Technical University of Košice

Faculty of Electrical Engineering and Informatics

Electrical Engineering and Informatics

16

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

ISBN 978-80-553-4854-4

2025

Electrical Engineering and Informatics 16
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 2025
Edition:	1^{st}
Available on:	eei.fei.tuke.sk
Languague:	English, Slovak
Pages:	658

Editors:
Karol Kyslan
Emília Pietriková
Lukáš Pancurák

Proceedings reviewers:
Ján Bačík
Michaela Bačíková
Ľubomír Beňa
Peter Bober
Lenka Bubeňková
Miloš Drutarovský
František Ďurovský
Daniel Dzivý
Jaroslav Džmura
Peter Feciľak
Pavol Galajda
Ján Genči
Peter Girovský
Filip Gurbáľ
Milan Guzan
Renát Haluška
Maroš Harahus
Daniel Hládek
Marek Horváth
Ján Hurtuk
Ondrej Kainz
Erik Kajáti
Slavomír Kardoš
Jozef Király
Štefan Korečko
Tomáš Kormaník
Eva Kupcová
Karol Kyslan
Milan Lacko
Tomáš Lenger
Ľubomír Livovský
Peter Lukács
Branislav Madoš
Ján Magyar
Dušan Medved
Miroslav Murin
Stanislav Ondáš
Jakub Oravec
Ján Papaj
Peter Papcun
Ján Paralič
Marek Pavlík
Daniela Perduková
Alena Pietriková
Emília Pietriková
Matúš Pleva
Marek Pástor
Viktor Šlapák
Branislav Sobota
Miroslav Sokol
Zuzana Sokolová
Ján Staš
Martina Szaboová
Mária Timková
Ján Vaščák
Igor Vehec
Iveta Zolotová
Norbert Ádám
Juraj Ďurišin
Jaroslava Žilková

Copyright © 2025 FEI TUKE
ISBN 978-80-553-4854-4

Contents

Hajnalka Haneszova
Development of the software simulation framework gSeaGen for neutrino telescope of the experiment KM3NeT 8

Radovan Zajac
Využitie technológií umelej inteligencie v prostredí webových aplikácií 15

Miroslav Jusko, Peter Lukács
Aplikácia grafénu vo flexibilnej elektronike 21

Tadeáš Kmecik, Peter Girovský, Daniel Gordan, Tomáš Basarik
Výpočtovo Efektívne Metódy na Identifikáciu a Kompenzáciu Cogging momentu 27

Ľuboš Ovseník, Jakub Oravec, Zuzana Liščinská, Dávid Andrišov
Návrh a realizácia firemnej metalickej počítačovej siete 35

Ľuboš Ovseník, Jakub Oravec, Zuzana Liščinská, Viliam Jenča
Optimalizácia viac-kanálového DWDM systému 43

Ľuboš Ovseník, Jakub Oravec, Zuzana Liščinská, Branislav Kačmarik
Komunikácia FSO systémov v UV spektre 50

Ľuboš Ovseník, Jakub Oravec, Zuzana Liščinská, Daniel León Labanc
Inteligentný psí obojok s monitorovaním jeho vitálnych funkcií 56

Ján Bačík
Využitie dronov a umelej inteligencie v bezpečnostných systémoch 62

Peter Nemergut, Alena Pietriková
Simulácia vplyvu defektov na distribúciu tepla vo veľkoplošných spojoch výkonových modulov 67

Ladislav Hric, Daniela Perduková
Real-time detekcia obrazcov pri obmedzených datasetoch pomocou YOLO11 75

Filip Mahdal, Peter Lukács, Miroslav Jusko, Tomáš Lenger
Využitie papiera ako substrátu v elektronike 80

Jerguš Ševec, Peter Lukács, Miroslav Jusko, Tomáš Lenger
Flexibilné mikrozberače energie z ľudského tela 87

Ján Kaňuch, Peter Girovský, Norbert Rešovský
Nové trendy vo vývoji a využití elektrických strojov s permanentnými magnetmi 94

Peter Nemergut, Peter Girovský
Návrh a vyhotovenie pracovného nástroja pre vykonávanie dotykov na dotykových obrazovkách . 102

Ondrej Kainz, Christián Gučik, Miroslav Murin, Ivana Nováková
Scalable Infrastructure Monitoring System for Corporate Environments 107

Nazarii Herasymchuk, Ondrej Kainz, Miroslav Murin, Ivana Nováková
Web Platform for Personal Finance Management 115

Filip Rzuhovský, Ondrej Kainz, Miroslav Murin, Ivana Nováková
Design and Implementation of an AI-Driven Mobile Fitness and Nutrition Assistant 122

Jakub Škulec, Ondrej Kainz, Miroslav Murin, Ivana Nováková
Web Platform for Report Generation and Support Request Management 130

Andrii Khainas, Ondrej Kováč, Antónia Kováčová
Kompresné snímanie EKG pomocou učenia slovníkov 137

Tomáš Kanálik, Ondrej Kováč, Antónia Kováčová
Základné vlastnosti signálov globálneho lokalizačného systému 142

Pavol Fedor, Daniela Perduková
Nízkoenergetický statický zdroj pre elektrosystémy 148

Branislav Sobota, Denis Redaj, Štefan Korečko
Digital Twin of a Humanoid Robot in a Virtual Reality Environment 152

Dominik Ivanoc, Branislav Madoš
Generátor internetových portálov s využitím Model Driven Architecture 159

Tomáš Vargoško, Miroslav Murin, Ondrej Kainz, Miroslav Michalko
Software for testing and managing HTTP requests 165

Matúš Penička, Miroslav Murin, Lenka Bubeňková, Miroslav Michalko
Finance Management Application 172

Matúš Vaško, Miroslav Murin, Lenka Bubeňková, Miroslav Michalko
Web Application with Booking System for Service Providers 181

Jozef Király, Peter Daniel Kelemen
Simulácia decentralizovaných zdrojov v rámci distribučnej siete 189

Tomáš Čigáš, Branislav Madoš
Procedural Generation of Polygonal Surface Models of Plants for Computer Games 194

Tomáš Lenger, Peter Lukács
Analýza stability elektrického odporu vnorených spájkovaných spojov 202

Richard Král, Tibor Vince
Optimal Haptic Feedback for Spatial Navigation and Orientation of Visually Impaired Individuals 207

Martin Nguyen
Agent Reasoning Capabilities of GPT-4.1 in Algorithmic Problem Solving on CodeWars Platform 213

Kamil Berecký, Miloš Drutarovský
Šifrovací mód AES-XTS pre šifrovanie obsahu diskov 218

Ján Kaňuch, Peter Girovský, Nikolas Kondáš
Nové trendy vo využití magnetických kvapalín v elektrotechnike 225

Leo Buček, Branislav Madoš
Návrh simulátora jednoduchých logických obvodov 239

Daniel Gordan, Marek Pástor, Tadeáš Kmecik, Tomáš Basarik
Efektívne fotovoltické aplikácie a úloha zvyšovacieho meniča 246

Jaroslav Pajkoš, Miloš Drutarovský
Verifikácia RSA certifikátov v IoT uzle s podporou TLS protokolu 250

Karolína Polačková, Miloš Drutarovský
Digitálne podpisovanie občianskym preukazom a bezpečnostné riziko rodného čísla 257

Veronika Kalejová, Miroslav Murin, Ondrej Kainz, Miroslav Michalko
Platform for sharing projects 264

Jozef Kovalčín, Miloš Drutarovský
Autentizovaná výmena symetrických kľúčov s doprednou ochranou 271

Erika Abigail Katonová, Gabriel Strela, Rastislav Petija, Peter Feciľak
Automated discovery of topology structure in private network environments 279

Erika Abigail Katonová, Adam Fabišík, Rastislav Petija, Peter Feciľak
Classification of Data Flows Based on Network Traffic Characteristics 285

Erika Abigail Katonová, Alex Szaffko, Rastislav Petija, Peter Feciľak
A Predictive System for Network Failure and Congestion Using Time Series Analysis 291

Zuzana Gibová
Dynamické Modelovanie v Prostredí Coach 299

Daniel Marcin, Milan Lacko
Konštrukcia aktívneho balanséra na báze spínaných kondenzátorov 307

Simona Michaliková, Marek Horváth
Mental Health Support Application Designed for Local Data Processing and User Autonomy ... 311

Erik Podracký, František Ďurovský
Simulácia mild hybridného pohonu 318

Jakub Oravec, Ľuboš Ovseník, Zuzana Liščinská
Sieťová steganografia s využitím protokolu ICMP 326

Jakub Oravec, Ľuboš Ovseník, Zuzana Liščinská
Využitie laboratória sieťovej steganografie pre účely výuky 333

Ľubomír Livovský
Vplyv antény transpondéra na RF parametre bezkontaktného systému 339

Samuel Hliva, Sergej Chodarev
Usability Evaluation and Optimization of University Websites in the Eastern Slovakia Region .. 344

Eduard Michalek, Marek Horváth
Capturing Typing Patterns in Web Applications for Author Identification Research 351

Jaroslava Žilková, Peter Girovský, Samuel Tomko
Využitie kamerového systému v paletizačno-montážnych robotických aplikáciách 355

Miroslava Matejová, Ján Paralič, Anna Leonenko
Detekcia nenávistných prejavov v slovenskom jazyku s využitím modelov BERT 360

Miroslava Matejová, Ján Paralič, Karin Majirská
Vysvetliteľné modely umelej inteligencie pre multimodálnu klasifikáciu toxického obsahu 366

Jozef Kijac, Zuzana Sokolová, Patrik Jurík, Eva Kupcová
Vývoj interaktívneho robota KEMŤÁKa s jazykovým rozhraním na báze LLM 372

Martin Lukáč, Matúš Pleva, Eva Kupcová, Kristián Sopkovič
Automatické rozpoznávanie emócií z rečového signálu implementované vo webovej aplikácii 379

Eva Kupcová, Matúš Pleva, Kristián Sopkovič, Tsimafei Lapanik
Biometrická autentifikácia ako moderný prístup v internete vecí 385

Tibor Olearník, Miloš Drutarovský
Senzorový uzol s TCP/IP konektivitou a vzdialenou aktualizáciou firmvéru 392

Tomáš Buček, Matej Eliaš, Ján Genči
Transfer of the Evaluation of the Copymaster assignment to the Arena system 400

Tomáš Buček, Alina Stankovych, Ján Genči
Creation of Practical Tasks to Practice Inter-process Communication 408

Mikuláš Kövesi, Ján Papaj
Vizualizácia MANET-IoT siete 417

Lukáš Proner, Marek Horváth
Digital System for Psychodiagnostics 423

Matej Hric, František Ďurovský
Digitálne dvojča mechatronického systému 428

Roman Hudák, Ján Magyar
Identifying Industrial Processes from Time Series Data 434

Igor Vehec, Lukáš Garnek
Vplyv rozhrania Cu/PTF a Ag/PTF na dlhodobú stabilitu vnorených a nevnorených PTF rezistorov 440

Tomáš Basarik, Milan Lacko, Tadeáš Kmecik, Daniel Gordan
Metódy odväzbenia MAB meniča 445

Oleksandr Liashenko, Branislav Madoš
Procedurálny generátor polygonálnych povrchových modelov skál 449

Dušan Kantor, Ján Papaj
Aplikácia pre potreby propagacných akcií na katedre KEMT 455

