)] = \frac{\psi[\theta, i(\tau)]}{i(\tau)}. \tag{6}$$

IV. Experimentálna zostava

Meranie prebiehalo na spínanom reluktančnom motore s výkonom 200 W. Jedná sa o trojfázový motor s počtom pólov na statore/rotore 12/8, ktorý je napájaný napätím 120 V a prúdom 2,5 A. Bloková schéma merania je na obr. 1 a). Jedna fáza statora je pripojená cez galvanicky oddelený výkonový spínač na jednosmerný zdroj s napätím 16 V. Osciloskop Tektronix TDS 2014C zaznamenáva časový priebeh napätia na svorkách fázového vinutia a prúdu, ktorý je snímaný prúdovou sondou HAMEG HZ56-2 (obr. 1. b). Mikropočítač generuje impulzy pre výkonový spínač o šírke 50 ms. Zároveň komunikuje so snímačom uhla natočenia rotora. Poloha rotora bola počas merania fixovaná s rozostupom jeden stupeň. Bolo vykonaných 95 meraní s dobou vzorkovania 40 μs. Fotografia meracieho pracoviska je na obr. 1 c).

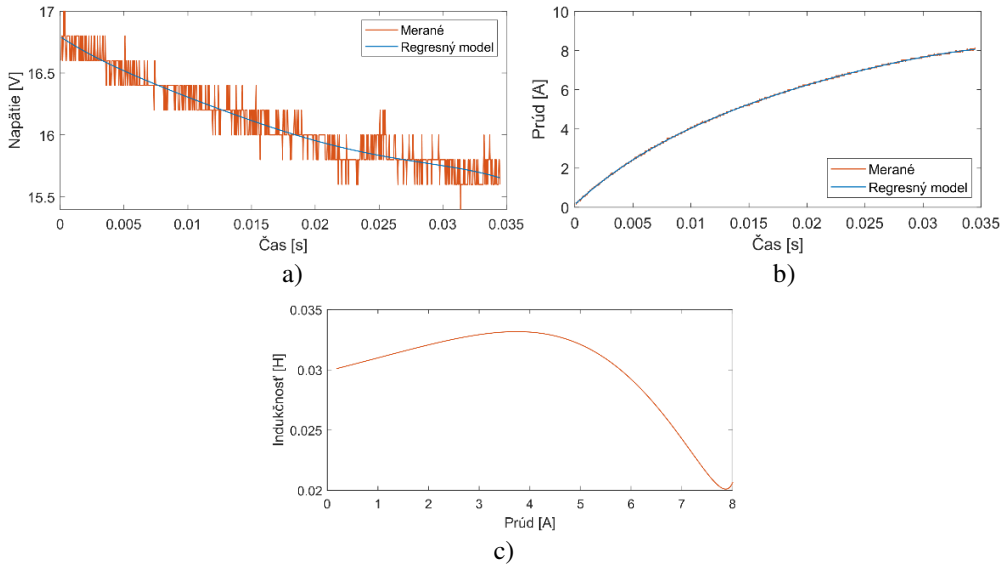


Obr. 1 Bloková schéma merania a); zaznamenané údaje na osciloskope b); meracie pracovisko c).

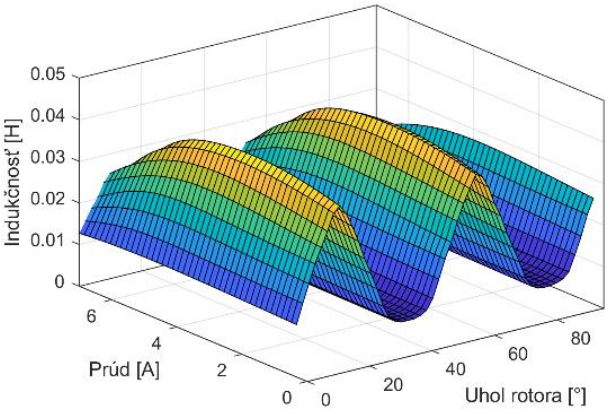
V. VÝSLEDKY MERANIA

Údaje z osciloskopu sú uložené v súboroch .csv a spracované pomocou Matlab script-u. Dáta boli aproximované polynómom piateho rádu pre napätie a šiesteho rádu pre prúd. Vzhľadom na použitú aproximáciu je možné určiť deriváciu prúdu a s dostatočnou presnosťou vypočítať indukčnosť z rovnice (4). Elektrický odpor bol určený ako priemer zo 47 nameraných hodnôt napätia a prúdu v ustálenom stave. Ukážka meraných a aproximovaných údajov a vypočítaná indukčnosť pre hodnotu uhla rotora 19° je na obr. 3. Priebeh vypočítanej indukčnosti na obr. 3 c) sa pre hodnoty prúdu blížiace k 8 A začína nelogicky zvyšovať. Jednou z možných príčin nesprávneho určenia indukčnosti môže byť pokles veľkosti derivácie prúdu, ktorá sa vyskytuje v menovateli rovnice (4). Preto z celého výpočtu budeme uvažovať len hodnoty indukčnosti pre hodnoty prúdu do 7 A, čo aj tak presahuje nominálny prúd 2,5 A uvedený na štítku motora.

Na obr. 4 je zobrazený celý profil nameranej indukčnosti jednej fázy v rozsahu uhlov 0° až 95°. To predstavuje štvrtinu mechanickej otáčky rotora, v ktorej sa oblasť maximálnej a minimálnej indukčnosti opakuje dva krát. Analýzou polohy maxím a miním sme zistili, že snímač polohy rotora neukazuje nulu pri nezarovnanej polohe s minimálnou indukčnosťou. Preto pre použitie v modeloch motora, ktoré majú takúto podmienku, je potrebné vyčísliť posun a transformovať dáta.



Obr. 3 Namerané údaje pre uhol 19°: a) napätie, b) prúd, c) vypočítaná indukčnosť.



Obr. 4 Profil indukčnosti jednej fázy v rozsahu uhlov od 0° do 95°.

VI. ZÁVER

Meranie je základom na vytvorenie modelov, ktoré sú dostatočne presné na to, aby sa závery z experimentov mohli aplikovať na reálne podmienky. Merania bývajú technicky a časovo náročné a je potrebné nájsť kompromis medzi presnosťou a časom a prostriedkami venovanými meraniu. Meranie indukčnosti uvedené v tomto článku predstavuje takýto kompromis, kde na meranie je využitý osciloskop so záznamom nameraných dát do súboru, je meraná jedna fáza v jednej štvrtine otáčky rotora.

Nameraný profil indukčnosti bude v najbližšej dobe porovnaný s údajmi získanými metódou konečných prvkov a po zostavení modelu motora bude možné porovnávať napríklad časové priebehy prúdov so skutočnými prúdmi v motore.

POĎAKOVANIE

Tento príspevok vznikol s podporou Vedeckej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie vied v rámci projektu VEGA 1/0363/23.

LITERATÚRA

[1] J. -W. Ahn and G. F. Lukman, “Switched reluctance motor: Research trends and overview,” in CES Transactions on Electrical Machines and Systems, vol. 2, no. 4, pp. 339-347, Dec. 2018, doi: 10.30941/CESTEMS.2018.00043.

[2] Rekha, P.S., Kumar, M., Panda, G., Vijayakumar, T. “Switched Reluctance Motor Drive Control for Electric Vehicle Application,” in: Panda, G., Alhelou, H.H., Thakur, R. (eds) Sustainable Energy and Technological Advancements. ISSETA 2023. Advances in Sustainability Science and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-99-4175-9_36.

[3] A. Peniak, J. Makarovic, P. Rafajdus et al., “Design and optimization of switched reluctance motor for electrical vehicles,” Electr Eng 99, 1393–1401 (2017). https://doi.org/10.1007/s00202-017-0621-4.

[4] C. Zwyssig, J. W. Kolar and S. D. Round, “Megaspeed Drive Systems: Pushing Beyond 1 Million r/min,” in IEEE/ASME Transactions on Mechatronics, vol. 14, no. 5, pp. 564-574, Oct. 2009, doi: 10.1109/TMECH.2008.2009310.

[5] E. W. Fairall, B. Bilgin and A. Emadi, “State-of-the-art high-speed switched reluctance machines,” 2015 IEEE International Electric Machines & Drives Conference (IEMDC), Coeur d'Alene, ID, USA, 2015, pp. 1621-1627, doi: 10.1109/IEMDC.2015.7409280.

[6] C. Gong and T. Habetler, “Design and Experimental Analysis of an Ultra-High Speed Switched Reluctance Machine With Aerostatic Bearings,” IEEE Transactions on Transportation Electrification, vol. 9, no. 2, pp. 2255-2266, June 2023, doi: 10.1109/TTE.2022.3222481

[7] P. J. Lawrenson, J. M. Stephenson, P. T. Blenkinsop, J. Corda and N. N. Fulton, “Variable-speed switched reluctance motor,”, Proc. Inst. Elect. Eng., vol. 127, pp. 253-265, Jul. 1980.

[8] J. A. Makwana, P. Agarwal, S. P. Srivastava, “Modeling and Simulation of Switched Reluctance Motor,” in Konkani, A., Bera, R., Paul, S. (eds) Advances in Systems, Control and Automation. Lecture Notes in Electrical Engineering, vol 442. Springer, Singapore. https://doi.org/10.1007/978-981-10-4762-6_52.

[9] Tektronix, “Capacitance and Inductance Measurements Using an Oscilloscope and a Function Generator,” 03/2023, [online], [cit. 30.5.2024], Dostupné na <https://www.tek.com/en/documents/application-note/capacitance-and-inductance-measurements-using-oscilloscope-and-function-ge>

[10] F. Alferin, “Measuring self-inductance & ESR,” [online] meettechniek.info, [cit. 30.5.2024], Dostupné na <https://meettechniek.info/passive/inductance.html>

[11] Ž. Ferková, Ľ. Suchý and J. Černohorský, “Measurement of switched reluctance motor parameters,” 2017 19th International Conference on Electrical Drives and Power Electronics (EDPE), Dubrovnik, Croatia, 2017, pp. 287-290, doi: 10.1109/EDPE.2017.8123247.

[12] P. Dúbravka, P. Rafajdus, P. Makyš, L. Szabó, “Control of switched reluctance motor by current profling under normal and open phase operating condition,” IET Electr Power Appl 11(4):548–556. https://doi.org/10.1049/iet-epa.2016.0543.

[13] P. Sovicka, P. Rafajdus, V. Vavrus, “Switched reluctance motor drive with low-speed performance improvement,” Electr Eng 102, 27–41 (2020). https://doi.org/10.1007/s00202-019-00839-0.

Akcelerometrická inerciálna jednotka v rôznych stupňoch voľnosti pre bezkontaktné spínacie pole

Slavomír KARDOŠ

Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky,
Technická univerzita v Košiciach, Slovenská Republika

slavomir.kardos@tuke.sk

***Abstrakt* — Predmetom príspevku je analýza multimodálnej 3-osovej akcelerometrickej inerciálnej jednotky vo funkcii bezkontaktného spínacieho prvku alebo jeho poľa. Bol implementovaný lineárny model využívajúci dynamické zrýchlenie a 8-prvkový inklinačný model využívajúci statické zrýchlenie. Začlenenie vhodnej mechanickej zostavy umožňuje spoľahlivé rozpoznanie jednotlivých stavov a odolnosť voči rušeniu a falošným detekciám. Účelom tejto štúdie je rozvinúť aplikačný potenciál inerciálnej meracej jednotky (IMU) z hľadiska jej robustnosti, spoľahlivosti a použiteľnosti pre pole bezkontaktných spínačov.**

***Kľúčové slová* — akcelerometer, IMU, MEMS, DOF, bezkontaktné spínacie pole**

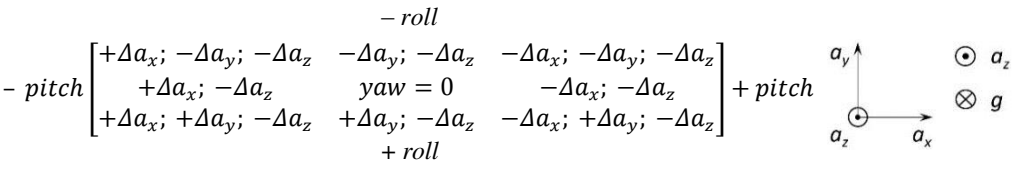
I. Úvod

Rozsah použitia inerciálnych meracích jednotiek (IMU) je rozsiahly kvôli interpretovateľnosti ich signálu v pohybujúcich sa systémoch. Využitím schopnosti snímania gravitačnej zložky zrýchlenia umožňujú aj poskytovanie informácií o zmenách orientácie [1], [2], [3]. Nesmierne využitie našli v medicínskych a fitness prístrojoch, ktoré umožňujú zisťovať a analyzovať pohybové aktivity či poruchy chôdze [4], [5], [6]. Tam, kde sa tradične používajú/používali mechanické spínače a potenciometre, našli priestor pre multimediálne systémy vo funkciách ako "tapping", pre detekciu voľného pádu a podobne. Takéto jednotky často podporujú polohovacie a bezpečnostné zariadenia [7], [8], [9], [10], [11]. Sú široko používané pri snímaní a analýze vibračných signálov v stavebnom priemysle, mechanike a seizmografii [12], [13]. V mnohých aplikáciách sa na zvýšenie presnosti merania používa fúzia akcelerometrických, gyroskopických a magnetometrických jednotiek. V špecifických prípadoch interpretovateľnosť bez tejto fúzie ani nie je možná [1], [3]. Ťažiac z vysokej informačnej hodnoty ich signálu, potenciál využitia spočíva aj v bezkontaktnom princípe snímania, výnimočnom počte pracovných cyklov a extrémnej odolnosti snímacej jednotky voči zaťaženiu, ktorá je až 10^4 g. Potenciálne využitie je vo funkcii bezkontaktného senzora alebo senzorového poľa v aplikáciách s extrémnym počtom pracovných cyklov, v drsnom prostredí, pre anti-vandal aplikácie a ďalšie. V takýchto prípadoch sa kladie dôraz na mechanickú a environmentálnu odolnosť celého elektronického systému [14].

II. Teoretické Pozadie

Predpokladajúc, že modelový systém pozostáva z troch ortogonálne orientovaných ideálnych akcelerometrov, ktorých ťažiská koincidujú v jednom bode s počiatkom súradnicového systému referenčného systému, potom pri lineárnom pohybe takejto sústavy pozdĺž osi kolmej na g vektor, ktorý je orientovaný proti osi z, má zmena lineárnej rýchlosti pozitívny zmysel. Za takýchto podmienok vrátane $a_z \parallel g$ sú prírastky statického zrýchlenia nulové, pretože nedochádza k žiadnej zmene orientácie vzhľadom na referenčný ortogonálny systém, dochádza len k jeho lineárnemu pohybu - zmene polohy.

V prípade rotačného pohybu jednotky okolo ťažísk ich seizmických hmôt sú prírastky statického zrýchlenia [15]:



Obr. 1 Prírastky statického zrýchlenia 3-osového akcelerometra, ktorý je vystavený rotačnému pohybu okolo ťažísk.

Uhol náklonu je vhodným parametrom pre popis predpokladaného stavu sústavy. Pre jednoosový náklon v <90°; -90°> rozsahu platnosti je možné vypočítať náklon na základe jednoosového merania. Povaha funkcie sin() však vyplýva v citlivosť merania, ktorá prudko klesá mimo <45°; -45°> rozsah. Preto je pre vyhodnotenie uhla náklonu spoľahlivejšie meranie v dvoch osiach [22]:

$$pitch = \theta = arctan\,(a_x/a_z)$$
$$roll = \Phi = arctan\,(a_y/a_z) \quad (1)$$

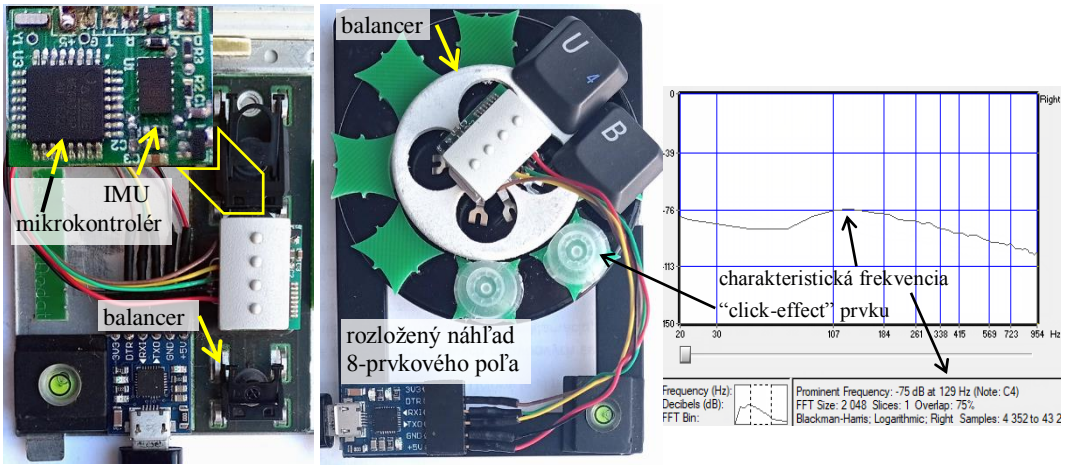
Ak je interval náklonu <45°; -45°>, údaje osi z prispievajú k citlivosti merania minimálne, preto v prípade malých uhlov náklonu postačuje jednoosové snímanie, pretože:

$$\lim_{\theta\to 0}\frac{da_x}{d\theta} = \lim_{\Phi\to 0}\frac{da_y}{d\Phi} = \infty, \quad \lim_{\theta\to 0}\frac{da_z}{d\theta} = 0 \quad (2)$$

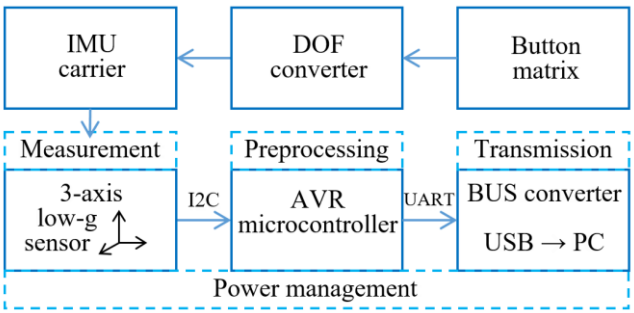
Ak sú súčasne použité oba stupne rotačnej pohybovej voľnosti, je vhodné použiť transformáciu rotačnými maticami [17]. Okrem toho má analýza časového profilu potenciál pre multifunkcionálnu interpretáciu pre rôzne aplikácie.

III. Testovacie Štruktúry

Po začlenení predchádzajúcich úvah boli navrhnuté a implementované dva modely snímania. 1-osový lineárny model (Obr. 2 vľavo) je založený na snímaní zložky dynamického zrýchlenia s jedným stupňom lineárnej pohybovej voľnosti (1LDOF). Obsahuje mechanický prevodník, "click-effect" prvok (komerčný typ), jednotku IMU integrovanú s mikrokontrolérom a mechanické rozhranie. Druhý, 3-osový inklinačný model so 4 stupňami rotačnej pohybovej voľnosti (4RDOF) (Obr. 2 v strede) je založený na statickej zložke zrýchlenia. Experimentálny model obsahuje 8 samostatných tlačidiel naviazaných na mechanický prevodník, ktorý transformuje lineárny pohyb s jedným lineárnym DOF na rotačný. Signál IMU akcelerometra ADXL345 [18] je predspracovaný mikrokontrolérom ATmega328P a odoslaný do PC cez USB na spracovanie (Obr. 3). Experimentálny model bol zarovnaný kolmo na vektor vector $\overrightarrow{F_g}$ a kalibrovaný pre každú zložku orientovanú $a_i \parallel g$ udržujúc $a_i = g$. Inklinačné režimy pitch a roll ovplyvňujú režim rotácie 2 osí a režim zmiešaného náklonu rotáciu všetkých 3 osí. Šírka pásma 1,6 kHz spĺňa požadovanú periódu merania 20 ms kompletného 3x16-bitového signálu. Kapacitný akcelerometer je vystavený mechanickému šumu, jeho signál je preto filtrovaný dolnopriepustným filtrom (LPF). „Click-effect" prvok (suplujúci efekt kliknutia) slúži na percepciu stlačenia tlačidla a má potenciál poskytnúť frekvenčne charakteristický signál.



Obr. 2 1LDOF (vľavo) a 4RDOF (v strede) pole bezkontaktných IMU tlačidiel a spektrálna analýza „click-effect" prvku s dominantnou zložkou (vpravo).



Obr. 3 Bloková schéma hlavných prvkov systému bezkontaktných IMU tlačidiel.

IV. Meracia Procedúra

Stavovej analýze predchádza kalibračná korekcia signálu, jeho úprava a transformácia. Použitý MEMS kapacitný akcelerometer má šírku pásma niekoľko kHz a jeho signál je filtrovaný dolnopriepustným filtrom typu FIR [4], zachovávajúcim dostatočne vysokú citlivosť na zmeny signálu, ale dostatočne eliminujúci mechanický šum v dôsledku zotrvačného pohybu seizmického prvku akcelerometra, ako aj zostavy a vonkajšími vibráciami. Transformácia údajov zahŕňa výpočet absolútneho zrýchlenia a dynamickej zložky, vyhodnotenie sklonu a náklonu, vyhodnotenie smernice a detekciu stavov.

Za predpokladu, že IMU je vodorovne zarovnaný s povrchom kolmým na *g*, stlačením ľavého alebo pravého tlačidla sa snímač otočí okolo osi y, čím je ovplyvnené meranie v osiach x a z. Stlačenie tlačidla hore alebo dole ovplyvňuje osi y a z a v ideálnom prípade nespôsobí žiadnu zmenu stavu na osi x, keď sa systém okolo nej otáča.

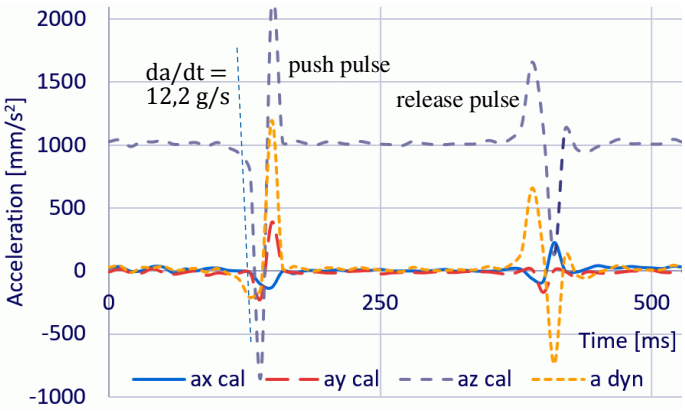
Je možné detegovať jednotlivé stavy, pričom odolnosť voči nepravej detekcii umožňujú charakteristické zložky transformovaných signálov [1], [2]. Sú to schopnosť detegovať statické aj dynamické ($a_{dyn} = |\overrightarrow{a_i}| - |\overrightarrow{a_g}| = \sqrt{a_x^2 + a_y^2 + a_z^2} - 1\,\mathrm{g}$) zložky zrýchlenia a ich extrahovateľnosť, charakteristický tvar impulzov stlačenia a uvoľnenia, charakteristická frekvencia dominantného signálu pochádzajúceho „click-effect“ prvku (Obr. 2 vpravo) a charakteristický uhol náklonu pri analýze zložky statického zrýchlenia.

V. Interpretácia Výsledkov

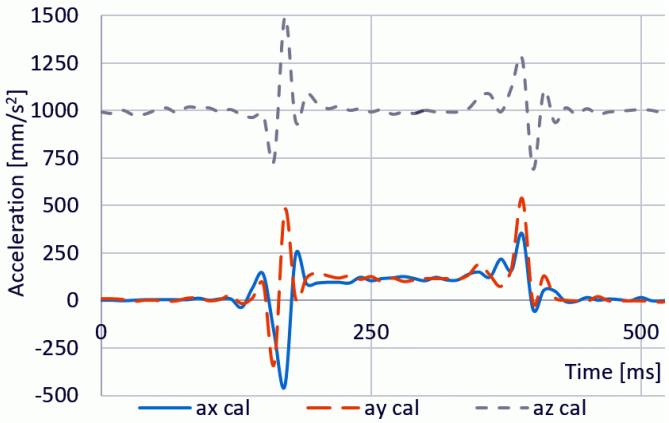
Spracované signály z meracích procedúr na základe predpísaných scenárov boli graficky interpretované a analyzované. Každá jednotlivá prechodová fáza stlačenia bezkontaktného tlačidla bola odlíšená od ostatných fáz ustálenou fázou pre ľahké rozpoznanie jednotlivých stavov. Ako je možné vidieť na Obr. 4 až Obr. 6, takáto prechodná fáza predstavuje signál s vyšším stupňom dynamiky. Teoretické predpoklady vlastností systému sa líšia od interpretovaných profilov, ktoré sú spôsobené zotrvačným pohybom seizmického prvku akcelerometra a vonkajšími faktormi ako mechanický šum a tiež charakterom aplikácie mechanického zaťaženia. Charakter oscilačnej krivky tlmenia vyplýva z charakteru akcelerometrickej jednotky so seizmickou hmotou zavesenou na pružných ramenách z polykryštalického kremíka MEMS štruktúry. Normalizovaná výstupná hodnota senzora umožňuje výrazné oddelenie dynamickej zložky zrýchlenia vzhľadom na vektor *g*. V prípade systému 4RDOF je neideálnosť profilov spôsobená aj nerovnomernou polohou osí otáčania vzhľadom na ťažiská seizmických prvkov akcelerometra.

Systém s jedným stupňom lineárnej pohybovej voľnosti, označovaný ako 1LDOF, má po kalibrácii dobre identifikovateľné stavy a ľahko rozlíšiteľné fázy stlačenia a uvoľnenia (push-release), ako aj evidentný vplyv „click-efekt“ prvku (Obr. 4). Cyklický proces stlačenia a uvoľnenia testovacej zostavy vykázal, že nie všetky elementárne cykly boli ekvivalentné, zakaždým však boli dobre rozlíšiteľné. Táto vlastnosť je spôsobená nedokonalosťou mechanickej zostavy, nerovnomerným rozložením kinetickej energie, ako aj mechanickým šumom vonkajšieho charakteru.

Systém so štyrmi stupňami rotačnej pohybovej voľnosti, označovaný ako 4RDOF, taktiež vykazuje dobre identifikovateľné stavy, podfázy stlačenia a uvoľnenia i vplyv „click-efekt“ prvku (obr. 5). Takto zhotovená zostava vykazovala nižšiu opakovateľnosť cyklov ako 1LDOF, čo je ovplyvnené prevodníkom primárneho pohybového prvku s 1LDOF na 1RDOF, teda prevodom lineárneho pohybu na rotačný.

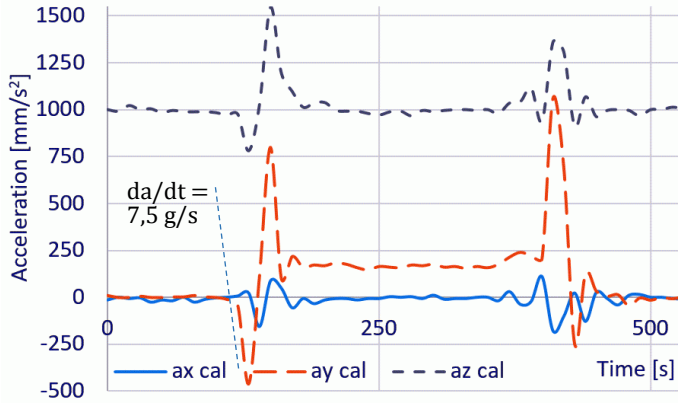


Obr. 4 Odpovedný signál 1LDOF bezkontaktného tlačidla na lineárny mód.



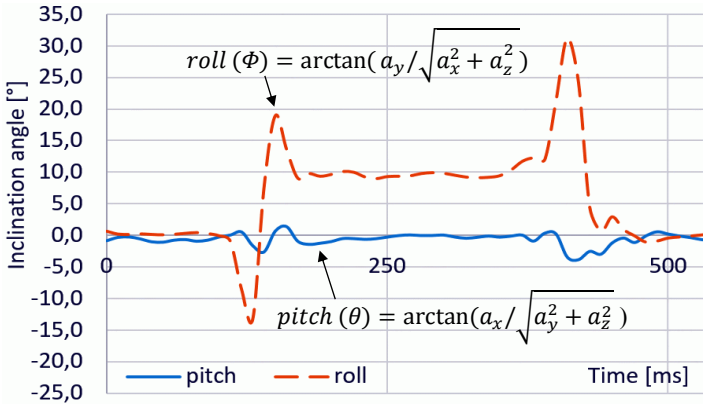
Obr. 5 Odpovedný signál 4RDOF bezkontaktného tlačidla na mód pitch a roll.

Medzifáza stlačenia a uvoľnenia vykazuje výrazne nenulovú hodnotu zrýchlenia pre osi x a y pre kladný zmysel náklonu pich aj roll (Obr. 5) a pre os y v prípade kladného náklonu roll (Obr. 6). Jeho hodnota je dominantne určená len príspevkom tiažového zrýchlenia (Obr. 7). Nenulové hodnoty pre os z možno ľahko interpretovať z vlastností aplikovanej funkcie arctan, respektíve arcsin (rovnica 2). Táto vlastnosť je dôvodom pre použitie dvojosového snímania náklonu pri väčších uhloch.



Obr. 6 Odpovedný signál 4RDOF bezkontaktného tlačidla na mód roll.

Na analýzu a identifikáciu stavov je možné využiť už spomínanú transformáciu signálu na náklon (alebo inklináciu) (Obr. 7), označovaný ako pitch and roll, pričom matematický aparát je využívaný z aplikácie tzv. rotačnej transformácie pomocou rotačných matíc [17]. Trojsignálová forma zrýchlenia pre jednotlivé ortogonálne prvky sa tak transformuje do formy interpretujúcej náklon okolo jednotlivých osí. Pri akcelerometrických meraniach majú pochopiteľne zmysel len pre pitch a roll, avšak nie pre yaw (okolo vektora g). V signáli analyzovanom pre pozitívny roll je tento dominantný náklon dobre rozlíšiteľný, zatiaľ čo parazitný signál pre pitch je spôsobený zotrvačnou povahou seizmickej hmoty akcelerometra a mechanickým šumom zostavy.



Obr. 7 Analýza inklinácie 4RDOF s pozitívnym náklonom roll.

Vzhľadom na charakteristiky aplikovaných goniometrických funkcií pre pracovný interval ±30° (Obr. 7) je snímanie v osiach x a y v oblasti najväčšieho účinku a snímanie v osi z sa na citlivosti podieľa minimálne. merania. Zmena zrýchlenia s časom závisí od charakteru pôsobenia mechanického zaťaženia. Strmosť jej smernice je ovplyvnená mechanickou zostavou, hlavne „click-effect“ prvku. Pre lineárny systém 1LDOF bola zmena zrýchlenia analyzovaná ako 120 m/s^3 (Obr. 4) alebo 12,2 g/s a pre rotačný systém 4RDOF ako 73,5 m/s^3 alebo 7,5 g/s. (Obr. 7).

VI. Záver

Na navrhnutých a implementovaných modeloch kinetického snímania bolo preukázané, že statické aj dynamické zložky zrýchlenia sú vhodné na analýzu jednotlivých stavov bezkontaktného tlačidla (alebo ich poľa). Bol realizovaný 1-prvkový lineárny (1LDOF) model využívajúci dynamickú akceleračnú zložku a 8-prvkový inklinačný (4RDOF) model využívajúci primárne statickú akceleračnú zložku. Začlenením vhodnej mechanickej zostavy oba modely demonštrujú potenciál pre spoľahlivé rozpoznanie jednotlivých stavov a dostatočnú odolnosť voči rušeniu a falošným detekciám, čo je výrazne podporené použitím prvku s efektom kliknutia. 8-tlačidlová konfigurácia je pre implementáciu najefektívnejšia, pretože v systéme 4RDOF vyžaduje diskrétne kombinácie náklonu v zmysle ±pitch a ±roll a ich kombinácie. Významným potenciálom využitia takéhoto bezkontaktného snímača alebo poľa snímačov IMU je tiež robustnosť a spoľahlivosť v aplikáciách s extrémnym počtom pracovných cyklov, v drsnom prostredí, anti-vandal aplikáciách a pod. Na multifunkcionálnu interpretáciu signálu jedinej jednotky možno navyše využiť frekvenčnú analýzu dynamického zrýchlenia, ako aj analýzu inklinačného profilu.

Poďakovanie

Príspevok bol vytvorený s podporou projektov KEGA 011TUKE-4/2023, VEGA 1/0049/24 a 1/0278/24 a ITMS2014+ 313011T557.

Literatúra

[1] A. Vitali, ‘Exploiting the gyroscope to update tilt measurement and eCompass’. STMicroelectronics, Jan. 2021.
[2] T. Zou, F. Ni, C. Guo, K. Li, and H. Liu, ‘A 6-DOF acceleration sensor with cylindrical configuration’, Sens. Actuators Phys., vol. 251, pp. 167–178, Nov. 2016, doi: 10.1016/j.sna.2016.10.005.
[3] P. Patonis, P. Patias, I. N. Tziavos, D. Rossikopoulos, and K. G. Margaritis, ‘A Fusion Method for Combining Low-Cost IMU/Magnetometer Outputs for Use in Applications on Mobile Devices’, Sensors, vol. 18, no. 8, p. 2616, Aug. 2018, doi: 10.3390/s18082616.
[4] J. Fridolfsson et al., ‘Effects of Frequency Filtering on Intensity and Noise in Accelerometer-Based Physical Activity Measurements’, Sensors, vol. 19, no. 9, p. 2186, May 2019, doi: 10.3390/s19092186.
[5] J. J. Kavanagh and H. B. Menz, ‘Accelerometry: A technique for quantifying movement patterns during walking’, Gait Posture, vol. 28, no. 1, pp. 1–15, Jul. 2008, doi: 10.1016/j.gaitpost.2007.10.010.
[6] S. Kardos, S. Slosarcik, and P. Balog, ‘Sensor Array for Evaluation of Gait Cycle’, Adv. Electr. Electron. Eng., vol. 17, no. 4, pp. 459–465, Dec. 2019, doi: 10.15598/aeee.v17i4.3352.
[7] Chunming Gao, R. Pastel, and Jindong Tan, ‘Yet another user input method: Accelerometer assisted single key input’, in 2010 8th World Congress on Intelligent Control and Automation, Jinan, China: IEEE, Jul. 2010, pp. 3949–3954. doi: 10.1109/WCICA.2010.5553840.
[8] A. J. Zwart, G. M. Derige, D. Effa, P. Nieva, and S. Lancaster-Larocque, ‘A Novel Virtual Button User Interface for Determining the Characteristics of an Impulse Input Based on MEMS Inertial Sensors’, Sens. Transducers J., vol. 7, no. Special Issue, pp. 179–190, Oct. 2009.
[9] Y. Zhao, X. Ren, C. Lian, K. Han, L. Xin, and W. J. Li, ‘Mouse on a Ring: A Mouse Action Scheme Based on IMU and Multi-Level Decision Algorithm’, IEEE Sens. J., vol. 21, no. 18, pp. 20512–20520, Sep. 2021, doi: 10.1109/JSEN.2021.3096847.
[10] D. Effa and A. Zwart, ‘Virtual Button® MEMS Accelerometer Design’. University of Waterloo, Faculty of Engineering, Jul. 23, 2010.

[11] V. Janousek, Z. Jirankova, and P. Stukjunger, 'Setting up single-tap and double-tap recognition with ST's MEMS accelerometers'. STMicroelectronics, May 2018.
[12] M. Looney, 'MEMS Vibration Monitoring: From Acceleration to Velocity'. Analog Devices, Inc., Jun. 2017.
[13] M. Looney, 'An Introduction to MEMS Vibration Monitoring. ADI Analog Dialogue'. Analog Devices, Inc., Jun. 2014.
[14] J. L. Evans, P. Lall, R. Knight, E. Crain, T. Shete, and J. R. Thompson, 'System Design Issues for Harsh Environment Electronics Employing Metal-Backed Laminate Substrates', IEEE Trans. Compon. Packag. Technol., vol. 31, no. 1, pp. 74–85, Mar. 2008, doi: 10.1109/TCAPT.2008.916792.
[15] S. Kardos, 'Utilization of a Digital Accelerometer Unit for a Switch Array Emulation', in 2022 45th International Spring Seminar on Electronics Technology (ISSE), Vienna, Austria: IEEE, May 2022, pp. 1–5. doi: 10.1109/ISSE54558.2022.9812712.
[16] M. Pedley, 'Tilt Sensing Using a Three-Axis Accelerometer'. Freescale Semiconductor, Inc., Mar. 2013.
[17] M. Khurma, 'Rotation Matrix'. Cuemath learning center.: https://www.cuemath.com/algebra/rotation-matrix/
[18] Analog Devoces, Inc., 'ADXL345 3-Axis, ±2 g/±4 g/±8 g/±16 g Digital Accelerometer, Rev. G'. 2022.

Využitie GAN sietí pre augmentáciu dát a zlepšenie klasifikácie dát

[1] Lenka Kališková, [2] Peter Butka

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

[1]lenka.kaliskova@tuke.sk, [2]peter.butka@tuke.sk

Abstrakt **– Predmetom tejto práce je využitie augmentácie založenej na báze GAN siete pre zlepšenie klasifikácie konvolučnej neurónovej siete. V práci sa využilo viacero druhov augmentácií založených na báze GAN sietí pre dátové množiny z astro/geo oblastí. Na porovnanie týchto výsledkov sa použili metódy klasickej augmentácie. Výsledky tejto práce ponúkajú prehľad augmentácií založených na báze GAN sietí a ich vplyvu na kvalitu klasifikačných modelov.**

Kľúčové slová **– augmentácia, generative adversarial network, konvolučná neurónová sieť, klasifikácia, hlboké učenie, Python**

I. Úvod

Neurónová sieť je model skladajúci sa z viacerých vrstiev, ktorých cieľom je simulovať správanie ľudského mozgu. Učenie neurónových sietí vyžaduje dátovú množinu s dostatočným množstvom dát. V prípade, ak model na vstupe nemá dostatočné množstvo dát, tak nemôže dosiahnuť svoj potenciál. Riešením daného problému je augmentácia dát, ktorá dokáže z existujúcej dátovej množiny vygenerovať nové dáta. Táto technika je populárna najmä v oblastiach, kedy získanie nových dát je v princípe náročné alebo nemožné. Veľmi často sa augmentácia používa v spracovaní obrázkových dát. Existuje viacero klasických augmentačných techník pre tento typ dát, ako napríklad otočenie, ktoré spĺňa svoju úlohu dobre, ale aj napriek tomu sa v súčasnosti skúmajú nové druhy augmentácií, ktoré by boli schopné výrazne zlepšiť výsledky v oblastiach, v ktorých nie je k dispozícii veľké množstvo dát. Najpopulárnejším typom novej augmentácie je využitie GAN siete, od ktorej sa sľubuje dosiahnutie lepších výsledkov modelu vďaka kvalitnejšej augmentácii. V práci využívame viacero typov GAN sietí na augmentáciu viacerých dátových množín obsahujúcich obrázky.

II. Dáta

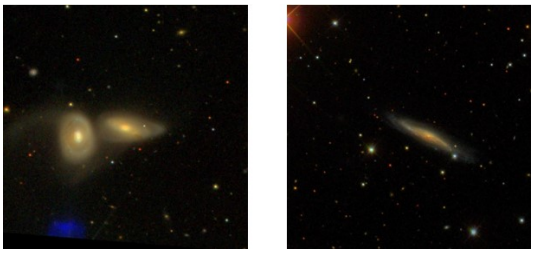
V práci boli použité tri dátové množiny, pričom dve z nich je možné zaradiť do astronomickej oblasti a jednu do geologickej oblasti.

A. Galaxy Mergers

Prvé dáta pochádzajúce z astronomickej oblasti sa nazývajú Galaxy Mergers a obsahujú snímky galaxií. Tieto snímky galaxií pochádzajú zo Sloan Digital Sky Survey (skr. **SDSS**) Data Release 7, pričom sa delia do dvoch tried, a to:

- Merger - trieda obsahuje snímky galaxií, ktoré spolu interagujú.
- Non-Interacting - trieda obsahuje snímky galaxií, ktoré neprichádzajú do spoločného kontaktu.

Celkový počet snímok v dátovej množine je 12 001, pričom 4 501 snímok patrí do triedy označovanej merger a 7 500 snímok patrí do triedy označovanej non-interacting. Ukážka príkladov z tejto dátovej množiny je zobrazená na obrázku 1.



Obrázok 1 Ukážka príkladov z dátovej množiny Galaxy Mergers

B. FIRST Data

Druhé dáta pochádzajúce z astronomickej oblasti nazývané FIRST Data obsahujú snímky rádiových galaxií. Dané snímky boli získané z katalógu Faint Images of the Radio Sky at Twenty Centimeters (skr. **FIRST**). Snímky tejto dátovej množiny sú morfologicky klasifikované do štyroch tried, a to:

- Fanaroff–Riley Class I (skr. **FRI**) - rádiové galaxie tohto typu obsahujú regióny s nízkym jasom, ktoré sa nachádzajú ďalej od centrálnej galaxie a regióny s vysokým jasom, ktoré sa nachádzajú v blízkosti centra galaxie.
- Fanaroff–Riley Class II (skr. **FRII**) - rádiové galaxie tohto typu sú zložené zo svetelných rádiových zdrojov, ktoré obsahujú aktívne body vo svojich lalokoch.
- Bent-Tailed (skr. **BENT**) - rádiové galaxie vyznačujúce sa ohýbaním, ktoré je ovplyvnené prostredím alebo vnútornými vlastnosťami, pričom sa delia do dvoch tried, a to na galaxie so širokouhlým chvostom a galaxie s úzkym chvostom.
- Compact (skr. **COMPT**) - kompaktné rádiové galaxie, zvyčajne zložené z jediného viditeľného objektu.

Celkový počet snímok v dátovej množine je 528, pričom z tohto počtu 125 snímok patrí do triedy FRI, 216 snímok patrí do triedy FRII, 104 snímok patrí do triedy BENT a 83 snímok patrí do triedy COMPT. Ukážka príkladov z tejto dátovej množiny je zobrazená na obrázku 2.



Obrázok 2 Ukážka príkladov z dátovej množiny FIRST Data

C. SVM Data

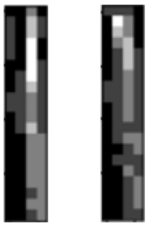
Dátová množina patriaca do geologickej oblasti sa nazýva SVM Data a obsahuje spektrogramy zo seizmografu, ktoré pochádzajú z Fínskej seizmografickej siete z rôznych vysokofrekvenčných staníc. Dátová množina obsahuje rozdelenie do dvoch tried, a to:

- Explózie - spektrogramy obsahujúce údaje o explóziách, označené ako trieda 0.
- Zemetrasenia - spektrogramy obsahujúce údaje o zemetraseniach, označené ako trieda 1.

Počet snímok pre triedu reprezentujúcu explózie je 25 852 a počet snímok reprezentujúcich zemetrasenia je 7 510. Dáta sú definované poliami, z ktorých sú vygenerované obrázky vo veľkosti 20 x 4 pixelov. Ukážka príkladov z tejto dátovej množiny je zobrazená na obrázku 3.

III. Konvolučné neurónové siete

Konvolučná neurónová sieť (skr. **CNN**) je považovaná za jeden z najpoužívanejších typov neurónových sietí, s primárnym zameraním na detekciu a klasifikáciu obrazov. Tento typ neurónovej siete pozostáva z troch hlavných typov vrstiev, ktorými sú konvolučná vrstva (angl. **convolutional layer**), vzorkovacia vrstva (angl. **pooling layer**), aktivačná vrstva (angl. **activating function layer**) a plne prepojená vrstva (angl. **fully connected layer**) [1]. Ako už naznačuje názov, hlavným stavebným blokom CNN je konvolučná vrstva, ktorá vykonáva operáciu nazývanú konvolúcia. Konvolúcia je lineárna operácia, ktorá zahŕňa násobenie váh so vstupom. Ak si predstavíme



Obrázok 3 Ukážka príkladov z dátovej množiny SVM Data

dvojrozmerný vstup, násobenie prebieha medzi maticou vstupných dát a dvojrozmernou maticou váh, ktorá sa nazýva filter alebo jadro. Tento proces násobenia sa opakuje, až kým jadro nepokryje celý obraz. Výstupom z konvolúcie je mapa príznakov [2]. Konvolúcia ako súčasť neurónovej siete je definovaná rovnicou (1).

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n) K(i-m, j-n), \tag{1}$$

kde I označuje vstupný obrázok, K označuje kernel, m a n sú indexy riadkov a stĺpcov kernela, i a j sú indexy stĺpcov a riadkov výslednej matice, ktorá predstavuje jadro konvolúcie [1].

Vzorkovacia vrstva slúži na redukciu rozmerov, a tiež pomáha znížiť počet parametrov na vstupe konvolučnej neurónovej siete, čo umožňuje zlepšiť efektívnosť a znižuje riziko preučenia siete. Existujú dva typy, a to maximálne vzorkovanie (angl. **max pooling**) a priemerné vzorkovanie (angl. **average pooling**).

Plne prepojená vrstva je umiestená vždy na konci konvolučnej neurónovej siete, pričom súčasťou plne prepojenej vrstvy je aj výstupná vrstva. Cieľom plne prepojenej vrstvy je klasifikovať vstupy na základe vlastností, ktoré boli extrahované zo vstupu pomocou predchádzajúcich vrstiev konvolučnej neurónovej siete [2].

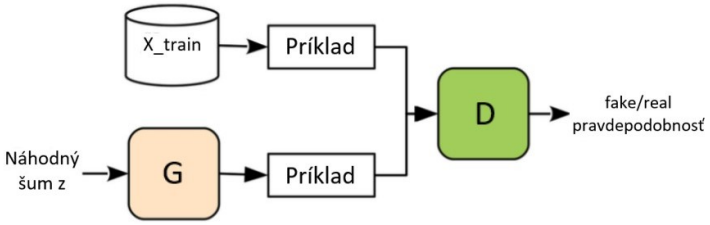
IV. Augmentácia založená na báze GAN siete

Predtým než sa dostaneme priamo ku konkrétnym variáciám GAN sietí na augmentáciu, je podstatné si najskôr priblížiť základnú GAN sieť.

A. *Generative Adversarial Network*

GAN sieť bola prvýkrát predstavená v práci [3]. Cieľom autorov bolo predložiť nový prístup ku generatívnym modelom. GAN sieť je generatívny model, ktorý simultánne trénuje dva typy modelov. Generatívne modely sa trénujú tak, aby popisovali ako sa generujú dáta pomocou pravdepodobnostného modelu. Vďaka vzorkovaniu z generatívneho modelu sme schopní generovať nové dáta. Generatívne modely sa často používajú na neoznačené dáta a môžeme ich považovať za nekontrolované učenie [3].

GAN pozostáva z generátora G, ktorý zachytáva distribúciu dát a diskriminátora D, ktorý určuje pravdepodobnosť, že vzorka dát pochádza z trénovacej množiny, a nie z generátora G. Cieľom trénovacej procedúry pre generátor G je maximalizovať pravdepodobnosť, že diskriminátor D spraví chybu. Tento framework je podobný jednému zo základných algoritmov umelej inteligencie nazývanému Minimax. Princípom tohto algoritmu je minimalizácia straty pre najhorší možný scenár, teda pre maximálnu možnú stratu [3]. Odtiaľ pochádza aj názov Minimax, teda minimalizácia maximálnej možnej straty. Štruktúra GAN siete je zobrazená na obrázku 4.



Obrázok 4 Štruktúra GAN siete

Štúdia [3] sa stala východiskom pre vytvorenie viacerých variácií GAN sietí, ktoré slúžia na riešenie rôznych typov problémov, pričom tu samozrejme patrí aj úloha, ktorou sa zaoberáme, teda augmentácia dát pre zlepšenie úspešnosti klasifikácie konvolučnej neurónovej siete.

B. Deep Convolutional Generative Adversarial Network

Prvá variácia GAN siete, ktorá sa používa na augmentáciu obrázkov sa nazýva Deep Convolutional Generative Adversarial Network (skr. **DCGAN**). DCGAN sieť bola prvýkrát predstavená v práci [4], pričom ide o GAN sieť, ktorá je upravená rozšírením generátora a diskriminátora o konvolučné vrstvy. Generátor v DCGAN sieti obsahuje namiesto vzorkovacích vrstiev transponovanú konvolúciu (angl. **transposed convolution**), vďaka čomu môže sieť meniť veľkosť vstupu a nemusí byť tak určená vopred. Vzorkovacia vrstva v diskriminátore je nahradená krokom konvolúcie (angl. **strided convolution**). Batch normalizácia (angl. **batch normalization**) pomáha stabilizovať trénovanie pomocou aktivácií, pričom je použitá na každej vrstve okrem výstupnej vrstvy pre generátor a vstupnej vrstvy pre diskriminátor. V generátore sa používa aktivačná funkcia ReLU okrem výstupnej vrstvy, kde je použitá výstupná funkcia tanh, vďaka ktorej model dokáže rýchlejšie pokryť priestor farieb trénovacej dátovej množiny [4].

C. Data Augmentation Generative Adversarial Network

Data Augmentation Generative Adversarial Network (skr. **DAGAN**) je sieť, ktorá bola navrhnutá v práci [5] špeciálne pre augmentáciu dát, pričom jej cieľom je vyriešiť problém s dátovými množinami, ktoré obsahujú nedostatočné množstvo dát. DAGAN sieť generuje nové obrázky na základe redukcie dimenzionality vstupných obrázkov. Generátor v DAGAN sieti funguje ako encoder, čo znamená, že vstupný obrázok zakóduje, pridá šum a následne ho rozkóduje. Diskriminátor dokáže rozlíšiť pôvodný vstupný obrázok a transformovaný obrázok, pričom taktiež vie rozlíšiť dvojicu obrázkov z rovnakej triedy. Vďaka tomu diskriminátor pomáha generátoru v lepšom učení transformácií, ktoré následne modifikujú obrázky. Vygenerované obrázky sa môžu charakterizovať tým, že nie sú príliš podobné pôvodným obrázkom, ale aj napriek tomu nemenia triedu, do ktorej daný obrázok patrí [5].

D. Auxiliary Classifier Generative Adversarial Network

Auxiliary Classifier Generative Adversarial Network (skr. **ACGAN**) sieť je rozšírením conditional GAN (skr. **cGAN**) siete. Predtým, ako si vysvetlíme ACGAN sieť je preto potrebné sa zamerať na cGAN sieť. Architektúra cGAN siete dokáže narozdiel od pôvodnej GAN siete využiť na vstupe pre generátor a diskriminátor informácie, ktoré sú k dispozícii, ako napríklad trieda, do ktorej daný obrázok patrí. Generátor v ACGAN má na vstupe náhodne body zo skrytého priestoru (angl. **latent space**), a taktiež triedu, do ktorej sa snaží vygenerovať nový obrázok. Vďaka tejto štruktúre generátora je tréning stabilnejší, keďže generuje obrázky pre konkrétnu triedu. Diskriminátor v ACGAN sieti sa následne nezaoberá len predikciou či obrázok je skutočný alebo nie, ale taktiež predikuje aj triedu, do ktorej daný obrázok patrí [6].

E. Wasserstein Generative Adversarial Network

Wasserstein Generative Adversarial Network (skr. **WGAN**) sieť bola predstavená v práci [7]. WGAN sieť narozdiel od iných GAN sietí nahrádza diskriminátor kritikom (angl. **critic**), ktorého cieľom je na základe skóre určiť, či obrázok je skutočný alebo nie je. Hlavným dôvodom tejto zmeny je matematický dôkaz, že trénovanie generátora by malo minimalizovať vzdialenosť medzi distribúciou dát v trénovacej dátovej množine a distribúciou vo vygenerovaných obrázkoch. Na daný matematický dôkaz WGAN sieť využíva Wassersteinovú vzdialenosť. Wassersteinova vzdialenosť zaručuje, že narozdiel od klasického diskriminátora dokáže poskytovať lineárny gradient aj potom, čo kritik dosiahol optimálne trénovanie. Strata diskriminátora sa taktiež môže podpísať pod slabšiu kvalitu vygenerovaných obrázkov, a keďže strata kritika je pri generovaní obrázkov výrazne nižšia, WGAN generuje kvalitnejšie obrázky ako pôvodná GAN sieť [7].

F. Balancing Generative Adversarial Network

Balancing Generative Adversarial Network (skr. **BAGAN**) sieť bola predstavená v práci [8]. Úlohou BAGAN siete je generovanie realistických obrázkov, vďaka ktorým sa obnoví rovnováha v nevybalansovaných dátových množinách. Nevybalansovaná dátová množina predstavuje takú množinu, v ktorej počet príkladov jednej triedy výrazne prevyšuje počet príkladov triedy druhej. BAGAN spája GAN sieť a autoencoding, pričom cieľom je využiť silné stránky obidvoch metód. Zatiaľ čo GAN dokáže generovať obrázky vo vysokej kvalite, tak autoencoder sa dokáže rýchlo a jednoducho priblížiť k optimálnemu riešeniu úlohy. Autoencoder spracováva všetky vstupné obrázky bez ohľadu na to, do ktorej triedy daný obrázok patrí. Narozdiel od autoencondera, generátor G a diskriminátor D v GAN sieti poznajú jednotlivé triedy, do ktorých obrázky patria. Cieľom generátora G je následne počas trénovania siete generovať nové obrázky, zatiaľ čo úlohou diskriminátora D je určiť, či daný obrázok je skutočný alebo nie. Autoencoder vstupuje do procesu v momente, kedy je inicializovaná sieť GAN, a to poskytnutím znalostí, ktoré nadobudol [8].

V. EXPERIMENTY

Naše experimenty sú zamerané na využitie rôznych typov GAN pre augmentáciu s hlavným cieľom zistiť, ktorá augmentácia na báze GAN je najvhodnejšia pre zlepšenie klasifikácie.

Dôležitou časťou experimentovania je navrhnutie architektúry konvolučnej neurónovej siete pre klasifikáciu obrázkov. Daná konvolučná neurónová sieť je použitá na klasifikáciu dátových množín obsahujúcich obrázky vygenerované GAN sieťou. Výsledky z daných sietí sú následne použité na určenie najvhodnejšej augmentácie na báze GAN pre zvolené dátové množiny. Okrem určenia samotnej najvhodnejšej augmentácie na báze GAN sme na porovnanie vytvorili aj kombinácie GAN augmentácií, ktoré predstavujú spojenie dvoch najlepších sietí na báze GAN, kombináciu najlepšej a najhoršej siete na báze GAN a spojenie dvoch najhorších sietí.

Vytvorená konvolučná neurónová sieť obsahuje tri konvolučné vrstvy. Prvá konvolučná vrstva obsahuje 128 filtrov, druhá a tretia konvolučná vrstva obsahuje 64 filtrov. Konvolučné vrstvy využívajú aktivačnú funkciu ReLU. Po každej konvolučnej vrstve je použitá vzorkovacia vrstva typu MaxPooling s veľkosťou 2 x 2. V závere je použitá metóda Dropout s hodnotou 0.3 a dopredná vrstva s počtom neurónov 128. Na výstupnej vrstve je použitá aktivačná funkcia Softmax. Na optimalizáciu je využitá metóda Adam.

A. Generative Adversarial Network

Architektúra generátora GAN je v našej práci zložená z plne prepojených vrstiev, jednej vrstvy typy Reshape slúžiacej na úpravu tvaru, vzorkovacích vrstiev typu UpSampling, normalizačných vrstiev typu BatchNormalization a konvolučných vrstiev. Medzi jednotlivými vrstvami je použitá aktivačná funkcia ReLU, pričom na výstupe sa používa aktivačná funkcia tanh.

Architektúra diskriminátora našej GAN siete je zložená z konvolučných vrstiev, normalizačných vrstiev typu BatchNormalization, jednej vrstvy typu ZeroPadding slúžiacej na úpravu tvaru a viacerých regularizačných vrstiev typu Dropout. Medzi jednotlivými vrstvami je použitá aktivačná funkcia LeakyReLU, pričom na výstupe sa používa aktivačná funkcia Sigmoid.

B. Deep Convolutional Generative Adversarial Network

DCGAN je variácia pôvodnej GAN siete, ktorá je modifikovaná len nahradením vzorkovacích vrstiev transponovanou konvolúciou.

Architektúra generátora DCGAN je v našej práci zložená z plne prepojených vrstiev, jednej vrstvy typy Reshape slúžiacej na úpravu tvaru, viacerých vrstiev transponovanej konvolúcie a konvolučných vrstiev. Medzi jednotlivými vrstvami je použitá aktivačná funkcia LeakyReLU, pričom na výstupe sa používa aktivačná funkcia tanh.

Architektúra diskriminátora našej DCGAN siete je zložená z konvolučných vrstiev a jednej regularizačnej vrstvy typu Dropout. Medzi jednotlivými vrstvami je použitá aktivačná funkcia LeakyReLU, pričom na výstupe sa používa aktivačná funkcia Sigmoid.

C. Data Augmentation Generative Adversarial Network

Sieť DAGAN, ktorá bola primárne navrhnutá na augmentáciu obrázkov je nepatrnou modifikáciou pôvodnej GAN siete.

Architektúra generátora a diskriminátora je v našej práci identická, a skladá sa výhradne z plne prepojených vrstiev. Aktivačná funkcia medzi jednotlivými vrstvami je LeakyReLU, pričom na výstupnej vrstve je v prípade generátora použitá funkcia tanh v kombinácii s vrstvou typu Reshape slúžiacou na úpravu tvaru. V prípade diskriminátora je na výstupnej vrstve použitá funkcia Sigmoid.

D. Auxiliary Classifier Generative Adversarial Network

Výhodou modifikácie pôvodnej architektúry GAN siete nazývanej ACGAN je možnosť využiť na vstupe všetky dostupné informácie. Príkladom takejto informácie je napríklad trieda, ktorú v pôvodnej architektúre GAN nie je možné využiť.

Architektúra generátora ACGAN je v našej práci zložená z plne prepojených vrstiev, jednej vrstvy typy Reshape slúžiacej na úpravu tvaru, vzorkovacích vrstiev typu UpSampling, normalizačných vrstiev typu BatchNormalization a konvolučných vrstiev. Medzi jednotlivými vrstvami je použitá aktivačná funkcia ReLU, pričom na výstupe sa používa aktivačná funkcia tanh.

Architektúra diskriminátora našej ACGAN siete je zložená z konvolučných vrstiev, normalizačných vrstiev typu BatchNormalization, jednej vrstvy typu ZeroPadding slúžiacej na úpravu tvaru a viacerých regularizačných vrstiev typu Dropout. Medzi jednotlivými vrstvami je použitá aktivačná funkcia LeakyReLU.

E. Wasserstein Generative Adversarial Network

WGAN sieť je variáciou architektúry GAN siete, ktorá nahrádza diskriminátor modelom nazývaným kritik. Daný model určuje pôvod obrázka na základe skóre.

Architektúra generátora WGAN je v našej práci zložená z plne prepojených vrstiev, jednej vrstvy typy Reshape slúžiacej na úpravu tvaru, vzorkovacích vrstiev typu UpSampling, normalizačných vrstiev typu BatchNormalization a konvolučných vrstiev. Medzi vrstvami je použitá aktivačná funkcia ReLU, pričom na výstupe sa používa aktivačná funkcia tanh.

Architektúra kritika našej WGAN siete je zložená z konvolučných vrstiev, normalizačných vrstiev typu BatchNormalization, jednej vrstvy typu ZeroPadding slúžiacej na úpravu tvaru a viacerých regularizačných vrstiev typu Dropout. Medzi vrstvami je použitá aktivačná funkcia LeakyReLU.

F. Balancing Generative Adversarial Network

Narozdiel od ostatných variácií GAN sietí, BAGAN obsahuje okrem generátora a diskriminátora aj modely nazývané decoder a encoder.

Generátor BAGAN v našej práci obsahuje jednoduchú architektúru. Architektúra diskriminátora BAGAN je v našej práci zložená z plne prepojených vrstiev a konvolučných vrstiev. Medzi jednotlivými vrstvami je použitá aktivačná funkcia LeakyReLU.

Architektúra decodera BAGAN je zložená z plne prepojených vrstiev, jednej vrstvy typy Reshape slúžiacej na úpravu tvaru, viacerých vrstiev transponovanej konvolúcie a normalizačných vrstiev typu BatchNormalization. Medzi vrstvami je použitá aktivačná funkcia LeakyReLU.

Architektúra encodera BAGAN je zložená výhradne z konvolučných vrstiev. Medzi vrstvami je použitá aktivačná funkcia LeakyReLU.

Na porovnanie augmentácie na báze GAN s metódami klasickej augmentácie sme použili rotáciu obrazu v troch uhloch, ktorými sú uhly 90 stupňov, 180 stupňov a 270 stupňov.

VI. VYHODNOTENIE

Pri vyhodnotení konvolučnej neurónovej siete a porovnaní vplyvu jednotlivých GAN augmentácií na klasifikáciu využívame viaceré metriky, ako napríklad úspešnosť (angl. **accuracy**), presnosť (angl. **precision**), návratnosť (angl. **recall**) a F1 skóre (angl. **f1 score**).

A. Galaxy Mergers

Základná konvolučná neurónová sieť bez akéhokoľvek rozšírenia dosiahla presnosť klasifikácie 89.71% Všetky variácie augmentácie založené na báze GAN pomohli zlepšiť presnosť klasifikácie o 3%. Najväčší vplyv mali rozšírené obrázky z pôvodnej GAN siete, ktoré rovnako pomohli dosiahnuť presnosť klasifikácie 92.92%, ale rozdiel medzi najlepšou a najhoršou variáciou augmentácie založenej na báze GAN bol len 0.58%. V porovnaní s tým, klasifikácia s obrázkami z klasickej augmentácie dokázala dosiahnuť oveľa výraznejšie zlepšenie na 95.46%. O niečo lepšie výsledky boli dosiahnuté aj pri kombinácii najlepšej augmentácie na báze GAN, teda pôvodnej GAN siete a klasickej augmentácie na 95.88%. Na základe toho je za najlepší model možné považovať spomenutú kombináciu. Podrobnejšie výsledky tohto modelu sú uvedené v tabuľke 1.

Tabuľka 1
Vyhodnotenie najlepšieho modelu pre dátovú množinu Galaxy Mergers

Trieda	Presnosť	Návratnosť	F1-skóre	Počet dát
merger	0.94000	0.95508	0.94748	935
non-interacting	0.97105	0.96112	0.96606	1466
Úspešnosť			0.95877	2401
Vážený priemer	0.95896	0.95877	0.95883	2401

B. FIRST Data

Základná konvolučná neurónová sieť bez akéhokoľvek rozšírenia dosiahla presnosť klasifikácie 83.02% Všetky variácie augmentácie založené na báze GAN pomohli zlepšiť presnosť klasifikácie v rozsahu 4 až 7%. Najväčší vplyv mali rozšírené obrázky z DCGAN a WGAN, ktoré rovnako pomohli dosiahnuť presnosť klasifikácie 90.57%. V porovnaní s tým, klasifikácia s obrázkami z klasickej augmentácie dokázala dosiahnuť dodatočné zlepšenie na 92.45%. Pretože sa nám nepodarilo dosiahnuť lepšie výsledky s kombináciou viacerých rozšírení založených na báze GAN, zvolili sme kombináciu DCGAN a ACGAN. Je dôležité spomenúť, že do jednej triedy boli pridané iba obrázky z jednej variácie augmentácie založenej na báze GAN. Táto kombinácia dosiahla výrazné zlepšenie na 98.11%. Podrobnejšie výsledky tohto modelu sú uvedené tabuľke 2.

Tabuľka 2
Vyhodnotenie najlepšieho modelu pre dátovú množinu FIRST Data

Trieda	Presnosť	Návratnosť	F1-skóre	Počet dát
BENT	1.00000	0.95238	0.97561	21
COMP	1.00000	1.00000	1.00000	17
FRI	1.00000	0.96000	0.97959	25
FRII	0.95556	1.00000	0.97727	43
Úspešnosť			0.98113	106
Vážený priemer	0.98197	0.98113	0.98114	106

C. SVM Data

Základná konvolučná neurónová sieť bez akéhokoľvek rozšírenia dosiahla presnosť klasifikácie 92.73%. Päť variácií augmentácií založených na báze GAN pomohlo mierne zlepšiť presnosť klasifikácie. Konkrétne variácia DAGAN bola o 0.3% horšia ako základný model. Najväčší vplyv mal ACGAN, ale v tomto prípade ide tiež len o mierne zlepšenie na 93.10% presnosť klasifikácie. Pretože sa nám nepodarilo dosiahnuť lepšie výsledky s kombináciou viacerých rozšírení na báze GAN, zvolili sme kombináciu BAGAN, GAN a DCGAN pre triedu explózií a WGAN, DAGAN a DCGAN pre triedu zemetrasení, ktoré dosiahli najvyšší počet správne klasifikovaných príkladov. Táto kombinácia dosiahla zlepšenie na 94.10%. Podrobnejšie výsledky tohto modelu sú uvedené v tabuľke 3. Dôležité je spomenúť, že keďže tento dataset neobsahuje obrázky, nepoužili sme klasickú augmentáciu.

Tabuľka 3
Vyhodnotenie najlepšieho modelu pre dátovú množinu SVM Data

Trieda	Presnosť	Návratnosť	F1-skóre	Počet dát
0	0.95188	0.97130	0.96145	4982
1	0.90416	0.84630	0.87427	1594
Úspešnosť			0.94100	6576
Vážený priemer	0.94026	0.94100	0.94032	6576

VII. ZÁVER

Vyhodnotenie všetkých spomenutých experimentov potvrdilo potenciál augmentácie na báze GAN sietí, keďže experimenty s využitím dátových množín obsahujúcich obrázky augmentované na báze GAN sietí dokázali dosiahnuť najvyššiu úspešnosť klasifikácie konvolučnej neurónovej siete spomedzi jednotlivých možností pre všetky dátové množiny s výnimkou jedného prípadu. Dátová množina Galaxy Mergers dosiahla najlepší výsledok klasifikácie pri použití kombinácie techniky klasickej augmentácie a augmentácie na báze GAN siete. Na základe vyhodnotenia je možné konštatovať výrazný prínos augmentácie na báze GAN, ktorý môže byť v budúcnosti užitočný pri výbere rozšírenia údajov pre podobné súbory.

POĎAKOVANIE

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0685/21.

LITERATÚRA

[1] Albawi, S., Abed Mohammed, T. and ALZAWI, S. (2017). Understanding of a convolutional neural network.
[2] O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks, arXiv preprint arXiv:1511.08458.
[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio: Generative Adversarial Networks. Journal **2**(5), 99–110 (2016)
[4] Radford, A., Metz, L. and Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks.
[5] Antoniou, A., Storkey, A. and Edwards, H. (2018). Data augmentation generative adversarial networks.
[6] Lu, C.-Y., Rustia, D. J. and Lin, T.-T. (2019). Generative adversarial network based image augmentation for insect pest classification enhancement, 52: 1–5.
[7] Arjovsky, M., Chintala, S. and Bottou, L. (2017). Wasserstein gan.
[8] Mariani, G., Scheidegger, F., Istrate, R., Bekas, C. and Malossi, C. (2018). Bagan: Data augmentation with balancing gan.

Využitie evolučných a heuristických princípov pri trénovaní konvolučných neurónových sietí

[1]Lenka Kališková, [2]Peter Butka

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

[1]lenka.kaliskova@tuke.sk, [2]peter.butka@tuke.sk

***Abstrakt* – Optimalizácia konvolučnej neurónovej siete je označovaná za jednu z najťažších úloh v oblasti hlbokého učenia, ktorá však dokáže výrazne ovplyvniť výkonnosť siete. Optimalizáciou modelu sme schopní nájsť najlepšie možné hyperparametre a tak dosiahnuť najlepší možný výkon. V tejto práci sa zameriavame na evolučné a metaheuristickej algoritmy. Použitím týchto algoritmov sa snažíme zlepšiť klasifikáciu konvolučnej neurónovej siete pre dátové množiny z astronomickej oblasti a geologickej oblasti. Pri vyhodnotení algoritmov sme sa zamerali predovšetkým na úspešnosť a čas optimalizácie.**

***Kľúčové slová* – optimalizácia, konvolučná neurónová sieť, hyperparameter, evolučné algoritmy, metaheuristické algoritmy**

I. Úvod

V súčasnosti je hlboké učenie považované za populárnu a rozšírenú oblasť patriaciu pod strojové učenie, ktoré sa zameriava na napodobňovanie ľudského spôsobu získavania vedomostí. Každý model strojového učenia má široký záber, no pre návrh a vytvorenie čo najlepšieho modelu je dôležité definovať určité vlastnosti modelu, ktoré sa nazývajú hyperparametre. Medzi dva najbežnejšie spôsoby nastavenia hyperparametrov modelu patrí manuálne nastavenie a použitie optimalizačných algoritmov. Manuálne nastavenie sa nepovažuje za optimálnu metódu na optimalizáciu hyperparametrov, pretože vyhodnotenie všetkých možných kombinácií, ktoré môžu existovať môže trvať týždne, prípadne mesiace. To je dôvod, prečo sú primárne na optimalizáciu využívané algoritmy. Medzi základne algoritmy na optimalizáciu patria Grid Search a Random Search. V tomto článku sa ale nezameriavame na tieto typy algoritmov, ale na algoritmy, ktoré sú definované ako evolučné a metaheuristické algoritmy a mali by poskytovať lepší spôsob hľadania hyperparmetrov ako spomínané algoritmy.

II. Dátové množiny

Experimenty v tejto práci sú realizované na troch dátových množinách z astro/geo domény, pričom všetky dátové množiny pozostávajú z obrázkov.

A. Dátová množina Galaxy Mergers

Dátová množina Galaxy Mergers obsahuje farebné snímky galaxií, ktoré pochádzajú z dát Sloan Digital Sky Survey Data Release 7[1]. Samotné snímky galaxií sú rozdelené do dvoch tried, ktoré sú vytvorené z interagujúcich galaxií (angl. **merger**) a neinteragujúcich galaxií (angl. **non-interacting**), pričom na určenie tried jednotlivých snímok autori dátovej množiny využili platformu Galaxy Zoo. Počet snímok v dátovej množine je 12 001, pričom z daného počtu patrí 4 501 snímok do triedy merger a 7 500 snímok do tried non-interacting.

[1]https://github.com/SpaceML/merger$_{t}ransfer_{l}earning$

B. Dátová množina FIRST Data

Dátová množina FIRST Data obsahuje archívne snímky rádiových galaxií, ktoré pochádzajú z astronomického prieskumu severnej hemisféry Faint Images of the Radio Sky at Twenty-Centimeters (skr. **FIRST**)[2]. Rádiové galaxie sú zvyčajné označované ako kompaktné alebo rozšírené. Snímky v dátovej množine sú rozdelené do štyroch tried, ktoré sú Fanaroff-Riley Class I (skr. **FRI**), Fanaroff-Riley Class II (skr. **FRII**), Bent-Tailed (skr. **BENT**) a Compact (skr. **COMPT**), pričom triedy FRI, FRII a BENT sú súčasťou rozšírených rádiových galaxií a trieda COMPT je súčasťou kompaktných rádiových galaxií. Celkový počet snímok v dátovej množine je 528, pričom rozdelenie do jednotlivých tried je 125 snímok pre triedu FRI, 216 snímok pre triedu FRII, 104 snímok pre triedu BENT a 83 snímok pre triedu COMPT.

C. Dátová množina explozií a zemetrasení

Dátová množina obsahuje dáta, ktoré nadobúdajú podobu spektogramov, ktoré pochádzajú zo seizmických meraní viacerých vysokofrekvenčných staníc nachádzajúcich sa vo Fínsku [3]. Celkový počet súborov v dátovej množine je 54, pričom rozdelenie triedy zemetrasenia a triedy explózií je rovnomerné, teda 27 súborov pre obe triedy.

III. Konvolučné neurónové siete

Konvolučná neurónová sieť (angl. **Convolutional neural network**, skr. **CNN**) je typ neurónovej siete navrhnutej pre spracovanie dát, ktoré nadobúdajú podobu obrázkov. Keďže konvolučné neurónové siete sú schopné rozpoznať vzory vstupných obrázkov na vysokej úrovni, ich využitie je rozšírené v odvetví počítačového videnia (angl. **computer vision**), a pri úlohách typu klasifikácie obrázkov, detekcie objektov, prípadne segmentácie obrázkov. Konvolučná neurónová sieť sa skladá z troch základných vrstiev, a to konvolučnej vrstvy (angl. **convolution layer**), vzorkovacej vrstvy (angl. **pooling layer**) a plne prepojenej vrstvy (angl. **fully connected layer**) [2].

Konvolučná vrstva je základnou vrstvou konvolučnej neurónovej siete, pričom na tejto vrstve prebieha extrakcia príznakov. Extrakcia príznakov je zvyčajne tvorená kombináciou konvolúcie a aktivačnej funkcie.

Konvolúcia je lineárna operácia, pri ktorej je číselne pole nazývané filter aplikované na vstup za účelom vytvorenia aktivačnej mapy (angl. **activation map, feature map**). Daný filter je aplikovaný na zvolenú časť obrázku, a je vypočítaný skalárny súčin vstupných pixelov obrázku a filtra. Skalárny súčin je uložený v aktivačnej mape. Následne prebehne aplikácia kroku posunutia (angl. **stride**) na filter a celý proces sa opakuje pokiaľ filter nespracuje celý obrázok. Výsledným produktom konvolúcie je vyplnená aktivačná mapa.

Vzorkovacia vrstva (angl. **pooling layer**) sa zameriava na redukciu počtu dimenzií aktivačnej mapy, čo vedie k zníženiu počtu parametrov, a teda vo výsledku znižuje aj výpočtovú náročnosť. Vzorkovacia vrstva zvyčajne ponúka na výber dva typy vzorkovania, a to maximálne vzorkovanie (angl. **max pooling**) a priemerné vzorkovanie (angl. **average pooling**) [2].

Úlohou plne prepojenej vrstvy je poskytnúť konečný výstup neurónovej siete. Daný výstup môže mať napríklad pri klasifikačných úlohách formát pravdepodobnosti zaradenia príkladu do jednotlivých tried. Plne prepojená vrstva má definovaný taký počet uzlov (angl. **nodes**), ktorý je rovný počtu tried. Vstupom pre plne prepojenú vrstvu je zvyčajne výstup z konvolučnej alebo vzorkovacej vrstvy, ktorý prešiel procesom vyhladzovania (angl. **flatten**) a je transformovaný na jedno dimenzionálne pole čísel [2].

IV. Optimalizácia hyperparametrov

Optimalizácia hyperparametrov je jednou z najťažších úloh v oblasti hlbokého učenia a strojového učenia. Hlavným cieľom optimalizácie je nájsť takú kombináciu hyperparametrov, ktoré zaručia najlepší výkon modelu. Pre optimalizáciu je potrebné definovať vyhľadávací priestor, ktorým je n-dimenzionálny priestor, kde každá n-dimenzia predstavuje jeden hyperparameter a rozsah dimenzie hodnoty, ktoré môže hyperparameter nadobudnúť. Hlavné hyperparametre, ktoré môžu byť súčasťou optimalizačného procesu, sú napríklad počet epoch, veľkosť vzorky, rýchlosť učenia, aktivačná funkcia, dropout alebo optimalizačný algoritmus. V tejto práci sme na optimalizáciu použili päť algoritmov, ktoré je možné definovať ako evolučné a metaheuristické algoritmy, pričom taktiež bol použitý aj algoritmus Random Search na porovnanie s týmito algoritmami.

[2]https://sundog.stsci.edu/index.html
[3]Dáta poskytnuté autormi článku [1]

A. Random Search

Random Search definuje náhodne kombinácie hyperparametrov. To znamená, že je potrebné definovať maximálny počet kombinácií pre algoritmus. Počet je zvyčajne založený na čase, ktorý je k dispozícii na optimalizáciu. Dôležité je spomenúť, že keďže všetky kombinácie sú náhodné, algoritmus pravdepodobne nedokáže nájsť lepšiu kombináciu ako Grid Search, ktorý prechádza všetkými možnými kombináciami [3].

B. Evolution Strategy

Evolution strategy (skr. **ES**) je optimalizačná metóda, ktorá je súčasťou podmnožiny evolučných algoritmov. Evolučné algoritmy sú populačne orientované optimalizačné algoritmy inšpirované a navrhnuté na základe princípu prirodzeného výberu (angl. **natural selection**) evolučnej teórie. Prirodzený výber je založený na zvolení tých jedincov v populácií, ktorých vlastnosti garantujú prežitie počas generácií a presun dobrých vlastnosti na ďalšiu generáciu [4].

C. Population Based Training

Metóda Population Based Training (skr. **PBT**) je kombináciou algoritmu Random Search a manuálnej optimalizácie. Táto metóda inicializuje súčasne viacero modelov s náhodnými hyperparametrami. Namiesto toho, aby sa trénovali modely nezávisle, sú schopné medzi sebou zdieľať informácie tak, aby sa výpočtové zdroje zamerali na vykonnejšie modely [5].

D. Particle Swarm Optimization

Particle Swarm Optimization (skr. **PSO**) je populačný metaheuristický algoritmus, ktorý je schopný jednoducho, ale efektívne vyhľadávať v masívnom vyhľadávacom priestore. Dôležité je však spomenúť, že nezaručuje nájdenie globálneho optima ako riešenia. Riešenie, ktoré algoritmus nájde, bude pravdepodobne blízko globálneho optima, ale nemalo by to byť globálne optimum [6].

E. Simulated Annealing

Simulated Annealing (skr. **SA**) je metaheuristický algoritmus vhodný pre diskrétny vyhľadávací priestor s vyšším počtom hyperparametrov, kde je schopný nájsť riešenie, ktoré sa blíži k optimálnemu riešeniu. Tento algoritmus je inšpirovaný procesom žíhania využívaným v metalurgii, kedy sa atómy v materiáli zahrievajú na vysokú teplotu, čo umožňuje ich pohyb a možnosť nájsť si vhodnejšiu polohu. Potom sa materiál pomaly ochladí, čo umožňuje atómom zostať v nových polohách.