Christopher Nagy, Ján Zbojovský
Návrh informačného systému pre tieniacu komoru 461

Renát Haluška, Matúš Pleva, Valér Jakubčo
Biometrický Prístupový Systém Riadený a Manažovaný cez Počítačovú Sieť 465

Renát Haluška, Matúš Pleva, Maroš Pavlus
Overenie Používateľa na Základe Dynamiky Stlačení Klávesov 471

Daniel Čorňák, Dušan Medveď
Možnosti regulácie výkonu v ostrovnej prevádzke 476

Matej Hirko, Dušan Medveď
Návrh funkčného modelu Stirlingovho motora 482

Pavlo Hryhoriev, Dušan Medveď
Vplyv elektromobility na mestskú elektrizačnú sieť a infraštruktúru 489

Yurii Hryhoriev, Dušan Medveď
Dynamické riadenie záťaže v inteligentných sieťach 496

Karina Savenko, Dušan Medveď
Optimalizácia ukladania elektrickej energie pre off-grid systém rodinného domu 501

Vladimír Szomosi, Dušan Medveď

Návrh zariadenia pre výrobu vodíka II 507

Artem Vozdrahanov, Dušan Medveď
Úloha prosumera vo flexibilnom riadení siete a jeho vplyv na energetickú bezpečnosť 513

Milan Lacko, Viliam Jakubec
Návrh zariadenia pre zapovanie 519

Kristián Sopkovič, Eva Kupcová, Matúš Pleva
Optimálne vytváranie vektorových reprezentácií vzhľadom na sémantiku slovenského jazyka 525

Erik Kajáti, Ľubomír Urblík, Veronika Kosiuk
Cybersecurity Implications of Pickle-based Model Serialization in Containerized Machine Learning Environments 531

Ján Kaňuch, Peter Girovský, Matej Onderko
Spätný vplyv LED žiaroviek na napájaciu sieť 536

Tomáš Buček, Martin Pacanovský, Ján Genči
Creation of a Tool for Extracting Metadata from Student Works 545

Peter Girovský, Ján Kaňuch, Tadeáš Kmecik, Pavol Mackanič
Návrh riadenia BLDC motora 550

Slavomír Kardoš, Dmytro Babitskyi
Implementácia inerciálnej jednotky pre zvýšenie bezpečnosti práce 554

Jozef Badár, Ján Papaj, Dávid Maček
Porovnanie smerovacích algoritmov v MANET pomocou metód posilňovaného učenia 560

Peter Girovský, Jaroslava Žilková, Tadeáš Kmecik, Daniel Kováč
Riadenie robotického ramena 566

Marek Pavlík, Maroš Bosák
Dopad variability obnoviteľných zdrojov energie na cenotvorbu elektriny: Kvantitatívna korelačná analýza pre slovenský energetický trh 571

Mykhailo Bakhir, Branislav Madoš
Procedurálny generátor polygonálnych povrchových modelov stromov s využitím L-systémov 577

Štefan Korečko, Jaroslav Mathes, Branislav Sobota
K slovenskému prekladu dotazníkov na meranie používateľskej skúsenosti vo virtuálnej realite .. 582

Ivana Nováková, Miroslav Michalko
Adaptive Multi-Physics vSLAM for Industrial AR: Prototype Implementation and Early Validation 588

Peter Bober
Meranie a modelovanie torzného kmitania v točivých zariadeniach poháňaných spínaným reluktančným motorom 593

Ján Vaščák
Využitie elektronického vodoznaku a siete ANFIS pri ochrane medicínskych dokumentov 600

Kevin Sendrovič, Ján Papaj
Zvyšovanie odolnosti smerovania v MANET – IoT 606

Ján Spišák, Peter Papcun

Distributed edge architecture for control of selected laboratory models 612

Ľubomír Urblík, Michal Ilek, Erik Kajáti, Iveta Zolotová
Security of Native, Docker, and Kubernetes Deployments: A Comparative Analysis for Edge and IoT Applications 620

Marek Sabol, Michaela Bačíková
Design and implementation of an automated timetable generation system using genetic and memetic algorithms 624

Raman Listarevich, Michaela Bačíková
Japanese Pronunciation Component for Self-Learners 630

Michal Popovič, Michaela Bačíková
Effective Collection, Structuring and Analysis of Medical Data 636

Marek Barnovský, Michaela Bačíková
Web Touch Component for Self-Learners of Japanese Hiragana 642

Andrea Kanóczová, Michaela Bačíková
Web Touch Component for Self-Learners of Japanese Katakana 648

Author's Index 655

Development of the software simulation framework gSeaGen for neutrino telescope of the experiment KM3NeT

[1]*Hajnalka HANESZOVÁ*

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

[1]hajnalka.haneszova@student.tuke.sk

Abstract **– Neutrinos are the most abundant elementary particles with non-zero mass in the Universe. However, their weak interaction with matter makes their detection particularly challenging, as they are highly elusive. To study the different types of neutrinos originating from diverse sources, various detection techniques have been developed. The goal of the KM3NeT collaboration is to construct a large-scale Cherenkov neutrino telescope in the Mediterranean Sea. It consists of two detectors: ARCA and ORCA, which are located near Sicily, Italy, and Toulon, France, respectively. The ARCA detector is designed to detect high-energy neutrinos of astrophysical origin, while ORCA is focused on studying neutrino oscillations with low-energy atmospheric neutrinos. Monte Carlo simulations are essential for neutrino telescope construction and data analysis. The `gSeaGen` simulation framework is used to model the neutrino event generation process. The objective is to extend the framework's capability to handle a variety of neutrino flux models without requiring modifications to the code. This enhancement allows for more accurate calculations and provides greater flexibility for researchers.**

Keywords **– gSeaGen, KM3NeT, neutrino flux input, simulation framework**

I. Introduction

Astronomy is a long-established field of science that has evolved with the advancement of observational techniques. While the study of photons has led to groundbreaking discoveries [1], the detection of these particles is often hindered by scattering and absorption [2]. In response, cosmic rays [3] and gravitational waves [4] have provided additional tools for exploring the Universe. Among these tools, neutrinos play a key role in multi-messenger astronomy, offering a unique perspective on astrophysical events [5].

The aim of the KM3NeT neutrino telescope, currently under construction in the Mediterranean Sea is to detect neutrinos from diffuse, point-like, or extended sources of atmospheric or astrophysical origin and subsequently to study their properties. In order to recognize neutrino interactions as accurately as possible, the simulation of these events and the detector response is essential. For this purpose, various simulation frameworks have been developed, including `gSeaGen`, which is a software simulation framework for generating neutrino events in the KM3NeT detector. This work enhances the `gSeaGen` simulation framework, allowing for more flexible neutrino flux modeling. The improvements made to `gSeaGen` contribute to more precise simulations, ultimately supporting more accurate predictions and research in neutrino astrophysics.

II. Neutrino Physics

A. Neutrino as elementary particle

Neutrinos are part of the Standard Model of Elementary Particles and are classified as leptons. Consequently, they are not subject to the strong interaction, but they do interact via the weak nuclear force . They are electrically neutral and have a very small mass, which is still not precisely known. They come in three flavors: electron neutrinos (ν_e), muon neutrinos (ν_μ), and tau neutrinos (ν_τ), whereas their antiparticles are subject to the same classification [6]. Neutrino oscillation is a phenomenon in which neutrinos change their flavor as they propagate through space. Neutrinos manifest themselves only through weak interactions [7].

B. Neutrino sources

Neutrinos cover a large energy range $E \in (10^{-6}, 10^{21})$ eV and they come from a wide variety of sources. They can have terrestrial or astrophysical origin [8]. Accelerated charged particle interactions, referred to as cosmic ray interactions, can produce neutrinos. If this process occurs in the Earths atmosphere, we call them atmospheric neutrinos. The characteristic energy range of atmospheric neutrinos is $E \in (10^{6}, 10^{12})$ eV. The production of atmospheric neutrinos involves *only the muon and electron (anti)neutrino flavors* [9]. The most energetic neutrinos are cosmogenic neutrinos of astrophysical origin and can indicate the existence of events and sources in the Universe that are not yet fully understood, such as active galaxies or gamma-ray bursts [10].

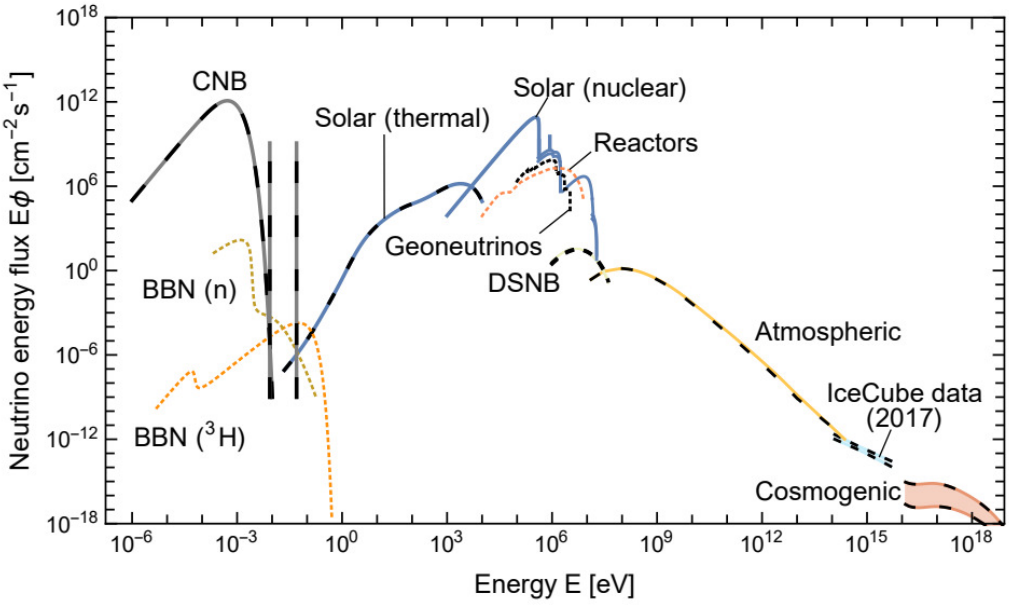


Fig. 1 GUNS at Earth, multiplied by energy E, summed over flavors and integrated over directions. Neutrinos are shown with solid lines, antineutrinos with dotted lines, and a combination of both with alternating dashed lines. Adapted from [11].

C. Neutrino detection

The neutrino interaction probability increases with energy, the neutrino flux declines at a significantly faster rate, surpassing the growth of the cross section by several orders of magnitude [2]. As a direct consequence, the detection of higher-energy neutrinos necessitates a larger target mass, which translates to the need for increasingly bigger detectors. Several detection techniques have been developed to detect neutrinos that address different energy ranges and research goals. Cherenkov radiation is one of the most widely used techniques for neutrino detection [12].

When a charged particle originating from neutrino interactions moves faster than the speed of light in a transparent medium, it emits Cherenkov radiation of visible blue light. In order to observe high- and ultra-high-energy neutrinos, km^3-scale detectors are needed. To achieve this cost-effectively, natural water or ice is used as the detection medium [13].