F. Parallel Tempering

Parallel Tempering (skr. **PT**) je optimalizačný algoritmus, ktorý môže byť považovaný za variáciu už spomenutého Simulated Annealing, ale na rozdiel od tohto algoritmu, Parallel Tempering nevyužíva proces pomalého chladenia teploty. Pri použití Parallel Tempering sa vytvorí viacero kópií systému, ktoré sa nazývajú repliky, pričom tieto repliky sú náhodne definované na rozdielnych teplotách. Následné je proces algoritmu totožný s algoritmom Simulated Annealing s tým rozdielom, že na základe špecifického protokolu môžu byť dve repliky systému vymenené medzi dvoma teplotami. Daná výmena pomáha riešiť situáciu, keď prehľadávanie uviazne v lokálnom optime [7].

V. Experimenty

Experimenty v tejto práci boli zamerané na optimalizáciu konvolučnej neurónovej siete na klasifikáciu troch súborov údajov. Prvým dôležitým krokom bolo navrhnúť architektúru konvolučnej neurónovej siete, vybrať hyperparameter pre optimalizáciu, a tiež manuálne vybrať hodnoty hyperparametra pre základný model.

Hyperparametre konvolučnej neurónovej siete s hodnotami, ktoré môžu nadobúdať pri optimalizácii, sú uvedené v tabuľke 1. Základné hyperparametre konvolučnej neurónovej siete sú v tabuľke 2.

Algoritmy použité na optimalizáciu boli vybrané z Python knižníc Hyperactive [4] a Parameter-Sherpa [5]. Všetky experimenty sa uskutočňovali v ekvivalentných podmienkach. V rámci experimentov sme vyskúšali aj kombináciu optimalizácie a augmentácie, aby sme dosiahli ešte lepšie výsledky.

[4]https://pypi.org/project/hyperactive/
[5]https://parameter-sherpa.readthedocs.io/en/latest/

Tabuľka 1
Hodnoty hyperparametrov pre optimalizáciu

conv2d_1, **conv2d_2**	[16, 32, 64, 128, 256]
conv2d_3, **conv2d_4**	[16, 32, 64, 128, 256]
dropout_1	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
layer_1	[16, 32, 64, 128, 256]
activation_1	['relu', 'tanh']
optimizer	['adam', 'adagrad', 'sgd', 'rmsprop']
batch_size	[8, 16, 32, 64]
epochs	[10, 20, 30, 40]

Tabuľka 2
Hodnoty hyperparametrov pre základný model

conv2d_1, **conv2d_2**	32
conv2d_3, **conv2d_4**	16
dropout_1	0.6
layer_1	16
activation_1	'relu'
optimizer	'adam'
batch_size	8
epochs	20

Pretože všetky algoritmy majú samé viacero hyperparametrov, pred optimalizáciou sme potrebovali nastaviť hodnoty pre tieto hyperparametre. Pre každý algoritmus boli zvolené hodnoty, ktoré sú v dokumentácií zadefinované ako optimálne.

Pre algoritmus Evolution Strategy sme definovali veľkosť populácie, pravdepodobnosť mutácie, pravdepodobnosť kríženia jedinca s najlepším jedincom v populácií a pravdepodobnosť skoku na náhodnú pozíciu v iterácií. Zvolené hodnoty hyperparametrov pre algoritmus Evolution Strategy sú uvedené v tabuľke 3.

Pre algoritmus Population Based Training sme definovali veľkosť populácie a počet generácií. Zvolené hodnoty týchto hyperparametrov pre algoritmus Population Based Training sú uvedené v tabuľke 4.

Tabuľka 3
Hodnoty parametrov algoritmu Evolution Strategy

veľkosť populácie	10
pravdepodobnosť mutácie	0.7
pravdepodobnosť kríženia	0.3
náhodna pozícia	0

Tabuľka 4
Hodnoty parametrov algoritmu Population Based Training

veľkosť populácie	10
počet generácií	10

Pre algoritmus Particle Swarm Optimization bolo potrebné definovať veľkosť populácie, zotrvačnosť pohybu jednotlivých jedincov v populácií, kognitívnu váhu definujúcu pohyb smerom k lokálne najlepšej pozícií jedinca v populácií, sociálnu váhu definujúcu pohyb smerom ku globálne najlepšej pozícií jedinca v populácií a pravdepodobnosť skoku na náhodnú pozíciu. Zvolené hodnoty parametrov pre algoritmus Particle Swarm Optimization sú uvedené v tabuľke 5.

Pre algoritmus Simulated Annealing bolo potrebné definovať hodnotu epsilon definujúcu vzdialenosť skoku z aktuálnej pozície na novú pozíciu, matematickú distribúciu dát, počet susedov preskúmaných z aktuálnej pozície pred skokom na najlepšiu pozíciu, pravdepodobnosť skoku na náhodnú pozíciu, hodnotu žíhania definujúcu začiatok znižovania teploty žíhania a počiatočnú teplotu, ktorá definuje pravdepodobnosť skoku na horšiu pozíciu. Zvolené hodnoty parametrov pre algoritmus Simulated Annealing sú uvedené v tabuľke 6.

Tabuľka 5
Hodnoty parametrov algoritmu Particle Swarm Optimization

veľkosť populácie	10
zotrvačnosť	0.4
kognitívná váha	0.7
sociálná váha	0.7
náhodná pozícia	0

Tabuľka 6
Hodnoty parametrov algoritmu Simulated Annealing

epsilon	0.05
distribúcia	'normal'
počet susedov	10
náhodná pozícia	0
hodnota žíhania	0.975
počiatočná teplota	1

Pre algoritmus Parallel Tempering, sme definovali hyperparametre veľkosť populácie, počet iterácií pred výmenou teploty medzi jedincami v populácii a pravdepodobnosť skoku na náhodnú pozíciu. Zvolené hodnoty parametrov pre algoritmus Parallel Tempering sú uvedené v tabuľke 7.

Tabuľka 7
Hodnoty parametrov algoritmu Parallel Tempering

veľkosť populácie	10
počet iterácií pred výmenou	10
náhodná pozícia	0

Pre algoritmus Random Search sme definovali len maximálny počet iterácií. Pri vyhodnotení sme sa zamerali predovšetkým na metriky accuracy a čas optimalizácie. Okrem toho sme použili aj precision, recall, f1-score a confusion matrix.

A. *Vyhodnotenie modelov pre dátovú množinu Galaxy Mergers*

Základný model konvolučnej neurónovej siete pre dátovú množinu Galaxy Mergers dosiahol úspešnosť klasifikácie 90.420%. Výsledky jednotlivých optimalizačných algoritmov pre dátovú množinu Galaxy Mergers sú uvedené v tabuľke 8.

Tabuľka 8
Úspešnosť a čas optimalizácie jednotlivých optimalizačných algoritmov pre dátovú množinu Galaxy Mergers

Optimalizačný algoritmus	**Úspešnosť**	**Čas v minútach**
Evolution Strategy	0.92621	45.9213
Population Based Training	**0.93722**	**169.4248**
Particle Swarm Optimization	0.90645	91.1214
Simulated Annealing	0.92546	46.5547
Parallel Tempering	0.91921	58.2920
Random Search	0.91721	321.1082

Na základe metriky úspešnosti boli dva najlepšie modely vytvorené pomocou optimalizačných algoritmov Evolution Strategy a Population Based Training. Rozdiel v úspešnosti medzi týmito modelmi je 1.101% v prospech algoritmu Population Based Training. Aj napriek dlhšiemu času optimalizácie je možné za najlepší model považovať ten, ktorý bol optimalizovaný algoritmom Population Based Training. Klasifikačný report a kontingenčná tabuľka pre model optimalizovaný algoritmom Population Based Training sú uvedené v tabuľke 9, respektíve v tabuľke 10.

Tabuľka 9
Klasifikačný report Population Based Training modelu pre dátovú množinu Galaxy Mergers

	Presnosť	**Návratnosť**	**F1-skóre**	**Počet dát**
merger	0.89942	0.93725	0.91795	1 498
non-interacting	0.96143	0.93720	0.94916	2 500
Úspešnosť			0.93722	3 998
Vážený priemer	0.93820	0.93722	0.93746	3 998

Tabuľka 10
Kontingenčná tabuľka Population Based Training modelu pre dátovú množinu Galaxy Mergers

	merger	non-interacting
merger	1 404	94
non-interacting	196	2 304

Za účelom zlepšenia úspešnosti, sme vyskúšali augmentáciu dát, ktorá zahŕňala rotáciu obrázkov v troch uhloch. Po použití vytvoreného modelu na rozšírených dátach sme dosiahli presnosť 94.832%. Potom sme vyskúšali optimalizáciu na rozšírených údajov pomocou Population Based Training, čím sa nám úspešnosť zvýšila na 94.922%. Hodnoty hyperparametrov pre algoritmus Population Based Training pôvodnej a rozšírenej dátovej množiny sú uvedené v tabuľke 11.

Tabuľka 11
Hodnoty hyperparametrov získané počas optimalizácie PBT pôvodnej a rozšírenej dátovej množiny Galaxy Mergers

Hyperparametre	**Pôvodná dátová množina**	**Augmentovaná dátová množina**
conv2d_1, conv2d_2	256	32
conv2d_3, conv2d_4	16	16
dropout_1	0.9	0.6
layer_1	128	16
activation_1	'relu'	'relu'
optimizer	'adam'	'adam'
batch_size	8	32
epochs	40	30

Ako si je možné všimnúť v tabuľke 11 pri použití augmentovanej dátovej množiny nastali väčšie zmeny v hodnotách hyperparametrov, čo je možné pripísať trojnásobne väčšiemu rozdielu v počte obrázkov v jednotlivých triedach oproti pôvodnej dátovej množine.

B. *Vyhodnotenie modelov pre dátovú množinu FIRST Data*

Základný model konvolučnej neurónovej siete pre dátovú množinu FIRST Data dosiahol úspešnosť klasifikácie 80.189%. Výsledky jednotlivých optimalizačných algoritmov pre dátovú množinu FIRST Data sú uvedené v tabuľke 12.

Na základe metriky úspešnosti boli dva najlepšie modely vytvorené pomocou optimalizačných algoritmov Particle Swarm Optimization a Random Search. Rozdiel v úspešnosti medzi týmito

Tabuľka 12
Úspešnosť a čas optimalizácie jednotlivých optimalizačných algoritmov pre dátovú množinu FIRST Data

Optimalizačný algoritmus	**Úspešnosť**	**Čas v minútach**
Evolution Strategy	0.91509	6.9181
Population Based Training	0.87736	9.0352
Particle Swarm Optimization	**0.94340**	**4.0835**
Simulated Annealing	0.89623	5.3367
Parallel Tempering	0.89623	2.3280
Random Search	**0.95283**	**17.0603**

modelmi je 0.934% v prospech algoritmu Random Search. Ako najlepší model sme označili v tomto prípade model optimalizovaný algoritmom Particle Swarm Optimization, keďže algoritmus Random Search slúži primárne na porovnanie s evolučnými a metaheuristickými algoritmami. Klasifikačný report a kontingenčná tabuľka tohto modelu sú uvedené v tabuľke 13, respektíve v tabuľke 14.

Tabuľka 13
Klasifikačný report Particle Swarm Optimization modelu pre dátovú množinu FIRST Data

	Presnosť	**Návratnosť**	**F1-skóre**	**Počet dát**
BENT	0.83333	0.71429	0.76923	21
COMP	0.80952	1.00000	0.89474	17
FRI	0.92857	0.52000	0.66667	25
FRII	0.75472	0.93023	0.83333	43
Úspešnosť			0.80189	106
Vážený priemer	0.82009	0.80189	0.79117	106

Tabuľka 14
Kontingenčná tabuľka Particle Swarm Optimization modelu pre dátovú množinu FIRST Data

	BENT	COMP	FRI	FRII
BENT	15	0	1	5
COMP	0	17	0	0
FRI	0	4	13	8
FRII	3	0	0	40

Za účelom zlepšenia úspešnosti, sme vyskúšali augmentáciu dát, ktorá zahŕňala rotáciu obrázkov v troch uhloch. Po použití vytvoreného modelu na rozšírených dátach sme dosiahli presnosť 96.226%. Potom sme vyskúšali optimalizáciu na rozšírených údajov pomocou Particle Swarm Optimization, čím sa nám úspešnosť zvýšila na 97.170%. Hodnoty hyperparametrov pre algoritmus Particle Swarm Optimization pôvodnej a rozšírenej dátovej množiny sú uvedené v tabuľke 15.

Tabuľka 15
Hodnoty hyperparametrov získané počas optimalizácie PSO pre pôvodnú a rozšírenú dátovú množinu FIRST Data

Hyperparametre	**Pôvodná dátová množina**	**Augmentovaná dátová množina**
conv2d_1, conv2d_2	64	64
conv2d_3, conv2d_4	64	64
dropout_1	0.6	0.5
layer_1	64	64
activation_1	'tanh'	'relu'
optimizer	'adam'	'adam'
batch_size	8	16
epochs	20	20

Ako si je možné všimnúť v tabuľke 15 pri použití augmentovanej dátovej množiny oproti pôvodnej dátovej množine nenastala výrazna zmena v hodnotách hyperparametrov, čo je možné pripísať menšiemu rozdielu v počte obrázkov v jednotlivých triedach.

C. Vyhodnotenie modelov pre dátovú množinu explozií a zemetrasení

Základný model konvolučnej neurónovej siete pre dátovú množinu explozií a zemetrasení dosiahol úspešnosť klasifikácie 91.265%. Výsledky jednotlivých optimalizačných algoritmov pre dátovú množinu explozií a zemetrasení sú uvedené v tabuľke 16.

Na základe metriky úspešnosti bol najlepší model vytvorený použitím optimalizačného algoritmu Particle Swarm Optimization, ktorého úspešnosť dosiahla 93.355%. Klasifikačný report a kontingenčná tabuľka pre daný model sú uvedené v tabuľke 17, respektíve v tabuľke 18.

Tabuľka 16
Úspešnosť a čas optimalizácie jednotlivých optimalizačných algoritmov pre dátovú množinu explozií a zemetrasení

Optimalizačný algoritmus	**Úspešnosť**	**Čas v minútach**
Evolution Strategy	0.92473	162.0150
Population Based Training	0.91363	217.2810
Particle Swarm Optimization	**0.93355**	**73.9222**
Simulated Annealing	0.92473	46.4707
Parallel Tempering	0.91986	68.0210
Random Search	0.92746	401.0599

Tabuľka 17
Klasifikačný report Particle Swarm Optimization modelu pre dátovú množinu explozií a zemetrasení

	Presnosť	**Návratnosť**	**F1-skóre**	**Počet dát**
explosion	0.94533	0.96829	0.95667	4 982
earthquakes	0.89274	0.82497	0.85752	1 594
Úspešnosť			0.93355	6 576
Vážený priemer	0.93258	0.93355	0.93263	6 576

Tabuľka 18
Kontingenčná tabuľka Particle Swarm Optimization modelu pre dátovú množinu explozií a zemetrasení

	explosion	earthquakes
explosion	4 824	158
earthquakes	279	1 315

Za účelom zlepšenia úspešnosti, sme vyskúšali augmentáciu dát pridaním šumu. Avšak touto augmentáciou sa nám nepodarilo zlepšiť úspešnosť. Hodnoty hyperparametrov pre algoritmus Particle Swarm Optimization pôvodnej a rozšírenej dátovej množiny sú uvedené v tabuľke 19.

Tabuľka 19
Hodnoty hyperparametrov získané počas optimalizácie PSO pre pôvodnú a rozšírenú dátovú množinu explozií a zemetrasení

Hyperparametre	**Pôvodná dátová množina**	**Augmentovaná dátová množina**
conv2d_1, conv2d_2	128	16
conv2d_3, conv2d_4	32	16
dropout_1	0.4	0.8
layer_1	16	16
activation_1	'relu'	'relu'
optimizer	'adam'	'adam'
batch_size	16	64
epochs	20	40

VI. Záver

Na základe výsledkov sme schopní preukázať rozdiely medzi optimalizačnými algoritmami v dvoch rôznych oblastiach dát, a zároveň sme dokázali výhody použitia optimalizácie na dosiahnutie lepších výsledkov modelov. Táto práca môže byť v budúcnosti využitá na rozhodnutie, ktorý algoritmus by mal byť použitý pre optimalizáciu, pretože na základe množstva uskutočnených experimentov ponúkame rôzne kombinácie, ktoré môžu byť vybrané pre rôzne dátové množiny.

Poďakovanie

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0685/21.

Literatúra

[1] Tiira, T., Kortström, J., and Uski, M., Automatic classification of seismic events within a regional seismograph network, 2015.
[2] Al-Saffar, A. A. M., Tao, H. and Talab, M. A. (2017). Review of deep convolution neural network in image classification, 2017 International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET), IEEE, pp. 26–31.
[3] Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization., Journal of machine learning research 13(2).
[4] Wierstra, D., Schaul, T., Glasmachers, T., Sun, Y. and Schmidhuber, J. (2011). Natural evolution strategies.
[5] Jaderberg, M., Dalibard, V., Osindero, S., Czarnecki, W. M., Donahue, J., Razavi, A., Vinyals, O., Green, T., Dunning, I., Simonyan, K. et al. (2017). Population based training of neural networks.
[6] Lorenzo, P. R., Nalepa, J., Kawulok, M., Ramos, L. S. and Pastor, J. R. (2017). Particle swarm optimization for hyper-parameter selection in deep neural networks, Proceedings of the genetic and evolutionary computation conference, pp. 481–488.
[7] Pushkarov, V., Efroni, J., Maksymenko, M. and Koch-Janusz, M. (2019). Training deep neural networks by optimizing over nonlocal paths in hyperparameter space.

Segmentácia koronálnych dier a aktívnych regiónov

[1] Lenka Kališková, [2]Peter Butka, [3]Adam Majirský

[1,2,3]Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]lenka.kaliskova@tuke.sk,[2] peter.butka@tuke.sk, [3]adam.majirsky@student.tuke.sk

Abstrakt **– Slnečná aktivita má vplyv na viaceré aspekty, medzi ktoré patria napríklad technológie alebo klíma. Je dôležité sledovať a identifikovať javy vyskytujúce sa na Slnku, aby sme ich mohli predvídať a lepšie pochopiť ich vplyv na nás. Táto práca sa zaoberá segmentovaním koronálnych dier a aktívnych regiónov na snímkach sondy SOHO, ktorá monitoruje Slnko od roku 1996. Na segmentáciu sme použili konvolučnú neurónovú sieť SCSS-Net. Výstupmi práce sú nové anotácie týchto dvoch javov a ich zobrazenie, ktoré môže slúžiť na ďalší výskum slnečnej aktivity.**

Kľúčové slová **– segmentácia, hlboké učenie, SCSS-Net, aktívne regióny, koronálne diery**

I. Úvod

Slnko je jedným z najdôležitejších faktorov ovplyvňujúcich život na Zemi. Jeho vplyv sa prejavuje v mnohých oblastiach napríklad klímy, rastlinnej produkcie, zdravia ľudí alebo aj pri výrobe elektrickej energie. Javy na slnku môžu však ovplyvniť aj technológie dôležité pre ľudstvo, ako napríklad satelity na obežnej dráhe Zeme, ktoré zabezpečujú telekomunikáciu a navigáciu alebo energetické siete a infraštruktúry, ktoré zabezpečujú dodávky elektrickej energie. Výpadok týchto technológii môže mať negatívny dopad na celú spoločnosť. Monitorovanie a skúmanie slnečnej aktivity je teda dôležité, aby sme sa mohli pripraviť na jej možné negatívne následky.

Algoritmy strojového a hlbokého učenia sa čoraz viac dostávajú do popredia a prinášajú pridanú hodnotu nielen výskumníkom a špecialistom na informačné technológie, ale aj bežným užívateľom produktov a služieb. Vďaka aplikovaniu matematických princípov a výpočtovej sile počítačov, sú tieto algoritmy schopné rozoznávať vzťahy a vzorce vo veľkom objeme dát, ktoré nám vedia poskytnúť dôležitý prehľad a informácie. Tieto algoritmy sa v súčasnosti využívajú v rôznych oblastiach, ako napríklad v medicíne, priemysle, marketingu alebo vede. Pri skúmaní slnečnej fyziky môžeme teda využiť tieto algoritmy na dostupných dátach. V tomto článku sme sa venovali využitiu hlbokého učenia na segmentáciu dvoch javov na slnečnej koróne. Teda ako využiť SCSS-Net model na segmentovanie koronálnych dier a aktívnych regiónov na snímkach slnečného disku zo sondy SOHO (angl. **Solar and Heliospheric Observatory**).

II. Dáta

Koronálna diera je tmavá oblasť na slnečnom disku, ktorá má tvar nepravidelného kruhu, prípadne oválu. Koronálna diera môže dosahovať niekoľkonásobnú veľkosť Zeme, pričom sa môže rozšíriť po veľkej časti slnečného disku. Ich výskyt je najčastejšie rozšírený na južnom, prípadne severnom póle, pričom majú nižšiu hustotu plazmy v porovnaní s okolitými oblasťami [1].

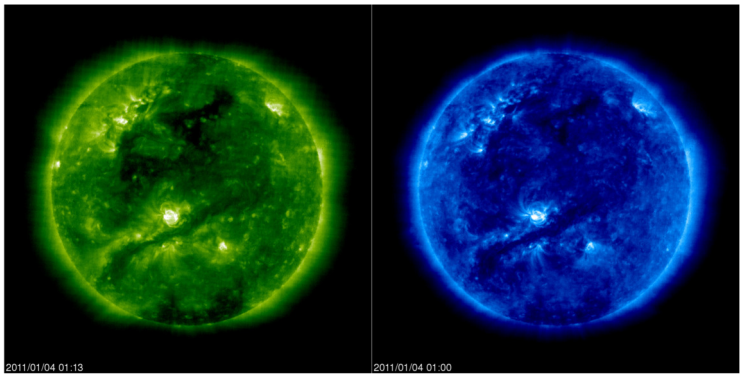
Aktívne oblasti sú dočasné oblasti, ktoré majú signifikantné silnejšie magnetické pole ako ich okolie. Počet a veľkosť týchto oblastí závisí od cyklu slnečnej aktivity. Tak ako koronálne diery, aj aktívne regióny sú schopné ovplyvniť život na Zemi. Procesy v aktívnych oblastiach môžu spôsobiť výron koronálnej hmoty, ktorý dokáže zasiahnuť Zem, a tým narušiť telekomunikačné technológie [2].

Satelit SOHO je vedecký satelit, ktorý bol vypustený v roku 1995 vďaka spolupráci agentúr European Space Agency (skr. **ESA**) a National Aeronautics and Space Administration (skr. **NASA**). Cieľom satelitu je skúmať Slnko, pričom disponuje 12 prístrojmi na pozorovanie a zachytávanie javov, ktoré sa vyskytujú nielen v slnečnej koróne, ale aj javov týkajúcich sa helioseizmológie a slnečného vetra.

Prístroj EIT na SOHO je zameraný na pozorovanie slnečnej koróny v krátkych vlnových dĺžkach elektromagnetického spektra. Je schopný vytvoriť obrazy vo vlnových dĺžkach 17.1 nm, 19.5 nm, 28.4 nm a 30.4 nm s rozlíšením 1024 x 1024 pixelov prostredníctvom CCD detektoru.

S cieľom vylepšiť segmentáciu koronálnych dier a aktívnych javov, inšpirovanú predchádzajúcimi prácami [3], [4] bol v práci trénovaný model na snímkach zo sondy SOHO od roku 1996.

V práci boli použité snímky slnečného disku a anotácie, ktoré vyznačujú pozíciu koronálnych dier na snímke disku. Pre segmentovanie koronálnych dier boli použité snímky slnečného disku s vlnovou dĺžkou 195Å, zatiaľ čo pre segmentáciu aktívnych regiónov boli použité snímky s vlnovou dĺžkou 171Å, keďže obsahujú rozličné javy s najlepšou viditeľnosťou. Obe typy snímok boli získané nástrojom EIT na satelite SOHO. Ukážky týchto dát sú zobrazené na obrázku 1.



Obrázok 1 Na ľavej strane je snímka s vlnovou dĺžkou 195Å a na pravej strane je snímka s vlnovou dĺžkou 171Å

Získanie/Spracovanie dát

Na anotácie boli použité SPoCA dáta, ktoré kategorizujú javy v slnečnej koróne a označujú ich ako koronálne diery alebo aktívne regióny. SPoCA dáta sú binárne masky s bielymi pixelmi pre miesta výskytu daných javov a čiernym pozadím [5]. Pre získanie SPoCA anotácií bol použitý nástroj Heliophysics Events Knowledgebase (skr. **HEK**) a databáza slnečných javov z pozorovaní satelitu SDO. HEK kategorizuje javy, ako sú slnečné erupcie a koronálne diery [6].

Získaných bolo 8 038 snímok s vlnovou dĺžkou 195Å vhodných pre segmentáciu koronálnych dier a 7 985 snímok s vlnovou dĺžkou 171Å vhodných pre segmentáciu aktívnych regiónov. Aj napriek faktu, že chýbali dáta z rôznych časových úsekov, bola dosiahnutá úroveň približne 85% ideálneho stavu. Počas skúmania problému neúplných dát boli identifikované poškodené alebo chýbajúce dáta v dôsledku porúch v prenose dát z technických problémov na sonde SOHO [7].

Pre snímky s vlnovou dĺžkou 195Å bolo odstránené čierne pozadie a nahrádené bielym, aby sa predišlo označeniu čierneho pozadia ako koronálnej diery počas segmentácie. Na zjednodušenie spracovania bola farebná škálu snímok zmenená na čiernobielu a rozmer bol upravený na 256 x 256 pixelov.

V rámci anotácií bolo získaných 3 278 anotácií pre koronálne diery a 3 277 pre aktívne regióny v časovom rozpätí od roku 2011 do roku 2021. Tieto anotácie boli upravené na obrázky s rozmermi 1024 x 1024 pixelov, ktoré reprezentovali javy na korešpondujúcich snímkach slnečného disku. Keďže snímky slnečného disku boli upravené na 256 x 256 pixelov, aj anotácie boli prispôbené na totožný rozmer.

III. Konvolučné neurónové siete

Konvolučná neurónová sieť (skr. **CNN**) je považovaná za jeden z najpoužívanejších typov neurónových sietí s primárnym zameraním na detekciu a klasifikáciu obrazov. Tento typ neurónovej siete pozostáva z troch hlavných typov vrstiev, ktorými sú konvolučná vrstva (angl. **convolutional layer**), vzorkovacia vrstva (angl. **pooling layer**), aktivačná vrstva (angl. **activating function layer**) a plne prepojená vrstva (angl. **fully connected layer**) [8]. Ako už naznačuje názov, hlavným stavebným blokom CNN je konvolučná vrstva, ktorá vykonáva operáciu nazývanú konvolúcia. Konvolúcia je lineárna operácia, ktorá zahŕňa násobenie váh so vstupom. Ak si predstavíme dvojrozmerný vstup, násobenie prebieha medzi maticou vstupných dát a dvojrozmernou maticou váh, ktorá sa nazýva filter alebo jadro. Tento proces násobenia sa opakuje, až kým jadro nepokryje celý obraz. Výstupom z konvolúcie je mapa príznakov [9]. Konvolúcia ako súčasť neurónovej siete je definovaná rovnicou (1).

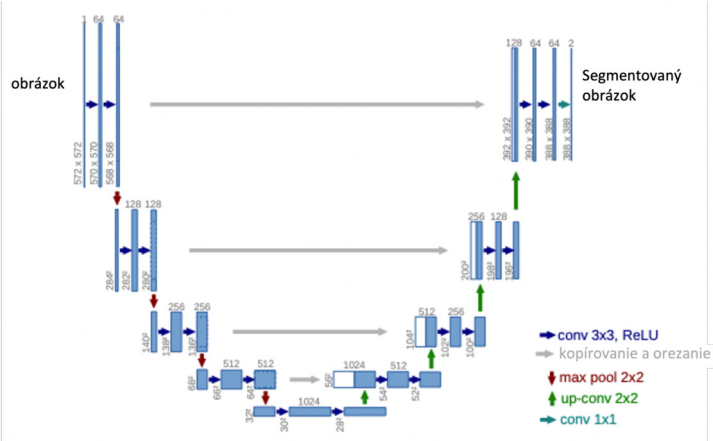
$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n) K(i-m, j-n), \tag{1}$$

kde I označuje vstupný obrázok, K označuje kernel, m a n sú indexy riadkov a stĺpcov kernela, i a j sú indexy stĺpcov a riadkov výslednej matice, ktorá predstavuje jadro konvolúcie [8].

Vzorkovacia vrstva (angl. **pooling layer**) sa zameriava na redukciu počtu dimenzií aktivačnej mapy, čo vedie k zníženiu počtu parametrov, a teda vo výsledku znižuje aj výpočtovú náročnosť. Vzorkovacia vrstva zvyčajne ponúka na výber dva typy vzorkovania, a to maximálne vzorkovanie (angl. **max pooling**) a priemerné vzorkovanie (angl. **average pooling**).

A. U-Net

U-Net je konkrétna architektúra konvolučnej neurónovej siete, ktorá bola navrhnutá v roku 2015 [10]. Architektúra U-Net, ktorá je zobrazená na obrázku 2 sa skladá z eknódera a dekódera. Enkóder využíva konvolučnú vrstvu s veľkosťou filtra 3 x 3 pixelov, za ktorou nasleduje aktivačná funkcia ReLU. Túto postupnosť opakuje dvakrát než rozmery obrázka zmenší pomocou vzorkovacej vrstvy o veľkosti 2 x 2 pixelov. Dekodér tento postup opakuje ale reverzne, čiže obrázok postupne zväčšuje. Enkóder a dekóder sú navzájom prepojení, aby sa príznaky z jednej veľkosti obrázka preniesli priamo do inej škály v dekódovacom procese. Na konci architektúry U-net je použitá konvolučná vrstva s veľkosťou 1 x 1 pixel, ktorá má na starosti segmentáciu.



Obrázok 2 Architektúra U-Net

B. SCSS-Net

Model SCSS-Net, predstavený v práci [3], ktorý je inšpirovaný architektúrou konvolučnej siete U-net. SCSS-Net sa skladá z enkodéra, ktorý zachytáva príznaky na snímkach a dekódera, ktorý podľa týchto príznakov klasifikuje jednotlivé pixely, či sú súčasťou daného javu alebo nie. Architektúra SCSS-Netu sa začína vstupom spracovanej snímky do enkodéra.

Enkodér je zložený z piatich konvolučných blokov. Jeden konvolučný blok je zložený z dvoch postupnosti, kde jedna postupnosť je zložená z konvolučnej vrstvy o veľkosti 3 x 3 pixelov, za konvolučnou vrstvou nasleduje batch normalizácia a za ňou ReLU aktivačná funkcia.

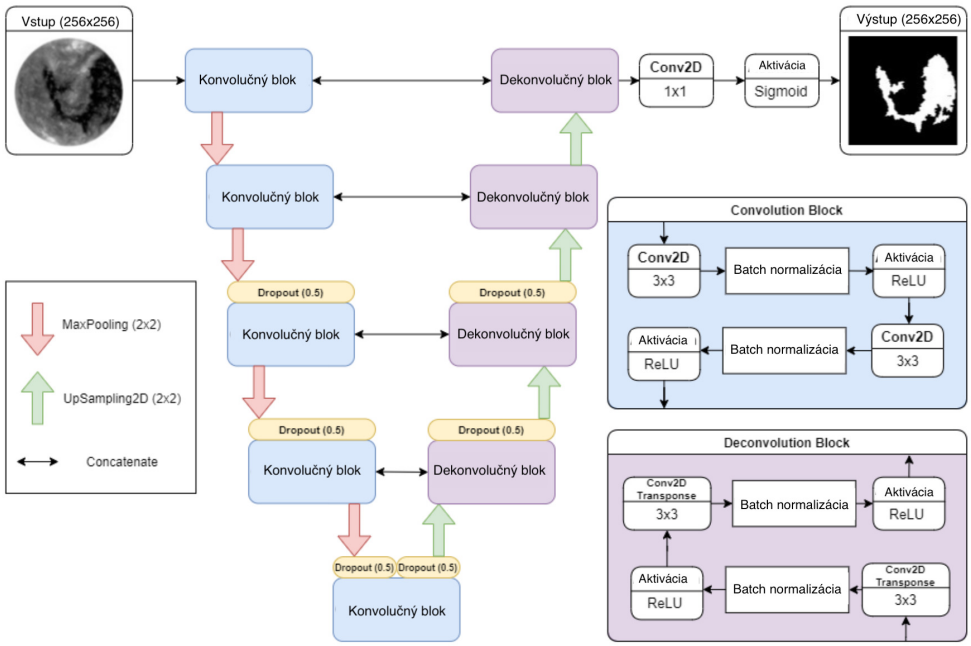
Medzi konvolučnými blokmi sa nachádza Max-pooling vrstva a v posledných troch konvolučných blokoch je využitá regulačná technika dropout. Po týchto piatich blokoch nasleduje dekóder.

Dekóder je zložený zo štyroch dekonvolučných blokov. Tento blok je takisto zložený z dvoch sérií, kde jedna séria obsahuje transponovanú konvolúciu, za ktorou nasleduje batch normalizácia a ReLU aktivačná funkcia.

Medzi dekonvolučnými blokmi sa využíva UpSampling vrstva a v prvých dvoch blokoch je použitý Dropout. Štyri konvolučné a dekonvolučné bloky sú prepojené funkciou concatenate, ktorá prepojí snímku s príznakmi. Architektúra sa končí konvolučnou vrstvou s veľkosťou 1 x 1 pixelov. Táto vrstva klasifikuje pixely pomocou aktivačnej funkcie Sigmoid a určuje pravdepodobnosť toho, či patria do danej triedy. Architektúra SCSS-Netu je zobrazená na obrázku 3.

IV. MODELOVANIE

Po získaní a spracovaní boli dáta rozdelené na trénovaciu, validačnú a testovaciu množinu. Pre dosiahnutie vyváženej množiny jasných a menej jasných snímok boli snímky náhodne rozdelené pre každý rok od roku 2011 do roku 2021 na trénovaciu, testovaciu a validačnú množinu v pomere 60:20:20.



Obrázok 3 Architektúra SCSS-Netu [3]

Na segmentovanie javov bol zvolený model SCSS-Net, ktorý bol vytvorený v práci [3]. Na segmentáciu boli použité tri postupy, a to využitie starých váh, prenesené učenie a výpočet nových váh.

Pri druhom a treťom postupe, teda pri doladení existujúceho modelu alebo trénovaní nového modelu na dátach SOHO, bolo potrebné zvoliť vhodné hyperparametre pre trénovanie neurónovej siete. Preto bolo vhodné použiť skúsenosti a poznatky z predchádzajúcej práce [3], a použiť ich výsledky ladenia hyperparametrov aj pre našu prácu.

1) Postupy pri vytváraní segmentácií: K dispozícii sme mali viacero postupov, ako vytvoriť segmentácie:

- Prvým postupom bolo použiť spracované SOHO dáta a využiť model, ktorý bol natrénovaný na SDO dátach [3]. Tento postup by mohol byť najrýchlejší, pretože priamo segmentujeme a netrénujeme žiadny model. Avšak jeho nevýhoda spočíva v nepresnosti, keďže SOHO a SDO dáta boli odlišné.
- Druhý postup predstavoval doladenie váh starého modelu SCSS-Net pomocou našej trénovacej množiny, inak nazvané aj prenesené učenie (angl. **transfer learning**). V tomto prípade by to znamenalo, že model natrénovaný na SDO dátach sa dotrénuje spolu s našimi spracovanými SOHO dátami. Tento prístup by mal byť presnejší ako prvý postup, avšak aj časovo náročnejší.
- Tretím a posledným postupom bolo za pomoci pripravených SOHO dát natrénovať model úplne odznova, teda začať s náhodnými váhami pre neurónovú sieť SCSS-Net.

Aby sa tieto postupy dali ľahko rozlíšiť, zaviedli sme pre nich konkrétne názvy:

- Využitie starých váh.
- Prenesené učenie.
- Výpočet nových váh.

Pri druhom a treťom postupe, teda pri doladení existujúceho modelu alebo trénovaní nového modelu na dátach SOHO, bolo potrebné zvoliť vhodné hyperparametre pre trénovanie neurónovej siete. Hyperparametre sú premenné modelu, ktoré sa neoptimalizujú počas trénovania, ale sú zvolené ešte pred trénovaním a ovplyvňujú kvalitatívny výsledok modelu. Proces hľadania najvhodnejších premenných modelu sa nazýva optimalizácia hyperparametrov a pozostáva zo skúsených odhadov a experimentov s rôznymi kombináciami týchto premenných. Tento proces však môže byť časovo náročný, pretože sa každý model s rôznymi hyperparametrami musí natrénovať a vyhodnotiť samostatne. Preto bolo pre nás vhodné použiť skúsenosti a poznatky z predchádzajúcej práce [3], a použiť ich výsledky ladenia hyperparametrov aj pre našu prácu.

V. Vyhodnotenie

Každý postup bol vyhodnotený na testovacej množine, pričom hlavnými metrikami boli Dice a IoU, ktoré sa využívali pri výpočte presnosti segmentácie. Pri prenesenom učení a výpočte nových

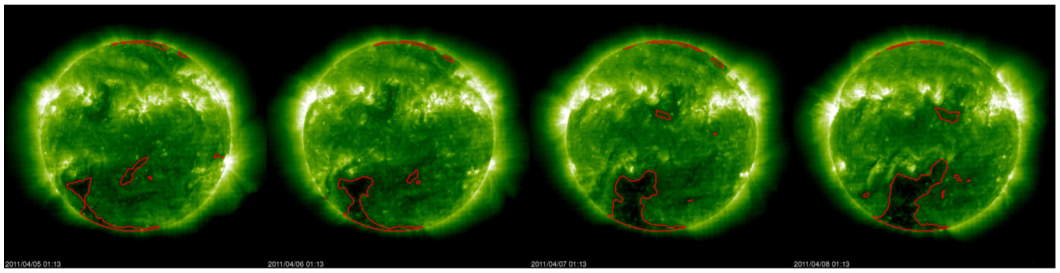
váh, kde sa trénoval model, boli zisťované metriky Dice, IoU a hodnota chybovej funkcie na validačnej množine. Vyhodnotenie na testovacej množine je uvedené v tabuľke 1 pre segmentovanie koronálnych dier a v tabuľke 2 pre segmentovanie aktívnych regiónov. Výsledok segmentácie koronálnych dier a aktívnych oblastí s využitím SCSS-netu je zobrazený na obrázku 4, respektíve 5.

Tabuľka 1
Vyhodnotenie segmentácie koronálnych dier na testovacej množine

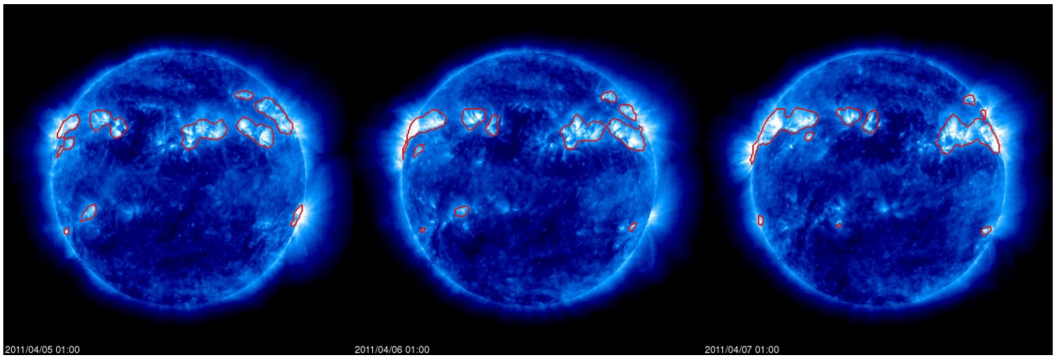
Typ učenia	Dice	IoU
Využitie starých váh	0.44	0.29
Prenesené učenie	0.78	0.64
Výpočet nových váh	0.80	0.69

Tabuľka 2
Vyhodnotenie segmentácie aktívnych regiónov na testovacej množine

Typ učenia	Dice	IoU
Využitie starých váh	0.50	0.33
Prenesené učenie	0.61	0.44
Výpočet nových váh	0.62	0.45



Obrázok 4 Segmentácie koronálnych dier vytvorené SCSS-Netom pre dni 5.4.2011 - 8.4.2011



Obrázok 5 Segmentácie aktívnych regiónov vytvorené SCSS-Netom pre dni 5.4.2011 - 7.4.2011

V porovnaní s predchádzajúcimi prácami dosiahol SCSS-Net pri segmentácii koronálnych dier mierne nižšie hodnoty metrík o 10% v prípade porovnania s Region Growth anotáciami a vyššie hodnoty metrík o 6% v porovnaní so SPoCA anotáciami. V prípade segmentácie aktívnych regiónov dosiahol zvolený model hodnotu Dice 0.61 a IoU 0.44, čo sú nižšie hodnoty o 7%, ako v predchádzajúcich prácach. Metriky, kde sa používali SPoCA anotácie, mohli klesnúť kvôli ich limitáciám, čo viedlo k väčšiemu výskytu falošne pozitívnych pixelov, čím mohlo nastať znižovanie metrík pre zlú anotáciu, aj keď segmentácia mohla byť správna.

VI. Záver

V tejto práci sme nadviazali na prácu [3] a využili sme model SCSS-Net na segmentovanie koronálnych dier a aktívnych regiónov na snímkach zo sondy SOHO. Získali sme potrebné dáta a spracovali ich v požadovanej kvalite, pre vstup do SCSS-Net modelu. Zistili sme najvhodnejší

postup pre trénovanie modelu na segmentáciu koronálnych dier a aktívnych regiónov na SOHO snímkach a segmentovali sme tieto javy v časovom rozmedzí rokov 1996 až 2021.

Týmto sme získali nové anotácie pre tieto dva javy, ktoré môžu slúžiť ako alternatívna voľba spomedzi existujúcich anotácií, ktoré majú svoje limitácie. Avšak, naše anotácie koronálnych dier z rokov 1996 až 2011 na snímkach SOHO sú prvé a jediné, ktoré existujú pre tento časový úsek. Týmto sme rozšírili dostupné anotácie o ďalších 15 rokov, ktoré môžu byť cenné pre výskumníkov pri získavaní nových vedomostí v tejto problematike.

Poďakovanie

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0685/21.

Literatúra

[1] Cranmer, Steven R. (2009). "Coronal holes", Living Reviews in Solar Physics, 2009, Springer, Volume 6, pp. 1-66.
[2] van Driel-Gesztelyi, Lidia and Green, Lucie May (2015). "Evolution of active regions", Living Reviews in Solar Physics, vol. 12, pp. 1-98, 2015.
[3] Mackovjak, Š., Harman, M., Maslej-Krešňáková, V., and Butka, P., "SCSS-Net: solar corona structures segmentation by deep learning", Monthly Notices of the Royal Astronomical Society, vol. 508, no. 3, OUP, pp. 3111–3124, 2021. doi:10.1093/mnras/stab2536.
[4] Illarionov, E. A. and Tlatov, A. G., "Segmentation of coronal holes in solar disc images with a convolutional neural network", Monthly Notices of the Royal Astronomical Society, vol. 481, no. 4, OUP, pp. 5014–5021, 2018. doi:10.1093/mnras/sty2628.
[5] Verbeeck, C. and Delouille, V. and Mampaey, B. and De Visscher, R., "The SPoCA-suite: Software for extraction, characterization, and tracking of active regions and coronal holes on EUV images", Astronomy & Astrophysics, vol. 561, 2014.
[6] Hurlburt, N., "Heliophysics Event Knowledgebase for the Solar Dynamics Observatory (SDO) and Beyond", Solar Physics, vol. 275, no. 1–2, pp. 67–78, 2012. doi:10.1007/s11207-010-9624-2.
[7] Vandenbussche, F. C., "Saving SOHO", in SOHO-17. 10 Years of SOHO and Beyond, 2006, vol. 617.
[8] Albawi, Saad and Mohammed, Tareq Abed and Al-Zawi, Saad., "Understanding of a convolutional neural network", 2017 international conference on engineering and technology (ICET), pp. 1-6, 2017.
[9] O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks, arXiv preprint arXiv:1511.08458.
[10] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv e-prints, 2015. doi:10.48550/arXiv.1505.04597.

Príklad využitia programovateľného drona pre monitorovanie nehodovej udalosti

[1]*Renát HALUŠKA,* [2]*Dalibor ŠINAĽ*

[1,2] Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Letná 9, 004001 Košice, Slovenská republika

[1]renat.haluska@tuke.sk, [2]dalibor.sinal@student.tuke.sk

***Abstrakt* — Tento článok sa zaoberá využitím dronov na rozpoznávanie a monitorovanie požiarov pri dopravných nehodách s cieľom poskytnúť pomoc hasičským a záchranárskym zložkám. Článok popisuje implementáciu neurónových sietí na rozpoznanie ohňa z obrazu natočeného dronom a navrhuje sa program na určenie optimálnej letovej trajektórie pre detekciu ohňa. Výsledky tejto práce majú zásadný význam pre efektívnejšie zvládanie požiarov a ochranu ľudských životov a majetku prostredníctvom využitia technológie dronov a strojového učenia.**

***Kľúčové slová* — dopravná nehoda, požiar, R-CNN, UAV, YOLOv8**

I. Úvod

Bezpilotné prostriedky (UAV) majú stále väčšie a väčšie uplatnenie pri zisťovaní a monitorovaní aktuálneho miesta dopravnej nehody. Dokážu rýchlo preskúmať škody preletom nad postihnutými oblasťami, pričom poskytujú detailné snímky, ktoré umožňujú ľuďom na mieste rýchlo získať prehľad o rozsahu škôd. Monitorujú horenie a šírenie ohňa a sledujú aj pohyb plameňov, čo umožňuje lepšiu predpoveď jeho správania sa a teda aj efektívne plánovanie a riadenie hasičských operácií. Sú nesmierne dôležité pri hľadaní ohnísk požiaru ktoré môžu vzniknúť pri nehode, kde s využitím termálnej kamery lokalizujú skryté ohniská požiaru, ktoré by mohli byť pre záchranárov nebezpečné alebo ťažko dostupné. UAV môžu identifikovať a definovať riziká ako sú horľavé plyny alebo nebezpečné materiály v potenciálne nebezpečných oblastiach, a umožniť hasičom zabezpečiť tieto oblasti pred prípadným šírením požiaru alebo ďalšími škodami.

Perspektíva využitia dronov nie je obmedzená len na tvorbu bežných fotografií a videí z výšky. Využívanie dronov v krajinnom výskume čoraz viac nahrádza časovo a finančne náročnejšie letecké snímkovanie. Medzi ich hlavné benefity oproti leteckému snímkovaniu patrí ľahká manipulácia, nízka hlučnosť, schopnosť rýchleho nasadenia v teréne a možnosť letu aj pri nízkej oblačnosti ako aj tvorba ortofotosnímok a trojrozmerných (3D) modelov s vysokým rozlíšením na stanovenie krajinného prostredia [1].

Väčšina dronov (UAV) je vyrobená z ľahkých kompozitných materiálov, ktoré dokážu dobre absorbovať vibrácie a musia mať aj malú hmotnosť pre lepšiu ovládateľnosť. Sú vybavené rôznymi technickými prostriedkami, ktorými sú predovšetkým GPS zariadenia, klasické, alebo multispektrálne kamery, prípadne môžu niesť rozprašovače chemických látok, alebo ťažné zariadenia.

II. Letové Vlastnosti Zvoleného Drona

Na túto prácu bol použitý Tello EDU, čo je špeciálna verzia drona, ktorá bola navrhnutá s dôrazom na vzdelávanie a výučbu programovania. Hoci Tello EDU nie je určený na pokročilé letové manévre, má niekoľko základných letových vlastností, ktoré ho robia vhodným nástrojom pre začiatočníkov a študentov. Má jednoduché ovládanie pomocou intuitívneho diaľkového ovládača alebo prostredníctvom smartfónu cez aplikáciu TELLO [2]. Tieto ovládacie možnosti robia ovládanie drona jednoduchým a prístupným. Je vybavený stabilizačnými senzormi, ktoré umožňujú stabilný let aj pri miernom vetre alebo pri zmene smeru. Je aj schopný udržiavať konštantnú výšku vďaka svojim výškovým senzorom. To umožňuje stabilný let aj pri zmene

nadmorskej výšky, taktiež má funkciu automatického vzlietania a pristátia, čo robí manipuláciu s ním jednoduchou a bezpečnou [3]. To znamená, že je ľahko ovládateľný a vhodný aj pre začiatočníkov. Jedným z hlavných rysov Tello EDU je jeho schopnosť programovateľnosti. Pomocou programovacieho jazyka Scratch alebo Python je možné vytvárať a upravovať letové manévre, zadávať pokyny pre dron a vytvárať interaktívne letecké projekty. Je navrhnutý tak, aby mohol lietať v interiéroch aj exteriéroch.

Nízky profil letu umožňuje presné ovládanie aj vo vnútorných priestoroch. Jeho základné vlastnosti sú:

- Maximálna vzdialenosť od ovládača 100m.
- Maximálna rýchlosť 8m/s.
- Maximálny čas letu 13min.
- Maximálna výska 30m [4].

A. Zabudované Manévre Dronu Tello

Dron sa vie vznášať v konštantnej výške a udržuje si tak stabilnú pozíciu pomocou kamier a infračervených senzorov umiestnených na spodnej časti. Dron má zabudovaných niekoľko inteligentných letových módov ako sú:

- **Bounce mode**, kde dron automaticky lieta hore a dole v rozmedzí medzi 0.5 a 1.2 m.
- **8D Flips dron** je schopný preklápať sa v ôsmich rôznych smeroch.
- **Throw and Go** je funkcia dronu, kde ho možno vypustiť jemným vyhodením do vzduchu.
- **Ez shots** je funkcia, kde dron nahráva krátke video, pričom sa otáča okolo vlastnej osi.

B. Dizajn a Konštrukcia

Dizajn drona je kompaktný a ľahký, umožňuje skladovateľnosť a ľahkosť prenášania zariadenia na miesta dopravnej nehody. Je vyrobený z odolných materiálov, ktoré zaisťujú trvanlivosť a odolnosť voči bežným poškodeniam, ale na druhej strane, nie sú to dostatočne odolné materiály voči vyšším teplotám a prachovým časticiam. Konštrukcia zahŕňa ochranné mriežky okolo vrtuliek, čo znižuje riziko zranenia alebo poškodenia pri kontakte s používateľmi alebo predmetmi v okolí.

C. Ovládanie

Ovládanie je intuitívne a jednoduché. Užívatelia nemusia mať žiaden špeciálny tréning na prácu s týmto zariadením. Dron je možné ovládať viacerými spôsobmi:

- **Mobilnou aplikáciou**: Tello Edu je možné ovládať pomocou mobilnej aplikácie TELLO, ktorá je dostupná pre platformy iOS a Android. Táto aplikácia ponúka užívateľsky prívetivé rozhranie s jednoduchými ovládacími prvkami, ako sú ikony a virtuálne páčky. Používatelia môžu sledovať živý prenos videa z kamery drona priamo na svojom smartfóne a jednoduchoriadiť jeho pohyb pomocou dotykového displeja. Vďaka tomu je možné nasadiť ho rýchlo a účinne v núdzových situáciách pri požiaroch s použitím akéhokoľvek mobilného zariadenia.
- **Vlastným programom**: Tello Edu je programovateľný pomocou rôznych programovacích jazykov, ako je Scratch alebo Python, kde je k dispozícii knižnica djitellopy [5]. Pomocou týchto jazykov môžu používatelia vytvárať a upravovať jednoduché programy a sekvencie príkazov, ktoré určujú správanie drona. To umožňuje záchranárom vypustiť dron na inkriminované miesto a nechať ho monitorovať situáciu bez zásahu človekom.

III. Tvorba Detekčného Systému

Pre detekciu požiaru sme v tejto práci použili YOLOv8 (You Only Look Once verzia 8), keďže predstavuje najnovšiu verziu v sérii modelov pre detekciu objektov, ktoré sú známe svojou rýchlosťou a presnosťou. YOLOv8 prináša niekoľko významných vylepšení oproti predchádzajúcim verziám, vrátane zlepšeného architektonického dizajnu a optimalizácií, ktoré zvyšujú jeho výkon v reálnom čase [6].

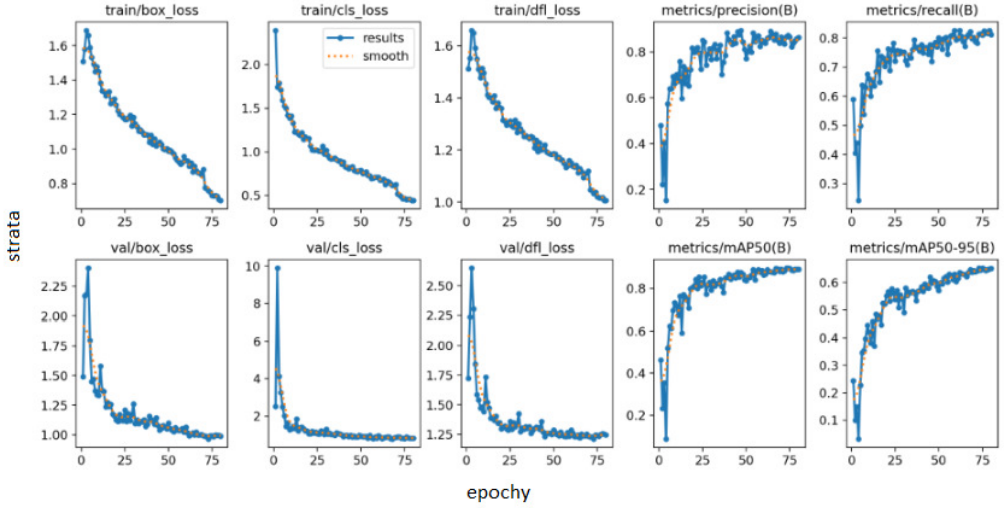
Na trénovanie Region-based Convolutional Neural Network (R-CNN), ktorú využíva YOLO algoritmus [7] sme zvolili Google Colab, čo nám umožnilo využiť bezplatnú akceleráciu GPU (Graphics Processing Unit). Pre vstupné dáta sme využili databázu Robotflow [8], obsahujúcu obrazové dáta s anotáciami ohňa a plameňov. V priebehu trénovania sme postupne a systematicky skúšali viacero prístupov. Prvé iterácie trvali približne dva dni, avšak kvôli technickým

komplikáciám nedosiahli uspokojivý výsledok. Neskôr sme identifikovali, že využívanie akcelerácie pomocou CPU (Central Processing Unit) je neefektívne, najmä pri spracovaní obrazových dát, ktoré sme používali na trénovanie algoritmus strojového učenia. Preto sme prešli na akceleráciu pomocou GPU, čo viedlo k výraznému urýchleniu procesu trénovania. Celkový čas tréningu bol približne 42 minút.

Detekčný systém využíva knižnicu UltraAnalytics, ktorá obsahuje implementácie rôznych algoritmov, vrátane YOLOv8 a knižnicu Roboflow, ktorá umožňuje spracovávať a anotovať dátové sady pre strojové učenie. Roboflow poskytuje nástroje na jednoduché nahrávanie, spracovanie a správu dátových sád, ako aj na trénovanie modelov strojového učenia.

Obrázok 1 zobrazuje grafy ukazujúce rôzne metriky týkajúce sa trénovania a validácie modelu strojového učenia. Dve najdôležitejšie metriky sú "metrics/mAP50(B)" a "metrics/mAP50-95(B)", ktoré zobrazujú Mean Average Precision (mAP). Táto metrika je kľúčová pri hodnotení presnosti detekcie objektov. Graf "mAP50" meria presnosť pri prahovej hodnote IoU (Intersection over Union) 0.50 (Prienik nad Zjednotenie), táto metrika používaná na vyhodnotenie presnosti detekcie objektov v oblasti počítačového videnia. IoU meria, do akej miery sa ohraničujúci rámec (bounding box) predikovaný modelom prekrýva so skutočným ohraničujúcim rámcom objektu v obraze. Zatiaľ čo "mAP50-95" priemeruje presnosť pri rôznych prahových hodnotách od 0.50 do 0.95. Rastúce hodnoty v týchto grafoch indikujú zlepšujúcu sa schopnosť modelu presne detegovať objekty, čo je kľúčový ukazovateľ výkonu.

Ďalšie kľúčové grafy sú "val/box_loss" a "val/cls_loss". "Val/box_loss" predstavuje straty v predikcii ohraničujúcich rámcov, čo znamená, že model má problémy s lokalizáciou objektov v obraze. "Val/cls_loss" predstavuje straty v predikcii tried, čo znamená, že model má problémy s klasifikáciou objektov na obraze. Tieto metriky sledujú chyby počas overovania modelu. Graf "train/box_loss" je ďalší dôležitý ukazovateľ, ktorý meria chybu v predikcii súradníc ohraničujúcich rámcov počas tréningu. Pokles tejto chyby indikuje, že model sa postupne učí predpovedať polohy ohraničujúcich rámcov presnejšie, čo je kritické pre úspešnú detekciu objektov.



Obr. 1 Výstupné grafy trénovania

Na obrázku 2 je zobrazený výstup detekčného algoritmu zameraného na identifikáciu plameňa pri dopravnej nehode a na obrázku 3 je zobrazený real-time výstup detekčného algoritmu z kamerového výstupu drona. V týchto výstupoch vidíme ohraničený plameň, ktorý je označený rámom (bounding box) a doplnený o percentuálnu hodnotu predstavujúcu mieru istoty algoritmu v identifikácii objektu ako plameň. Tento percentuálny údaj vyjadruje pravdepodobnosť, s akou model klasifikuje detegovaný objekt ako plameň, pričom vyššia hodnota naznačuje väčšiu dôveru v správnosť detekcie. Výstup tak poskytuje nielen vizuálnu informáciu o polohe plameňa, ale aj kvantitatívnu mieru istoty, čo je kľúčové pre následné rozhodovacie procesy v reálnom čase.



Obr. 2 Detekcia ohňa z obrázka



Obr. 3 Detekcia ohňa z kamerového výstupu drona

IV. Záver

V tejto práci sme predstavili systém na detekciu požiaru pre UAV, ktorý využíva najnovší model strojového učenia YOLOv8. Systém sme úspešne natrénovali na súbore obrázkov s anotáciami ohňa a plameňov a dosiahli sme vysokú presnosť detekcie. Systém je schopný detegovať oheň v reálnom čase z kamerového výstupu drona a poskytovať záchranárom vizuálnu informáciu o polohe a rozsahu požiaru, ako aj kvantitatívnu mieru istoty detekcie.

Tento systém má veľký potenciál pre zlepšenie bezpečnosti a efektivity záchranných prác pri dopravných nehodách. Ďalší výskum by sa mal zamerať na integráciu systému do platformy UAV a na testovanie jeho výkonu v reálnych podmienkach.

Pre pokračovanie vo výskume by sme chceli integrovať systém detekcie požiaru do platformy UAV a otestovať jeho výkon v reálnych podmienkach. Vyvinúť algoritmy na automatické riadenie drona na základe informácií z detekčného systému. Preskúmať možnosti využitia systému detekcie požiaru pre iné typy núdzových situácií a zvýšiť presnosť a robustnosť detekčného systému pomocou pokročilejších techník strojového učenia.

Poďakovanie

Táto práca bola čiastočne podporená Vedeckou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky a Slovenskou akadémiou vied v rámci projektu VEGA 2/0165/21, Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky cez projekt KEGA 049TUKE-4/2024.

Referencie

[1] P. Araújo, J. Fontinele and L. Oliveira, "Multi-Perspective Object Detection for Remote Criminal Analysis Using Drones," in IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 7, pp. 1283-1286, July 2020, doi: 10.1109/LGRS.2019.2940546.

[2] O. Pohudina, M. Kovalevskyi and M. Pyvovar, "Group Flight Automation Using Tello EDU Unmanned Aerial Vehicle," 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), LVIV, Ukraine, 2021, pp. 151-154, doi: 10.1109/CSIT52700.2021.9648704.

[3] A. Kasraoui, K. Bingi, R. Ibrahim, M. Omar, P. A. Mozhi Devan and B. R. Prusty, "Tracking Control of Tello EDU Quadrotor Drone Using Image Thresholding," 2023 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia, 2023, pp. 1-6, doi: 10.1109/i-PACT58649.2023.10434743.

[4] M. Iskandar, K. Bingi, R. Ibrahim, M. Omar and P. A. Mozhi Devan, "Hybrid Face and Eye Gesture Tracking Algorithm for Tello EDU RoboMaster TT Quadrotor Drone," 2023 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia, 2023, pp. 1-6, doi: 10.1109/i-PACT58649.2023.10434449.

[5] N. Virbora, U. Sokoeun, M. Saran, S. Sovicheyratana, S. Channareth and S. Saravuth, "Implementation of Matrix Drone Show Using Automatic Path Generator with DJI Tello Drones," 2022 International Conference on Engineering and Emerging Technologies (ICEET), Kuala Lumpur, Malaysia, 2022, pp. 1-5, doi: 10.1109/ICEET56468.2022.10007090.

[6] S. Reddy, N. Pillay and N. Singh, "Comparative Study of Convolutional Neural Network Object Detection Algorithms for Image Processing," 2023 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 2023, pp. 1-5, doi: 10.1109/ICECET58911.2023.10389186.

[7] J. Tao, H. Wang, X. Zhang, X. Li and H. Yang, "An object detection system based on YOLO in traffic scene," 2017 6th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2017, pp. 315-319, doi: 10.1109/ICCSNT.2017.8343709.

[8] "Roboflow: Computer vision tools for developers and enterprises." https://roboflow.com/

Analýza smerovacích protokolov pre MANET v kontexte rôznych modelov mobility

[1]Natalia KURKINA, [2]Ján PAPAJ

[1,2]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

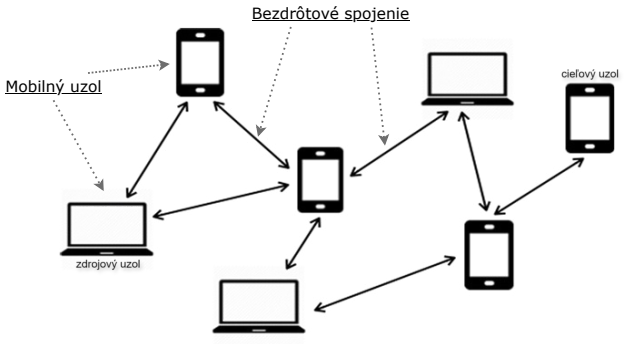
[1]natalia.kurkina@tuke.sk, [2]jan.papaj@tuke.sk

Abstract **– Mobilné ad-hoc siete umožňujú bezdrôtovú komunikáciu medzi mobilnými zariadeniami bez potreby centralizovanej infraštruktúry, čo je obzvlášť užitočné v situáciách, kde pevná infraštruktúra chýba alebo je poškodená. V práci sú analyzované dva hlavné smerovacie protokoly: Ad hoc On-Demand Distance Vector a Optimized Link State Routing. Simulačné experimenty boli vykonané pomocou simulátora NS-3, pričom sa skúmali rôzne modely mobility, vrátane Random Waypoint, Random Direction, Gauss-Markov a Manhattan Grid. Výsledky ukazujú, že výkonnosť smerovacích protokolov závisí na použitom modeli mobility a počte uzlov v sieti. Na vyhodnotenie boli použité rôzne parametre vratané pomeru doručených paketov, end-to-end oneskorenia a priepustnosti. AODV dosahoval lepšie výsledky v prípadoch, kde bolo potrebné minimalizovať šírenie služobných paketov, zatiaľ čo OLSR poskytoval menšie end-to-end oneskorenie vďaka neustálej aktualizácii smerovacích tabuliek.**

Keywords **– MANET, smerovacie protokoly, modely mobility, AODV, OLSR**

I. ÚVOD

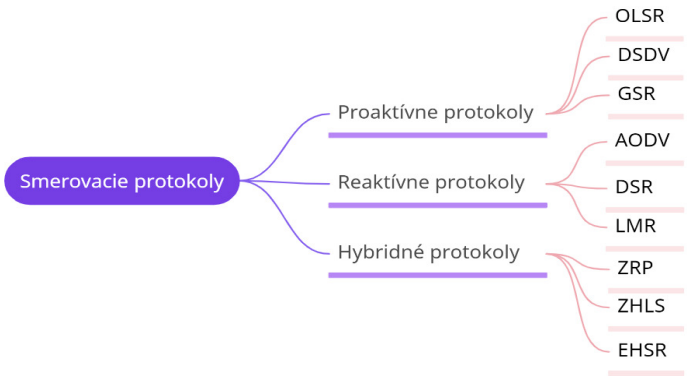
V modernom svete neustále rastie počet mobilných zariadení a v dôsledku toho rastie záujem o mobilné ad-hoc siete (MANET) [1]. MANET možno opísať ako siete, ktoré je možné vytvárať bez použitia štandardnej centralizovanej infraštruktúry. Uzly v MANET sa môžu voľne pohybovať rôznymi smermi, pripájať sa alebo odpájať od siete. Každý uzol v tejto sieti vykonáva funkciu smerovača aj funkciu koncového bodu, preto môžu uzly komunikovať priamo alebo nepriamo (odosielaním informácií cez iné uzly). Používanie takýchto sietí môže byť veľmi užitočné v oblastiach, kde neexistuje pevná infraštruktúra, alebo kde bola infraštruktúra poškodená v dôsledku prírodných katastrof alebo iných podobných situácií. MANET sa dajú jednoducho zorganizovať a môžu pomôcť v prípadoch, keď je urgentne potrebná komunikácia medzi rôznymi zariadeniami alebo tímami ľudí na vykonanie napríklad záchranných prác v horách alebo po zemetrasení. V dôsledku dynamickej povahy uzlov v sieti nie je možné použiť štandardné smerovacie protokoly a preto v tejto súvislosti boli vyvinuté nové prístupy pre smerovanie [2].



Obr. 1 Zakladná schema MANET siete.

II. TYPY SMEROVACÍCH PROTOKOLOV PRE MANET

Smerovacie protokoly pre MANET možno kategorizovať rôznymi spôsobmi: podľa dizajnu, architektúry siete, dostupných funkcií a podobne. Obrázok 2 predstavuje klasifikáciu založenú na dizajne siete. Proaktívne protokoly si pravidelne vymieňajú riadiace pakety, aby boli cesty aktuálne. Tieto protokoly neustále zaplavujú sieť aktualizáciami, čím zaisťujú, že každý uzol má aktuálny stav siete vo svojej smerovacej tabuľke, čo umožňuje každému uzlu vypočítať najlepšiu cestu. Oproti tomu reaktívne protokoly neudržiavajú aktuálny záznam o stave siete. V prípade potreby uzly si zisťujú cestu do cieľa pomocou špeciálnych paketov-požiadaviek, ktoré posielajú všetkým svojim susedom. Tento prístup často vedie k oneskoreniu v dôsledku času potrebného na počiatočné vytvorenie cesty. Hybridné protokoly kombinujú proaktívny aj reaktívny prístup [3]. Existuje mnoho navrhnutých smerovacích protokolov pre MANET siete; zameriame sa na tie najznámejšie.



Obr. 2 Klasifikácia smerovacích protokolov pre MANET.

A. Ad hoc On-demand Distance Vector protokol

Ad hoc On-Demand Distance Vector (AODV) je reaktívny smerovací protokol, ktorý zisťuje cestu od zdroja do cieľa iba v prípade potreby [4]. Keď chce uzol komunikovať s iným uzlom, zdrojový uzol posiela paket-požiadavku (od ang. Route Request, RREQ). Ktorýkoľvek uzol, ktorý dostane RREQ ho buď prepošle, alebo ak má cestu do cieľa, odpovie správou Route Reply (RREP).

Po vytvorení spojenia, cesta sa udržiava tak dlho, ako to zdrojový uzol potrebuje. Ak na aktívnej ceste nastane nejaký problém, zdrojovému uzlu sa odošle správa o chybe (od ang. Route Error, RERR). Zdrojový uzol potom môže v prípade potreby spustiť ďalší proces zisťovania novej komunikačnej cesty. Na zabezpečenie aktuálnosti smerovacích informácií a na vylúčenie smerovacích slučiek AODV používa poradové čísla. Každý uzol udržiava smerovaciu tabuľku, kde ukladá informácie o aktívnych cestách. Položky tabuľky obsahujú podrobnosti, ako je cieľová IP adresa, ďalší skok, počet skokov do cieľa a poradové číslo. Keďže trasy sa vytvárajú na požiadanie, AODV znižuje počet preposlaných služobných paketov a tým šetrí šírku pásma a výkon siete [4].

B. Optimized Link-State Routing protokol

Optimized Link State Routing protokol (OLSR) je proaktívny smerovací protokol [5]. OLSR si pravidelne vymieňa informácie o stave prepojenia so všetkými uzlami v sieti. Uzly pravidelne posielajú HELLO správy, aby zistili a zachovali informácie o svojich priamych susedoch a stave ich spojenia. Každý uzol udržiava kompletnú mapu topológie, ktorá mu umožňuje vypočítať optimálne cesty k akémukoľvek cieľu pomocou algoritmov, ako je Dijkstrov algoritmus.

Na zníženie počtu služobných paketov, OLSR používa koncept MPR uzlov (od ang. Multi Point Relay, MPR), podla ktoreho každý uzol si vyberie množinu susedných uzlov ako MPR. Tieto uzly sú zodpovedné za preposielanie riadiacich paketov, čo minimalizuje počet prenosov potrebných na šírenie informácií o stave spojenia cez sieť [5].

OLSR je navrhnutý tak, aby bol škálovateľný pre veľké a husté siete. Použitie MPR výrazne znižuje počet prenosov služobných paketov, čím sa stáva efektívnejším vo väčších sieťach. OLSR sa dobre hodí pre siete, kde sa uzly často pohybujú, a kde je kľúčové zisťovanie cesty s nízkou latenciou.

III. TYPY MODELOV MOBILITY

Model mobility je súbor pravidiel, ktoré riadia pohyb mobilného uzla. Rôzne modely mobility určujú, ako sa v priebehu času mení rýchlosť a umiestnenie uzla. Tieto modely mobility sú potrebné na účely simulácie pri skúmaní nových smerovacích protokolov a techník [6]. V súčasnosti bolo navrhnutých veľa modelov mobility, v tomto článku sú analyzované najpoužívanejšie typy modelov mobility.

A. Random Waypoint model mobility

Model mobility náhodného bodu (od ang. Random Waypoint model) zahŕňa časy prestávok medzi zmenami smeru a/alebo rýchlosti. Uzol sa náhodne umiestní v oblasti simulácie. Pohyb každého uzla je nezávislý od iného uzla. Pohyb uzla sa začína po uplynutí určitého časového obdobia (pauza), keď si uzol vyberie náhodný cieľ v oblasti simulácie a rýchlosť, ktorá je rovnomerne rozdelená medzi minimálnou a maximálnou hodnotou. Mobilný uzol potom cestuje smerom k cieľu zvolenou rýchlosťou. Po príchode sa uzol zastaví na určitý čas. Po uplynutí tohto času sa proces znova spustí [6].

B. Random Direction 2D model mobility

2D model mobility náhodného smeru (od ang. Random Direction 2D model) bol vyvinutý z Random Waypoint modelu mobility, aby riešil problém koncentrácie uzlov v centrálnej oblasti. Na rozdiel od Random Waypoint, Random Direction vyberie cieľové miesto na hraniciach simulačnej oblasti. Potom sa zastaví a znova si vyberie nové náhodné cieľové miesto.

Random Direction 2D model mobility poskytuje jednoduchý, ale efektívny spôsob simulácie mobility uzlov v MANET sieťach, čo umožňuje výskumníkom získať prehľad o správaní a výkone týchto sietí v rôznych podmienkach [7].

C. Gauss-Markov model mobility

Gauss-Markov model mobility pridáva do procesu pohybu časovo-závislú charakteristiku, keď sa mobilný uzol pohybuje podľa scenára, čo mu umožňuje napodobňovať obmedzenia pohybu v reálnom svete. Pri Gauss-Markov modelu mobility uzly sú od začiatku umiestnené na náhodných miestach v oblasti pohybu. Pohyb uzla je nezávislý od iného uzla v sieti. Každý uzol dostane najprv počiatočnú rýchlosť a smer. Uzly sa pohybujú aktualizáciou ich príslušných rýchlostí a smerov v nastavených intervaloch. Konkrétne sa to dosiahne určením rýchlosti a smeru v časovom intervale n na základe rýchlosti a smeru v časovom intervale $(n-1)$ a náhodnej premennej pomocou nasledujúcich rovníc [7]:

$$\begin{aligned} v_n &= \alpha v_{n-1} + (1-\alpha)\bar{v} + \sqrt{1-\alpha^2} v_{x_{n-1}}, \\ d_n &= \alpha d_{n-1} + (1-\alpha)\bar{d} + \sqrt{1-\alpha^2} d_{x_{n-1}} \end{aligned} \tag{1}$$

kde v_n a d_n sú nová rýchlosť a smer uzla v časovom intervale n; s a d sú stredné hodnoty rýchlosti a smeru; $v_{x_{n-1}}$ a $d_{x_{n-1}}$ sú náhodné premenné z Gaussovho rozloženia. Parameter α určuje dôležitosť starých hodnôt pri výpočte nových. V každom časovom intervale sa vypočíta ďalšia poloha na základe aktuálnej polohy, rýchlosti a smeru pohybu. Konkrétne v časovom intervale n je poloha uzla daná rovnicami:

$$\begin{aligned} x_n &= x_{n-1} + v_{n-1}\cos d_{n-1}, \\ y_n &= y_{n-1} + v_{n-1}\sin d_{n-1} \end{aligned} \tag{2}$$

kde x_n, y_n a x_{n-1}, y_{n-1} sú súradnice x a y polohy uzla v n a $n-1$ časových intervaloch, a s_{n-1} a d_{n-1} sú rýchlosť a smer uzla v časovom intervale $n-1$ [7].

D. Manhattan Grid model mobility

Pri použití Manhattan Grid modelu mobility celá oblasť pohybu uzlov je rozdelená na mriežku a uzol sa pohybuje od jednej križovatky do druhej. Tento model mobility bol navrhnutý tak, aby simuloval pohyb vozidla v mestskom prostredí. Mobilné uzly sa pohybujú horizontálne alebo vertikálne v zadanej oblasti pohybu. Tento model využíva náhodný prístup k výberu pohybu uzlov, pretože vozidlo si môže vybrať, či bude pokračovať v jazde rovnakým smerom alebo odbočí na križovatke [8].

Tabuľka 1
Parametre simulácie

Simulačná oblasť $[m]$	800×800
Čas simulácie [s]	100
Počet simulácií pre každý scenár	10
Počet uzlov v siete	25, 50, 100
Max. prenosový dosah uzla [m]	200
Komunikačný protokol	IEEE 802.11b
Model mobility	Random WayPoint, Random Direction 2D, Gauss-Markov, Manhattan Grid
Rýchlosť pohybu uzlov [m/s]	2 – 6

IV. SIMULÁCIA A ANALÝZA VÝSLEDKOV

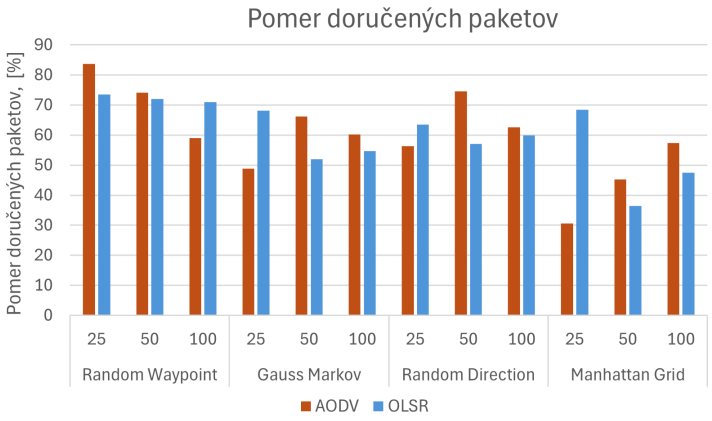
Táto časť predstavuje analýzu výsledkov rôznych simulačných scenárov pri použití vyššie popísaných modelov mobility. Na tento účel boli vybraté AODV a OLSR smerovacie protokoly. Na simuláciu MANET siete bol použitý simulačný nastroj NS-3. Hlavné parametre simulácie sú uvedené v Tabuľke 1.

Na vyhodnocovanie boli použité nasledujúce parametre [9]:

- Pomer doručených paketov. Tento parameter zahŕňa celkový počet úspešné prijatých dátových paketov v percentách.
- End-to-end oneskorenie. End-to-end oneskorenie je definované ako čas potrebný na prenos dátového paketu cez sieť zo zdroja do cieľa.
- Priepustnosť. Tento parameter je definovaný ako priemerný počet dátových paketov, ktoré sú úspešne prenesené počas komunikácie. Popisuje vzťah celkového počtu paketov prijatých cieľovým uzlom k celkovému počtu odoslaných paketov.

Na obrázkoch 3 až 5 môžeme vidieť výsledky simulácie pre všetký vyššie uvedené parametre pre rôzne typy modelov mobility uzlov.

Pri analýze pomeru doručených paketov v prípade keď v zadanej oblasti sa pohybuje 25 uzlov AODV je lepší len pri využití Random Waypoint modelu. V ostatných prípadoch je lepší OLSR protokol. Ale keď sa počet uzlov zvýši na 50 alebo 100 OLSR je lepší len pri použití Random Waypoint modelu. V ostatných prípadoch naopak je lepší AODV protokol.



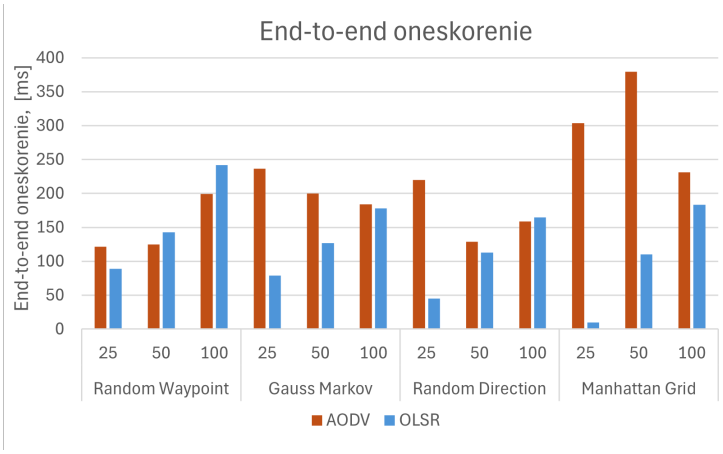
Obr. 3 Pomer doručených paketov pre rôzne typy modelov mobility.

Pri analýze end-to-end oneskorenia vo väčšine prípadov menšie hodnoty má OLSR protokol. Tieto dosiahnuté výsledky sú očakávane lebo mobilne uzly pri smerovaní s využitím OLSR protokolu nestrácajú čas získaním cesty a všetky informácie majú vo svojej smerovacej tabuľke od začiatku komunikácie.