In these transparent media, interactions can have track-like or shower-like signatures. Track-like events are induced by muons, which are capable of traveling long distances in the detector medium, leaving behind a track of Cherenkov light. Shower-like events are produced by all types of neutrinos (including muon neutrinos), where the distinction between each type cannot be made based on the event signature alone. However, at higher energies above a few PeV, the decay of a τ lepton resulted from a tau neutrino interaction can be observed as a second cascade resulting in a double-bang signature [14].

In natural water and ice environments, there can be a significant amount of background light, which can be caused by bioluminescence of living organisms, radioactivity, especially the decay of potassium-40 ^{40}K, atmospheric neutrinos created in the Earth's atmosphere that have efficient energy to reach the detector, and in case of focusing on high-energy astrophysical neutrinos, atmospheric neutrinos [15]. The first two can be filtered out through local coincidences of photo sensors and time and space correlations, while for atmospheric muon background is eliminated by using the Earth as a shield. Atmospheric neutrinos constitute an irreducible background that

must be handled statistically. Key examples of Cherenkov neutrino telescopes of natural media are Baikal-GVD in Lake Baikal, Russia, KM3NeT in the Mediterranean Sea, and IceCube in the Antarctic ice.

III. The KM3NeT collaboration

The KM3NeT collaboration aims to construct a multi-km^3, deep-sea research infrastructure with detectors designed for neutrino detection. The project is divided into two main detectors: ARCA (Astroparticle Research with Cosmics in the Abyss) in the Ionian Sea near Sicily, Italy, focused on high-energy neutrinos and ORCA (Oscillation Research with Cosmics in the Abyss) in the Ligurian Sea near Toulon, France, aimed at studying neutrino oscillations [16]. The detectors are arrays of vertical strings connected to the seabed, stretched upwards by buoys. The strings are equipped with digital optical modules (DOMs) containing 31 photomultiplier tubes (PMTs) that detect Cherenkov light emitted by charged particles produced in neutrino interactions. They are under construction and at the end of the current phase, the ARCA detector will consist of 230 strings with 18 DOMs each, while ORCA will consist of 115 strings with 18 DOMs each. The number of strings deployed in February 2025 is 33 for ARCA and 24 for ORCA. When PMTs detect a photon, they generate a signal that is transmitted to the shore station via optical fibers. This is known as *all-data-to-shore* approch [17]. Subsequently, the data is stored in a data center, where they are processed and analyzed to identify neutrino events [18].

IV. Simulation process chain

Reconstruction is a key process for extracting meaningful scientific information from the detector's data. It involves fitting models to the recorded data, using simulations that replicate detector and data-taking conditions. Monte Carlo (MC) methods are employed, utilizing random sampling to solve complex problems. The simulation begins with event generation, followed by particle propagation and the simulation of Cherenkov photons reaching the PMTs. These simulated events are processed using the same reconstruction algorithms as real data, enabling accurate comparison. This process involves event generators like `GENHEN [19]`, `gSeaGen` [20], and `KM3BUU` [21], along with particle propagators and light simulation tools such as `KM3Sim` and `JSirene`, which enables precise modeling of particle interactions and ends with simulating the detector response by `JTriggerEfficiency`. Event reconstruction uses specialized algorithms to determine the physical properties of interactions, including energy and particle direction.

V. Neutrino flux input to gSeaGen

A. *gSeaGen*

The `gSeaGen` open-source simulation framework is designed for large-volume underwater/ice neutrino telescopes. The software can simulate events from multiple (anti)neutrino flavors, including both CC and NC interactions, within a single execution. Highly configurable, `gSeaGen` allows users to define parameters such as the surrounding medium (sea water, rock, mantle, or core) and choose between Cartesian or UTM coordinate systems. The framework defines three active volumes, instrumented, can, and interaction volumes, ensuring accurate simulation of neutrino interactions. Event weights are applied to scale fluxes and ensure statistical significance across a wide energy range. The framework integrates `GENIE` for neutrino interactions, allowing simulation of neutrinos up to 1 EeV [22]. Furthermore, `gSeaGen` can be linked with several external tools, including `CORSIKA` for cosmic ray simulations with atmosphere propagated to sea level, muon propagators like `MUSIC` and `PROPOSAL` and tau propagators like `TAUOLA` and `TAUSIC` [22]. Figure 2 illustrates the `gSeaGen` software and its linkable tools.

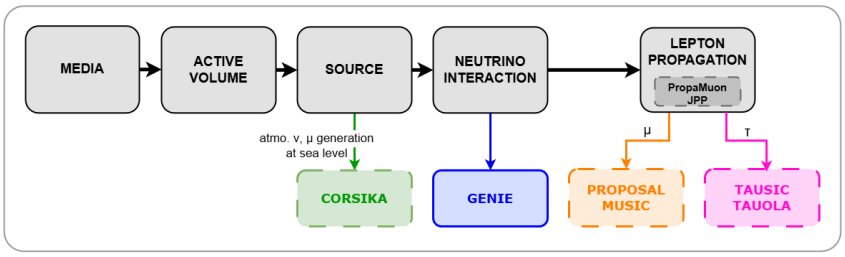


Fig. 2 The connected black blocks show the sequence of event generation in `gSeaGen`. Dashed frames indicate optional packages. Additional information is provided above certain tools. Abbreviations used: atmo. - atmospheric, ν - neutrino, μ - muon, τ - tau lepton. The diagram was created using draw.io.

B. Neutrino flux input to gSeaGen

Since gSeaGen is highly configurable, it allows for customization at compile time and at runtime. Executing the application with the -f option defines the neutrino flux model to be used to weight the generated events. The paper focuses on the neutrino flux input to gSeaGen and its extension to support a wider range of flux models. The feature allows for the use of new neutrino flux files without requiring code modifications.

The general syntax of the neutrino flux is defined as follows:

```
-f "SIMUL:FLUX,...[NU_CODE],FLUX...[NU_CODE];SIMUL:...[NU_CODE]"    (1)
```

where SIMUL represents the name of the flux model, FLUX specifies the flux data distribution, and NU_CODE refers to the required neutrino flavor PDG code for event generation. The valid NU_CODE values are: 12 for electron neutrinos, 14 for muon neutrinos, and 16 for tau neutrinos. For antineutrinos, the values are -12, -14, and -16, respectively. The available SIMUL and FLUX options are listed in Table 1.

Table 1
SIMUL and FLUX options possible values, where the new implementation extension is highlighted with an orange color.

No.	Type of SIMUL	SIMUL values	FLUX
1.	Function	FUNC1, FUNC2, FUNC3	Mathematical expression
2.	Flux models in ASCII format	FLUKA, BARTOL, HONDA	File name
3.	CORSIKA output	EVT, BIN	File name
4.	Flux models in ROOT format	TGRAPH, TGRAPH2D	File name with .root extension

The key point is that all flux models can be used in ROOT format as in ASCII format, while not all flux models are available in ASCII format.

In the previous implementation, the neutrino flux was read from an ASCII file based on the 2nd option of the SIMUL parameter. (Tab. 1). In this case, the structure of FLUKA, BARTOL, HONDA flux models is defined in the gSeaGen code, so it can accurately read the data from the file and store it in histograms. The flux is then evaluated using the interpolation methods of the histograms at the generated neutrino energy and direction.

C. Extension of neutrino flux input

The new feature enhances the gSeaGen software, enabling the input of neutrino flux data in ROOT file format. This allows for the correct reading and loading of data into TGraph/TGraph2D objects and uses graph interpolation methods for flux evaluation based on neutrino energy and direction.

The syntax for the -f option remains the same, but now supports TGRAPH for angle-independent fluxes and TGRAPH2D for zenith-angle-dependent fluxes. Note that there is no option to store azimuth angle-dependent fluxes within these ROOT data structures. The SIMUL argument now specifies the data format instead of the flux model, providing more flexibility.

In the gSeaGen, several modifications needed to be done to achieve the required functionality, including defining the methods for reading the ROOT files and extracting the right graphs. It is essential to determine the right name of the required graph object. For this purpose, a naming convention was established, summarized in Tab. 2.

Table 2
Naming convention for TGraph and TGraph2D graphs, based on the neutrino type defined with its PDG code.

Physics name	Symbol	Code	TGraph name	TGraph2D name
electron neutrino	ν_e	12	gflux1D_nu_e	gflux2D_nu_e
electron antineutrino	$\overline{\nu_e}$	-12	gflux1D_nu_e_bar	gflux2D_nu_e_bar
muon neutrino	ν_μ	14	gflux1D_nu_mu	gflux2D_nu_mu
muon antineutrino	$\overline{\nu_\mu}$	-14	gflux1D_nu_mu_bar	gflux2D_nu_mu_bar
tau neutrino	ν_τ	16	gflux1D_nu_tau	gflux2D_nu_tau
tau antineutrino	$\overline{\nu_\tau}$	-16	gflux1D_nu_tau_bar	gflux2D_nu_tau_bar

The process is essentially the same for both TGraph and TGraph2D cases, so to avoid redundancy, the LoadFluxGraph function is defined as a template. Templates are instantiated with specific types when needed, therefore, the compiler generates the code for the specific type of graph at compile time.

VI. Validation and comparison

To validate the new implementation, that it does not produce any errors, does not affect the already implemented functionality, and that the results are consistent with the previous implementation, the Honda et al. flux model was used as a test case. In order to be able to use the Honda flux model in `ROOT` format, the corresponding `ROOT` file needed to be created to `ASCII` format. For this purpose, a simple `ROOT` macro called **`hondaTGraph2D.C`** was created, which reads the zenith-dependent but azimuth-averaged `ASCII` file, passed as argument to the macro. It creates `TGraph2D` objects for each neutrino type, iterating through the file line by line to extract the required values and populate the corresponding graphs. While each graph has to follow the naming convention defined in Tab. 2. This macro can be used for visualization and plotting purposes, as shown in Fig. 3, and can be used as a reference for other flux models conversion to `ROOT` format.

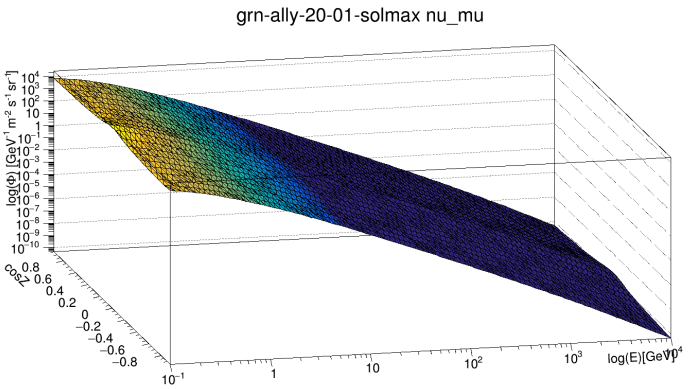


Fig. 3 `TGraph2D` visualization of HAKKM 2014 azimuth angle-averaged muon neutrino flux for Gran Sasso.

For the output comparison, two simulation runs were conducted, utilizing the same flux data but loaded from different formats, i.e. first from the `ASCII` file and then from the `ROOT` file. In both runs, the same parameters were used, with the only difference being the `-f` option. To achieve statistically significant results, one million neutrino events were generated. The initial inspection was based on the terminal output, which serves as a log of the simulation process. Both runs completed without errors and operated as expected, correlating well with each other. For comparing the events from the output files, the `compare-output-files.py` Python script was used. After opening the `ROOT` output files from both runs, relevant information was extracted, such as event weights and energies. Subsequently, the results were visualized as shown in Fig. 4 using a histogram with 50 bins, the weighted event density for both runs was compared.