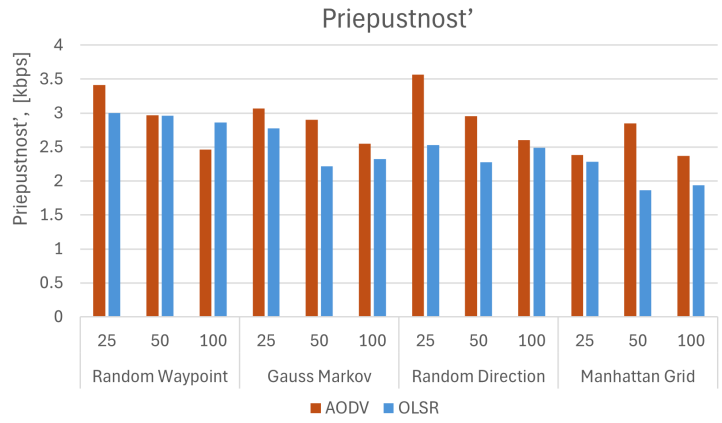
Pri analýze priepustnosti vo väčšine prípadov väčšie hodnoty má AODV protokol. Kvôli svojej reaktívnej povahe AODV protokol nezahlcuje sieť služobnými informáciami a preto priepustnosť siete je väčšia ako pri použití OLSR protokolu.

V. ZÁVER

Tento článok skúma vplyv modelov mobility na výkon smerovacích protokolov AODV a OLSR pomocou simulačného prostredia NS3. Taktiež táto štúdia obsahuje komplexné vyhodnotenie štyroch bežne používaných modelov mobility: Random Waypoint, Gaussov-Markov, Random Direction 2D a Manhattan Grid. Výkon protokolov bol analyzovaný pomocou rôznych metrík (pomer doručených paketov, end-to-end oneskorenie, priepustnosť a počet uzlov). V prípadoch, kde bolo potrebné minimalizovať šírenie služobných paketov AODV preukázal lepšie výsledky. To



Obr. 4 End-to-end oneskorenie pre rôzne typy modelov mobility.



Obr. 5 Priepustnost' pre rôzne typy modelov mobility.

sa odrazilo na vyššej priepustnosti siete, keď že AODV využíva šírku pásma efektívnejšie tým, že sa smerovacie informácie získavajú len v prípade potreby. Na druhej strane, OLSR protokol, ktorý je proaktívny, poskytoval menšie end-to-end oneskorenie vďaka neustálej aktualizácii smerovacích tabuliek, čo je výhodné najmä v situáciách, kde je kľúčová rýchla odozva siete.

Simulácie ukázali, že model mobility ovplyvňuje výkonnosť protokolov. Napríklad, pri použití Random Waypoint modelu mobility bol AODV výkonnostne lepší pri menšom počte uzlov, zatiaľ čo pri väčšom počte uzlov sa jeho výkonnosť zhoršila v porovnaní s OLSR. Naopak, pri použití ostatných modelov mobility sa AODV ukázal byť lepší aj pri vyššom počte uzlov.

POĎAKOVANIE

Tento výskum bol financovaný agentúrou na podporu výskumu a vývoja SR, výskumný grant č. APVV-17-0208 a VEGA 1/0260/23.

LITERATÚRA

[1] S. Yogarayan, "Wireless ad hoc network of manet, vanet, fanet and sanet: a review," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 13, no. 4, pp. 13–18, 2021.
[2] D. Kanellopoulos, "Congestion control for manets: An overview," *ICT Express*, vol. 5, no. 2, pp. 77–83, 2019.
[3] S. K. Sarkar, T. G. Basavaraju, and C. Puttamadappa, *Ad hoc mobile wireless networks: principles, protocols and applications*. Auerbach publications, 2007.
[4] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (aodv) routing," Tech. Rep., 2003.
[5] R. Jain, "Ant colony inspired energy efficient olsr (ac-olsr) routing protocol in manets," *Wireless Personal Communications*, vol. 124, no. 4, pp. 3307–3320, 2022.
[6] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, 08 2002.
[7] N. Abdelwahab, T. A. Mounir, and S. Boumerdassi, "Impact of mobility model on lorawan performance," *Journal of Communications*, vol. 19, pp. 7–18, 01 2024.
[8] V. Siva and M. Hemalatha, "Simulation and evaluation of different mobility models in ad-hoc sensor network over dsr protocol using bonnmotion tool," vol. 335, 10 2012, pp. 157–167.
[9] N. Shah, H. El-Ocla, and P. Shah, "Adaptive routing protocol in mobile ad-hoc networks using genetic algorithm," *IEEE Access*, vol. 10, pp. 132 949–132 964, 2022.

Návrh letových vlastností drona pre monitorovanie miesta vzniknutého požiaru

[1]*Renát HALUŠKA,* [2]*Filip PAJKOŠ*

[1,2] Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Letná 9, 004001 Košice, Slovenská republika

[1]renat.haluska@tuke.sk, [2]filip.pajkoš@student.tuke.sk

***Abstrakt* — Tento článok sa zaoberá návrhom letových a komunikačných vlastností programovateľného dronu (UAV) určeného na monitorovanie miesta vzniku požiaru. Cieľom tejto práce je navrhnúť dron, ktorý bude schopný rýchleho a presného prieskumu postihnutých oblastí, čo umožní rýchlu reakciu záchranných tímov a minimalizáciu škôd. V rámci práce bol využitý bezpilotný dron DJI Ryze Tello s pridaním mikrokontroléru ESP32, ktorý slúži ako riadiaca jednotka v kombinácii s GPS modulom NEO-7M. Na detekciu objektov bola využitá metóda YOLO.**

***Kľúčové slová* — dron, ESP32, protipožiarna bezpečnosť, Raspberry Pi, YOLO**

I. Úvod

Použitie dronov (UAV) je veľmi dôležité aj pri rôznych druhoch požiarov. Pohľady zhora umožňujú hasičom sledovať smer a vývoj požiaru a taktiež aj zmenu jeho intenzity. Pri požiaroch je veľmi kľúčové monitorovanie ohnísk požiaru a rozširovanie ohňa, prípadne detekciu osôb v bezprostrednej blízkosti požiaru [1]. Drony zo vzduchu monitorujú šírenie ohňa a s použitím kamery upozorňuje na miesta, na ktoré je potrebné sa pri likvidácii požiaru zamerať. Pomocou záberov vytvorených z rôznych uhlov a dát, napr. kamery, dokážu hasiči určiť najefektívnejší spôsob uhasenia požiaru [2].

Drony používané pri rôznych záchranných prácach sú rôznej váhovej kategórii a rôzneho druhu vybavenia. Profesionáli využívajú v prevažnej miere drony okolo jedného kilogramu. Je potrebné uviesť, že v ich výbave sa nájdu aj drony vo váhovej kategórii 40 kilogramov a viac. Čo sa týka kamerového vybavenia je na výber nespočetné množstvo druhov. Niektoré lacnejšie typy majú iba základné rozlíšenie. Pre profesionálne použitie je potreba výberu kamier s lepším a vysokým rozlíšením. Drony sa líšia nielen s ohľadom na svoju hmotnosť, veľkosť a vybavenosť, ale aj svojou operatívnosťou. To, že sa drony dostanú do miest a priestorov, v ktorých by sa nedali snímať objekty v dobrej kvalite, je ich najväčšou výhodou [3].

V tejto práci sa zameriavame na testovanie nasledovných hypotéz:

Presnosť detekcie: Predpokladáme, že dron vybavený kamerou a algoritmom YOLO dokáže detegovať požiare s vysokou presnosťou, ktorá sa bude pohybovať nad 70%.

Vplyv poveternostných podmienok: Hypotéza znie, že poveternostné podmienky, ako je dážď, sneh alebo silný vietor, nebudú mať značný vplyv na presnosť detekcie požiarov systémom dronu a algoritmu YOLO.

Výhody oproti tradičným metódam: Predpokladáme, že použitie dronu s algoritmom YOLO na monitorovanie požiarov ponúka oproti tradičným metódam, ako sú veže a hliadky, viacero výhod, vrátane flexibility, efektívnosti a nižších nákladov.

II. Požiadavky na Prostriedky Použité pre Monitorovanie Požiaru

Výhodou bezpilotných leteckých prostriedkov je to, že sa dokážu dostať na rôzne miesta, či už v objektoch, ale aj na voľnom priestranstve. Na využitie v civilnej sfére, ale aj vo verejnej sfére je najvhodnejšie použiť drony. Dron, je možné ho využiť na rôzne účely napr. na mapovanie, fotografovanie, výrobu videí, topografiu, v policajných akciách na vyhľadávanie nezvestných osôb, doprave, či hasení požiarov a monitorovanie miest požiaru.

Tieto drony si získavajú stále väčšiu popularitu práve kvôli svojim výhodám. Svoje nezastupiteľné miesto si získavajú pri použití nielen na rôznych záchranných a pátracích

akciách, ale aj pri použití v rôznych nebezpečných prostrediach vďaka tomu, že sa dajú naprogramovať na sledovanie konkrétnej dráhy letu. Hlavnou výhodou týchto dronov je ich pomerne jednoduché ovládanie. Sú dodávané s ovládačom, čo umožňuje zadávanie špecifických príkazov. Ovládanie tradičných dronov môže byť niekedy náročné a vďaka svojim výhodám sú programovateľné drony pohodlnejšie. Mnoho dronov zo skupiny programovateľných je navyše vybavených GPS funkciami, pomocou ktorých sa ľahko sledujú a lokalizujú pri prípadnej strate [4].

Najlepšími programovateľnými dronmi sú zvyčajne tie, ktoré majú dobrý letový čas, rýchlosť, dosah, kvalitu fotoaparátu, hmotnosť a veľkosť. K obľúbeným modelom patria modely DJI Tello, Tello Edu, Parrot Anafi, Sky Viper E1700. Tieto drony disponujú malými rozmermi, relatívne dobrým letovým časom a ľahkou konštrukciou.

A. Tello Edu

Programovateľný dron Tello Edu je veľmi pôsobivý a ideálny pre vzdelávanie. Využíva programovacie jazyky Scratch, Python a Swift. Spolu s technológiou riadenia letu od DJI má možnosť podpory elektornickej stabilizácie obrazu. Pri tomto type je možnosť naprogramovať viac dronov, ktoré môžu lietať v roji a vykonávať určité akrobacie. Tello Edu využíva výukový program Tello Space Travel, ktorý naučí svojho používateľa použitie programovacieho jazyka Swift na kódovanie. Prostredníctvom jedného zariadenia je možné naprogramovať viacero dronov súčasne. Drony Tello Edu používajú Mission Pads, ktoré slúžia pre zvýšenú presnosť jeho programovania. Sú to nielen vodiace body, ale aj spúšťacie mechanizmy. Na základe napísaného kódu, ktorý umožňuje dronu rozpoznať ID každého Padu môže dron Tello Edu vykonávať aj rôzne vzdušné akrobacie. Tello Edu podporuje blokové programovanie jednoduchým potiahnutím blokov kódovania na obrazovke [7].

B. Parrot Anafi

Od samého začiatku je spoločnosť Parrot kľúčovým hráčom vo vývoji dronov nielen pre profesionálne, ale aj pre rekreačné použitie. Táto spoločnosť vyrobila niekoľko modelov dronov a každý z nich je techologicky vyspelejší, ako predchádzajúci. Parrot Anafi je najnovším modelom tejto spoločnosti. Je vybavený funkciami, ktoré mu zabezpečujú miessto pre široké spektrum používateľov. Tým, že tento dron má dlhú dobu letu, predĺžený dolet a vysokú rýchlosť, dáva svojmu užívateľovi možnosť lietať bez obmedzení. Disponuje kamerou s rozlíšením 4K a horizontálnym zorným poľom. Doba jeho letu je 25 minút, čo poskytuje dostatočný čas na preskúmanie okolia. Obrovskou výhodou tohto drona je aj schopnosť rýchlo meniť smer a s jeho rýchlosťou 15 m/s je jedným z najrýchlejších dronov. Anafi je taktiež jedným z najľahších a najmenších dronov. Váži 320 gramov, meria 175 x 240 x 65 mm. Vďaka týmto parametrom si ho používateľ môže so sebou vziať kamkoľvek [4].

C. Sky Viper E 1700

Tento model dronu je pomerne málo známy. Je malý a je vhodný pre tých, ktorí sa chcú venovať programovaniu. Môže robiť najrôznejšie triky a prevracanie. Má kratšiu dobu letu. Kamera nie je súčasťou, čo je určitá nevýhoda, ale každý užívateľ si ju vie pridať sám podľa vlastných kritérií a možností dronu. Maximálna vzdialenosť, kedy už stráca spojenie, je 100 metrov a dosahuje rýchlosť 8 m/s. Sky Viper E 1700 je malý dron s rozmermi 35,56 x 7,62 x 30,48 cm, váži 75 gramov. Vďaka tomu je vo vzduchu obratnejší a vzlet mu trvá pomerne krátky čas [4].

III. Programové Prostriedky pre Monitorovanie Miesta Požiaru

V rámci výskumu bol k dronu DJI Tello pridaný mikrokontrolér ESP32 v kombinácii s modulom GPS NEO-7M na ovládanie dronu. Mikrokotolér ESP32 spolu s GPS modulom používaný na riadenie dronu s použitím GPS súradníc. Na rozpoznanie objektu – ohňa v obraze bol použitý algoritmus YOLOv5 (You Only Look Once).

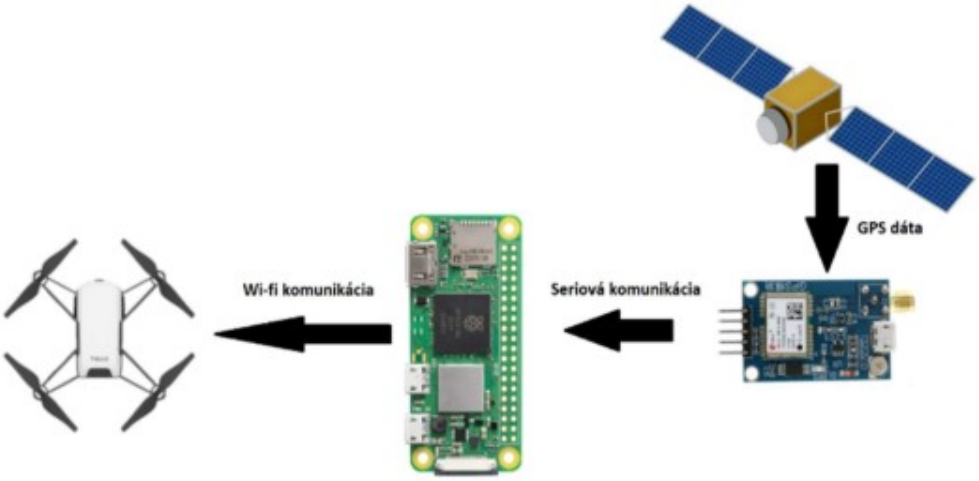
Navrhnutý systém pozostáva z troch hlavných komponentov: dronu DJI Tello, mikrokontroléra ESP32 a GPS modulu NEO-7M. Dron je vybavený kamerou s rozlíšením 5 MP a zorným uhlom 84°, ktorá sníma obrazy požiaru. Obrazy sú prenášané do mikrokontroléra ESP32 cez Wi-Fi pripojenie s priemernou rýchlosťou 20 snímok za sekundu.

Mikrokontrolér ESP32 je programovateľný mikrokontrolér s dvoma jadrami a taktom 240 MHz. Mikrokontrolér ESP32 spracováva obrazy z kamery a vykonáva detekciu objektov

pomocou algoritmu YOLOv5s. Mikrokontrolér ESP32 tiež komunikuje s GPS modulom NEO-7M a posiela informácie o polohe dronu do riadiaceho centra.

Na získavanie polohy sme využili GPS modul NEO-7M. Pre čo najpresnejšie získavanie polohy sme k GPS modulu pripojili externú anténu a modul pripojili k mikrokontroléru ESP32. Prostredníctvom ESP32 sme dekódovali informáciu, prichádzajúcu z GPS modulu cez sériovú komunikáciu. Pomocou toho bola zistená zemepisná dĺžka a zemepisná šírka polohy miesta, kde sa GPS modul nachádza (Obr. 1).

Ovládacia časť dronu je rozdelená na dve časti. Prvou časťou je GPS koordinácia, ktorá dovoľuje dronu preletieť väčšie vzdialenosti bez zásahu používateľa a umožňuje dronu efektívny dolet na miesto prostredníctvom GPS koordinátov. Druhá časť pozostáva z ovládania dronu použitím klávesnice. Toto ovládanie použitím klávesnice slúži na kvalitnejšie a detailnejšie prehliadky miesta požiaru alebo stabilizáciu drona v zlých poveternostných podmienkach. Tieto dve časti ovládania sa počas vykonávania programu dajú prepínať bez zastavenia detekcie, alebo pristátia drona. Vďaka tomu môže byť detekcia pre používateľa efektívnejšia.



Obr. 1 Schéma komunikácie

IV. Detekcia Požiaru

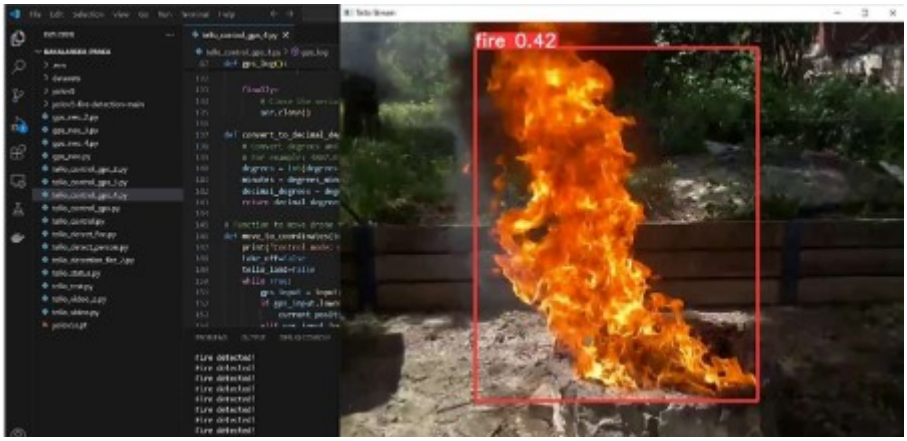
Dôležitou časťou pri monitorovaní miest požiaru je obhliadka (Obr. 2). Keď vznikne požiar v bytových priestoroch alebo v areáloch fabrík, firiem a pod., nie vždy je viditeľné alebo ľahko nájditeľné miesto požiaru. Po prílete dronu na miesto požiaru, môže operátor prepnúť program do módu ovládania cez klávesnicu, kde sa mu následne zobrazia informácie a klávesy na ovládanie. Vďaka tomu si môže operátor prezrieť podrobne miesto požiaru, miesto odkiaľ mohol potenciálne požiar vzniknúť alebo prekonať nejaké výškové prekážky, oblety budov a miesta, kde už GPS pozíciu nie je možné správne identifikovať. Ovládanie umožňuje ovládať dron vo všetkých smeroch, čím umožňuje presnú identifikáciu miesta požiaru, rozsiahlosť a prípadne obhliadku napáchaných škôd.



Obr. 2 Obhliadka miesta požiaru

Detekciu požiaru bola riešená pomocou modelu algoritmu YOLOv5s, ktorý bol natrénovaný z datasetu získaného z internetu [22]. Model spracovával obraz v reálnom čase v každom móde. Časť kódu, ktorá zabezpečovala stream videa z dronu a detekciu, beží v samostatnom jadre, takže je v rámci programu nezávislá od úkonov ostatných funkcií. Jadro sa uzatvára až po kompletnom zastavení práce s dronom. Jadro, do ktorého vstupuje funkcia na detekciu sa zapína po zavolaní v main funkcii. Detekcia sa uskutočňuje v reálnom čase (Obr. 3), takže operátor vidí výsledky okamžite. Konfigurácia algoritmu YOLOv5s použitá v tejto štúdii je nasledovná:

- Počet tried: 3 (oheň, dym)
- Veľkosť vstupu: 640 x 480 pixelov
- Počet konvolučných vrstiev: 18
- Počet dekonvolučných vrstiev: 3
- Počet výstupných prvkov: 2 (počet objektov v každej triede)



Obr. 3 Príklad detekcie založeného požiaru

V. ZÁVER

Monitorovanie požiarov a miest vzniknutého požiaru je v súčasnosti veľmi dôležité. Problémom je, že požiare vznikajú na akomkoľvek mieste a preto je potrebné zabezpečiť monitorovanie ohnísk kdekoľvek, aby sa zabránilo stratám na životoch a majetku. Častokrát požiare vznikajú v dôsledku nezodpovedného konania ľudí a nie vždy vzniknuté ohniská môžu kontrolovať priamo hasiči. Preto je dôležité, aby záchranné a monitorovacie tímy získavali čo najpresnejšie, najaktuálnejšie a najkvalitnejšie informácie. K získaniu týchto informácií prispievajú v nemalej miere práve bezpilotné letecké zariadenia, teda programovateľné drony.

Kamera, ktorou bol vybavený dron, mala dostačujúcu kvalitu na to, aby YOLOv5 algoritmus

dokázal správne vyhodnotiť snímky z dronu. Zvolená metóda GPS súradníc je vhodná pre veľké vzdialenosti a GPS modul komunikoval spoľahlivo pri prenose dát do ESP32, pričom bolo možné použiť sériovú komunikáciu na GPIO pinoch. Pri niektorých testoch bolo potrebné využívať externú GPS anténu, aby sme dostali presnejšie výsledky a rýchlejšiu konektivitu. Dosiahli sme veľkú kompatibilitu a efektivitu v rámci ovládania a detekcie požiaru, ale sme aj otvorili nové možnosti vývoju a zdokonaľovaniu programu, ktoré môžu pomôcť pri zhotovení plnohodnotného autonómneho dronu, ktorý bude schopný detegovať požiar na vyššej úrovni.
Systém dosiahol priemernú presnosť detekcie 90%. Presnosť detekcie sa nelíšila v závislosti od poveternostných podmienok. V nočných podmienkach bola presnosť detekcie mierne nižšia (80%), ale stále dostatočná na použitie v praktických aplikáciách.

Pri testovacích a simulačných procesoch sme zistili, že dron má veľmi krátku výdrž batérie, čo neumožňuje let na veľké vzdialenosti. Taktiež jeho rozmery a nosnosť obmedzili montáž súčiastok.

Poďakovanie

Táto práca bola čiastočne podporená Vedeckou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky a Slovenskou akadémiou vied v rámci projektu VEGA 2/0165/21, Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky cez projekt KEGA 049TUKE-4/2024.

Referencie

[1] Z. Jiao et al., "A YOLOv3-based Learning Strategy for Real-time UAV-based Forest Fire Detection," 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 2020, pp. 4963-4967, doi: 10.1109/CCDC49329.2020.9163816.

[2] X. Xie, J. Wang, Y. Zhang, J. Xin, L. Mu and X. Xue, "A UAV Forest Fire Detection Method Based on Dual-Light Vision Images," 2023 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), Yibin, China, 2023, pp. 1-6, doi: 10.1109/SAFEPROCESS58597.2023.10295827.

[3] G. D. Georgiev, G. Hristov, P. Zahariev and D. Kinaneva, "Forest Monitoring System for Early Fire Detection Based on Convolutional Neural Network and UAV imagery," 2020 28th National Conference with International Participation (TELECOM), Sofia, Bulgaria, 2020, pp. 57-60, doi: 10.1109/TELECOM50385.2020.9299566.

[4] H. Dang-Ngoc and H. Nguyen-Trung, "Aerial Forest Fire Surveillance - Evaluation of Forest Fire Detection Model using Aerial Videos," 2019 International Conference on Advanced Technologies for Communications (ATC), Hanoi, Vietnam, 2019, pp. 142-148, doi: 10.1109/ATC.2019.8924547.

[5] A. Kasraoui, K. Bingi, R. Ibrahim, M. Omar, P. A. Mozhi Devan and B. R. Prusty, "Tracking Control of Tello EDU Quadrotor Drone Using Image Thresholding," 2023 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia, 2023, pp. 1-6, doi: 10.1109/i-PACT58649.2023.10434743.

[6] E. Ackerman, "Parrot's new drone reclaims aniche: The Anafi marks the company's return to theconsumerspace - [Resources_Review]," in IEEE Spectrum, vol. 55, no. 9, pp. 21-21, Sept. 2018, doi: 10.1109/MSPEC.2018.8449041.

[7] S. Cherupally, “Multi Robot Coordination In Unstructured Environment” (Doctoral dissertation, Arizona State University).

Detekcia meteorov na rádiových spektrogramoch

[1] *Lenka Kališková,* [2]*Peter Butka,* [3]*Kamila Jenčiková*

[1,2,3]Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]lenka.kaliskova@tuke.sk,[2] peter.butka@tuke.sk, [3]kamila.jencikova@student.tuke.sk

Abstrakt **– V dnešnej dobe už existujú meracie aparatúry, ktoré zameriavajú meteory na obrovskú vzdialenosť. Meraniami sa získavajú dáta, ktorých analýza poskytuje veľmi dôležité a zaujímavé informácie. Dát je ale veľké množstvo a na ich spracovanie častokrát nepostačujú len ľudské sily, preto je dôležité využitie techník strojového učenia. Cieľom tejto práce bolo vytvorenie modelu neurónovej siete, ktorá na základe učenia, trénovania a testovania dokázala detegovať a následne klasifikovať existenciu meteorických javov na oblohe.**

Kľúčové slová **– neurónová sieť, detekcia meteorov, YOLO, anotovanie, rádiový spektrogram, ozvena, headecho**

I. Úvod

Článok sa zameriava na detekciu meteorov na rádiových spektrogramoch, využívajúc neurónovú sieť You Only Look Once (skr. YOLO). Pozostáva z dvoch častí, a to získavania dát prostredníctvom anotácií a použitia týchto dát na trénovanie detekčnej siete YOLO, ktorá dokáže automaticky detegovať a klasifikovať meteory do dvoch skupín.

Detekcia meteorov je dôležitá pre ich stále nedostatočné prebádanie. S veľkým množstvom dát je nevyhnutné hľadať efektívne spôsoby ich spracovania. Hlboké učenie umožňuje neurónovým sieťam rozpoznať ozveny a iné javy, čo je využiteľné aj v iných oblastiach, ako je rozpoznávanie tvárí, či detekcia zvierat v automobilovom priemysle.

Všetky kroky sú vykonané na zozbieraných dátach, ktoré anotovali študenti Technickej univerzity v Košiciach v rámci crowdsourcingového projektu. Dáta boli spracované v Pythone, očistené a upravené na ďalšie spracovanie v prostredí YOLO.

Článok sa zameriava na meteorické stopy, nazývané ozveny, zachytené rádiovým rozptylom. Táto efektívna technika sledovania meteorov zahrňuje odraz rádiových vĺn od ich stôp pri vstupe do atmosféry. Sledovanie charakteristík prijatého signálu nám poskytuje informácie o fyzikálnych parametroch meteorickej udalosti. Zachytený odraz je zrkadlový, čo znamená, že meteory prinášajú odrazy len v obmedzenom pásme oblohy.

II. YOLO

V roku 2015 bol prvýkrát prezentovaný algoritmus YOLO, ktorý sa stal známy najmä na základe svojej presnosti a rýchlosti [1]. Bol vyvinutý s cieľom zahrnutia jednokrokového procesu detekcie a následnej klasifikácie. Tento algoritmus slúži na detekciu jednotlivých objektov na obrázkoch v reálnom čase a prvýkrát ho vo svojom článku popísal Joseph Redmon s kolektívom [2]. Jeho využitie sa našlo vo viacerých smeroch, či už pri detekciách ľudí, zvierat, dopravy, alebo napríklad pri parkovacích automatoch. Algoritmus YOLO sa líši od ostatných najmä v tom, že sa predikcie ohraničení súčasne vytvárajú s predikciami tried [3].

YOLO vzniklo ako najlepší predstaviteľ konvolučných neurónových sietí (angl. *Convolutional Neural Network*, skr. CNN), ktorý prišiel na základe naliehavej požiadavky na zlepšenie detekcie prostredníctvom jej zrýchlenia. Výkon CNN však nepostačoval na riešenie komplexnejších úloh, preto sa začalo pracovať na lepšom modeli, ktorý by zvládol riešiť úlohy efektívnejšie. YOLO nielen deteguje obrázky automaticky, ale taktiež poskytuje hodnotu pravdepodobnosti toho, ako si je istý detekovanými objektami [4].

Algoritmus rozdeľuje obrázky do mriežok, ktoré väčšinou nadobúdajú rozmery 19×19 [2]. To zabezpečuje predikciu určitého počtu boxov od každej jej mriežky, vďaka čomu je algoritmus predikcií tak rýchly. Následne je každej bunke priradená najpravdepodobnejšia trieda.

YOLO už dnes obsahuje niekoľko rôznych verzií, medzi ktoré patria:

- `YOLOv2`: základná verzia YOLO má dve podstatné chyby, a to nepresné určovanie polohy (angl. *inaccurate positioning*) a nízku mieru návratnosti (angl. *low recall rate*). Táto verzia sa preto venuje práve zlepšeniu týchto dvoch aspektov, ktoré zabezpečujú väčšiu rýchlosť a kvalitu. Nie je to na základe prehlbovania, či rozširovania siete, ale práve prostredníctvom jej zjednodušenia.
- `YOLOv3`: s príchodom každej novej verzie sa tvorcovia zameriavali na dosah presnejších výsledkov. Táto tretia verzia využíva na detekciu objektov siete pyramíd prvkov (angl. *Feature Pyramid Networks*, skr. FPN).
- `YOLOv4`: na rozdiel od verzie YOLOv3 prichádza opäť s presnejšími výsledkami detekcií. Taktiež zlepšuje chybovosť, ktorá je na rozdiel od predošlej verzie nižšia. Pre detekciu objektov používa sieť PANet (angl. *Path Aggregation Network for Instance Segmentation*), ktorá poskytuje rôzne úrovne detekcií.
- `YOLOv5`: prichádza s lepším výkonom a presnosťou, preto sme sa v tejto práci rozhodli pokračovať s danou verziou YOLO. Jej bližší popis je znázornený v podkapitole nižšie.
- `iné`: okrem týchto verzií existujú aj ďalšie, limitované verzie, ako napríklad YOLO-LITE.

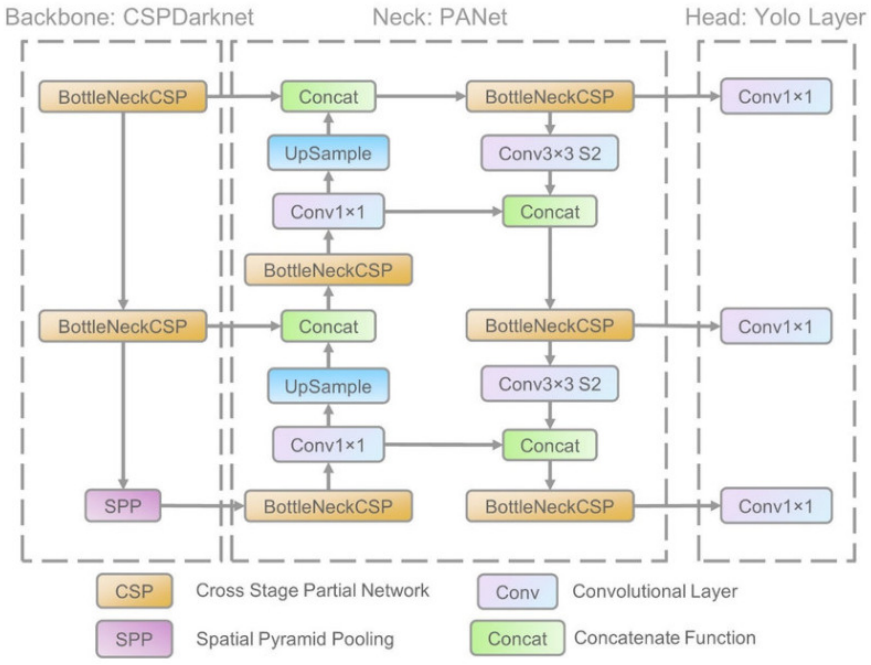
A. *YOLOv5*

Ako bolo spomínané vyššie, YOLOv5 je jednou z generácií YOLO a prichádza s lepším výkonom, presnosťou a rýchlosťou. Pôvodne vznikla ako vylepšená verzia YOLOv3 pre PyTorch a jej prvú publikáciu uverejnila firma Ultralytics LLC v roku 2020 [4].

Architektúra YOLOv5 je zložená z troch základných častí, a to:

- `Backbone`: vloženie dát do CSP Darknet (angl. *Cross Stage Partial*) kvôli extrakcií funkcií. Práve tu sa extrahujú vlastnosti objektov [5].
- `Neck`: v tejto časti dochádza k prevodu dát z CSP Darknet do PANet (angl. *Path Aggregation Network for Instance Segmentation*) [6]. PANet zabezpečuje ukladanie vlastností detekcií do tzv. pyramíd, vďaka čomu sú identifikované objekty s inými vlastnosťami.
- `Head`: v poslednej časti sa prostredníctvom *Yolo Layer* vydávajú výsledky detekcií, ako napríklad trieda, lokácie, skóre, veľkosť či iné.

YOLOv5 architektúra je bližšie vyobrazená na Obrázku 1, ktorý popisuje jej vyššie uvedené časti.



Obrázok 1 Architektúra YOLOv5. Zdroj: [7]

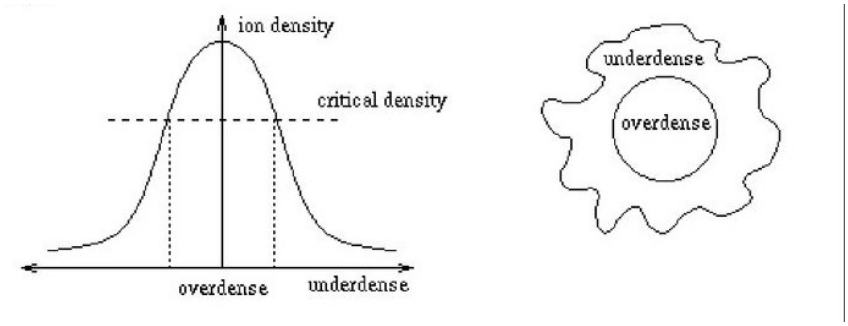
III. Dáta

A. *Rozdelenie meteorov*

Odrážanie rádiových vĺn meteorom záleží od hustoty voľných elektrónov v jeho stope. Vo všeobecnosti sú známe dva základné prípady, a to veľmi nízke a veľmi vysoké elektrónové hustoty

meteoru. Meteory s nízkou hustotou budeme nazývať *Underdense* a naopak, tie s vysokou hustotou budú ďalej spomínané ako *Overdense*.

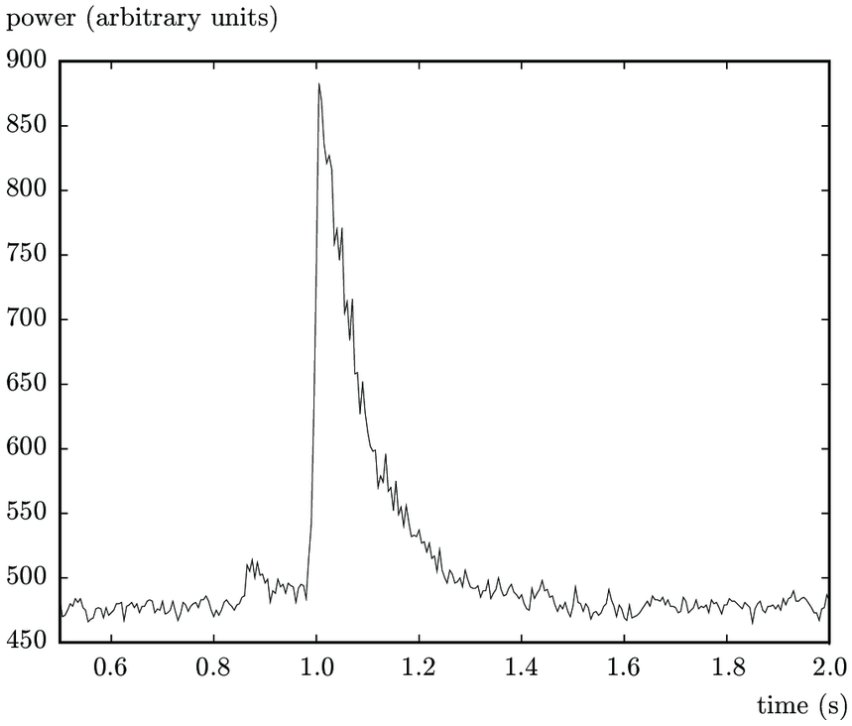
Plazmová frekvencia, ktorá je úmerná hustote iónov, vykazuje normálne rozdelenie. Stred stopy, kde je plazmová frekvencia vyššia ako použitá rádiová frekvencia, je *Overdense* plazma, viď Obrázok 2.



Obrázok 2 Rozloženie hustoty iónov v stope meteoru. Zdroj: [8]

1) Nízkohustotné – Underdense: Nízkohustotné meteory predstavujú slabé, teda menej výrazné meteory. Ich hornou hranicou je približne 2×10^{14} elektrónov na meter. To značí hustotu čiar všeobecne spojenú s meteormi piatej magnitúdy [8].

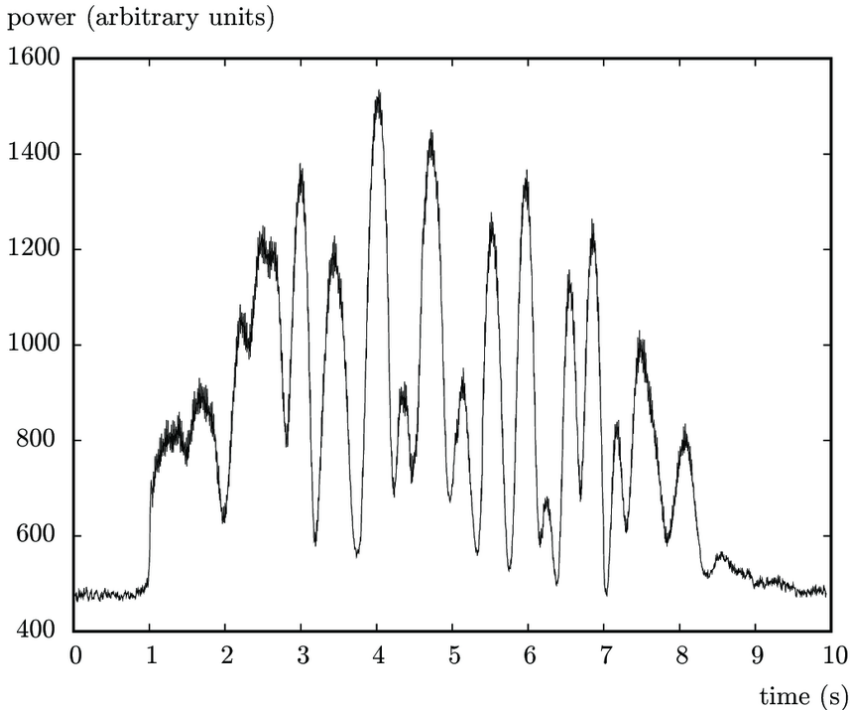
Na Obrázku 3 môžeme vidieť zachytený signál prijatý odrazom na nedostatočne hustých stopách. Signály z týchto nedostatočne hustých meteorov v skutočnosti netrvajú dlhšie ako niekoľko desatín sekundy. Tieto javy sú veľmi početné, čo umožňuje použitie štatistických redukčných techník.