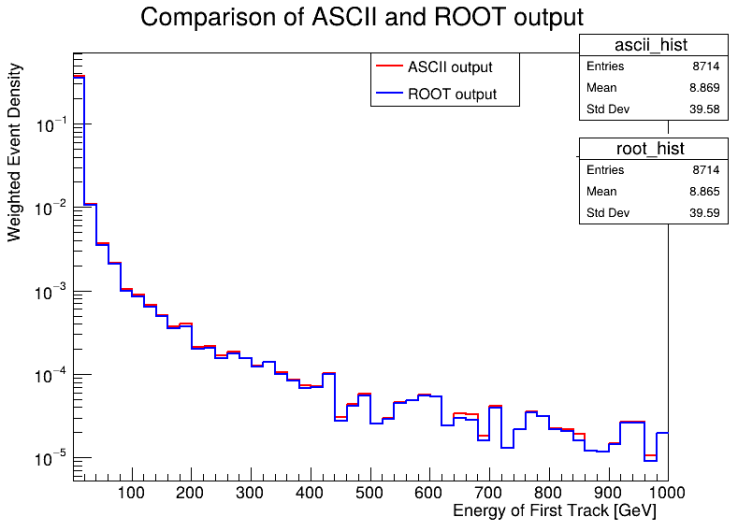


Fig. 4 Comparison of the `gSeaGen` runs. Data from `ASCII` files shown with a red and from `ROOT` files with a blue line.

While the correlation is excellent, minor discrepancies can be observed. With the `ASCII` input files, the neutrino flux data is represented and stored in histograms throughout the simulation.

However, correctly defining bins to represent the flux as a central value can be complex, especially near the boundary values. This issue is mitigated when using `TGraph/TGraph2D` input, as the values stored in `gSeaGen` match those in the tables precisely. This inconsistency in the histogram approach may lead to small differences in results, and the fact that the MC simulations are handled with some degree of randomization may also have an effect.

The evaluation process ended with performing benchmark simulations using the internal `gSeaGenCheck` project to evaluate the system's efficiency after implementing the new feature. Benchmark simulations were performed with both the locally compiled `gSeaGen` and the containerized version `v7.5.0`. The good correlation between the runs indicates that the implementation of the new feature does not have a negative impact on the system and function as expected.

VII. Conclusion

The new implementation for flux input in `ROOT` format offers several advantages over the previous `ASCII` method. It accommodates various flux types without modifying the `gSeaGen` code, offers better data handling efficiency, and ensures greater scalability with large datasets due to its structured binary format. While `ROOT` files are space-efficient and include data consistency checks, they do have a limitation in that they do not support a third dimension, limiting angle-dependent flux models to the zenith angle. Despite this, the implementation provides more precise data representation through `TGraph/TGraph2D` structures compared to histograms used with `ASCII` files, and the required preprocessing of flux data is manageable.

Acknowledgment

I would like to thank my supervisor RNDr. Blahoslav Pastirčák, CSc., and my consultant Mgr. Rastislav Dvornický, PhD., for their patience, invaluable guidance, and support throughout my thesis. Their expertise and constructive feedback were instrumental in shaping this work. I am also grateful to my consultant doc. Ing. Ján Genči, PhD., for offering me the opportunity to choose my thesis topic and for his guidance in shaping my research.

References

[1] M. J. Rycroft, *Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing*. Cham: Springer International Publishing, 2017, pp. 833–846. [Online]. Available: https://doi.org/10.1007/978-3-319-23386-4_38
[2] K. Zuber, *Neutrino Physics*.
[3] M. Boezio, R. Munini, and P. Picozza, "Cosmic ray detection in space," *Progress in Particle and Nuclear Physics*, vol. 112, p. 103765, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0146641020300120
[4] B. Abbott *et al.*, "Observation of gravitational waves from a binary black hole merger," *Physical Review Letters*, vol. 116, no. 6, 2016. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.116.061102
[5] I. Bartos and M. Kowalski, *Multimessenger Astronomy*, ser. 2399-2891. IOP Publishing, 2017. [Online]. Available: https://dx.doi.org/10.1088/978-0-7503-1369-8
[6] M. K. Gaillard, P. D. Grannis, and F. J. Sciulli, "The standard model of particle physics," *Rev. Mod. Phys.*, vol. 71, pp. S96–S111, 1999. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.71.S96
[7] Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.81.1562
[8] S. Hannestad, "Neutrinos in cosmology," *New Journal of Physics*, vol. 6, no. 1, p. 108, 2004. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/6/1/108
[9] K. Melis, "Studying the universe from -3000m n.a.p."
[10] C. Bogazzi, "Search for cosmic neutrinos with antares."
[11] E. Vitagliano, I. Tamborra, and G. Raffelt, "Grand unified neutrino spectrum at earth: Sources and spectral components," *Reviews of Modern Physics*, vol. 92, no. 4, 2020. [Online]. Available: https://doi.org/10.1103%2Frevmodphys.92.045006
[12] J.-C. L. M. Wurm, F. von Feilitzsch, "Neutrino detectors," in *Handbook of Particle Detection and Imaging*, 2nd ed., ser. Physics and Astronomy, Reference Module Physical and Materials Science, Reference Module Chemistry, Materials and Physics, I. Fleck, M. Titov, C. Grupen, and I. Buvat, Eds. Springer Cham, 2021, pp. 372–411, published: 13 December 2021 (Hardcover), 12 December 2021 (eBook). [Online]. Available: https://doi.org/10.1007/978-3-319-93785-4
[13] J. S. Blair Tarcliff, "Cherenkov radiation," in *Handbook of Particle Detection and Imaging*, 2nd ed., ser. Physics and Astronomy, Reference Module Physical and Materials Science, Reference Module Chemistry, Materials and Physics, I. Fleck, M. Titov, C. Grupen, and I. Buvat, Eds. Springer Cham, 2021, pp. 584–608, published: 13 December 2021 (Hardcover), 12 December 2021 (eBook). [Online]. Available: https://doi.org/10.1007/978-3-319-93785-4
[14] S. Hallmann, "Sensitivity to atmospheric tau-neutrino appearance and all-flavour search for neutrinos from the fermi bubbles with the deep-sea telescopes km3net/orca and antares," Ph.D. dissertation, Erlangen - Nuremberg U., 2021. [Online]. Available: https://ecap.nat.fau.de/wp-content/uploads/2021/02/Dissertation_HallmannSteffen_Opus.pdf
[15] E. Resconi, "Status and prospects of the icecube neutrino telescope," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 602, no. 1, pp. 7–8, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.nima.2008.12.013
[16] P. Bagley *et al.*, "KM3NeT: Technical Design Report for a Deep-Sea Research Infrastructure in the Mediterranean Sea Incorporating a Very Large Volume Neutrino Telescope," 2009.
[17] M. R. Musone and the KM3NeT Collaboration, "Outlook of the km3net neutrino telescope and latest results," in *Andromeda Proceedings BSM2023*, 2023. [Online]. Available: http://www.andromedapublisher.com/media/img/confprocimg/5/Pdf/14_Maria_Rosaria_Musone.pdf

[18] S. Adrián-Martínez *et al.*, "Letter of intent for km3net 2.0," *Journal of Physics G: Nuclear and Particle Physics*, vol. 43, no. 8, p. 084001, 2016. [Online]. Available: http://dx.doi.org/10.1088/0954-3899/43/8/084001

[19] A. Albert *et al.*, "Monte carlo simulations for the antares underwater neutrino telescope," *Journal of Cosmology and Astroparticle Physics*, vol. 2021, no. 01, p. 064–064, Jan. 2021. [Online]. Available: http://dx.doi.org/10.1088/1475-7516/2021/01/064

[20] C. Distefano, A. A. G. Soto, and P. Kalaczyński, "gseagen," 10 2024. [Online]. Available: https://git.km3net.de/opensource/gseagen

[21] J. Schumann and Rodri, "Km3net/km3buu: Km3buu v1.8.2," Jul. 2024. [Online]. Available: https://doi.org/10.5281/zenodo.12679605

[22] A. Garcia Soto, C. Distefano, and P. Kalaczyński, "Upgrading gseagen: from mev to pev neutrinos," *Journal of Instrumentation*, vol. 16, no. 09, p. C09008, 2021. [Online]. Available: http://dx.doi.org/10.1088/1748-0221/16/09/C09008

Využitie technológií umelej inteligencie v prostredí webových aplikácií

Radovan ZAJAC

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

radovan.zajac.2@student.tuke.sk

Abstrakt **– V tejto práci je predstavený návrh a implementácia chatbota, ktorý slúži ako pomocník pre študentov a uchádzačov o štúdium na Technickej univerzite v Košiciach. Systém funguje vo forme webovej aj mobilnej aplikácie. Odpovede sú spracovávané pomocou umelej inteligencie na základe dát z univerzitných systémov a boli počas vývoja testované z hľadiska presnosti a zrozumiteľnosti.**

Kľúčové slová **– AI, aplikácia, chatbot, umelá inteligencia, univerzitný asistent**

I. Úvod

Univerzity produkujú veľké množstvo informácií týkajúcich sa štúdia, predmetov, rozvrhov, prijímacieho konania či organizačných záležitostí. Hoci sú tieto údaje zverejnené online, často sa nachádzajú na rôznych stránkach alebo sú súčasťou samostatných systémov.

Z tohto dôvodu vznikol nápad vytvoriť riešenie, ktoré by sprístupnilo všetky dôležité informácie prostredníctvom jedného rozhrania. Motiváciou pre vznik chatbota bolo najmä uľahčiť prístup k informáciám novým študentom a uchádzačom o štúdium na Technickej univerzite v Košiciach(TUKE), ktorí často nevedia, kde konkrétne informácie hľadať.

Cieľom práce je preto návrh a implementácia inteligentného chatbota, ktorý prostredníctvom prirodzenej komunikácie odpovedá na otázky používateľov, pričom čerpá z existujúcich univerzitných dát a využíva technológie umelej inteligencie na ich spracovanie.

II. Chatbot

Chatbot je nástroj určený na automatizovanú komunikáciu, ktorý odpovedá na používateľské dopyty simuláciou písanej alebo hovorenej konverzácie. Zatiaľ čo jednoduchšie chatboty využívajú rozhodovacie stromy a fungujú ako systémy s pevne zadefinovanými odpoveďami, pokročilejšie riešenia využívajú umelú inteligenciu (AI) a spracovanie prirodzeného jazyka (NLP), aby porozumeli vstupu používateľa a zvládli aj komplexnejšie interakcie.[1]

III. Veľké jazykové modely (LLM)

Veľké jazykové modely predstavujú vysoko pokročilé nástroje spracovania prirodzeného jazyka, ktoré disponujú obrovským počtom parametrov a výnimočnou schopnosťou učiť sa. Vychádzajú z architektúry Transformer, kde kľúčovú úlohu zohráva mechanizmus self-attention. Ten umožňuje efektívne spracovávať sekvenčné dáta a zachytávať dlhodobé vzťahy v texte. Na zlepšenie kvality výstupov sa často využívajú techniky ako prompt engineering, doladenie modelu (fine-tuning)[2] alebo učenie s pomocou spätnej väzby od ľudí (RLHF).[3]

Najmodernejšie chatboty ako ChatGPT sú práve založené na LLM, ktoré umožňujú generovať prirodzene znejúce a relevantné odpovede [4].