Obrázok 3 Typický meteor s nízkou hustotou, pozorovaný zariadením RAMSES. Zdroj: [8]

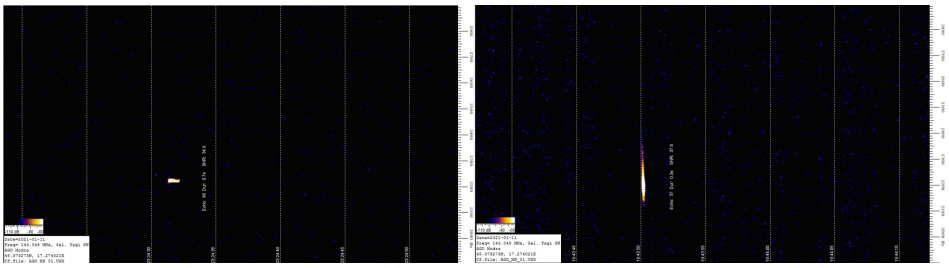
2) Prehustené – Overdense: Tieto meteory majú nižšiu početnosť, ale predstavujú jasnejšie meteory. Ich odrazy môžu trvať až niekoľko sekúnd. To ale spôsobuje problémy, pretože počas dlhého časového obdobia môže dochádzať k deformáciám spôsobeným vetrom. To má za následok to, že príliš hustá stopa môže odrážať rádiové vlny do miesta prijímača, aj keď pôvodne nebola vhodne orientovaná na zrkadlový odraz. Spôsobuje to aj fakt, že vývoj výkonu signálu je väčšinou nepredvídateľný a často aj veľmi nepravidelný [8].

Na Obrázku 4 je zobrazená typická stopa pozorovania takéhoto typu meteoru. Silné oscilácie na obrázku, čiže hlboké vyblednutie sú dôsledkom skreslenia stopy vetrom na rôznych častiach jeho trasy.



Obrázok 4 Typický meteor s prehustením, pozorovaný zariadením RAMSES. Zdroj: [8]

3) Hlava ozveny – Headecho: Niektoré meteory, ktoré prejavujú veľmi jasnú stopu, môžu vykazovať *headecho*, teda hlavu ozveny[1]. Dochádza k tomu vtedy, keď sa od plazmového oblaku, ktorý obklopuje meteorit vstupujúci do atmosféry, odrazia rádiové vlny. Vyskytujú sa najčastejšie vtedy, keď majú meteority veľmi nízky sklon dráhy k atmosfére. Plazmový oblak sa pohybuje spolu s meteoritom jeho rýchlosťou. Na Obrázku 5 je znázornené porovnanie ozveny bez headecha vľavo a ozveny s headechom v pravo, ktorý sa na dátach prejavuje zvislým natiahnutím po osi y [8].



Obrázok 5 Porovnanie meteoru bez *headecha* s meteorom s headechom

Všetky získané prípady, s ktorými sa v tomto článku pracuje boli zachytené prijímačom, ktorý sa nachádza na observatóriu *AGO Modra-Piesok*. Toto observatórium je profesionálnym výskumným, a taktiež vzdelávacím pracoviskom, ktoré spadá pod Fakultu matematiky, fyziky a informatiky Univerzity Komenského v Bratislave [9].

V Európe taktiež existuje sieť rádio-meteorických pozorovaní, známa ako RMOB (angl. *Radio Meteor Observing Bulletin*), kde by sa mohlo využiť aj naše finálne riešenie. Viac o tejto sieti je možné dočítať sa na tomto linku[2].

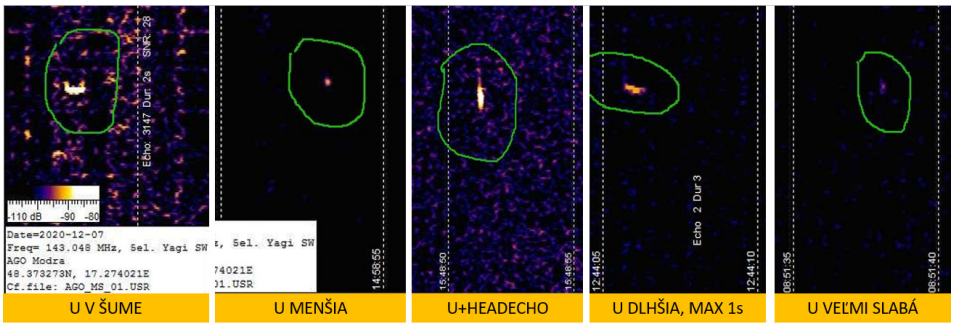
B. Pochopenie dát

Dáta obsahujú informácie o ozvenách, vrátane trvania, dátumu, frekvencie, parametroch a polohy zhotovenia snímky. Pre nás sú kľúčové informácie o dĺžke trvania, ktoré analyzujeme podľa pozície ozveny voči x-ovej osi s pravidelnými intervalmi 5 sekúnd na obrázku. Frekvencie ozvien sa jednoducho určujú na y-ovej osi s hodnotami od 1 900Hz do 2 900Hz.

[1]V ďalšom texte budeme používať anglický ekvivalent – *headecho*.
[2]https://www.rmob.org/index.php

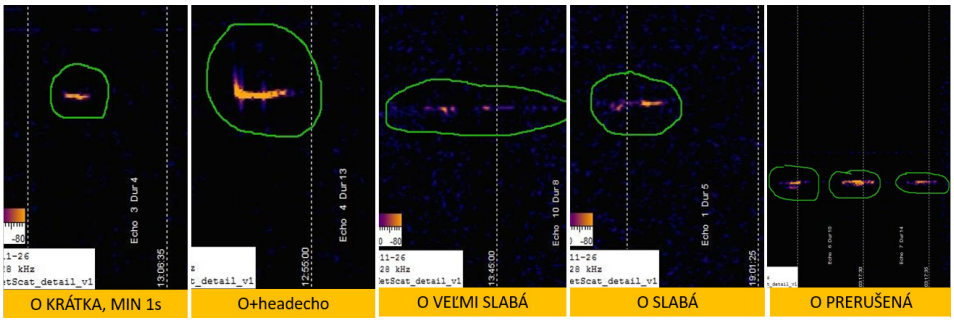
Dáta sme rozdelili do štyroch skupín podľa dĺžky ozvien a existencie headecha, v rozmedzí frekvencií 2 200Hz až 2 300Hz. Napriek tomu sme identifikovali falošné javy, ako umelý signál a družice, ktoré spôsobili problémy pri detekcii ozvien, najmä v zvýšenom šume.

Výskyt headecha na meteoroch sa pozoruje cez ich zvislé natiahnutie na osi y. Prvý druh meteorov sú krátke, tiež označované ako ozveny *typu U* (angl. *underdense*), ktoré sa môžu vyskytovať aj s obsahom headecha. Ich maximálna dĺžka trvania je do jednej sekundy. Pri krátkych a slabých ozvenách, ako je možné vidieť na obrázku 6, môže dochádzať k nesprávnym označeniam, či k ich zámene so šumom. Preto bolo potrebné venovať ich výskytu väčšiu pozornosť.



Obrázok 6 Druhy krátkych ozvien U. Zdroj: P.Zigo, AGO Modra

Druhou skupinou meteorov sú dlhé, nazývané tiež ozveny *typu O* (angl. *overdense*). Tie sa v mnohých prípadoch taktiež vyskytujú aj s headechom. Ako je znázornené na obrázku 7, ich minimálna dĺžka je jedna sekunda.



Obrázok 7 Druhy dlhých ozvien O. Zdroj: P.Zigo, AGO Modra

Dáta boli anotované na platforme Zooniverse študentmi Technickej univerzity v Košiciach s cieľom označiť ozveny meteorov na obrázku. Študenti odpovedali na otázku o prítomnosti ozveny a v prípade kladnej odpovedi označili ozvenu červeným rámčekom pre krátke ozveny a bielym rámčekom pre dlhé ozveny. Potom zvolili vlastnosť ozveny, či obsahuje alebo neobsahuje headecho.

Výstupom bolo 5 757 anotácií, ktoré boli uložené a očistené od nepotrebných dát. Následne boli spracované, kontrolované a odstránené chybné údaje, identifikované ručným prehľadávaním dát na predchádzanie chybného učenia neurónovej siete. Kód na odstránenie duplicitných záznamov znížil počet jedinečných pozitívne označených javov na 2 023 obrázkov.

Dáta boli rozdelené na trénovaciu, validačnú a testovaciu množinu v pomere 70:10:20.

IV. Modelovanie

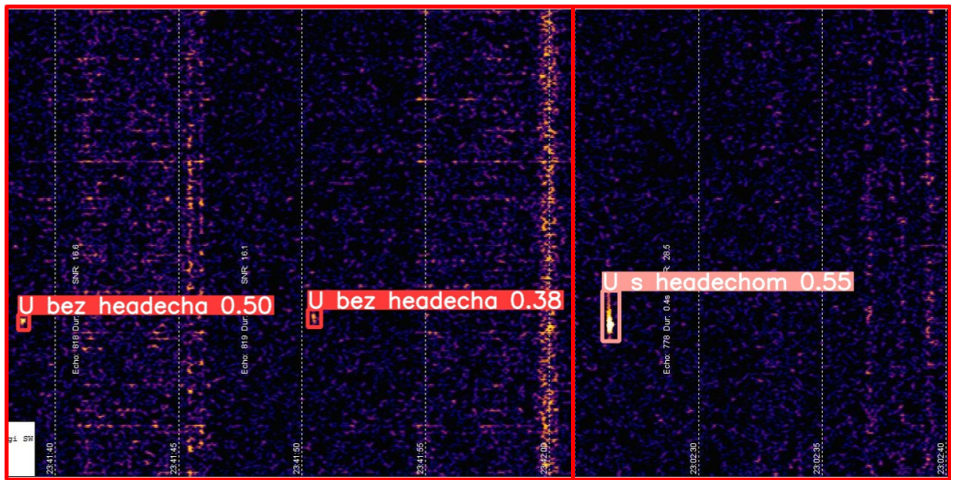
Po spracovaní dát a ich úspešnej príprave na modelovanie prostredníctvom detekčnej neurónovej siete YOLO sa v úvode použil programovací jazyk Python.

Na začiatku učenia modelu bolo potrebné prostredníctvom trénovacej množiny neurónovú sieť natrénovať. Boli vyskúšané mnohé verzie trénovania s cieľom nájsť také, ktoré dokáže neurónovú sieť naučiť detegovať jednotlivé prípady s čo najnižšou chybovosťou. Tieto kombinácie sa týkali použitia rôznych hodnôt parametrov `IoU` a `conf`. *IoU* parameter sa zameriava na porovnávanie jednotlivých boxov tried daných predikcií. Vypočíta sa ako podiel prieniku boxov a ich zjednotenia. V prípade *conf* sme si nastavili minimálnu akceptovateľnú hodnotu pravdepodobnosti zaradenia do triedy.

Samotnému trénovaniu dát predchádzalo vytvorenie yaml súboru potrebného pre uloženie jednotlivých ciest k detailom o trénovacích, validačných a testovacích dátach. Yaml súbor obsahoval informácie aj o počte tried, ktorý bol zapísaný prostredníctvom `nc: 4`, a taktiež o ich názvoch, znázornených ako `names: ['U bez headecha','U s headechom','O bez headecha','O s headechom']`. Výsledné trénovanie sa zastavilo na počte 198 zo 400 epoch. Zastavené bolo z toho dôvodu, že za posledných 100 epoch model nezaznamenal žiadne zlepšenie pri učení. Daný počet epoch sa trénoval vyše 2.5 hodiny. Trénovanie prebiehalo na GPU Tesla K40c.

V. Vyhodnotenie

Pri testovaní rôznych kombinácií nastavení kódu sa najlepšie výsledky dosiahli na dátach s nastavením parametrov IoU na 0.3 a conf na 0.24. Toto nastavenie ukázalo optimálne porovnanie skutočných pozitívnych, falošne pozitívnych, skutočných negatívnych a falošne negatívnych prípadov. Experimenty s rôznymi kombináciami parametrov IoU a conf, spolu s testovaním rôznych trénovacích množín, ukázali, že príliš vysoký parameter conf viedol k nižšiemu počtu detekovaných ozvien. Pri finálnych nastaveniach model správne identifikoval 392 z celkových 408 prípadov. Celkovo bolo 512 pozitívnych javov, 22 negatívnych, pričom model predpovedal 508 pozitívnych a 26 negatívnych prípadov. Príklad detekcie meteorov na rádiových spektrogramoch je zobrazený na obrázku 8.



Obrázok 8 Príklad detekcie meteorov na rádiových spektrogramoch

VI. Záver

Je pre nás veľkým prínosom pracovať na práci, ktorej idea vznikla vďaka spolupráci so Slovenskou akadémiou vied a Fakultou matematiky, fyziky a informatiky Univerzity Komenského v Bratislave.

Vďaka tejto spolupráci vzniká model umelej inteligencie, ktorá sa každým dňom vo svete zlepšuje, vyvíja a učí. Jej využitie je dnes možné pozorovať už naozaj v takmer všetkých oblastiach. Prichádza možnosť prispieť k tomuto učeniu aj vo smere vedy, ktorý doposiaľ možno nebol dostatočne rozvinutý. To nielenže uľahčí prácu pracovníkom z Univerzity Komenského, ale jeho detailnosť môže prispieť k zlepšeniu samotnej detekcie, a to napríklad vďaka zachyteniu javov, ktoré by ľudské oko ináč nezaznamenalo. Taktiež môže byť daný model nasadený na spracovanie veľkého množstva vstupov, čo by nebolo možné len za použitia ľudských síl.

Výhodami tohto modelu je vysoká presnosť detekcií. Model sa dokázal až na zopár chýb naučiť rozoznávať to, či daný objekt je, alebo nie je meteor. Dokáže určiť jeho typ z pohľadu dĺžky, taktiež z pohľadu existencie headecha. Výstup poskytuje používateľovi zhrnutie dát v podobe tabuľky, ktorá obsahuje všetky dôležité informácie o danej ozvene, či už sa jedná o časovú informáciu, dátum výskytu, konkrétne parametre *bounding boxu*, či presnosť, akou si je model istý, že daný objekt je ten, ktorý detegoval.

Poďakovanie

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva Slovenskej republiky a SAV v rámci projektu VEGA 1/0685/21.

LITERATÚRA

[1] S. Wang, L. Niu, and N. Li, "Research on image recognition of insulators based on yolo algorithm," in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 3871–3874.
[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
[3] D. Thuan, "Evolution of yolo algorithm and yolov5: the state-of-the-art object detection algorithm," 2021.
[4] M. H. Yap, R. Hachiuma, A. Alavi, R. Brüngel, B. Cassidy, M. Goyal, H. Zhu, J. Rückert, M. Olshansky, X. Huang *et al.*, "Deep learning in diabetic foot ulcers detection: a comprehensive evaluation," *Computers in Biology and Medicine*, vol. 135, p. 104596, 2021.
[5] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 390–391.
[6] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759–8768.
[7] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, p. 217, 02 2021.
[8] J.-M. Wislez, "Forward scattering of radio waves off meteor trails," in *Proceedings of the International Meteor Conference, 14th IMC, Brandenburg, Germany, 1995*, 1996, pp. 99–117.
[9] P. Zigo, "Kolonica-modra: the new system for forward-scatter radio meteor observations," in *Proceedings of the International Meteor Conference*, 2018, pp. 76–78.

Príklad aplikácie pre podporu logopedických hier

[1]*Renát HALUŠKA,* [2]*Adam REŇÁK,* [3]*Matúš PLEVA,* [4]*Eva KUPCOVÁ*

[1,2,3,4] Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Letná 9, 004001 Košice, Slovenská republika

[1]renat.haluska@tuke.sk, [2]adam.renak@student.tuke.sk,
[3]matus.pleva@tuke.sk, [4]eva.kupcova@tuke.sk

***Abstrakt* — Tento článok sa zaoberá vývojom webovej aplikácie v prostredí Electron na podporu logopedických hier a cvičení, ktoré sa zameriavajú primárne na rozpoznávanie reči. Článok poskytuje stručný prehľad k problematike logopedických hier a významu interaktívnych aplikácií v terapii a identifikuje potreby detí s logopedickými výzvami a ich rodičov prostredníctvom analýzy cieľovej skupiny.**

***Kľúčové slová* — Logopedické hry, Webová aplikácia, Web Speech API**

I. Úvod

Reč považujeme za komunikačný prostriedok, pričom ide o schopnosť, ktorá je daná iba človeku. Ak dôjde k narušeniu tejto schopnosti, nastávajú problémy vo viacerých oblastiach života. Dôsledkom je aj sťažený vzdelávací proces. Príčiny porúch reči môžu byť organické, sem patrí napríklad krvácanie do mozgu, trombóza, nádorové choroby alebo funkčné, kde spadajú napríklad neurotizácia zo strany prostredia, nesprávny rečový vzor. Narušenou komunikačnou schopnosťou jedinca sa zaoberá logopédia. Medzi najbežnejšie poruchy reči u detí patrí oneskorený vývin reči, dyslália, zajakavosť, mutizmus [1].

Logopedické terapie sú veľmi dôležité pri liečbe detí s poruchami reči. Liečbou je možné úplne odstrániť alebo aspoň zmierniť následky narušenej komunikačnej schopnosti. Dieťa absolvuje dlhé logopedické terapie, ktoré môžu trvať niekoľko mesiacov až rokov v závislosti od úrovne postihnutia. Terapie sú zvyčajne raz do týždňa, pričom na dosiahnutie pokroku je rozhodujúce časté a pravidelné cvičenie, ktoré napomáha k osvojeniu si nových zručností a návykov [2]. To však môže byť časovo náročné a dieťa môže po určitej dobe strácať chuť na precvičovanie. Okrem toho, nie každé dieťa má v blízkosti svojho bydliska kliniku, ktorá sa špecializuje na logopedické problémy a cestovanie do vzdialenejšieho mesta niekoľkokrát do mesiaca môže byť unavujúce, vyčerpávajúce, ale aj finančné náročné. Preto je vhodné terapiu doplniť aj o domáce cvičenia a tak urýchliť a zefektívniť liečbu [3].

Pri domácom cvičení existuje niekoľko problémov. Jeden z nich je, že deti musia byť motivované cvičiť, keďže cvičenia sú mnohokrát monotónne a opakujúce sa. Druhý problém je, že pri vykonávaniach logopedických cvičení musia byť rodičia prítomní a dohliadať a zároveň poskytovať deťom spätnú väzbu. Tieto faktory znižujú účinnosť a zároveň predlžujú celkovú dĺžku liečby [4].

II. Špecifikácie Cieľovej Skupiny pre Aplikáciu

Každá hra má špecifickú cieľovú skupinu užívateľov. Logopedické hry môžu zlepšiť rôzne aspekty reči a môžu byť zamerané na rôzne logopedické problémy. Hry môžu byť zamerané na výslovnosť slov, výslovnosť písmen alebo doladenie výslovnosti špecifických slov a viet. Aplikácie zároveň musia byť prispôsobené slovnej zásobe dieťaťa a musia byť prispôsobené aj ich úrovni technologickej gramotnosti [5].

Aplikácia sa zameriava na deti s narušenou zvukovou rovinou reči (dyslália), ktorá sa dnes ale bližšie špecifikuje ako artikulačná alebo fonologická porucha, prípadne ako vývinová dyspraxia. Pod pojmom narušená zvuková rovina reči rozumieme neschopnosť používať

jednotlivé hlásky či skupiny hlások. Narušenie pôsobí rušivo na komunikačný zámer a môže sa prejavovať znížením zrozumiteľnosti reči. Pri tejto rečovej chybe sa stretávame s problémom, že dieťa vynecháva, poprípade nahrádza hlásku inou alebo ju tvorí nesprávne. Medzi najproblematickejšie hlásky v slovenčine patria L, R, C, S, Z, DZ, Č, Š, Ž, DŽ. Pri odstraňovaní nesprávnej výslovnosti je potrebné dodržiavať určité zásady, ako je napríklad zásada krátkodobého cvičenia. Odporúča sa 3-4 minúty, niekoľkokrát denne. Týmto cvičením sa docieli, že z podmieneného reflexu sa vytvorí nepodmienený. Pri pravidelnom cvičení podľa pokynov logopéda je možné dosiahnuť efekt správnej výslovnosti hlásky po 3-4 mesiacoch cvičenia. Správna výslovnosť má byť ukončená pred začatím školskej dochádzky. Prax ukazuje, že čím je pacient starší, tým je problematickejšie zapojenie správnej výslovnosti do bežnej reči [6].

Na základe toho vieme, že cieľová skupina pre aplikáciu sú deti vo veku 3-6 rokov, ktoré majú problémy s vyslovovaním určitých slabík. Aplikácia by mala obsahovať hru o dĺžke troch až štyroch minút, kde dieťa precvičuje vyslovovanie slov s problematickými slabikami.

III. Príklady Exitujúcich Projektov v Oblasti Vzdelávania

Príklady existujúcich projektov v oblasti riešenia logopedických problémov detí predstavujú dôležitý zdroj inšpirácie a poznania existujúcich prístupov v tejto problematike. V súčasnosti existuje množstvo hier, ktoré sa zameriavajú na logopedické problémy u detí. Mnohým z nich však chýba spätná väzba. V tejto kapitole sa pokúsim ukázať príklady najlepších aplikácií, ktoré sa zameriavajú na tento problém. Jedná aplikácia je zameraná na anglicky hovoriace deti a druhá na slovenské deti s problémom rozpoznávania hlások.

A. Apraxia World

Logopedická hra Apraxia World je postavená na báze plnohodnotnej dvojrozmernej plošinovky, ktorá by mala byť neskôr spojená s modulom automatického rozpoznávania reči, schopným poskytovať spätnú väzbu v reálnom čase. Hra je farebná adventúra, vyvinutá v hernom engine Unity. V hre Apraxia World má dieťa za úlohu ovládať avatara (malú postavu opice), vo viacúrovňovom svete, v ktorom sa musí dostať do cieľa cez rôzne plošiny a jaskyne, pričom musí zbierať mince a zároveň sa vyhýbať nepriateľom a prekonávať prekážky na mape. Dieťa ovláda avatara pomocou štandardných vstupov, čo sú virtuálne tlačidlá a smerový ovládač. Rečové vstupy sú viazané na predmety, ktoré musí hráč počas hry zbierať. Tieto predmety sú povinné aby hráč mohol postúpiť do vyššej úrovne. Predmety, ktoré vyvolávajú rečové vstupy sú rozlíšiteľné od ostatných predmetov, preto hráč dokáže predvídať a kontrolovať, kedy sa rečové cvičenia objavia a tak rečové cvičenia neodvádzajú pozornosť od hrania a ani neprerušujú hráča pri vykonávaní zložitejších pohybov. Pre každú úroveň je definovaný potrebný počet rečových cvičení. Slová v úlohách sa náhodné vyberajú zo zoznamu slov tak, že sa neopakujú, kým sa nevystriedajú všetky slová zo zoznamu [7].

B. Squak

Squak je aplikácia určená pre slovenské deti, ktorá pomáha s rozoznávaním podobných slovíčok. Ide o cvičenia na zvukovú diferenciáciu u detí. Hlavnou postavou hry je papagáj Squak, ktorý vyslovuje podobné slová a dieťa má na základe prehraných nahrávok určiť, či išlo o rovnaké slová alebo sa slová len podobali. Cieľom hry je naučiť deti rozlišovať hlásky.

Priebeh hry je vcelku jednoduchý. Papagáj vysloví dve slová, ktoré sú veľmi podobné. Potom má dieťa za úlohu určiť, či tieto slová boli identické alebo sa líšili. Ak správne odpovie, získa pierko. V prípade nesprávnej odpovede sa slová zopakujú a dieťa môže znova reagovať. Cieľom hry je získať 3 pierka každej farby.

V hre je dostupná možnosť výberu medzi použitím skutočných slov alebo generovaním vymyslených slabík a slov. Tento výber poskytuje široké spektrum cvičení a umožňuje prispôsobiť hru podľa individuálnych preferencií a potrieb. Hra bola vyvinutá v hernom engine Unity [8].

IV. Návrh Aplikácie a funkcií

Pri navrhovaní webovej aplikácie pre deti sa stanovil hlavný cieľ, vytvoriť prostredie, ktoré bude nielen jednoduché a intuitívne, ale aj dostatočne atraktívne, aby zaujalo našu cieľovú skupinu a zároveň, ktoré berie ohľad na potreby a schopnosti cieľovej skupiny. Aplikácia ma osloviť deti a preto je dizajn viac zameraný na kreslený svet, ktorý je pre deti viac atraktívnejší a príťažlivejší.

Samotná webová aplikácia pozostáva z dvoch hlavných stránok, ktoré sú navrhnuté tak, aby užívateľovi poskytovali plynulý a príjemný zážitok. Prvý kontakt s aplikáciou užívateľ získava

na úvodnej stránke, ktorá slúži ako vstupný bod do aplikácie. Tu sa nachádzajú prvky, ako tlačidlá na spustenie hry alebo získania informácií. Úvodná stránka je navrhnutá tak, aby užívateľ okamžite dokázal nájsť všetky potrebné informácie a zároveň aby dieťa vedelo hru spustiť aj samostatne bez pomoci rodiča. Druhou časťou aplikácie je stránka, na ktorej sa už nachádza samotná hra. Táto stránka predstavuje hlavný obsah a zábavnú časť celého prostredia.

A. Princíp Hry

Princíp hry je v podstate veľmi jednoduchý a kladie sa dôraz hlavne na splnenie jednotlivých rečových cvičení. V hre dieťa prevezme úlohu baníka, ktorý sa vydáva do hlbín podzemia a snaží sa nájsť a vykopať rôzne poklady. Jeho úlohou je zbierať diamanty, zlato a ďalšie vzácne drahokamy. Ovládanie je intuitívne a jednoduché, čo umožňuje hráčovi sústrediť sa na samotnú hru. Pomocou šípiek na klávesnici, poprípade klávesnicami WASD, sa hráč pohybuje postavou po svete, pričom sa snaží nájsť cestu k pokladom. Kopanie prebieha stlačením klávesy medzerník.

V hre sa okrem kameňa nachádzajú dokopy tri druhy vzácnych drahokamov, ktoré hráč musí vykopať aby dosiahol svoj cieľ a vyhral hru. Najbežnejším drahokamom v hre je zlato, ktoré hráč získa dvojitým stlačením medzerníka a tým ho vykope. Ďalším predmetom, ktorý musí hráč vykopať je diamant. Na jeho vykopanie však dieťa musí splniť rečové cvičenie. V prípade, že rečové cvičenie nesplní, môže sa o vykopanie diamantu pokúšať znova, no opäť musí splniť rečové cvičenie. Posledným drahokamom v hre je fialový kryštál. Tento predmet indikuje posluchové cvičenie a funguje na rovnakom princípe ako diamant.

Aby dieťa úspešne dokončilo level musí získať všetky drahokamy a tým pádom dokončiť všetky rečové cvičenia. Počet jednotlivých drahokamov potrebných na dokončenie hry je vyznačený v bočnom informačnom panely. Tento stavový panel slúži ako ukazovateľ pokroku ale aj ako motivácia pre hráča. Každý úspešne vykopaný drahokam sa pridáva do zbierky.

B. Generovanie Sveta a Herných Prvkov

Celý proces generovania herného sveta a objektov v hre ako aj funkčnosť samotnej hry má na starosť JavaScript. Prvý krok pri vytváraní hry predstavuje inicializácia herného sveta. Sú definované základné parametre, ako je veľkosť mapy a veľkosť jednotlivých blokov, ktoré majú vypĺňať herný priestor. Tieto údaje slúžia ako východiskové body pre následne procesy generovania herného prostredia.

Generovanie terénu je kľúčovým prvkom vytvárania herného sveta. Použitý je algoritmus generovania náhodných čísel na umiestnenie rôznych objektov po celej hernej mape. Každý objekt má pridelené špecifické parametre a súradnice, ktoré ovplyvňujú jeho výskyt či význam. Implementácia hlavnej postavy, hráča, zahŕňa tvorbu herného ovládania a pohybu postavy po mape. Pomocou klávesnice má hráč možnosť ovládať pohyb postavy a interakciu s objektmi. Kód hry myslí na detekciu kolízií a interakciu s objektmi. Sú použité mechanizmy, ktoré umožňujú hráčovi reagovať na predmety v hre. Napríklad, ak hráč narazí na prekážku, kód reaguje zabraňuje hráčovi v prejdení cez ňu.

Vizuálna reprezentácia v hre je dôležitá pre užívateľský zážitok a preto je každý objekt vyobrazený pomocou obrázkov, ktoré jasne identifikujú ich typ a význam. V aplikácií sú použité aj zvukové efekty, čo prispieva k zlepšeniu herného zážitku. Zvukové signály sú spúšťané pri rôznych udalostiach v hre. Srdcom hry je hlavný herný cyklus, ktorý zabezpečuje plynulý chod. Tento cyklus prebieha nepretržite a herné prostredie tak dokáže reagovať na vstupy hráča a aktualizovať stav herného sveta.

C. Rečové a Posluchové Cvičenia

V hre sa vyskytujú dve základne formy cvičenia, ktoré sú zamerané na rozvoj jazykových a rečových schopností ale aj na rozvoj posluchových zručností dieťaťa. Cvičenie zamerané na reč sa sústreďuje na správnu výslovnosť slov a spúšťa sa pri vykopaní diamantov. Druhým cvičením je posluchové cvičenie, ktoré sa zameriava na schopnosť rozoznávať podobné slová, slabiky a písmena. Toto cvičenie sa spúšťa pri vykopaní kryštálu. Aplikácia je navrhnutá tak, aby dieťa muselo pre dokončenie hry absolvovať tieto cvičenia.

Rečové cvičenia sa spúšťa v momente, keď sa hráč pokúsi vykopať diamant. Následne sa zobrazí dialógové okno na ktorom sa nachádza obrázok slova s príslušným textom. V okne (Obr. 1) je tlačidlo, ktoré slúži na nahratie zobrazeného slova. Tlačidlo je v bežnom, neaktívnom, móde zelené s bielym textom.



Obr. 1 Príklad rečového cvičenia

Po vyslovení slova a vyhodnotení nahrávky sa v dialógovom okne zobrazí výsledok. V prípade nesprávneho vysloveného slova sa zobrazí červené „X“, indikujúce zlú odpoveď. Hráč má tri pokusy na správne vyslovenie slova. Ak sa mu to nepodarí, dialógové okno sa zatvorí a hráč musí diamant vykopať znova. V prípade správne vysloveného slova sa zobrazí zelená kvačka a cvičenie pokračuje ďalším slovom. Cvičenie pozostáva z dvoch kôl, čo znamená, že hráč musí správne vysloviť dve slová, aby sa diamant úspešne vykopal.

Pri posluchovom cvičení sa hráč pokúša správne určiť, či vyslovené slová boli rovnaké alebo rozdielne. Ide o jedno z úvodných cvičení v logopédii, ktoré musí dieťa zvládnuť pre ďalší pokrok vo vývoji reči. Samotné cvičenie sa spúšťa pri pokuse hráča vykopať fialový kryštál. Priebeh cvičenia je podobný ako pri rečovom cvičení. Takisto sa otvorí dialógové okno, v ktorom sú však dve tlačidlá. Jedno označuje „rovnaké“ a druhé zase „rozdielne“ slová. Hráčovi sa pri tomto cvičení prehrávajú slová. Najprv jedno a hneď po ňom druhé. Následne dieťa musí určiť, stlačením na tlačidlo, či boli slová rovnaké. Slová sa prehrávajú pomocou knižnice Howl. Hráč musí z piatich pokusov minimálne trikrát správne odpovedať. V prípade že sa mu to nepodarí, cvičenie musí zopakovať.

V. Grafický Dizajn Hry

Grafický dizajn webovej aplikácie je prispôsobený cieľovej skupine a má osloviť deti. Preto je dizajn viac zameraný na kreslený svet, ktorý je pre deti viac atraktívnejší a príťažlivejší. Snahou bolo vytvoriť jednoduchý a intuitívny dizajn, v ktorom sa bude dieťa vedieť zorientovať.

A. Úvodná stránka

Vstupným bodom webovej aplikácie je úvodná stránka (Obr. 2), ktorá slúži na získanie informácií o samotnej hre. Takisto sa cez úvodnú stránku spúšťa aj samotná hra. Celkový dizajn je jednoduchý
a dieťa si dokáže spustiť hru aj samo, nepotrebuje pomoc rodiča. Na úvodnej stránke sa nachádzajú len dve tlačidlá. Jedným sa spúšťa hra, druhým sa otvára dialógové okno, ktoré zobrazuje informácie o hre, pravidlá či návod na ovládanie postavy v hernom svete. Dizajn je prispôsobený aj pre mobilné zariadenia.



Obr. 2 Úvodná stránka webovej aplikácie

B. Stránka s hrou

Stránka s hrou je rozdelená do dvoch hlavných blokov, ktoré poskytujú užívateľovi prehľadné a funkčné prostredie. Na ľavej strane sa nachádza informačný panel, ktorý obsahuje informácie o priebehu hry. Tento panel slúži na zobrazovanie informácií o hre, ako je napríklad počet vyťaženého zlata. Na pravej strane sa nachádza herné pole o maximálnej veľkosti 800x800px. Hráčova postava je zobrazená ako malý, ale detailne vypracovaný avatar, ktorý je umiestnený na vrchu poľa. Pri kopaní prebieha animácia hernej postavy, čo pridáva hre dynamiku. Hracie pole vypĺňajú sivé kamene. Zlato, modré diamanty a fialové kryštály sú v poli náhodne umiestnené a pri každom generovaní sveta sa nachádzajú na inej pozícii. Tieto predmety sú zobrazené v dobrej kvalite a hráč ich dokáže jednoducho a jasne identifikovať. Pozadie herného poľa naznačuje podzemné prostredie alebo jaskyňu. Hnedé tóny majú pripomínať pôdu. Tento dizajn sa snaží vytvárať správnu atmosféru a vtiahnuť hráča do hry. Hra ponúka hráčovi aj zvukové efekty kopania, rozbitia kameňa alebo zvukový efekt pri získaní drahokamu.



Obr. 3 Stránka s hrou

C. Mobilná verzia

Dizajn stránky s hrou je prispôsobený pre aj mobilné zariadenia (Obr. 4). V menšom rozlíšení, typickom pre mobilné zariadenia, sa bočný panel zmenší a presunie na vrchnú časť obrazovky. Zmenší sa aj herné okno, ktoré sa naopak, presunie nižšie. Herné okno zostáva dostatočne veľké na to, aby hráčom poskytovalo jasný a detailný pohľad na hru. Jedným z najdôležitejších aspektov mobilného dizajnu sú ovládacie prvky, ktoré sú zobrazené priamo na obrazovke. V mobilnom rozlíšení sú tieto prvky umiestnené v spodnej časti obrazovky, čo umožňuje hráčom ľahký prístup palcami pri vertikálnom držaní zariadenia. Šípky na ovládanie pohybu a ďalšie interaktívne tlačidlá sú dostatočne veľké, aby ich bolo možné pohodlne používať aj na dotykových obrazovkách menších rozmerov.



Obr. 4 Hra prispôsobená pre mobilné rozlíšenie

VI. ZÁVER

Tento článok sa zameriava na vývoj multiplatformovej desktopovej webovej aplikácie využívajúcej prostredie Electron a Web Speech API s cieľom podporiť logopedické cvičenia pre deti s rečovými poruchami. Prostredníctvom analýzy potrieb cieľovej skupiny sme identifikovali kľúčové aspekty, ktoré sú nevyhnutné pre efektívnu logopedickú terapiu a implementovali sme ich do návrhu užívateľského rozhrania a funkcií aplikácie.

Hlavným prínosom tejto práce je vytvorenie platformy, ktorá kombinuje technologické riešenia s logopedickými potrebami detí. Výsledná aplikácia poskytuje intuitívne a interaktívne prostredie, ktoré motivuje deti k pravidelnému cvičeniu. Okamžitá spätná väzba umožňuje dosahovať v hre pokrok a tým udržiavať pozornosť detí. Vďaka vhodne zvoleným technológiám pri vývoji je aplikácia dostupná na rôznych zariadeniach, čo zabezpečuje jej širokú dostupnosť a jednoduché používanie pre deti aj rodičov.

Budúci vývoj by sa mal zamerať na zlepšenie a spresnenie rozpoznávania reči napríklad konvolučnými neurónovými sieťami. Ďalšie navrhované aktivity zahŕňajú rozsiahlejšie testovanie aplikácie v reálnom prostredí, zber spätnej väzby od používateľov a jej následné implementovanie do vylepšení aplikácie alebo vytváranie nových rečových cvičení pre rôzne stupne postihnutia reči.

POĎAKOVANIE

Táto práca bola čiastočne podporená Vedeckou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky a Slovenskou akadémiou vied v rámci projektu VEGA 2/0165/21, Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu,

vývoja a mládeže Slovenskej republiky cez projekt KEGA 049TUKE-4/2024, čiastočne z projektu FakeDetect-Early Fake Profile Detection FBR-PDI-019 financovaného z Nórskych fondov a EHP 2014-2021, a čiastočne v rámci projektov podporených Agentúrou na podporu výskumu a vývoja APVV-22-0261 a APVV-22-0414.

REFERENCIE

[1] E. Q. Rivas and E. S. Molina, "A proposal for a virtual world that supports theraphy of dyslalia," 2012 6th Euro American Conference on Telematics and Information Systems (EATIS), Valencia, Spain, 2012, pp. 1-4, doi: 10.1145/2261605.2261663.

[2] M. Santiputri, E. B. Sembiring, N. Z. Janah and M. K. Mufida, "Abangmanis: Speech Theraphy for Autism Monitoring Mobile Application," 2019 2nd International Conference on Applied Engineering (ICAE), Batam, Indonesia, 2019, pp. 1-6, doi: 10.1109/ICAE47758.2019.9221735.

[3] N. Nasiri, S. Shirmohammadi and A. Rashed, "A serious game for children with speech disorders and hearing problems," 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), Perth, WA, Australia, 2017, pp. 1-7, doi: 10.1109/SeGAH.2017.7939296.

[4] M. Karthan, D. Hieber, R. Pryss and J. Schobel, "Developing a Gamification-Based mHealth Platform to Support Orofacial Myofunctional Therapy for Children," 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), L'Aquila, Italy, 2023, pp. 169-172, doi: 10.1109/CBMS58004.2023.00211.

[5] Y. Getman et al., "Developing an AI-Assisted Low-Resource Spoken Language Learning App for Children," in IEEE Access, vol. 11, pp. 86025-86037, 2023, doi: 10.1109/ACCESS.2023.3304274.

[6] S. N. Sarah Abu Bakar, N. Hafizah Mahamarowi and S. Mustapha, "Game-Based Learning as a Teaching and Learning Tool for Dyslexic Children," 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia, 2022, pp. 50-55, doi: 10.1109/ICSPC55597.2022.10001824.

[7] A. Hair, P. Monroe, B. Ahmed, J. K. Ballard, R. Gutierrez-Osuna, (2018, June). Apraxia world: A speech therapy game for children with speech sound disorders. In Proceedings of the 17th ACM Conference on Interaction Design and Children (pp. 119-131).