IV. Existujúce riešenia

V úvodnej fáze vývoja bol realizovaný prieskum univerzitných webov na Slovensku a v Českej republike s cieľom identifikovať existujúce chatbotové riešenia. Na žiadnej z analyzovaných univerzít (napr. STU, UK, UNIZA, VUT, ČVUT) nebola implementovaná funkčná forma interaktívneho asistenta umožňujúca prirodzenú komunikáciu.

Zahraničným príkladom pokročilejšieho riešenia je systém *Ask Pounce* na University of Wisconsin–Milwaukee. Umožňuje komunikáciu v prirodzenom jazyku.V prípade presne položenej otázky poskytuje konkrétny odkaz na časť webu. V ostatných prípadoch odpovedá pomocou generatívnej umelej inteligencie [5].

Zaujímavou skutočnosťou je, že v záverečnej fáze vývoja nášho riešenia predstavila vlastného univerzitného chatbota aj TUKE. Tento bol zverejnený začiatkom roka 2025.[1]

V. Návrh používateľského rozhrania

Pri návrhu sme dbali na to, aby chatbot poskytoval presné a spoľahlivé odpovede na otázky týkajúce sa štúdia, rozvrhov, miestností, ako aj praktických aspektov študentského života na univerzite a v jej okolí.

A. Webová aplikácia

Pôvodná verzia chatbota bola navrhnutá ako webová aplikácia s možnosťou jednoduchého nasadenia do univerzitného webového prostredia. Pri návrhu rozhrania sa kládol dôraz na súlad s vizuálnou identitou univerzity – použitá bola farebná schéma, ktorá vychádza z oficiálneho dizajn manuálu TUKE.

Chatbot je umiestnený v pravom dolnom rohu obrazovky tak, aby bol používateľom k dispozícii bez narušenia hlavnej štruktúry stránky. V prípade potreby si môže používateľ zmeniť veľkosť rozhrania – zväčšiť ho pre pohodlnejšie čítanie rozsiahlejších odpovedí alebo ho minimalizovať pri nečinnosti.

Komunikácia prebieha prostredníctvom textového vstupu alebo hlasového ovládania. Výstupy systému sú dostupné nielen vo forme textu, ale aj ako syntetizovaný hlas, čo zvyšuje dostupnosť aplikácie aj pre používateľov so špecifickými potrebami.

Rozhranie podporuje tri jazykové verzie – slovenskú, anglickú a ukrajinskú, čo zvyšuje mieru personalizácie a prístupnosť pre zahraničných študentov a uchádzačov o štúdium.

B. Mobilná aplikácia

Na základe súčasných trendov bolo zrejmé, že väčšina študentov uprednostňuje vyhľadávanie informácií prostredníctvom mobilných zariadení. Aj preto bola popri webovej aplikácii navrhnutá mobilná verzia chatbota, ktorá zachováva rovnaké základné funkcie, no zároveň prináša aj nové možnosti.

Mobilná aplikácia umožňuje používateľovi komunikovať s chatbotom rovnako ako vo webovej verzii – prostredníctvom textového alebo hlasového vstupu s výstupom vo forme textu aj syntetizovaného hlasu. Systém je dostupný aj bez registrácie, čím sa zabezpečuje rýchly a jednoduchý prístup k informáciám.

Pre používateľov, ktorí sa rozhodnú vytvoriť si účet, je k dispozícii rozšírená funkcionalita. Po prihlásení si môžu vyplniť svoj profil s údajmi o štúdiu. Tieto informácie sú následne automaticky zohľadňované pri generovaní odpovedí. Mobilná aplikácia navyše umožňuje označovanie správ ako obľúbené.

VI. Získavanie a spracovanie dát

Dôležitou súčasťou návrhu systému bolo aj získanie dát, ktoré by chatbot mohol využívať na presné zodpovedanie otázok používateľov. Okrem neštruktúrovaných textov z webových stránok Technickej univerzity v Košiciach sa dôraz kládol najmä na spracovanie štruktúrovaných údajov dostupných prostredníctvom portálu `maisportal.tuke.sk`.

A. Neštruktúrované dáta

Pôvodný zámer počítal aj s ukladaním neštruktúrovaných textových dát zo stránok TUKE, najmä z webových podstránok fakúlt, kde sa nachádzajú harmonogramy, aktuálne informácie či dokumenty. V rámci prípravy boli vytvorené skripty na automatizovaný zber týchto údajov, ktoré využívali Python knižnice ako `BeautifulSoup` na extrakciu konkrétnych častí HTML (napr. nadpisov, odsekov).

Z dôvodu nejednotnej štruktúry stránok a hlavne veľkého objemu dát sme sa však rozhodli neukladať tieto dáta priamo do databázy, ale ponechať len možnosť ich budúceho získania pomocou pripravených nástrojov.

[1] Oficiálny TUKE chatbot: https://chatbot.tuke.sk/

B. Štruktúrované dáta

Portál MAIS poskytuje údaje o študijných programoch, predmetoch, rozvrhoch, miestnostiach, akademických rokoch a ďalších entitách, ktoré sú kľúčové pre fungovanie univerzitného asistenta. Pri analýze sme zistili, že každá z týchto entít obsahuje jednoznačný identifikátor, ktorý je priamo prítomný v HTML kóde. Na základe týchto identifikátorov je možné zostaviť presnú URL adresu, pomocou ktorej sa dá spätne získať podrobná informácia alebo navigovať používateľa priamo na požadovanú stránku.

Cieľom preto nebolo uložiť všetky textové výstupy, ale zamerať sa na získanie základných identifikátorov a názvov entít, ktoré postačujú na efektívne vyhľadávanie a prácu so systémom. Tieto údaje boli ukladané do databázy s použitím originálnych ID hodnôt z portálu, nie automaticky generovaných primárnych kľúčov.

Pre každý typ údajov bol vytvorený samostatný Python skript, ktorý iteroval cez všetky kombinácie fakúlt, akademických rokov, semestrov a foriem štúdia. Skripty vyhľadávali konkrétne HTML prvky obsahujúce požadované ID a pomenovania(napr. názov predmetu, akademický rok) a tieto údaje boli následne ukladané do databázy.

Vzhľadom na dobre štruktúrovaný HTML formát (napr. tabuľkové riadky) bola extrakcia údajov technicky nenáročná a umožnila jednoduchú automatizáciu. Výsledkom bol konzistentný základ štruktúrovaných univerzitných dát pripravených na ďalšie použitie v chatbote.

VII. Implementácia

Navrhovaný systém univerzitného asistenta je tvorený z troch hlavných častí – webovej aplikácie, mobilnej aplikácie a serverovej časti. Každá z nich bola vytvorená pomocou vhodne zvolených technológií a nástrojov.

A. Webová aplikácia

Používateľské rozhranie webovej aplikácie bolo implementované pomocou technológie `React`, s využitím nástroja `Vite`. Aplikácia komunikuje so serverom prostredníctvom REST API a jej komponenty boli navrhnuté pomocou CSS šablón.

Aplikácia okrem textového vstupu podporuje aj zadávanie otázok pomocou hlasu. Po aktivovaní mikrofónu sa hlasový vstup konvertuje do textovej podoby a následne spracuje. Odpovede chatbota sú okrem textovej formy taktiež dostupné aj pomocou hlasového výstupu.

Súčasťou riešenia bolo aj vytvorenie vlastného konfiguračného súboru `webpack.config.js`, ktorý zabezpečuje zostavenie aplikácie do jedného JavaScript súboru. Výsledný súbor je možné jednoducho vložiť do ktorejkoľvek webovej stránky.

Obr. 1 Chatbot na webovej stránke

B. Mobilná aplikácia

Mobilná verzia bola vytvorená pomocou multiplatformového frameworku `Flutter`, ktorý umožňuje vývoj pre systémy Android aj iOS z jedného zdrojového kódu. Obsahuje aj logiku registrácie, prihlasovania a personalizácie profilu.

Na hlavnej obrazovke sa nachádza samotný chat, ktorý slúži ako primárne rozhranie na komunikáciu. V hornej časti aplikácie je k dispozícii prepínač svetlého a tmavého režimu, možnosť zmeny jazykovej verzie a prístup do používateľského profilu.

Používateľ má možnosť po prihlásení vyplniť študijné údaje ako fakulta, program, ročník či forma štúdia, ktoré sa následne automaticky zohľadňujú pri generovaní odpovedí. Aplikácia taktiež umožňuje označiť vybrané odpovede ako obľúbené, ktoré sú následne dostupné priamo v profile.

C. Serverová časť

Serverová časť systému je postavený na Python frameworku `Django`, ktorý zabezpečuje správu používateľských účtov, autentifikáciu, prístup k databáze a spracovanie požiadaviek z webovej a mobilnej aplikácie. Server vystavuje REST API rozhranie, prostredníctvom ktorého prebieha všetka komunikácia so systémom.

Na zber a spracovanie dát boli využité Python knižnice ako `requests`, `BeautifulSoup`, `Scrapy`, `pdfplumber` a `PyMuPDF`, ktoré umožnili extrakciu informácií z HTML aj PDF formátov.

D. Externé služby

Do systému boli integrované aj externé služby:

- `OpenAI API` – na generovanie odpovedí prostredníctvom jazykového modelu, ktorý spracováva otázky používateľa.
- `Google Places API` – na získavanie informácií o okolitých lokalitách, ako sú reštaurácie, kaviarne, športoviská a iné objekty v Košiciach.
- `Google Custom Search API` – na vyhľadávanie všeobecných informácií z tuke podstránok.

VIII. Výber jazykového modelu

Počas návrhu systému sa ukázalo, že kľúčom k presnosti odpovedí nie je výhradne výber najvýkonnejšieho modelu, ale najmä to, aké údaje model spracúva. Preto sa dôraz kládol na vyhľadanie správnych a relevantných informácií, ktoré slúžia ako vstup pre generovanie odpovedí.

Zvolený model tak neplní len úlohu jazykového rozhrania, ktoré formuluje odpovede, ale zároveň pomáha s analýzou otázok a ich zaradením do kategórií, čo je dôležité pre výber vhodnej metódy spracovania.

V rámci testovania boli zvážené viaceré moderné jazykové modely (Claude 3.5 Haiku, Gemini 2.0 Flash, GPT-4o mini). Pre finálnu implementáciu bol zvolený `GPT-4o mini` od spoločnosti OpenAI, hlavne kvôli pomeru cena/výkon.

IX. Spracovanie otázok

Spracovanie otázok predstavuje najdôležitejšiu časť systému. Po prijatí vstupu prechádza otázka viacerými krokmi v rámci reťazca(pipeline), ktorý zabezpečuje jej porozumenie, klasifikáciu a vytvorenie relevantnej odpovede. Počas tohto procesu je využívaný jazykový model `GPT-4o mini`, ktorý pomáha pri úprave otázky, rozpoznaní zámeru používateľa a sformulovaní výstupu.