[8] "SPEEKLE," Speekle. https://www.speekle.sk/#about

Segmentácia erupčných vlákien

[1] Lenka Kališková, [2]Peter Butka, [3]Lívia Potočňáková

[1,2,3]Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]lenka.kaliskova@tuke.sk, [2]peter.butka@tuke.sk,[3] livia.potocnakova@student.tuke.sk

Abstrakt **– Dôsledky slnečných erupcií majú výrazný vplyv na našu Zem. Ich skúmanie je dôležité pre ochranu našich technológií a bezpečnosť. Erupčné vlákna pozorované počas ich priebehu predstavujú možný kľúč k pochopeniu vzniku tých najintenzívnejších erupcií. Štúdia je zameraná na využitie metód hlbokého učenia v kombinácii so štandardnými technikami spracovania obrazu. Navrhnutý prístup spája využitie konvolučnej neurónovej siete a metód počítačového videnia. Vytvára tak riešenie pozostávajúce z automatickej segmentácie s následnou detekciou paralelných erupčných vlákien, typických pri najintenzívnejších slnečných erupciách. Detegované udalosti budú použité pre ďalší výskum a lepšie pochopenie slnečných erupcií.**

Kľúčové slová **– konvolučné neurónové siete, počítačové videnie, slnečné erupcie, erupčné vlákna**

I. Úvod

Slnečné erupcie (angl. **solar flares**) predstavujú najintenzívnejšie udalosti v blízkom vesmíre. Ich dopady môžu mať výrazný vplyv na Zem, vrátane narušenia satelitných a komunikačných systémov, či možných výpadkov elektrických sietí. Erupčné vlákna (angl. **flare ribbons**) sú jasne pozorovateľné útvary v spodnej časti slnečnej atmosféry, ktoré sa vyskytujú počas slnečných erupcií. Ich paralelný tvar je spojený s najintenzívnejšími slnečnými erupciami [1]. Cieľom výskumu bolo vyvinúť prístup využívajúci hlboké učenie na hľadanie a rozpoznanie paralelného tvaru erupčných vlakien zo satelitných snímok Slnka, keďže tradičné metódy, ako napríklad manuálne pozorovanie expertov, sú časovo náročné a subjektívne. Navrhnutá metóda bola kombináciou konvolučných neurónových sietí pre účely segmentácie slnečných erupcií a tradičných metód spracovania obrazu na efektívne spracovanie veľkého objemu dát.

V rámci erupčných vlakien je dôležita misia Solar Dynamics Observatory (skr. **SDO**), ktorá začala svoje vesmírne pozorovania 11. februára 2010. Je to poloautonómna kozmická loď nepretržite smerovaná k Slnku, ktorá sa zameriava na tri vedecké experimenty.

V kontexte výskumu bol dôležitý experiment Atmospheric Imaging Assembly (skr. **AIA**), ktorý vytvára celodiskové Extreme Ultraviolet (skr. **EUV**) snímky každých desať sekúnd. Sú to snímky rôznych kanálov, ktoré snímajú povrch a atmosféru Slnka, pričom vytvárajú snímky pomocou filtrov v desiatich vlnových dĺžkach – sedem Extreme ultraviole (skr. **XUV**), dve Ultraviolet (skr. **UV**) a jeden pás viditeľného svetla na odhalenie kľúčových aspektov slnečnej aktivity. Každé pásmo vlnovej dĺžky zvýrazňuje inú časť slnečnej atmosféry, a boli zvolené tak, aby pokrývali rozpätie od 6000 K do 3×10^6 K. Snímky kanálu 1600 Å zvýrazňujú hornú fotosféru a prechodovú oblasť nachádzajúcu sa medzi chromosférou a korónou, erupčné vlákna sú na nich viditeľnejšie v porovnaní s inými kanálmi.

Slnečné erupcie zvyčajne v prekurzívnom štádiu spôsobujú vznik erupčných vlákien, ktoré majú špecifický tvar. Tvar, ktorý často nadobúdajú, je nazývaný ako paralelný, pretože je voľným okom viditeľné, že ho tvoria rovnobežky. Matematicky to však presne rovnobežky byť nemusia, môžu sa nachádzať v rôznej vzdialenosti od seba, môžu mať rozdielne sklony a byť rôznej dĺžky. Pre potreby hľadania zadaného tvaru je potrebné získať segmentované slnečné erupcie. Z tohto dôvodu sa výskum skladal z dvoch hlavných častí, a to vytvárania segmentácií a hľadania paralelného tvaru [2].

Pre získanie potrebných snímok bol použitý open-source vizualizačný nástroj Helioviewer. Databáza RibbonDB poskytla zoznam jednotlivých erupcií intenzity triedy C a vyššie, pričom v rámci výskumu boli najdôležitejšie informácie čas celého priebehu slnečných erupcií, ich poloha a trieda.

Priebeh erupcií je v katalógu zaznamenaný troma stĺpcami, ktoré obsahujú dátum a čas pre začiatok erupcie, vrchol erupcie a koniec erupcie. Poloha erupcií je udávaná prostredníctvom

heliografickej šírky (angl. **heliographic latitude**) a heliografickej dĺžky (angl. **heliographic longitude**), pričom obe sú zadané v stupňoch. Trieda erupcie je zapísaná spojením dátumu a času začiatku erupcie, čísla aktívnej oblasti, a konkrétnej triedy.

Získané dáta pôvodne obsahujúce celý disk boli prispôsobené optimálnemu rozostupu a orezané na štvorcový tvar pre vstup do konvolučných neurónových sietí (angl. **Convolutional neural network**, skr. **CNN**). Následne boli vytvorené objekty na reprezentáciu 2D obrazu Slnka, a taktiež na zvýšenie kontrastu a zlepšenie viditeľnosti detailov na snímkach bola použitá normalizácia intenzity obrazu. Napriek nastaveniu rozmerov vytvárajúcich štvorec, boli snímky nepresných tvarov, preto boli dodatočne orezané, aby tvorili presné štvorce. Po orezaní boli obrázky konvertované na čiernobiele a zmenšili sa rozmery na 512×512 pixelov pre lepšiu viditeľnosť tvaru erupčných vlákien. Týmto spôsobom bolo upravených 52 465 snímok, pričom vstupná množina bola náhodne premiešaná. Posledným krokom prípravy dát bolo normalizovanie obrázkov na hodnoty pixelov v rozsahu $< 0, 1 >$ a ich uloženie do poľa.

II. Konvolučné neurónové siete

Konvolučné neurónové siete sú jedným z najpoužívanejších typov hlbokých neurónových sietí. Sú modelované podľa prirodzených biologických nervových systémov a využívajú matematickú lineárnu operáciu konvolúcie, podľa ktorej sú aj pomenované.

Každá CNN obsahuje tri hlavné typy vrstiev, a to vstupnú vrstvu, skryté vrstvy, ktorých môže byť viacero, a výstupnú vrstvu.

Typicky je prvou skrytou vrstvou CNN konvolučná vrstva, ktorá aplikuje na vstupný obrázok filtre pre extrakciu dôležitých príznakov, ako napríklad hrany, čiary, alebo tvary. Konvolúcia je matematická operácia používaná pri spracovaní signálov, alebo obrazov v hlbokom učení. Používa filter (angl. **kernel**), teda dvoj-dimenzionálne pole určitých hodnôt. Tento filter sa počas procesu konvolúcie postupne posúva po vstupnom obraze o hodnotu kroku (angl. **strides**). V každom kroku sa hodnoty pixelov pokrytej časti obrazu násobia príslušnými váhami filtra a tieto produkty sa sčítajú, čím sa vytvorí vážený súčet. Postupne prejde celým vstupom a vytvorí tak nový zmenšený obraz.

Ďalšími typmi vrstiev sú vzorkovacie (angl. **pooling)** a plne prepojené. Výstup konvolučnej vrstvy sa ďalej posúva do vzorkovacej vrstvy, ktorá zmenšuje a zredukuje rozmery príznakovej mapy. Ponecháva najdôležitejšie informácie získané z konvolučnej vrstvy. Dopomáha tomu, aby sieť nebola príliš citlivá na umiestnenie nájdených príznakov. Vstup rozdeľuje na neprekrývajúce sa časti. Vzorkovacie vrstvy bývajú väčšinou typu **max**, pri ktorom sa ponechajú najvyššie hodnoty v rámci jednotlivých častí, alebo **average**, pri ktorom sa hodnoty spriemerujú.

Postup konvolúcie a vzorkovania môže byť niekoľkokrát zopakovaný. Následne sa výstup upraví na jednorozmerný a použije sa ako vstup plne prepojenej vrstvy. Účelom plne prepojenej vrstvy je spočítavanie váh predchádzajúcich vrstiev siete. Táto vrstva indikuje, aké príznaky je potrebné hľadať vo vstupných obrázkoch pre ich správnu klasifikáciu. Okrem spomenutých vrstiev model musí obsahovať aj výstupnú vrstvu, ktorá obsahuje zvolenú aktivačnú funkciu, v závislosti od typu siete a počtu hľadaných tried [3].

A. Architektúra U-Net

U-Net je populárny typ architektúry konvolučných neurónových sietí, ktorý bol pôvodne navrhnutý na segmentáciu biomedicínskych snímok. Bol predstavený v roku 2015 autormi [4]. Odvtedy bol využitý pri rôznych úlohách spracovania obrazu, ako napríklad segmentačné a klasifikačné úlohy, alebo detekcia objektov. Názov U-Net vznikol na základe architektúry takýchto sietí, ktorá pozostáva zo zmršťovacej a zväčšujúcej cesty a pripomína tvar písmena U.

U-Net má symetrickú enkóder-dekóder štruktúru. Časť enkóder získava vysoko-úrovňové vlastnosti zo vstupov aplikovaním konvolučných a vzorkovacích operácií, zatiaľ čo dekóder vykonáva operácie dekonvolúcie. Vďaka dekonvolúcii dokáže obnoviť informácie o mieste a rozmiestnení objektov, inak povedané priestorové informácie vstupného obrazu. Hlavnou inováciou tejto architektúry je doplnenie spojení (angl. **skip connections**), ktoré prepájajú korešpondujúce enkóder a dekóder vrstvy pre zachovanie priestorových informácií a lepšie šírenie vlastností [4].

III. Segmentácia slnečných erupcií

Pre trénovanie segmentačnej neurónovej siete boli vytvorené binárne segmentačné masky použitím metódy prahovania s tromi rôznymi úrovňami prahu v závislosti od veľkosti segmentovanej oblasti. Na prahované snímky bola aplikovaná morfologická operácia zatvorenia. Niektoré snímky s malými alebo žiadnymi segmentáciami boli odstránené z datasetu.

Testovacia množina tvorila 20% vstupných dát a zvyšných 80% bolo rozdelených na trénovaciu a validačnú množinu v pomere 9:1.

Cieľom bolo overiť univerzálnosť modelu SCSS-Net [5], ktorý využíva architektúru U-Net, jeho využitím pre segmentáciu iných oblastí, než sú aktívne oblasti a koronálne diery, na ktoré bol predtrénovaný.

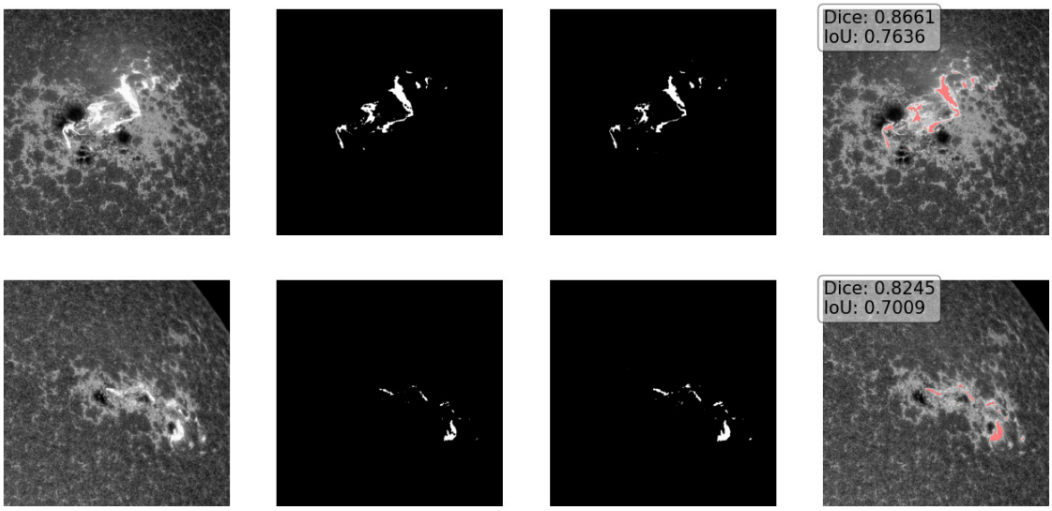
Výsledky segmentácie predtrénovaného modelu neboli uspokojivé, pretože bol pôvodne trénovaný na iné typy oblasti. Masky použité na porovnanie neboli optimalizované, čo mohlo ovplyvniť presnosť, avšak výsledky segmentácie boli veľmi nepresné a model vyžadoval pretrénovanie.

Model bol trénovaný na päťdesiatich epochách s trénovacou množinou a validáciou. Pre vyššiu výpočtovú náročnosť boli vytvorené dva modely. Prvý model bol trénovaný na vstupných dátach o rozmeroch 256×256. Druhý model bol trénovaný na vstupe o veľkosti 512×512.

Výsledky ukázali, že výstupy modelov a vytvorené masky boli takmer totožné, čo naznačuje úspešné naučenie modelov rozpoznávať a segmentovať erupčné vlákna na základe manuálne vytvorených segmentácií.

Modely boli vyhodnotené aj pomocou metrík IoU a Dice, pričom prvý model dosiahol IoU 97.66% a Dice koeficient 95.42%, zatiaľ čo druhý model dosiahol hodnoty nižšie o približne 10%. Porovnanie prahovaných výstupov druhého modelu s manuálnymi segmentáciami vidíme na obrázku 1.

Vzhľadom na dôležitosť zachovania tvaru erupčných vlákien bol zvolený druhý model, ktorý používa vstup o veľkosti 512×512. Výstupy segmentácie modelu boli použité na hľadanie najintenzívnejších slnečných erupcií, počas ktorých majú erupčné vlákna typicky paralelný tvar.



Obrázok 1 Výsledky druhého modelu. Zľava: vstup, manuálna segmentácia, segmentácia modelu, prekrytie a metriky.

Po získaní segmentovaných snímok modelom SCSS-Net bolo ďalším krokom hľadanie konkrétneho tvaru erupčných jadier. Túto úlohu sme sa rozhodli riešiť navrhnutím vlastného algoritmu pracujúceho s výstupmi modelu, ktorý determinuje, či segmentačné masky zobrazujú paralelný tvar slnečnej erupcie alebo nie.

Navrhnutý postup sa skladal z nasledovných krokov:

1) skeletonizácia masiek,
2) pravdepodobnostná Houghova tranformácia aplikovaná na skeletony,
3) úprava výstupov Houghovej transformácie,
4) rozhodnutie o paralelnom tvare na základe navrhnutých výpočtov.

A. Skeletonizácia masiek

Prvým krokom v navrhnutom algoritme je úprava segmentácie pomocou skeletonizácie, ktorá je vhodná pre Houghovu transformáciu. Segmentácia obsahovala drobné výčnelky narušajúce ich tvar a šum, preto bolo nutné odstrániť tieto nerovnosti bez straty tvarovej informácie. Použili sme funkciu $skeletonize()$ z knižnice Scikit-image, ale tá vytvorila mnoho malých vetiev. Preto sme prešli na funkciu $skeletonize_3d()$, ktorá produkovala čistejší skeleton s menej vetvami. Pre odstránenie zvyšných malých vetiev sme použili knižnicu $DSE - skeleton - pruning$, ktorá implementuje metódu zjednodušeného prerezávania vetiev. Funkcia $distance_transform_edt()$ vypočítala vzdialenosti vetiev a $skel_pruning_DSE()$ich následne očistila o malé nepotrebné vetvy, keďže nás zaujímal len dominantný tvar erupcií. Použitím spomínaných funkcií vznikol čistejší skeleton, lepšie vystihujúci hľadaný základný tvar danej erupcie.

B. Houghová transformácia

Po získaní skeletonu slnečných erupcií bola možná aplikácia pravdepodobnostnej Houghovej transformácie (skr. **PHT**). Táto transformácia musí byť prispôsobená na geometrický tvar, ktorý chceme nájsť, v našom prípade to predstavuje hľadanie čiar.

Vzhľadom na výstup skeletonizácie, ktorý obsahoval veľa nedokonalostí, sme po vykonaní PHT získali množstvo krátkych čiar, ktoré obkresľovali skeleton. Takéto čiary však nebolo možné vyhodnotiť, či tvoria paralelný tvar alebo nie, keďže by sme dospeli k veľa falošne pozitívnym výsledkom. Pri veľkom množstve krátkych čiar by bolo jednoduché nájsť dve paralelné v akomkoľvek vstupnom tvare, čo nie je žiadaný stav. Výstup transformácie bolo teda ešte potrebné dodatočne upraviť pre získanie čiar, ktoré by dokázali presnejšie reprezentovať hlavný tvar erupcií.

C. Úprava výstupov Houghovej transformácie

Úlohu úpravy výstupov PHT sme sa rozhodli riešiť spájaním jednotlivých čiar, ktoré táto transformácia našla. Ich spájaním sme chceli docieliť menšie množstvo čiar, ktoré by neboli presnou kópiou skeletonu, ale opisovali by vyhladené hlavné línie jadier erupcií. Na základe koordinátov týchto čiar sme vytvorili nové, ktoré spájali blízke čiary s podobným sklonom. Použili sme rekurzívnu funkciu, ktorú sme pomenovali `cleanLines()`, ktorá prechádzala všetky čiary a nahradila dvojice spĺňajúce stanovené kritériá podobnosti (maximálny rozdiel sklonov jedna, rovnaké znamienka hodnôt sklonu a maximálna povolená vzdialenosť čiar) novou čiarou, spájajúcou najvzdialenejšie body pôvodných čiar. To znamená, že ak našla takéto dve vertikálne orientované čiary, zvolila bod s najmenšou hodnotou koordinátu y a najväčšou hodnotou koordinátu y. Obdobne, pri horizontálne orientovaných čiarach, vybrala ako nové body tie, ktoré mali najmenšiu a najväčšiu hodnotu koordinátu x. Tento proces sa opakoval, kým nebolo možné nájsť ďalšie vhodné dvojice čiar na spojenie. Po aplikovaní opísanej funkcie, bol výstup Houghovej transformácie vhodnejším pre naše účely. Čiary, kopírujúce zjednodušený tvar slnečnej erupcie, na ňom boli viditeľnejšie a bolo možné s nimi ďalej pracovať.

Obrázok 2 obsahuje ukážku úprav aplikovaných na štyroch rôznych snímkach obsahujúcich paralelný tvar erupcie. Pre lepšiu viditeľnosť zmien sú tieto snímky o rozmeroch 1024×1024. Na ľavej strane obrázka sa nachádza neupravený výstup PHT. Pravá strana zobrazuje upravené výsledky. Obe časti obsahujú skeleton, na ktorom sú zelenou farbou znázornené nájdené čiary. Je viditeľné, že je po aplikovanej úprave nájdených čiar menej, sú dlhšie a nekopírujú presný tvar skeletonu, čo bolo aj našim zámerom. Na zložitejšom tvare erupcie, aká je vykreslená na treťom príklade, je vo výstupe stále veľké množstvo čiar. Ak by sme však nastavili podmienky ich spájania tak, aby ich zostalo menej, v menších erupciách by sa ľahko stratil hľadaný tvar.

D. Overenie paralelnosti čiar

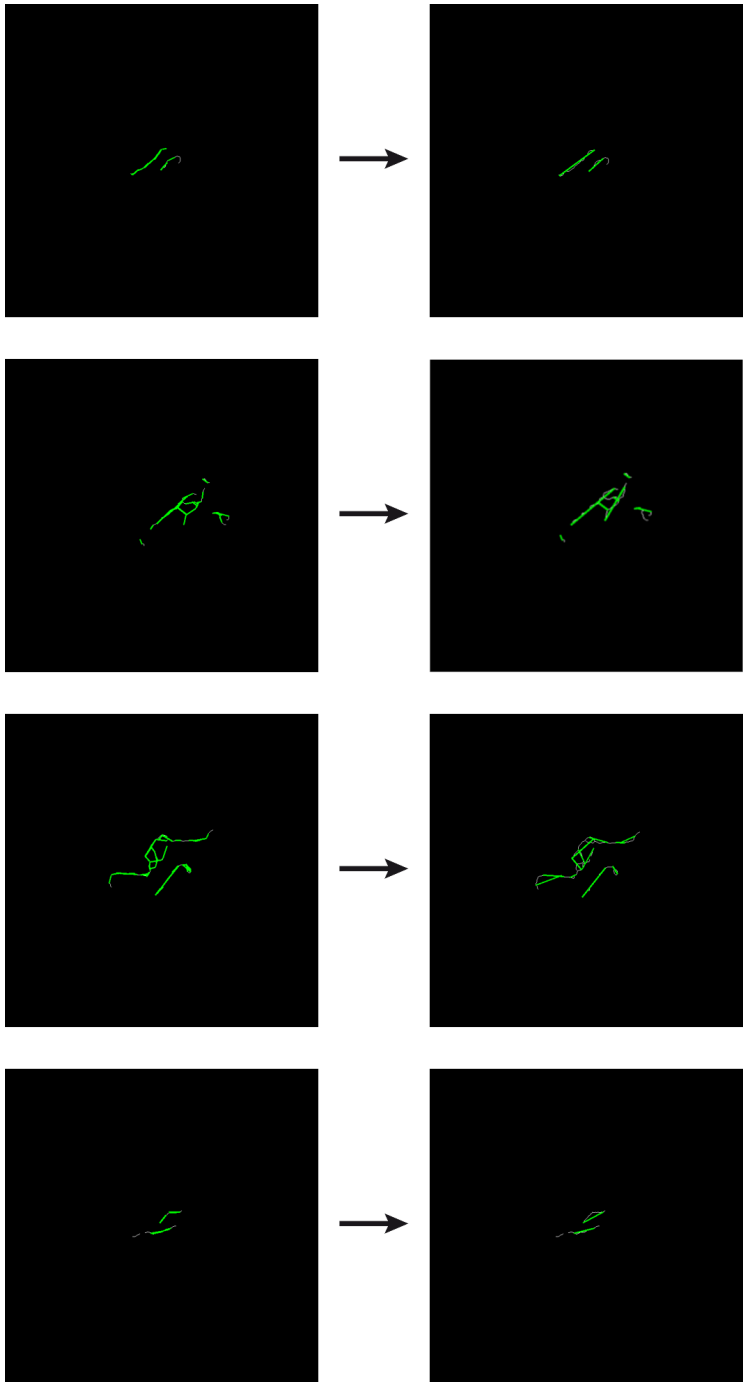
V tejto časti práce sme sa zameriavali na overenie paralelného tvaru pomocou nami navrhnutých výpočtov aplikovaných na upravené výstupy PHT. Predtým, než sme sa mohli pustiť do overovania paralelnosti čiar bolo nevyhnutné porozumieť tomu, v akých prípadoch sa čiary nepovažujú za paralelné. Identifikovali sme päť hlavných situácií, kedy sa dá jednoznačne povedať, že čiary nie sú paralelné:

1) výrazne rozdielny sklon čiar,
2) príliš malá vzdialenosť čiar,
3) príliš veľká vzdialenosť čiar,
4) čiary sa križujú,
5) čiary sa nenachádzajú nad sebou, respektíve vedľa seba.

Tieto prípady sú graficky znázornené na obrázku 3, kde číselné označenia korešpondujú so zoznamom vyššie. Jedným z dôležitých aspektov je overenie sklonu čiar – z matematického hľadiska by rovnobežky mali mať rovnaký sklon, avšak obdobne ako pri vo funkcii `cleanLines()`, sme vzhľadom na nepresné tvary, nastavili maximálnu odchýlku rozdielu sklonov, ktorý môžu porovnávané čiary mať. Tento sklon však nesmie byť príliš rozdielny, čo vyjadruje situácia v 1. bode zoznamu.

Body 2 a 3 vyjadrujú príliš malú a príliš veľkú vzdialenosť čiar. Ak sa čiary nachádzajú vo veľmi malom odstupe, nejedná sa o rovnobežky, aké hľadáme. Sú buď výsledkom intenzívnej časti erupcie, kde bolo skeletonizáciou vytvorených viacero vedľa seba idúcich vetiev, alebo je to len jedna časť erupcie, ktorá sa kvôli šumu javí nepresne, dôvodom čoho v nej vzniklo viacero vetiev. Ak sa naopak nachádzajú v príliš veľkej vzdialenosti, sú to len dve časti erupcie, v ktorých vznikol paralelný tvar, ktorý však nebol spôsobený vznikom erupčných vlákien.

V bode 4 vidíme prípad križovania čiar. V prípade, že sa čiary križujú, tak to samozrejme rovnobežky nie sú. Vzhľadom na povolený rozdiel sklonov sme museli ošetriť aj takéto prípady.

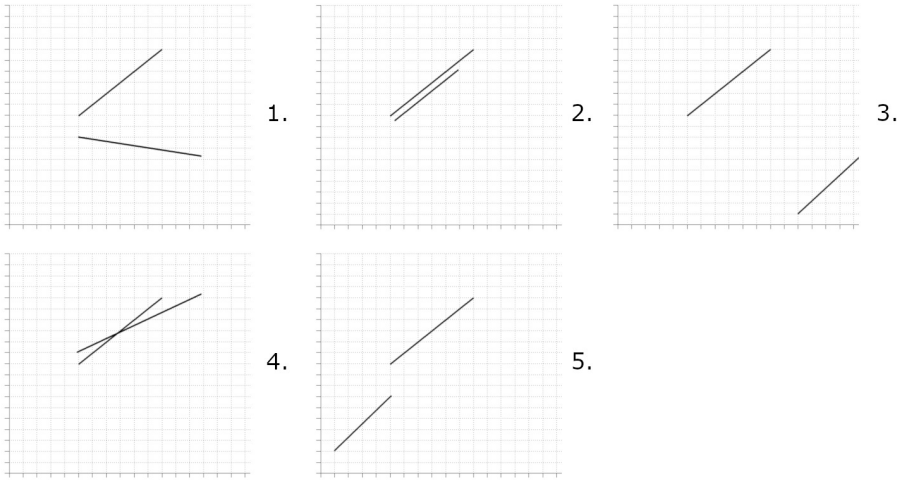


Obrázok 2 Ukážky upravených výstupov PHT

Posledným stavom, ktorému sme sa potrebovali vyhnúť, je zakreslený v 5. bode. Táto situácia nastala, keď sa čiary nenachádzali jedna nad druhou, v prípade horizontálnych čiar, alebo vedľa seba, v prípade vertikálnych čiar. Vždy sme mali dva body tvoriace prvú čiaru vyjadrené ako $(x_1,\ y_1)$ a $(x_2,\ y_2)$, a dva body tvoriace druhú čiaru ako $(x_3,\ y_3)$ a $(x_4,\ y_4)$. Ak sme našli dve vertikálne položené čiary, tak prvá čiara musela mať aspoň jednu z hodnôt y_1 a y_2 z rozpätia $< y_3,\ y_4 >$, prípadne aspoň jedna z hodnôt y_3 a y_4 druhej čiary musela byť z rozpätia $< y_1,\ y_2 >$. Ak by to boli horizontálne položené čiary, platila by rovnaká podmienka, ale pre koordináty osi x.

Čo sa týka samotného algoritmu, v zásade predstavoval zloženie rôznych podmienok, ktorými musela dvojica čiar prejsť, aby bola považovaná za paralelné. Vstupom do hlavnej funkcie je výstup funkcie `cleanLines` – vnorené pole koordinátov nájdených čiar. V kóde prechádzame toto pole, pričom porovnávame každú možnú dvojicu čiar, čo zahŕňa zistenie uhla, sklonu, dĺžky čiar, najväčšej a najmenšej možnej vzdialenosti čiar.

Overením rovnakého znamienka sklonu čiar a maximálneho povoleného rozdielu medzi nimi



Obrázok 3 Prípady, kedy sa čiary nepovažujú za paralelné

sme sa vyhli čiaram, spĺňajúcim 1. bod zo zoznamu neželaných pozícií. Výskytu 2. bodu sme čiastočne zabránili už pri očistení výstupov PHT funkciou `minDistance()`, kde sme všetky čiary, ktoré sa nachádzali veľmi blízko seba a spĺňali aj ostatné podmienky spojili. Ich vzdialenosť sme rovnakou funkciou overili aj v rámci tohto algoritmu. Príliš veľkej vzdialenosti čiar sme sa vyhli obdobne použitím funkcie `minDistance()`. 4. bod bol ošetrený funkciou `intersects()`, ktorá kontrolovala, či sa čiary v nejakom bode pretínajú. Piaty bod overený dvomi podmienkami kontrolujúcimi, či má jedna z čiar aspoň jednu hodnotu koordinátu x alebo y z rovnakého rozsahu, ako koordináty druhej čiary.

Okrem popísaných prvkov je dôležitá aj dĺžka čiar. Vzhľadom na rôzne tvary skeletonov aj samotných erupčných vlákien, dvojica čiar nemusí byť rovnakej dĺžky, avšak čiary nemôžu byť príliš krátke. Pri príliš krátkych nájdených čiarach to boli spravidla dve náhodne nájdené paralelné čiary, ktoré netvorili základný tvar erupcie. Minimálnou povolenou dĺžkou čiar sme ošetrili aj takéto prípady.

IV. Vyhodnotenie

Vyhodnotenie výsledkov detekcie paralelných tvarov erupčných vlákien nebolo možné vyjadriť kvantitatívnymi metrikami, keďže neexistoval žiaden zoznam snímok, obsahujúcich hľadaný tvar. Výsledky preto môžeme overiť len kvalitatívne vizuálnou kontrolou získaných výsledkov.

Výsledky detekcie sme poskytli doménovým expertom, ktorí budú udalosti vyhodnocovať na základe fyzikálnych poznatkov. Vzhľadom na skutočnosť, že model detegoval len malý počet erupčných vlákien, rozhodli sme sa pre ďalšie overenie výsledkov na novej množine dát, ktorá obsahovala 32 465 snímok. Týmto spôsobom sme identifikovali ďalších 122 paralelných erupčných vlákien, čím sme celkovo získali 131 snímok, ktoré boli vyhodnotené ako paralelné. Ukážka nájdených paralelných prípadov je zobrazená na obrázku 4.

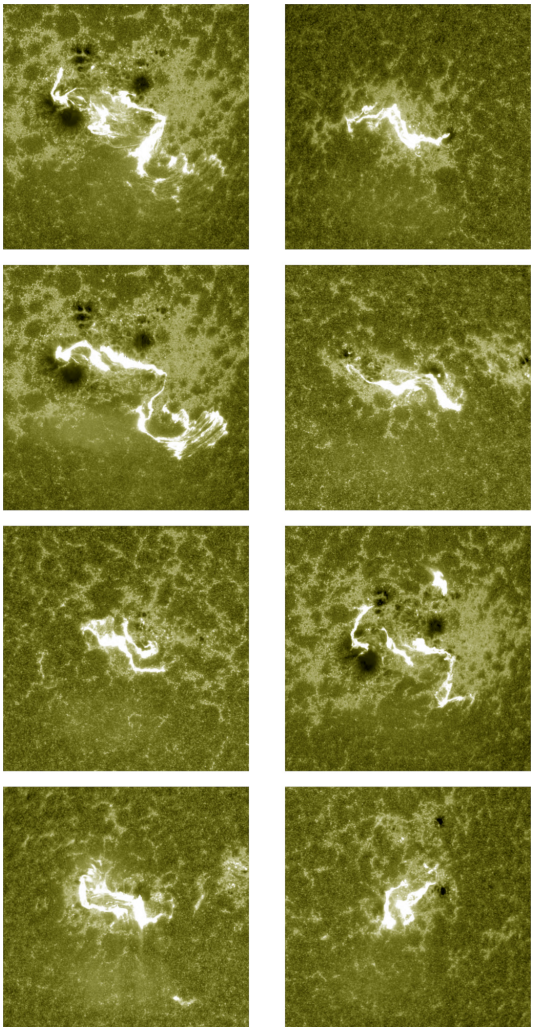
Experimenty s rôznymi modelmi, veľkosťami snímok a použitím manuálne vytvorených segmentácií naznačujú, že druhý model, ktorého vstup nadobúda veľkosť 512×512, bez prahových výstupov je najvhodnejším riešením pre segmentáciu erupčných vlákien. Tieto zistenia potvrdzujú presnosť a účinnosť navrhnutého modelu.

V. Záver

V práci sme sa venovali segmentácii erupčných vlákien na snímkach Slnka s cieľom detegovať paralelné tvary v týchto udalostiach. Tento výskum je dôležitý pre lepšie pochopenie slnečnej aktivity.

Využili sme techniky hlbokého učenia v podobe modelu konvolučnej neurónovej siete, a metódy počítačového videnia. Natrénovaním segmentačného modelu, využívajúceho architektúru SCSS-Net, sa nám podarilo overiť jeho univerzálnosť.

Vzhľadom na rôznorodý tvar a veľkosť vzniknutých erupčných vlákien, nie je ich spracovanie pomocou tradičných metód spracovania obrazu jednoduché. Tieto metódy sa preukázali ako pomerne citlivé na zmeny v obraze, a preto nemusia vždy správne identifikovať hľadaný tvar erupčných vlákien v rámci celého priebehu slnečnej erupcie.



Obrázok 4 Ukážky detegovaných paralelných slnečných vlákien

Z tohto dôvodu by bolo efektívnejším riešením vytvorenie modelu konvolučnej neurónovej siete, ktorá by vedela automaticky klasifikovať erupčné vlákna, v závislosti od ich tvaru. Nami vytvorená množina paralelných erupcií, ktorú je potrebné ešte dodatočne overiť expertom a rozšíriť pomocou augmentačných techník, by mohla slúžiť ako vstupná množina pre takýto model.

Poďakovanie

Táto práca vznikla s podporou Slovenskej grantovej agentúry Ministerstva školstva a SAV Slovenskej republiky v rámci projektu VEGA 1/0685/21.

Literatúra

[1] Janvier, M. and Savcheva, A. and Pariat, E. and Tassev, S. and Millholland, S. and Bommier, V. and McCauley, P. and McKillop, S. and Dougan, F., "Evolution of flare ribbons, electric currents, and quasi-separatrix layers during an X-class flare", 2016. doi:0.1051/0004-6361/201628406.

[2] Minaee, Shervin and Boykov, Yuri and Porikli, Fatih and Plaza, Antonio and Kehtarnavaz, Nasser and Terzopoulos, Demetri, "Image Segmentation Using Deep Learning: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, pp. 3523-3542, 2022. doi:10.1109/TPAMI.2021.3059968.

[3] Albawi, Saad and Mohammed, Tareq Abed and Al-Zawi, Saad., "Understanding of a convolutional neural network", 2017 international conference on engineering and technology (ICET), pp. 1-6, 2017.

[4] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv e-prints, 2015. doi:10.48550/arXiv.1505.04597.

[5] Mackovjak, Š., Harman, M., Maslej-Krešňáková, V., and Butka, P., "SCSS-Net: solar corona structures segmentation by deep learning", Monthly Notices of the Royal Astronomical Society, vol. 508, no. 3, OUP, pp. 3111–3124, 2021. doi:10.1093/mnras/stab2536.

Analýza elektromagnetického poľa v agrokomplexe

Daniel HALAS, Dušan MEDVEĎ

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

daniel.halas@student.tuke.sk, dusan.medved@tuke.sk

***Abstrakt* — Tento článok sa zaoberá analýzou rozloženia elektromagnetického poľa v agrokomplexe určenom na dojenie kráv a jeho súlad s aktuálnymi bezpečnostnými normami. Výskum sa zameral na meranie intenzity elektrického poľa a magnetickej indukcie na rôznych miestach komplexu s použitím špecializovaných meracích prístrojov. Výsledky boli porovnané s limitnými hodnotami stanovenými v príslušných legislatívnych predpisoch s cieľom identifikovať potenciálne riziká pre zdravie dobytka a pracovníkov. Meraním sa zistilo, že všetky merané hodnoty sú v rámci povolených limitov, čo poukazuje na dodržanie bezpečnostných štandardov a minimalizáciu rizika pre zdravie. Príspevok taktiež diskutuje o metodológii merania a navrhuje možnosti pre zlepšenie monitorovania elektromagnetických polí v podobných inštaláciách.**

***Kľúčové slová* — elektromagnetické pole, dojárne, bezpečnostné normy, zdravie dobytka, poľnohospodárske prevádzky**

I. Úvod

Elektromagnetické pole (EMP) je neoddeliteľnou súčasťou moderného životného prostredia, vyžarované množstvom technologických zariadení, od domácich spotrebičov po priemyselné zariadenia. V poľnohospodárskom sektore, kde technológia zohráva čoraz väčšiu úlohu pri zvyšovaní produktivity a efektivity, sa kládli otázky týkajúce sa potenciálneho vplyvu EMP na zdravie dobytka, ktoré sú vystavené týmto poliam kontinuálne. Štúdie sú zamerané na vyhodnotenie, či intenzity EMP presahujú stanovené bezpečnostné limity, a tým predstavujú riziko pre zvieratá a ľudí v ich bezprostrednej blízkosti.

Tento článok predstavuje výsledky merania elektromagnetických polí v agrokomplexe určenom na dojenie kráv. Merania boli realizované s cieľom identifikovať úrovne elektromagnetických polí v rôznych častiach agrokomplexu a porovnať ich s legislatívnymi limitmi. Vzhľadom na rozšírené používanie elektrických zariadení v takýchto komplexoch, ako sú napríklad dojacie stroje, ventilátory a svietidlá, je dôležité monitorovať úrovne EMP a zabezpečiť, že sú v súlade s predpismi na ochranu zdravia [1].

Článok tiež poukazuje na význam pravidelného monitorovania a evaluácie elektromagnetických polí v poľnohospodárskom prostredí, aby sa zabezpečilo, že dlhodobá expozícia týmto poliam neovplyvní zdravie dobytka ani pracovníkov v týchto inštaláciách [4]. Výsledky z tohto článku môžu poslúžiť ako základ pre ďalšie výskumy a možné regulácie v poľnohospodárstve, ako aj pre technologické inovácie zamerané na zníženie elektromagnetického žiarenia.

II. Hodnotenie elektromagnetického poľa v dojárňach

Táto kapitola sa zameriava na štúdium elektromagnetického poľa v prostredí určenom pre dojenie kráv. Vzhľadom na významné množstvo silnoprúdových zariadení používaných v dojárňach existuje možnosť, že dobytok ako aj personál môžu byť vystavení zvýšeným úrovniam elektromagnetického poľa [2]. Cieľom tohto článku je identifikovať a vyhodnotiť tieto riziká porovnaním nameraných úrovní magnetickej indukcie a intenzity elektromagnetického poľa s limitmi stanovenými v legislatíve, konkrétne vyhláškou č. 534/2007 Z.z. [3].

A. Metodológia výskumu

Pre efektívne posúdenie elektromagnetického poľa sme dojaci komplex rozčlenili na štyri hlavné pracovné oblasti. V každej z týchto oblastí sme zmerali intenzitu elektrického poľa

a magnetickej indukcie pri všetkých elektrických zariadeniach generujúcich elektromagnetické pole. Metodika práce zahŕňala podrobný popis každej z oblastí, ktorý umožňuje precízne meranie a následné hodnotenie [5].