V úvodnej fáze sa otázka pomocou modelu optimalizuje – odstránia sa preklepy, zjednoduší sa jazyk a extrahujú sa kľúčové výrazy. Následne ju model zaradí do jednej zo štyroch kategórií:

- `Štúdium a univerzitné dáta`: Otázky týkajúce sa predmetov, rozvrhov, miestností, študijných programov alebo stravovania sú spracovávané na základe štruktúrovaných údajov uložených v databáze. Po prijatí otázky model najprv automaticky klasifikuje, o aký typ požiadavky ide – napríklad či sa týka konkrétneho predmetu, rozvrhu v zadanom semestri. Následne sa vykonáva vyhľadávanie požadovanej entity v databáze. Ak nie je presná zhoda názvu nájdená (napr. používateľ použije skrátený alebo nepresný názov), systém využije `fuzzy porovnávanie`, ktoré porovnáva reťazce na základe podobnosti, aby sa zistila najpravdepodobnejšia zhoda.
  Po identifikovaní požadovanej entity sa zostaví presná URL adresa do systému MAIS, ktorá slúži ako overiteľný zdroj. Odpoveď je následne jazykovo spracovaná a prezentovaná používateľovi vo forme prirodzeného textu.
- `Lokálne otázky mimo univerzity`: napríklad o reštauráciách, pamiatkách, sú spracovávané pomocou samostatného modelu.
  Jazykový model najprv analyzuje vstupnú otázku a extrahuje jej kľúčové prvky – typ požadovaného miesta (napr. „kaviareň“) a prípadne lokalitu alebo orientačný bod (napr. „pri Staničnom námestí“). Na základe týchto údajov sa zostaví štruktúrovaný dotaz, ktorý je následne odoslaný

službe Google Places API. Tá poskytne presné údaje o miestach. Získané výsledky sú následne spracované a zhrnuté do odpovede.

- `Všeobecné informácie o TUKE`: Otázky týkajúce sa napríklad harmonogramu akademického roka, prijímacieho konania alebo kontaktov sú vyhľadávané prostredníctvom `Google Custom Search API`, ktoré umožňuje prehľadávať obsah verejných univerzitných stránok. Návratové stránky sú následne analyzované pomocou vlastných skriptov na spracovanie neštruktúrovaných dát.
  V prípade, že medzi výsledkami vyhľadávania sa nachádzajú aj PDF dokumenty, systém využíva nástroj `pdfreader`, ktorý umožňuje extrahovať a analyzovať textový obsah týchto súborov.
- `Otázky mimo akademického kontextu`: model rozpozná ako irelevantné a odpovie s odporúčaním obrátiť sa na iný zdroj.

Celý reťazec spracovania riadi centrálna funkcia `pipeline()`, ktorá zabezpečuje správne smerovanie otázok, výber vhodného zdroja a koordináciu modelov pri generovaní odpovede.

X. Vyhodnotenie a testovanie

Vzhľadom na to, že oficiálny TUKE chatbot bol zverejnený v záverečnej fáze vývoja nášho riešenia, naskytla sa vhodná príležitosť na porovnanie kvality a rýchlosti odpovedí oboch systémov.

A. Metodika testovania

Na účely vyhodnotenia bola zostavená testovacia sada otázok, ktoré simulujú reálne požiadavky študentov a uchádzačov. Otázky boli úmyselne formulované neformálne – bez diakritiky, so skratkami a hovorovým štýlom, tak ako by ich zadala cieľová skupina používateľov.

Testovanie bolo vykonané na webovej verzii aplikácie vyvinutej v rámci tejto práce. Pre každú otázku sme zaznamenali:

- čas spracovania odpovede,
- počet vstupných a výstupných tokenov,
- subjektívne hodnotenie presnosti a relevantnosti odpovede,
- porovnanie s výstupom oficiálneho TUKE chatbota.

B. Zhodnotenie presnosti odpovedí

Z hľadiska presnosti poskytoval navrhnutý chatbot vo väčšine prípadov relevantné, presné a overiteľné informácie, často vrátane priamych odkazov na oficiálne univerzitné systémy ako MAIS.

`Otázky týkajúce sa štúdia a univerzitných údajov`: Náš systém bol schopný generovať presné odpovede na základe dát uložených v databáze a identifikátorov z MAIS. TUKE chatbot často poskytol všeobecnejšiu, menej konkrétnu odpoveď bez spätnej väzby na používateľa alebo bez odkazu na oficiálny zdroj.

Príklad 1 – Kedy je prednáška z funkcionalneho programovania v ZS?

Náš chatbot: Prednáška sa koná v utorok od 09:10 do 10:40 v miestnosti L9-A504 alebo online. Odpoveď obsahovala aj hypertextový odkaz na rozvrh v systéme MAIS.

TUKE chatbot: Uviedol nesprávny deň (streda) a miestnosť (B524), bez akéhokoľvek odkazu na oficiálny zdroj.

Príklad 2 – Kde sa nachádza miestnosť P004 / P25?

Náš chatbot: Poslucháreň sa nachádza v hlavnej budove B na adrese Letná 9, 040 20 Košice. Odpoveď obsahovala aj počet miest na sedenie a odkaz na detailné informácie v MAIS.

TUKE chatbot: Namiesto lokalizácie miestnosti opísal predmet, ktorý sa v nej vyučuje.

`Lokalizačné otázky`: Náš chatbot úspešne využil externé zdroje na zodpovedanie otázok o okolí univerzity, čím poskytol konkrétne odporúčania.

Príklad – Kde nájdem basketbalové ihrisko pri Staničnom námestí?

Náš chatbot: Pri Staničnom námestí sa nachádza niekoľko basketbalových ihrísk. Playko – Mestský park, 040 01 Košice...

TUKE chatbot na podobné otázky nedokázal odpovedať. Počas testovania bola zaznamenaná výrazne dlhá odozva, pričom odpoveď nebola vygenerovaná ani po viac než troch minútach.

`Všeobecné otázky o TUKE`: V tejto kategórii boli odpovede oboch systémov porovnateľné. TUKE chatbot poskytoval kvalitné výstupy pri dotazoch na harmonogram, prihlášky či základné kontakty.

`Mimotematické otázky`: Náš systém cielene odmietol reagovať na otázku o politickej osobe, čím dodržal zásadu neutrality. Oproti tomu univerzitný odpovedal bez výhrad, vrátane detailov o politickej kariére.

C. Zhodnotenie výkonnosti a nákladov

Na základe vykonaného testovania možno konštatovať, že náš navrhnutý systém dosahoval okrem vysokej presnosti aj veľmi dobré časy odpovedí. V porovnaní s oficiálnym chatbotom boli odpovede generované rýchlejšie, ale testovanie prebiehalo v prostredí bez paralelného zaťaženia. V produkčných podmienkach sa však čas odozvy môže meniť v závislosti od vyťaženia.

Najnáročnejšou kategóriou z pohľadu počtu spracovaných tokenov boli všeobecné otázky, ktoré vyžadovali analýzu viacerých externých zdrojov. Naopak, mimotematické otázky mali najnižšiu spotrebu, keďže boli riešené jednoduchým zamietnutím.

Tabuľka 1
Priemerné hodnoty podľa typu otázky

Typ otázky	Čas (s)	Input tokeny	Output tokeny
Databázové	8,60	3192	274
Lokalizačné	8,97	1986	356
Webové vyhľadávanie	8,72	3945	322
Mimotematické	3,29	1069	106

Model `GPT-4o mini` predstavuje cenovo efektívne riešenie. Vzhľadom na nízku cenu modelu spracovania, rýchlu odozvu možno považovať navrhované riešenie za vhodné pre nasadenie v univerzitnom prostredí.

XI. Záver

Výsledkom tejto práce je chatbot určený pre študentov TUKE, ktorý poskytuje presné a overiteľné odpovede na akademické aj praktické otázky. V porovnaní s existujúcim riešením preukázal vyššiu presnosť najmä pri otázkach týkajúcich sa predmetov, rozvrhov, miestností a ďalších údajov získaných z portálu MAIS. Zároveň dokáže odpoveď na otázky s geografickým kontextom a má väčšiu kontrolu nad obsahom.

Systém bol vyvinutý vo forme webovej aj mobilnej aplikácie a podporuje viacero spôsobov interakcie, vrátane hlasového ovládania či personalizácie odpovedí podľa vyplneného profilu používateľa pri mobilnej verzii.

Chatbot má potenciál zjednodušiť komunikáciu medzi študentmi a univerzitou a v budúcnosti môže byť zaujímavé preskúmať možnosť jeho prepojenia s existujúcim systémom TUKE chatbot, čím by sa dali skombinovať výhody oboch riešení.

References

[1] Teaganne Finn, "6 types of chatbots and how to choose the right one for your business," https://www.ibm.com/think/topics/chatbot-types, 2025, cit. 2025-04.

[2] X. Lin, W. Wang, Y. Li, S. Yang, F. Feng, Y. Wei, and T.-S. Chua, "Data-efficient fine-tuning for llm-based recommendation," in *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, 2024, pp. 365–374.

[3] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang *et al.*, "A survey on evaluation of large language models," *ACM transactions on intelligent systems and technology*, vol. 15, no. 3, pp. 1–45, 2024.

[4] S. K. Dam, C. S. Hong, Y. Qiao, and C. Zhang, "A complete survey on llm-based ai chatbots," *arXiv preprint arXiv:2406.16937*, 2024.

[5] University of Wisconsin-Milwaukee and University of Wisconsin Board of Regents, "Ask pounce!" https://uwm.edu/finances/finances/ask-pounce/, 2024, cit. 2025-04.

Aplikácia grafénu vo flexibilnej elektronike

[1] *Miroslav JUSKO,* [2]*Peter LUKACS*

Katedra Technológií v Elektronike, Fakulta Elektrotechniky a Informatiky, Technická univerzita v Košiciach, Slovenská republika

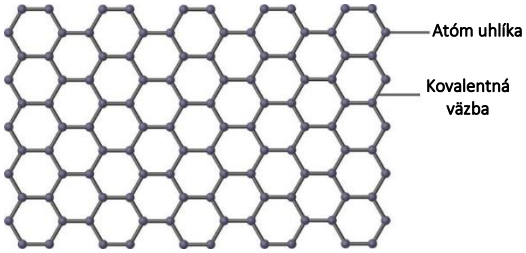
[1]miroslav.jusko@tuke.sk, [2]peter.lukacs@tuke.sk

***Abstrakt* — Táto práca sa zaoberá grafénom – materiálom s výnimočnými mechanickými, elektrickými a chemickými vlastnosťami, ktorý má veľký potenciál vo flexibilnej elektronike. Stručne opisuje jeho štruktúru, spôsoby výroby a možnosti využitia. Vďaka svojim unikátnym vlastnostiam umožňuje grafén vývoj nových typov zariadení aj zlepšenie existujúcich technológií. Od jeho objavenia v roku 2004 vzniklo množstvo výskumov, z ktorých niektoré zaujímavé štúdie sú v tejto práci zhrnuté a analyzované v kontexte moderných aplikácií v oblasti senzorov, zdrojov energie a elektroniky.**

***Kľúčové slová* — Flexibilná elektronika, grafén, senzory, zdroje energie**

I. Úvod

Grafén je dvojrozmerná štruktúra grafitu. Teoretické úvahy o vzniku jednoatómovej štruktúre uhlíka siahajú až do počiatku 40 rokov minulého storočia. Po prvýkrát sa grafén podarilo objaviť a jednoznačne opísať tímu, ktorý viedli profesori Andre Geim and Kostya Novoselov na univerzite v Manchesteri v roku 2004. Neskôr za túto prácu získali v roku 2010 Nobelovu cenu za fyziku. Tento materiál má jedinečné mechanické, elektrické a aj optické vlastnosti, čo ho predurčuje ako ideálneho kandidáta pre moderné elektronické zariadenia.



Obr. 1 Štruktúra grafénu

II. Popis grafénu

A. Jedinečné vlastnosti grafénu

Chemické vlastnosti

Grafén je tvorený jednoatómovou plenárnou vrstvou uhlíka, ktorý je usporiadaný do pravidelných šesťuholníkov. Dĺžka väzieb medzi jednotlivými atómami je 0,142 nm. Grafén je základnou štruktúrou grafitu a uhlíkových nanorúrok.

Ide o jedinú forma uhlíka, pri ktorej je každý z jeho atómov schopný z obidvoch strán tvoriť väzby s ďalšími molekulami. Na rohoch plátu grafénu vykazujú atómy uhlíka vysokú chemickú reaktivitu. Chyby v dvojrozmernej mriežke grafénu túto reaktivitu ďalej zvyšujú.

Mechanické vlastnosti

Ďalšou z jeho výnimočných vlastností je jeho pevnosť. Vďaka tomu, že jednotlivé atómy majú medzi sebou malú vzdialenosť, je grafén najpevnejší materiál na svete. Jeho pevnosť v ťahu sa pohybuje na hranici až 130 GPa. Pre porovnanie pevnosť ocele je iba 400 MPa a kevlaru 375 MPa.

Okrem vysokej pevnosti má grafén aj extrémne nízku hmotnosť, 0.77 mg na 1 m^2 [1]. Pre predstavu, kancelársky papier o rozmere 1m^2 má 1000x väčšiu hmotnosť. Jedna vrstva grafénu, ktorá by bola schopná pokryť futbalový štadión by potom vážila menej ako jeden gram.

Elektrické vlastnosti

Grafén sa radí k polovodičom bez zakázaného pásma. Atóm uhlíka má vo valenčnom pásme 4 elektróny, pričom jeden atóm je voľný a ostatné slúžia na spojenie s ďalšími atómami pomocou kovalentnej väzby. Tieto voľné elektróny potom zabezpečujú elektrickú vodivosť grafénu. Nazývajú sa aj pi (π) elektróny a nachádzajú sa nad a pod štruktúrou grafénu.

Pohyblivosť elektrónov grafénu je veľmi vysoká a pohybuje sa na 1.5 $m^2/V{\cdot}s$ pričom teoretický limit je rovný až 20 $m^2/V{\cdot}s$ [2].

B. Spôsoby výroby grafénu

Mechanická exfoliácia

Prvým grafén sa podarilo získať pomocou opakovanej mechanickej exfoliácie. Tento spôsob funguje vďaka slabej van der Waalsovej sile medzi vrstvami grafitu. Pri tomto experimente sa na kus lepiacej pásky pomocou trenia naniesla vrstva grafitu. V ďalšom kroku bola na ňu z vrchu prilepená druhá vrstva pásky. Následným oddelením došlo ku exfoliácii grafitu, čím bolo dosiahnuté zmenšenie jeho hrúbky. Niekoľkonásobným opakovaním tohto postupu bolo docielené vytvorenie dvojrozmernej štruktúry grafénu.

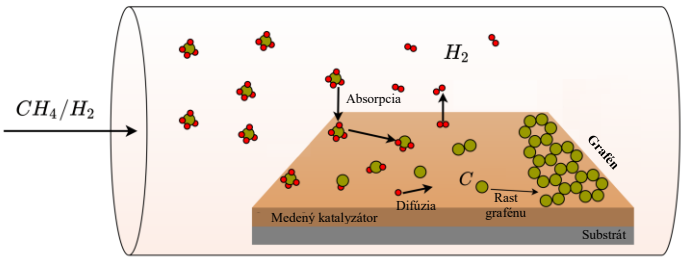
Tento spôsob napriek jednoduchému princípu fungovania dokáže vytvárať grafén s pomerne vysokou kvalitou. Nevýhodou je však jeho nízka výnosnosť

Chemická exfoliácia

Ide o dvojkrokový proces. V prvom kroku je potrebné znížiť van der Waalsové sily medzi jednotlivými vrstvami, čím sa zároveň zväčší aj vzdialenosť medzi nimi. To je docielené pomocou oxidácie grafénu silnými kyselinami ako napríklad manganistan draselný ($KMnO_4$) alebo kyselina sírová (H_2SO_4). V ďalšom kroku sú potom buď rapídnou zmenou teploty alebo pomocou ultrazvuku tieto vrstvy od seba oddelené [3]. Nevýhodou tejto metódy je znečistenie výsledného grafénu.

Chemická depozícia z plynnej fázy

V súčasnosti je to najčastejšie používaná metóda. Do komory je vháňaný plyn, ktorý obsahuje uhlík (napr. metán). Ten sa pri vysokých teplotách po dotyku s niektorými kovmi, ktoré pôsobia ako katalyzátor (napr. meď), rozloží. Následne sa voľné uhlíkové atómy usporiadajú na povrchu kovu do šesťuholníkovej grafénovej mriežky. Nakoniec je meď rozpustená a výsledkom je grafénová vrstva, ktorú je možné umiestniť na ľubovoľnú podložku podľa potreby.



Obr. 2 Chemická depozícia grafénu

III. Aplikácie grafénu vo flexibilnej elektronike

A. Flexibilné OLED displeje

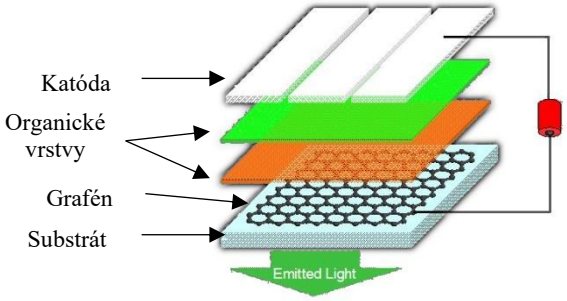
OLED displeje pozostávajú z elektroluminiscenčného organického materiálu umiestneného medzi dvoma elektródami. Ak je na tieto elektródy priložené jednosmerné napätie, vzniknutý náboj nabije elektroluminiscenčný materiál, v ktorom sa elektróny a diery navzájom rekombinujú, a tým vyžiaria svetlo. Aby bolo možné vzniknuté svetlo pozorovať, musí byť elektróda anódy zhotovená z transparentného, elektricky vodivého materiálu.

Najčastejšie používaným materiálom na tieto účely je oxid india a cínu (ITO). Tento materiál má nízky odpor a zároveň je dostatočne priehľadný. V súčasnej dobe, je však potrebné nájsť zaň náhradu. Hlavnou nevýhodou je jeho narastajúca cena, keďže sa pri výrobe využívajú vzácne

materiály a taktiež jeho slabá mechanická pevnosť. Aj pri malom ohnutí má totiž ITO tendenciu tvoriť malé praskliny, ktoré sa časom môžu spojiť a vytvoriť jednu veľkú prasklinu. Tento problém je obzvlášť závažný pri moderných ohybných OLED displejoch.

Grafén sa ukazuje ako vhodný materiál, ktorý by dokázal prekonať tieto nedostatky. Ten má veľmi vysokú pohyblivosť elektrónov, mechanickú pevnosť a tepelnú vodivosť. Oproti ITO má však jeden veľký nedostatok. Jeho odpor je viac ako 300 Ω/□. Pre porovnanie u doteraz používaných materiálov je to menej ako 10 Ω/□. Ďalšou nevýhodou je pomerne veľká energia potrebná na prekonanie bariéry medzi anódou a organickou vrstvou. Z tohto dôvodu majú OLED displeje s anódou vytvorenou z grafénu nižšiu účinnosť než tie s anódou z ITO.

Na riešenie tohto problému je možné nahradiť atómy uhlíka v mriežke čistého grafénu prímesami, ktoré do značnej miery dokážu upraviť jeho elektrické vlastnosti. Ako vhodné materiály na substitúciu sa ukázali byť atómy dusíka a bóru, pretože je ich veľkosť porovnateľná s atómom uhlíka. Pri použití CVD technológie je túto zámenu možné urobiť priamo v plynnej fáze vháňaním amoniaku alebo kyseliny boritej



Obr. 3 Štruktúra OLED displeja

V práci [4] bolo ukázané, že ak je grafén dopovaný kyselinou dusičnou a chloridom zlatistým tak dôjde ku zníženiu plošného odporu z pôvodných 300 Ω/□ na iba 34 Ω/□. Modifikáciou povrchu grafénu vodivým polymérom bolo taktiež dosiahnuté zníženie energie potrebnej na prekonanie bariéry medzi anódou a organickým materiálom. Takto upravený grafén má pri porovnaní s ITO podobné elektrické vlastnosti, pričom si zachováva svoje skvelé optické a mechanické atribúty.

B. Flexibilne organické solárne články

Vzrastajúci dopyt po prenosnej a nositeľnej elektronike akými sú napríklad biomedicínske prístroje a inteligentné hodinky, nesie so sebou potrebu pre ľahký, ohybný a výkonný zdroj energie. Z tohto pohľadu sa ako vhodný kandidát ponúkajú flexibilne solárne články. Momentálne produkované ohybné solárne články majú účinnosť iba 16% a maximálne je ich možné ohnúť o 60°. Takýto solárny panel je zložený z troch vrstiev. Kovová katóda, anóda tvorenou priehľadnou vrstvou ITO a organický materiál, ktorý zabezpečuje generovanie elektrickej energie.

Podobne ako pri ohybných OLED displejoch opísaných v predchádzajúcej časti, aj tu medzi hlavné nedostatky ITO patria vysoká cena a zlá mechanická pevnosť. Pre zvýšenie plošného odporu je grafén dopovaný kyselinou dusičnou.

V štúdii [5] bolo ukázané, že solárny článok, ktorého anóda bola tvorená grafénom, bolo možné ohnúť až o 78° viac v porovnaní s článkom s ITO.

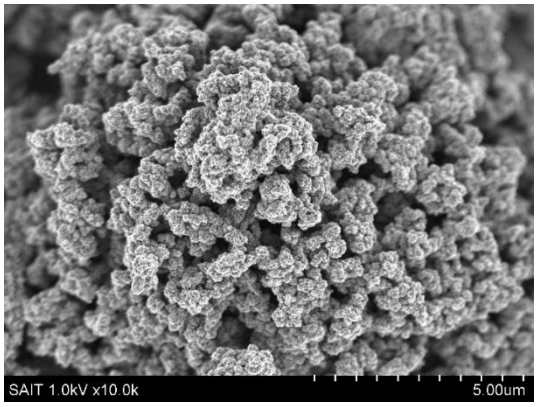
C. Flexibilne batérie

Vzrast nových flexibilných elektronických zariadení akými sú napríklad smart oblečenie, implantované medické zariadenia a podobne nesie so sebou aj potrebu flexibilného zdroja energie. Ako najlepším kandidátom na riešenie tohoto problému sa ukazujú flexibilné batérie. Problémom s existujúcimi flexibilnými batériami je ich veľkosť, váha a taktiež nízka miera ohybnosti.