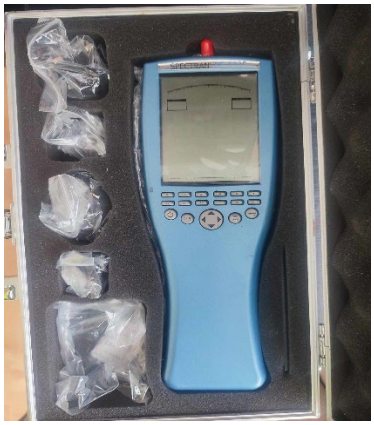
Popis pracovného postupu

Pracovný postup obsahoval:

- Pôdorysy každej z miestností,
- Fotografie a štítkové údaje jednotlivých zariadení,
- Tabuľky s nameranými hodnotami,
- Grafy znázorňujúce závislosti nameraných hodnôt,
- Analýzu a hodnotenie výsledkov pre každú miestnosť.

Použitý merací prístroj

Pre merania bolo využité zariadenie Spectran NF-5035, ktoré je schopné merať intenzitu elektrického poľa, magnetickej indukcie a magnetického poľa v rozsahu frekvencií od 1 Hz do 1 MHz, s možnosťou rozšírenia až do 30 MHz. Tento prístroj poskytuje vysokú presnosť a umožňuje adekvátne posúdiť prítomnosť elektromagnetického poľa v dojárňach.



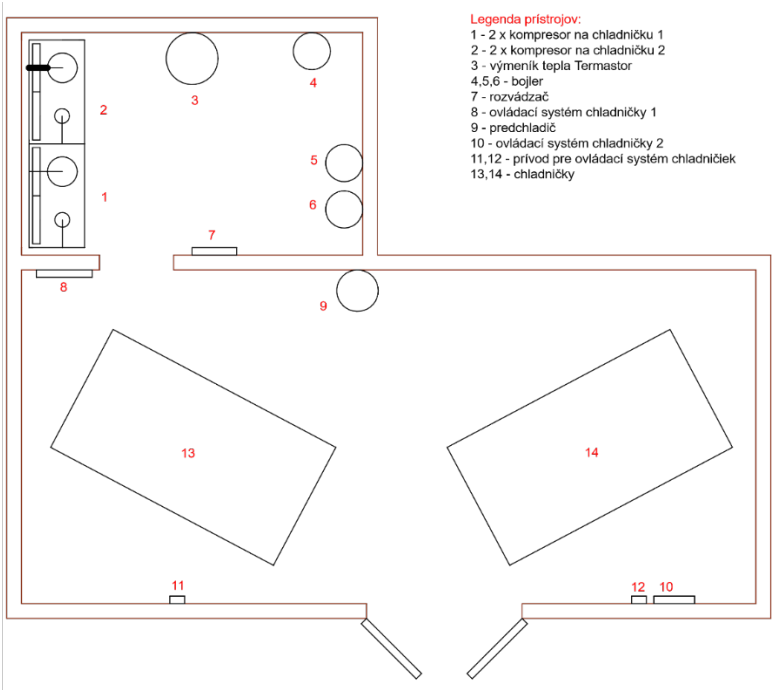
Obr. 1 Merací prístroj Spectran NF-5035

B. Rozdelenie jednotlivých miestností

Miestnosť mliečnych nádrží

Miestnosť, v ktorej sa skladuje vydojené mlieko.

Pôdorys miestnosti



Obr. 2 Pôdorys miestnosti mliečnych nádrží

Predná časť miestnosti mliečnych nádrží



Obr. 3 Mliečna nádrž, čerpadlo na opláchnutie nádrže a predchladič



Obr. 4 Čerpadlo predchladiča

Zadná časť miestnosti mliečnych nádrží



Obr. 5 Kompresor (4ks)

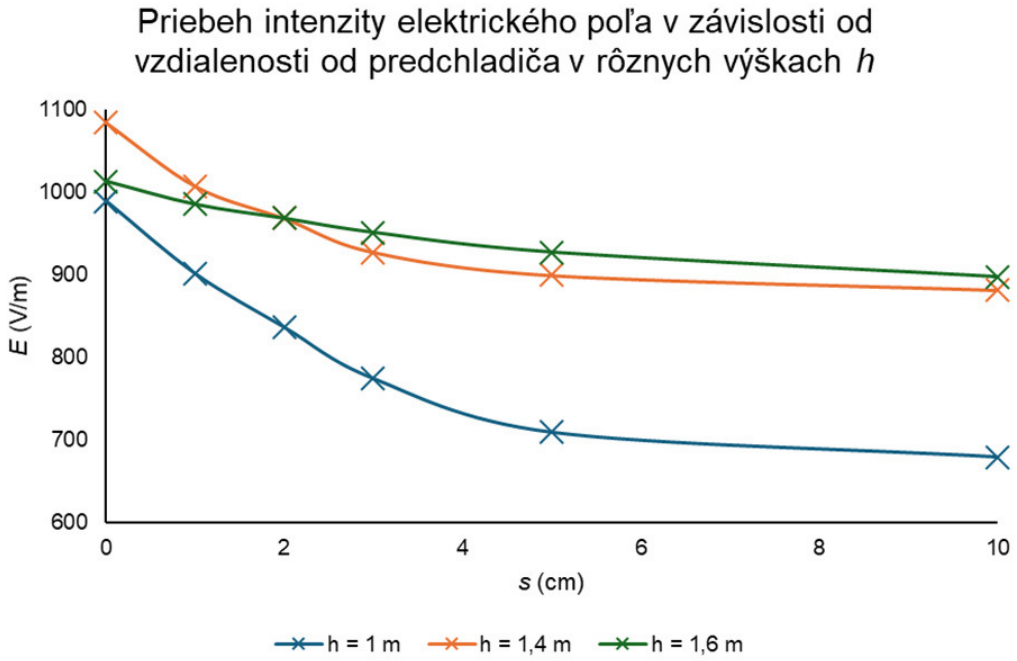


Obr. 6 Bojler (3ks)

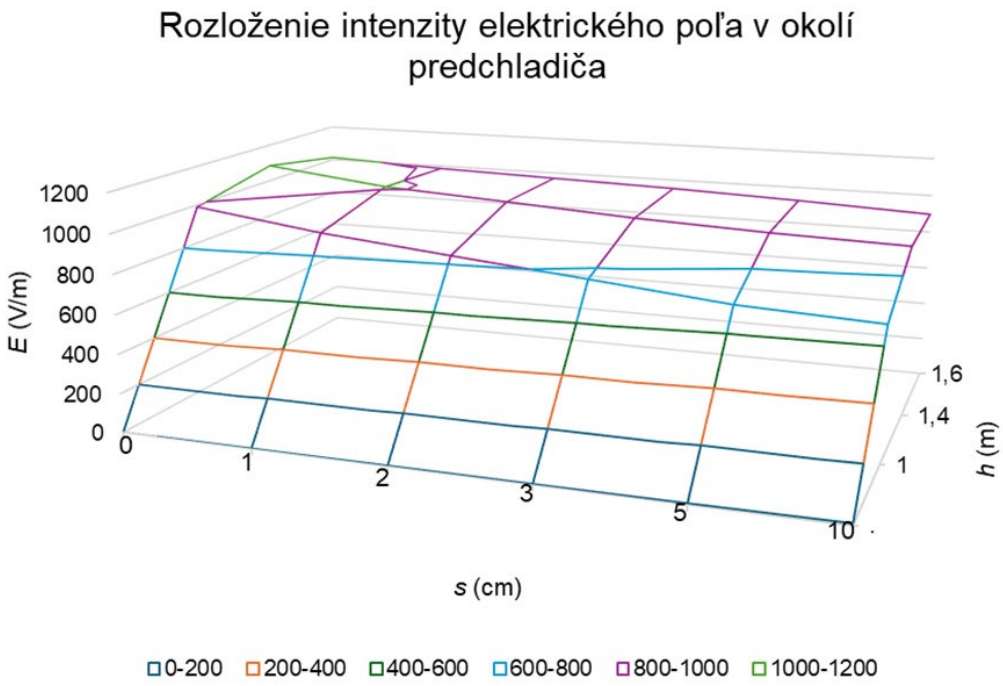
Tabuľka nameraných hodnôt v okolí elektrických zariadení v miestnosti mliečnych nádrží

Tabuľka 1
Namerané hodnoty magnetickej indukcie a intenzity elektrického poľa v miestnosti mliečnych nádrží

Predchladič				Kompresor				Čerpadlo na opláchnutie mliečnej nádrže			
E_{max} = 1090 V/m				E_{max} = 5100 V/m				E_{max} = 5700 V/m			
B_{max} = 11,15 μT				B_{max} = 2,537 μT				B_{max} = 12,9 μT			
h (m)	*s* (cm)	*E* (V/m)	*B* (nT)	*h* (m)	*s* (cm)	*E* (V/m)	*B* (nT)	*h* (m)	*s* (cm)	*E* (V/m)	*B* (nT)
1	0	989	46	1	0	1093	35,02	1	0	1006	1140
	1	901	41		1	1021	32,05		1	989	1031
	2	836	36		2	992	29,61		2	971	941
	3	774	33		3	978	27,69		3	659	883
	5	709	31		5	952	24,67		5	943	821
	10	679	29		10	929	21,43		10	931	738
1,4	0	1084	17	1,4	0	908	9,641	1,4	0	933	81
	1	1007	15,2		1	898	8,993		1	912	79
	2	968	14,2		2	883	7,248		2	890	77
	3	927	13,7		3	871	6,012		3	873	76
	5	899	12,8		5	857	3,992		5	843	73
	10	881	11		10	848	2,907		10	803	71
1,6	0	1013	23	1,6	0	901	4,747	1,6	0	748	30
	1	985	21,6		1	872	4,691		1	730	27
	2	968	20		2	860	4,601		2	718	25
	3	951	19		3	850	4,528		3	704	23
	5	927	17		5	830	4,501		5	691	21
	10	897	15		10	810	4,440		10	671	19
Bojler (4)				E_{max} = 1500 V/m				B_{max} = 0,3 μT			
Bojler (5)				E_{max} = 2500 V/m				B_{max} = 0,159 μT			
Bojler (6)				E_{max} = 1900 V/m				B_{max} = 0,123 μT			



Obr. 7 Priebeh intenzity elektrického poľa v závislosti od vzdialenosti od predchladiča



Obr. 8 Rozloženie intenzity elektrického poľa v okolí predchladiča

III. Záver

V tomto článku sme sa venovali hodnoteniu elektromagnetického poľa v miestnosti určenej pre uchovávanie mlieka počas rôznych dojácich cyklov. Mlieko z večerného a ranného dojenia sa tu uchováva do času, kým ho nevyzdvihne cisternový kamión. Hlavným procesom je predchladenie a následné uskladnenie mlieka v nádržiach, ktoré sú čistené pomocou čerpadiel a chladené štyrmi kompresormi. V tejto miestnosti sa taktiež nachádzajú tri bojlery zabezpečujúce ohrev vody pre rôzne potreby.

Merania intenzity elektrického poľa a magnetickej indukcie boli vykonané pri každom zariadení vo výškach 1 meter, 1,4 metra a 1,6 metra nad zemou podľa normy STN IEC 833, pričom sme postupne zvyšovali vzdialenosť od zariadení od 0 cm až po 10 cm. Červená farba v grafických zobrazeniach slúžila na lepšie vizuálne odlíšenie pozícií zariadení.

Z výsledkov merania vyplýva, že najvyššie hodnoty intenzity elektromagnetického poľa a magnetickej indukcie sú priamo pri zariadeniach a s rastúcou vzdialenosťou klesajú. Najvyššie merané hodnoty boli pri kompresore a čerpadle na čistenie chladničky, kde bola hodnota intenzity elektrického poľa mierne nad limitom stanoveným vyhláškou. Avšak pri štandardných výškach používaných pre merania neboli hodnoty prekročené.

Tento výsledok potvrdzuje, že zamestnanci pracujúci v tejto časti dojacieho komplexu sú vystavení elektromagnetickým poliam v rámci bezpečnostných limitov stanovených legislatívou. Tento záver umožňuje uistiť sa o bezpečnom pracovnom prostredí, pričom zároveň poskytuje dôležité informácie pre ďalšie zlepšenie pracovných postupov a technologických opatrení na zníženie expozície elektromagnetickým poľom.

Poďakovanie

Tento príspevok vznikol vďaka Agentúre na podporu výskumu a vývoja na základe zmluvy č. APVV-19-0576 a APVV-21-0312, ako aj Vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporená grantom VEGA 1/0627/24.

Referencie

[1] J. Myslík, „*Elektromagnetické pole*“, Praha: BEN - technická literatúra, 1998. 80-86056-43-0.
[2] A. N. Vasiljev, „*Klasická elektrodynamika – konspekt prednášok*“, Peterburg: BCHV, 2009. 978-80-89284-98-6.
[3] J. Veselský, „*Fyzika, - Elektrina a magnetizmus*“. Trnava: Univerzita sv. Cyrila a Metoda v Trnave, 2007. 978-80-89220-74-8.
[4] P. Čičmanec, „*Elektrina a magnetizmus*“, Bratislava: ALFA, vydavateľstvo technickej a ekonomickej literatúry, n.p., Bratislava, 1980. 63-560-80.
[5] D Halas, „*Analýza elektromagnetického poľa v elektroenergetike*“, Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2024.

Unveiling the Power of Capsule Networks: A Comparative Study with Convolutional Neural Networks in Feature Extraction and Transfer Learning

Dominik VRANAY, Laura PITUKOVÁ, Peter SINČÁK

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

dominik.vranay@tuke.sk, laura.pitukova@student.tuke.sk, peter.sincak@tuke.sk

Abstract **– This paper presents a comprehensive evaluation of Capsule Networks (CapsNets) and Convolutional Neural Networks (CNNs) in feature extraction tasks using the Fashion-MNIST dataset, providing crucial insights for optimizing deep learning architectures. We explored various configurations, including basic, pretrained, and non-pretrained feature extractors. CapsNets consistently outperformed CNNs with a basic extractor, achieving higher accuracy (0.9063 vs. 0.8960). CapsNets excelled with extended training (130 epochs), while CNNs demonstrated faster convergence (50 epochs). Experiments with swapped feature extractors showed that CapsNets significantly benefited from the flexibility to fine-tune CNN extractors. Using pretrained ResNet-18 extractors further enhanced performance, with CapsNets achieving slightly higher metrics (accuracy of 0.9299). These findings underscore CapsNets' robustness and adaptability for high-precision tasks, while CNNs remain highly efficient with shorter training times. This study provides valuable insights into optimizing deep learning architectures for various applications.**

Keywords **– Capsule Networks, Convolutional Neural Networks, Feature Extraction, Transfer Learning**

I. Introduction

Capsule Networks (CapsNets) represent a significant leap forward in the field of deep learning, offering a novel architectural paradigm that addresses some of the limitations inherent in traditional Convolutional Neural Networks (CNNs). Unlike CNNs, which rely on scalar outputs for feature detection, CapsNets encapsulate features in vector forms, allowing them to capture intricate details such as pose, orientation, and spatial hierarchies more effectively. This innovative approach enhances the network's ability to generalize from training data, making CapsNets particularly resilient to variations in input transformations.

Since their introduction, CapsNets have garnered some attention for their potential applications across a variety of domains, including computer vision, natural language processing, and robotics. Their architecture, which includes dynamic routing between capsules, provides a robust framework for understanding complex data relationships and achieving superior performance in tasks that require nuanced feature extraction. Despite these promising capabilities, a detailed comparative analysis between CapsNets and CNNs is crucial for understanding the specific strengths and weaknesses of each architecture [1].

Transfer learning, where a pretrained model is fine-tuned for a new but related task, has become an essential technique in deep learning due to its ability to leverage existing knowledge and reduce the need for extensive labelled data. CapsNets, with their unique feature encapsulation and dynamic routing mechanisms, offer a promising alternative to CNNs for transfer learning applications. This study aims to investigate how well CapsNets perform in transfer learning settings compared to CNNs, thereby offering a deeper understanding of their potential advantages and limitations [2], [3].

This paper aims to contribute to the understanding of feature extraction capabilities by evaluating CapsNets and CNNs. By systematically exploring the performance of these architectures, the study seeks to provide insights into their relative efficacy, particularly in the context of transfer learning. Through a thorough comparative analysis, this research focuses on their feature extraction and transfer learning capabilities, shedding light on the performance dynamics of these two architectures. Ultimately, this research aims to guide future developments in deep learning, fostering the advancement of more robust and efficient models for a wide range of applications.

II. Literature Review

CapsNets offer a unique architecture designed to preserve spatial hierarchies through dynamic routing between capsules, which contrasts sharply with the max-pooling operations of CNNs that often discard valuable positional information. This dynamic routing mechanism enables CapsNets to maintain more detailed feature representations, which has been shown to improve performance in various tasks [1].

Several comparative studies have evaluated the performance of CapsNets and CNNs, often highlighting the superior generalization capabilities of CapsNets. Ref. [4] conducted an evaluation on image segmentation tasks and reported that CapsNets outperformed CNNs, particularly in scenarios with limited training data. Their findings emphasized the ability of CapsNets to learn from fewer examples, owing to their efficient feature extraction processes. Similarly, ref. [5] found that CapsNets performed better than CNNs in recognizing rotated and overlapping objects, attributing this to the capsules' capacity to capture and retain spatial hierarchies more effectively.

The feature extraction capabilities of CapsNets have also been a focal point in the literature. Ref. [3] expanded on the original CapsNet architecture by introducing the Matrix Capsule Network, which utilized matrix transformations to enhance the robustness of feature extraction. This extension aimed to improve the network's ability to handle complex data structures compared to traditional CNNs. Ref. [6] explored the efficacy of CapsNets in transfer learning scenarios, demonstrating that CapsNets could leverage pretrained features more effectively than CNNs, achieving superior performance with minimal additional training.

Transfer learning has traditionally been dominated by CNNs, particularly due to the availability of pretrained models on large datasets such as ImageNet. However, recent studies have shown promising results for CapsNets in this domain. Ref. [7] demonstrated that CapsNets could be fine-tuned for specific applications, such as medical image analysis, with fewer training examples than typically required by CNNs. Ref. [8] explored the use of CapsNets in natural language processing tasks and found that they outperformed CNNs in sentiment analysis and text classification, leveraging the hierarchical structure of textual data effectively.

Despite these advantages, CapsNets face significant challenges related to computational efficiency and training time. Ref. [9] pointed out that the dynamic routing process, while beneficial for maintaining spatial hierarchies, introduces considerable computational complexity. Efforts to optimize this process are ongoing, with ref. [10] suggesting various approaches to reduce the computational overhead, aiming to make CapsNets more practical for large-scale applications.

The literature indicates that CapsNets provide significant improvements in feature extraction and generalization over CNNs, especially in tasks involving complex spatial relationships and limited training data. Their potential in transfer learning further enhances their applicability across diverse domains. However, addressing the computational challenges associated with CapsNets remains crucial for their broader adoption and practical implementation.

III. Methodology

This section details the experimental setup used to evaluate the performance of CapsNets and CNNs in feature extraction tasks. The goal is to understand the strengths and weaknesses of each architecture, particularly in the context of transfer learning.

A. Feature Extractors

Three distinct feature extractors were used in this study, which were then varied in the experiments with the models used:

- Basic CNN-based Extractor: This extractor includes two convolutional layers with ReLU activation functions and max-pooling. The first convolutional layer has 16 filters, and the second has 32 filters, both with a kernel size of 3x3. This basic structure serves as a benchmark for comparing more complex feature extractors.
- Pretrained ResNet-18: ResNet-18, a residual network, was utilized in two configurations: pretrained and non-pretrained. The pretrained ResNet-18 model, trained on the ImageNet

dataset, was incorporated to leverage its learned representations for transfer learning. The network's initial layers were used to extract features, which were then fed into the CapsNet and CNN models [11].

- Non-pretrained ResNet-18: This version of ResNet-18 was trained from scratch on the dataset used in this study. It allowed for a comparison of how much pretraining on a large dataset like ImageNet impacts performance in feature extraction tasks.

B. Models

Two primary classification architectures were evaluated in this study: simple fully connected classifier and capsule classifier.

The fully connected classifier architecture comprised the chosen feature extractor followed by a flattening layer and two fully connected layers. The first fully connected layer had 512 neurons, and the second matched the number of output classes. ReLU activation functions were applied after each fully connected layer. The models were optimized using the Adam optimizer and the Cross-Entropy loss function.

The capsule architecture incorporated the chosen feature extractor followed by a primary capsule layer and two routing layers. The primary capsule layer transformed the output of the feature extractor into 72 capsules with a dimensionality of 4. The routing layers applied dynamic routing to establish connections between capsules, enhancing the network's ability to capture spatial hierarchies. The routing layers transformed the input capsules first into 24 capsules of dimensionality 10 and finally into 10 class capsules with dimensionality of 10. Dropout layers were included between routing layers to prevent overfitting. The CapsNet models were optimized using the Adam optimizer and the Margin Loss function.

C. Variations and Comparisons

To comprehensively evaluate the feature extraction capabilities of CapsNets and CNNs, various configurations were explored. These configurations involved different combinations of feature extractors and models, with some being pretrained on large datasets while others were trained from scratch. The aim was to understand how different architectural choices and pretraining strategies impact performance. The simplest CNN feature extractor using fully connected classifier can be seen in Fig. 1.

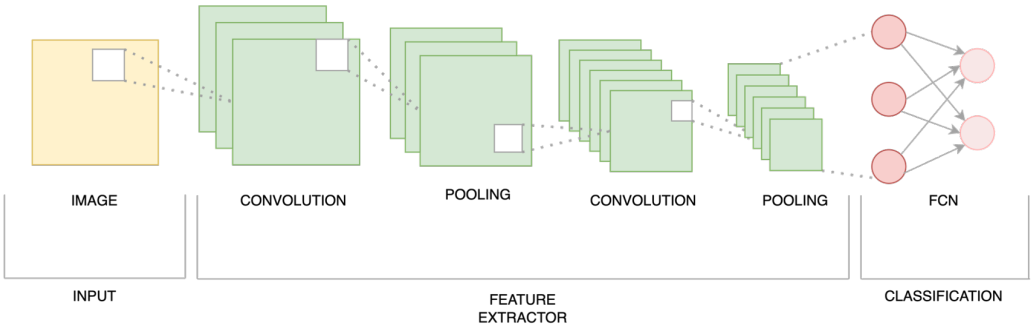


Fig. 1 Scheme of a simple CNN

By examining the effects of using feature extractors from different models, the study assessed the transferability and robustness of features across architectures. Incorporating pretrained feature extractors evaluated how leveraging existing knowledge from large datasets influences model performance, while using non-pretrained extractors provided insights into the benefits of learning features directly from task-specific data.

Additionally, some configurations involved freezing the parameters of the feature extractors during training. This approach aimed to isolate the impact of the higher layers in the network, focusing on the significance of the feature extraction layers without the influence of parameter updates.

IV. Experiments

This section details the experimental setup used to evaluate the performance of CapsNets and CNNs in feature extraction tasks. The experiments involve a comprehensive analysis of different configurations. Through these experiments, we aim to gain insights into the comparative strengths

and weaknesses of CapsNets and CNNs, providing a robust assessment of their performance across multiple scenarios. The results of each experiment will be presented and discussed to highlight key findings and implications.

A. Dataset

The experiments utilized the Fashion-MNIST dataset, a benchmark in the machine learning community, comprising 70,000 grayscale images of fashion items. Each 28x28 pixel image is categorized into one of 10 classes, with 7,000 images per category. The dataset is divided into 60,000 training images and 10,000 test images, offering a rich representation of fashion items for evaluating feature extraction capabilities. The dataset can be found on https://www.kaggle.com/datasets/zalando-research/fashionmnist [12].

Several preprocessing steps were applied to enhance model performance. First, a 2-pixel padding was added around each image to ensure features near the edges were captured, followed by random cropping back to 28x28 pixels to introduce variability. The images were then converted to PyTorch tensors for efficient processing and normalized using predefined mean and standard deviation values. These steps ensured consistency and robustness in the training data, providing a solid foundation for the experiments.

B. Experimental Setup

The models were trained using the PyTorch library, leveraging its efficient computational framework. For training, the Adam optimizer was employed with a learning rate of 0.0035, chosen for its adaptive learning capabilities. The models were trained for 130 epochs to ensure sufficient learning time if it is not mentioned otherwise. The performance of the models was evaluated using several metrics, including accuracy, balanced accuracy, precision, recall, and F1-score. These metrics provided a comprehensive assessment of the models' ability to correctly classify images and generalize to unseen data.

C. CapsNet and CNN with Basic Feature Extractor

Our focused examination of CapsNet and CNN performance using a basic feature extractor reveals notable differences, as summarized in Table 1. The table compares key performance metrics, including accuracy, balanced accuracy, and F1-score, for both models.

Table 1
Comparison of CapsNet and CNN Performance Metrics

Metric	**CNN**	**CapsNet**
Accuracy	0.8960	0.9063
Balanced Accuracy	0.8973	0.9055
F1-score	0.8953	0.9063

Using the basic feature extractor, CapsNet consistently outperforms CNN across all evaluated metrics. Specifically, CapsNet achieves an accuracy of 0.9063 compared to CNN's 0.8960, indicating better overall classification performance. The balanced accuracy, which accounts for class imbalances, is also higher for CapsNet (0.9055) compared to CNN (0.8973), suggesting more reliable performance across different classes.

The F1-score, which combines precision and recall into a single metric, further demonstrates CapsNet's superiority with a score of 0.9063 compared to CNN's 0.8953. This higher F1-score implies that CapsNet not only makes accurate predictions but also maintains a good balance between precision (the ability to avoid false positives) and recall (the ability to detect true positives).

These results highlight the advantages of CapsNet in feature extraction and classification tasks when using a basic feature extractor. The architectural differences between CapsNet and CNN, particularly CapsNet's ability to capture spatial hierarchies and maintain positional information through dynamic routing, likely contribute to its superior performance.

D. CapsNet vs CNN with Different Epochs

We compared the performance of CNN and CapsNet trained with 130 epochs versus 50 epochs to understand the impact of training duration, as shown in Table 2. Both models used a basic feature extractor for this experiment.

The results indicate that CapsNet outperforms CNN when both models are trained for 130 epochs, achieving higher scores across all evaluated metrics. This suggests that with extended training, CapsNet's architecture allows it to capture and classify features more effectively.

Table 2
Comparison of CNN and CapsNet Trained with Different Epochs

Metric	130 Epochs (CNN)	130 Epochs (CapsNet)	50 Epochs (CNN)	50 Epochs (CapsNet)
Accuracy	0.8960	**0.9063**	0.9025	0.8959
Balanced Accuracy	0.8973	**0.9055**	0.9034	0.8960
F1-score	0.8953	**0.9055**	0.9017	0.8960

However, when the models are trained for only 50 epochs, CNN demonstrates better performance metrics than CapsNet. These results suggest that CNN converges faster than CapsNet, reaching competitive performance in a shorter training time.

This experiment highlights the efficiency of CNNs in terms of training duration while maintaining competitive performance levels. On the other hand, CapsNets may require longer training periods to fully leverage their architectural advantages, which allow them to outperform CNNs in feature extraction and classification tasks given sufficient training time. Thus, the choice between CapsNet and CNN may depend on the specific requirements of the task, such as the availability of computational resources and the need for rapid model deployment.

E. CapsNet and CNN with Swapped Feature Extractors

We examined the performance of CapsNet and CNN models with swapped feature extractors, where the CNN utilized a pre-trained basic extractor from CapsNet and CapsNet utilized a pre-trained basic extractor from CNN. Additionally, we evaluated both models with their feature extractors' parameters frozen and not frozen. The results are summarized in Table 3.

Table 3
Performance with Swapped Feature Extractors

Metric	CNN classifier Frozen extractor	CNN classifier Not Frozen extractor	CapsNet classifier Frozen extractor	CapsNet classifier Not Frozen extractor
Accuracy	**0.8993**	0.8980	0.8417	0.8599
Balanced Accuracy	**0.8987**	0.8974	0.8364	0.8564
F1-score	**0.8990**	0.8970	0.8364	0.8564

The performance metrics reveal several key insights. When the CNN uses a frozen CapsNet feature extractor, it achieves slightly higher accuracy, balanced accuracy, and F1-score compared to when the CapsNet extractor is not frozen. This indicates that the CNN can effectively utilize the CapsNet feature extractor, with minimal performance loss when the extractor's parameters are frozen.

Conversely, CapsNet with a frozen CNN feature extractor shows lower performance metrics. However, when the CNN extractor is not frozen, CapsNet's performance improves significantly. This improvement suggests that CapsNet benefits more from the flexibility to update the feature extractor's parameters during training.

These results highlight that while CNNs can leverage CapsNet extractors effectively, even when frozen, CapsNets perform better with the ability to fine-tune the CNN extractors. This suggests that CapsNet's architecture, which relies on dynamic routing and complex feature relationships, benefits from the adaptability and updateability of the feature extractor's parameters.

F. CapsNet and CNN with Pretrained and Non-pretrained ResNet-18

We compared the performance of CapsNet and CNN models using both pretrained and non-pretrained ResNet-18 feature extractors. The results are summarized in Table 4.

Table 4
Performance with Pretrained and Non-pretrained ResNet-18

Metric	Pretrained (CNN)	Pretrained (CapsNet)	Non-pretrained (CNN)	Non-pretrained (CapsNet)
Accuracy	0.9273	**0.9299**	0.9144	0.9128
Balanced Accuracy	0.9272	**0.9300**	0.9132	0.9131
F1-score	0.9269	**0.9300**	0.9142	0.9131

The results indicate that CapsNet with a pretrained ResNet-18 feature extractor achieves the highest performance across all evaluated metrics, followed closely by CNN with a pretrained ResNet-18. This suggests that pretrained feature extractors significantly enhance the capability of both models to extract and classify features. The advantage of transfer learning is evident, as pretrained models perform better than their non-pretrained counterparts.

Specifically, the accuracy, balanced accuracy, and F1-score for CapsNet with a pretrained ResNet-18 are slightly higher than those for CNN with the same pretrained extractor. This indicates that CapsNet can more effectively leverage the rich feature representations learned by ResNet-18 during pretraining.

In contrast, models using non-pretrained ResNet-18 extractors perform slightly worse. Both CNN and CapsNet with non-pretrained feature extractors show lower accuracy, balanced accuracy, and F1-scores compared to their pretrained versions, highlighting the importance of pretraining in boosting model performance.

This experiment underscores the impact of pretrained feature extractors on model performance. CapsNet's superior performance with a pretrained ResNet-18 demonstrates its ability to utilize advanced feature representations effectively. Meanwhile, the performance of CNNs also benefits significantly from pretrained extractors, although to a slightly lesser extent than CapsNet.

V. Conclusion

In this study, we conducted a comprehensive evaluation of CapsNets and CNNs across various experimental setups to understand their performance in feature extraction tasks. Our experiments involved comparing basic feature extractors, examining the impact of training duration, swapping feature extractors between models, and utilizing pretrained and non-pretrained ResNet-18 feature extractors.

Our results highlight several key findings. First, CapsNets consistently outperformed CNNs when both models used a basic feature extractor. CapsNets demonstrated higher accuracy, balanced accuracy, and F1-scores, indicating their superior ability to capture and correctly classify features. This suggests that even with simple feature extraction methods, CapsNets are more effective in feature extraction and classification tasks.

Second, the impact of training duration revealed that CapsNets excel with longer training periods, outperforming CNNs when both models were trained for 130 epochs. However, CNNs achieved competitive performance with only 50 epochs, showcasing their faster convergence and efficiency in shorter training periods. This finding suggests that while CapsNets may require more training time to leverage their architectural advantages fully, CNNs can achieve good performance more quickly.

Third, when examining swapped feature extractors, we found that CapsNets with CNN feature extractors showed significant performance improvements, especially when the feature extractor parameters were not frozen. Conversely, CNNs effectively leveraged CapsNet feature extractors even with frozen parameters, though they performed slightly better with the ability to fine-tune. This demonstrates CapsNet's flexibility in utilizing diverse feature extraction methods and CNN's robustness in adapting to different extractors.

Lastly, the use of pretrained ResNet-18 feature extractors significantly enhanced the performance of both CapsNets and CNNs. CapsNets, in particular, showed a slight edge in leveraging these pretrained features, highlighting the benefits of transfer learning. Pretrained models provided richer feature representations, which greatly improved the classification capabilities of both architectures compared to their non-pretrained counterparts.

Overall, our study demonstrates that CapsNets offer substantial advantages in feature extraction tasks, particularly when using pretrained feature extractors and longer training durations. CapsNets' ability to preserve spatial hierarchies and dynamically route information enables them to outperform traditional CNNs in various scenarios. However, CNNs remain highly efficient and effective, especially with shorter training times and frozen feature extractors, showcasing their robustness and adaptability. It is important to note that the results are not statistically significant, and more experiments should be conducted on a more difficult dataset, which would likely amplify the differences between models.

These findings provide valuable insights into the comparative strengths of CapsNets and CNNs. CapsNets, with their advanced architectural features, hold significant promise for applications requiring high precision and robust feature extraction. Meanwhile, CNNs continue to be a reliable and efficient choice for many machine learning tasks. Future research could further explore optimizing CapsNet architectures and training strategies to fully harness their potential in diverse applications.

Acknowledgment

This work was supported by the Slovak National Science Foundation project "Basic Research of Deep Learning for Image processing - DL4VISION" supported during 2022–2025 under registration code 1/0394/22, and additionally also by the LIFEBOTS Exchange project funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824047.

REFERENCES

[1] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *Advances in neural information processing systems*, vol. 30, 2017.
[2] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
[3] G. E. Hinton, S. Sabour, and N. Frosst, "Matrix capsules with em routing," in *International conference on learning representations*, 2018.
[4] R. LaLonde and U. Bagci, "Capsules for object segmentation," *arXiv preprint arXiv:1804.04241*, 2018.
[5] E. Xi, S. Bing, and Y. Jin, "Capsule network performance on complex data," *arXiv preprint arXiv:1712.03480*, 2017.
[6] A. Gritsevskiy and M. Korablyov, "Capsule networks for low-data transfer learning," *arXiv preprint arXiv:1804.10172*, 2018.
[7] P. Afshar, A. Mohammadi, and K. Plataniotis, "Capsule networks for brain tumor classification based on mri images and coarse tumor boundaries. 2 018 25th ieee international conference on image processing (icip), athens," 2018.
[8] W. Zhao, J. Ye, M. Yang, Z. Lei, S. Zhang, and Z. Zhao, "Investigating capsule networks with dynamic routing for text classification," *arXiv preprint arXiv:1804.00538*, 2018.
[9] M. U. Haq, M. A. J. Sethi, and A. U. Rehman, "Capsule network with its limitation, modification, and applications—a survey," *Machine Learning and Knowledge Extraction*, vol. 5, no. 3, pp. 891–921, 2023.
[10] D. Rawlinson, A. Ahmed, and G. Kowadlo, "Sparse unsupervised capsules generalize better," *arXiv preprint arXiv:1804.06094*, 2018.
[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[12] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

Author's index

A
Avuk Jozef 317

B
Bačík Ján 7
Baláž Anton 121, 127
Basarik Tomáš 142, 147, 155
Beňa Ľubomír 251
Bober Peter 444
Butka Peter 454, 461, 468, 488, 502

Č
Čorňák Daniel 116

D
Dimunová Nikoleta 343
Drutarovský Miloš 202, 336
Durisin Juraj 285
Dzivý Daniel 246
Dzurnak Pavol 99

E
Eliáš Kristián 251

F
Falis Dominik 99
Fedor Marek 36
Fedor Pavol 7, 42
Ferenc Lukáš 190
Ferková Želmíra 444

G
Gajdoš Patrik 381
Gans Šimon 414
Garnek Lukáš 270
Girovský Peter 147, 275, 289, 304, 429
Gordan Daniel 142, 147, 155
Grodecká Lucia 32
Gula Julián 48
Guskov Stanislav 296
Guzan Milan 393, 420

H
Halas Daniel 509
Haluška Renát 381, 474, 483, 495
Hasin Martin 99
Herbert Miroslav 196
Hozova Diana 437
Hričák Štefan 184
Hurný Patrik 54

J
Janitor Martin 336
Jenčiková Kamila 488

K
Kainz Ondrej 72, 78, 296, 329
Kališková Lenka 454, 461, 468, 488, 502
Kaňuch Ján 275, 289, 304, 429
Kardoš Slavomír 448
Khorshidiyeh Heidar 257, 324
Kirešová Simona 393, 420
Klein Andrej 386
Kmecik Tadeáš 142, 147, 155
Knieža Dominik 135
Kompan Jozef 121
Korečko Štefan 234, 343
Kováč Adam 227
Kováč Ondrej 184, 190, 196
Kožík Štefan 11
Krasnova Yana 166
Križalkovičová Eva 72
Kupcová Eva 381, 495
Kurkina Natalia 478

L
Lacko Dávid 88
Lacko Milan 155, 241, 386
Lazorčák Štefan 355
Lechner David 362
Lenger Tomáš 178
Liščinská Zuzana 374
Lohaj Oliver 264
Lukács Peter 160, 215

Ľ
Ľubomír Livovský 209

M
Madoš Branislav 48, 54, 60, 135, 171, 257, 324
Mach Marián 67, 88, 94, 166, 350
Machová Kristína 11, 17, 23, 28, 32
Majirsky Adam 468
Maniak Tomáš 424
Marcin Daniel 241
Matejová Miroslava 227, 401
Mati Martin 215
Medveď Dušan 111, 116, 509
Michalko Miroslav 72, 78, 296, 329
Molnár Ján 414
Murin Miroslav 72, 78, 296, 329

N
Ngoc Hai Tran 311
Novák Daniel 350
Novysedlák Martin 78

O
Ofčák Ján 17
Ondo-Eštok Martin 67
Onofrej Daniel 264
Oravec Jakub 355, 362, 368, 374
Ovseník Ľuboš 355, 362, 368, 374

P
Pajkoš Filip 483
Pajkoš Jaroslav 202
Papaj Ján 311, 408, 478
Paralič Ján 227, 264, 401
Pastor Marek 142
Pavlik Marek 424
Perduková Daniela 7, 36
Perignáth Martin 94
Petrilák Tomáš 408
Pietrikova Alena 222
Pituková Laura 515
Pleva Matúš 381, 495
Pončáková Laura 23
Popík Peter 60
Potočňáková Lívia 502
Pustý Michal 234

R
Radvani Peter 42
Reňák Adam 495

S
Sanislová Viktória 401
Sinčák Peter 515
Sivák Tomáš 127
Sklenáriková Alžbeta 171
Slanina Daniel 83
Sobota Branislav 234, 343
Solanik Michal 83
Staš Michal 106
Stutiak Kristián 324
Sulír Matúš 317
Szomosi Vladimír 111

Š
Šimoňák Jakub 28
Šinaľ Dalibor 474
Štosel Slavomír 368
Šulek Filip 275

T
Tovstolis Anastasiia 257

V
Vehec Igor 270
Viski Barbora 160
Vranay Dominik 515

Y
Yatskiv Ivan 329

Z
Zbojovsky Jan 437

Ž
Žilková Jaroslava 289, 304, 429

Electrical Engineering and Informatics 15

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, 524 pages

Date of publication: July 2024
Edition: 1^{st}
Pages: 524
Available on: eei.fei.tuke.sk

ISBN 978-80-553-4682-3