Pre zlepšenie týchto nedostatkov sa ako ideálne riešenie opäť ponúka grafén. Ten je vo flexibilných batériách používaný v niekoľkých formách. Ako jednorozmerné grafénové vlákna, dvojrozmerné grafénové filmy a taktiež ako aj trojrozmerné grafénové siete.

Grafén je v tomto prípade používaný buď ako povrchová úprava hliníkových a medených elektród alebo dokonca aj ako ich úplná náhrada. Štandardné elektródy majú pri dlhodobom namáhaní tendenciu oddeliť sa od aktívneho materiálu. To vedie ku nenávratnej strate energie, ktorú je v batérii možné uložiť.

V roku 2017 kórejská firma Samsung vytvorila materiál nazývaný grafénové guličky. Vďaka tomuto materiálu bola schopná zvýšiť kapacitu batérie až o 45% a taktiež päťnásobne zrýchliť nabíjanie takéhoto to článku.



Obr. 4 Grafénové guličky

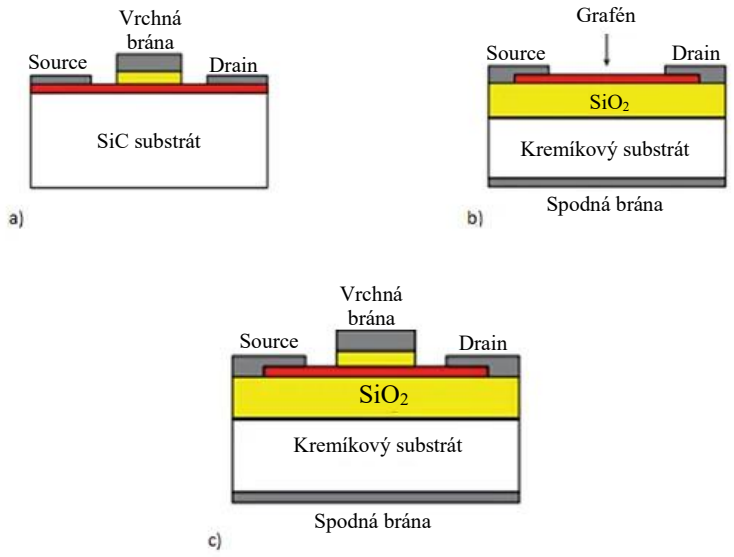
D. Flexibilné FET tranzistory

Tranzistory sú v dnešnej dobe považované za jeden z najdôležitejších stavebných blokov v celej elektronike. Väčšina dnes vyrobených tranzistorov je na báze kremíka, hlavne vďaka svojej malej veľkosti stabilite účinnosti a výkonu. Jednou z výhod kremíka je jeho prirodzená schopnosť oxidovať. Pre ďalší vývoj tranzistorov podľa zákona je však potrebné túto vrstvu každou generáciou zmenšovať. Prakticky však oxid kremíka o hrúbke 1-2 atómov nie je možné vyrobiť. Taktiež pre potreby flexibilnej elektroniky nie sú tieto tranzistory vhodné z dôvodu ich neohybnosti.

Na prekonanie týchto nedostatkov sa ukazuje ako možné riešenie použitie grafénu. Výroba tranzistorov na báze grafénu ako aktívneho člena štruktúry dokáže podstatne ovplyvniť evolúciu flexibilnej elektroniky.

Existujú tri typy grafénových tranzistorov:1.) s bránou vo vrchnej vrstve
2.) s bránou vo spodnej vrstve
3.) s bránou vo spodnej a vrchnej vrstve

Tieto konfigurácie sú zobrazené na Obr. 5.



Obr. 5 Tranzistor FET s a) bránou vo vrchnej vrstve, b) bránou vo spodnej vrstve a c) bránou po oboch stranách tranzistora

Ako pri tradičných FET tranzistoroch tak aj pri grafénových tranzistoroch ovplyvňuje potenciál na bráne tok elektrónov alebo dier cez vodivý kanál. Takýto tranzistor je v konečnom dôsledku veľmi citlivý, keďže hrúbka tohto kanálu je len jeden atóm.

Pri kremíkových tranzistoroch sú nosičmi náboja buď elektróny alebo diery. U grafénových tranzistoroch sa na tom v rovnakej miere môžu podieľať ako elektróny, tak aj diery. Tieto tranzistory majú takzvaný ambipolárny transport náboja. Pri negatívnom budení brány sa na transporte náboja podieľajú diery a pri pozitívnom elektróny.

Grafénové tranzistory majú oproti kremíkovým nižšie tepelné straty a taktiež lepší odvod tepla. Vďaka tomu dosahujú vyššie výkony s lepšou účinnosťou. Vďaka vysokej citlivosti sú vhodné pre použitie v biomedicínskych a chemických senzoroch. Ich nevýhodami sú hlavne vysoké výrobné náklady spojené s precíznou a komplexnou výrobou a absencia zakázaného pásma. Znamená to, že valenčné vodivostné pásmo sa prekrývajú na úrovni 0V. To v konečnom dôsledku spôsobí, že takýto tranzistor je veľmi ťažké uviesť do otvoreného stavu.

E. Senzory vitálnych funkcií

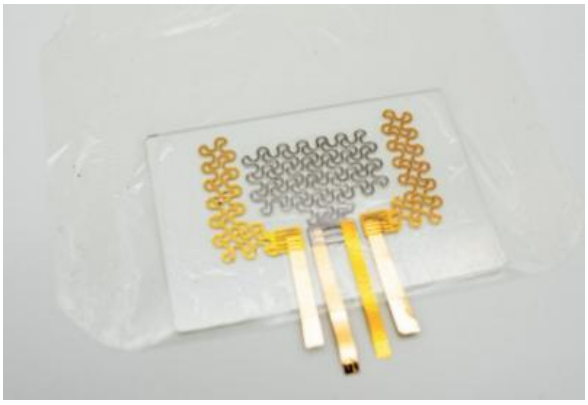
So zvyšovaním populácie rastie aj potreba na rýchlu a presnú diagnostiku vitálnych funkcií ľudského tela. Preventívna a personalizovaná medicína dokáže včasne predísť niektorým závažným chorobám. Stálym meraním biomarkerov ľudského tela je možné predpovedať zdravotný stav pacienta čo môže viesť ku skoršej diagnóze, určeniu choroby a taktiež ku skráteniu rehabilitácie. Hlavným komponentom zdravotného monitorovania sú nositeľné a implantovateľné senzory. Keďže ľudské telo je možné považovať za flexibilný mechanizmus, musia mať aj senzory ktoré sú naň pripojené rovnaké alebo podobné vlastnosti. Ďalšími požiadavkami sú biokompatibilita, spoľahlivosť, stabilita, komfort a v neposlednom rade aj cena.

Vďaka výnimočným vlastnostiam, ako sú vynikajúca elektrická a tepelná vodivosť, vysoká pohyblivosť nosičov náboja, veľký špecifický povrch a vysoká mechanická pevnosť, sa grafén javí ako mimoriadne perspektívny materiál pre vývoj moderných senzorov. Umožňuje konštrukciu širokej škály nositeľných aj implantovateľných zariadení vhodných na dlhodobé monitorovanie vitálnych funkcií..

Senzory na báze grafénu majú oproti klasický senzorom výhodu v tom, že majú vďaka veľkej špecifickej ploche veľmi vysokú citlivosť, Taktiež kvôli svojej flexibilite a nízkej hrúbke sú pri dlhodobom nosení pohodlnejšie.

Vo všeobecnosti ich vieme rozdeliť na invazívne a neinvazívne. Do neinvazívnej kategórie patria nositeľné a e-skin senzory. V druhej kategórii sa nachádzajú implantáty nervového, kardiovaskulárneho tráviaceho a pohybového ústrojenstva.

Ľudské telo produkuje periodické opakované signály známe ako aj biosignály. Tie je možné pomocou elektród zozbierať, odfiltrovať a previesť na elektrokardiogram (EKG/ECG), elektroencefalogramm (EEG), elektrookulogram (EOG) alebo elektromyogram (EMG). Pre zozbieranie presných a spoľahlivých výsledkov by tieto elektródy mali mať skvelé mechanické a elektrické vlastnosti. Všetky vlastnosti grafénu dokážu splniť tieto požiadavky. Na obr. 6 je možné vidieť príklad senzora na meranie EKG.



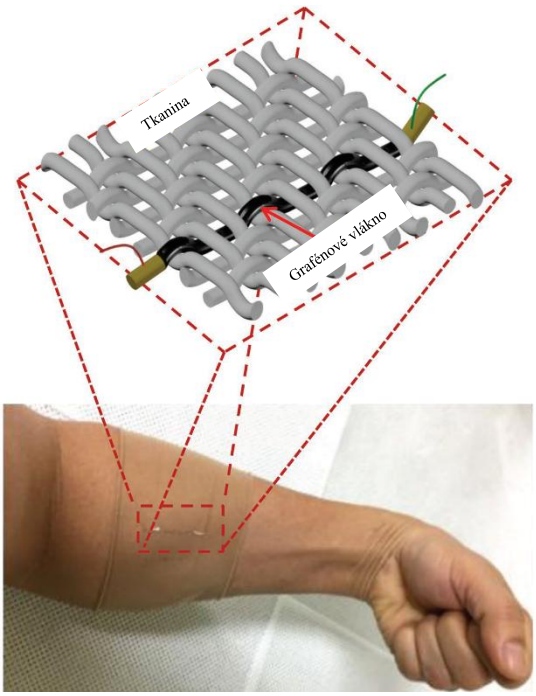
Obr. 6 Ukážka senzora pre meranie EKG na báze grafénu.

Skúmaním pohybu ľudského tela je možné získať niekoľko dôležitých signálov pre monitorovanie dýchacích a pohybových porúch, krvného tlaku a taktiež pre monitorovanie atletickej výkonnosti. Tlak vygenerovaný ľudským telom je možné rozdeliť na nízkotlakové, stredné (krvný tlak a tlkot srdca) a vysokotlakové (tlak spôsobený váhou človeka). Takéto senzory by mali pracovať v širokom rozsahu meraných hodnôt, mali by mať vysokú presnosť, nízku váhu, dobrú linearitu, robustnosť, rýchlu odozvu a taktiež aj dobrú opakovateľnosť.

Grafén je vhodný pre tieto účely kvôli značnej piezorezistivite. Existujú rôzne modifikácie štruktúr grafénu pre meranie tepu na ruke, zvukových vĺn, ohybu článkov prstov, mimiky tváre a ďalších aplikácií.

Ďalšou zaujímavou aplikáciou pre grafénové senzory sú snímače teploty. Teplota tela predstavuje skvelý indikátor stavu ľudského metabolizmu. Kontinuálnym meraním teploty je možné predikovať rôzne typy ochorení.

Tepelná vodivosť grafénu zaujala v posledných rokoch pozornosť mnohých vedcov. V práci [6] bol navrhnutý takýto senzor, ktorý pracoval na báze termorezistivity grafénu a dokázal pracovať v teplotách od 30 až po 80°C.



Obr. 7 Ukážka grafénového senzora teploty

Pre zvýšenie presnosti a komplexnosti monitorovania orgánov v ľudskom tele je potrebné použitie invazívnych senzorov. Podobne ako pri neinvazívnych senzoroch aj tu je potrebné riešiť problémy spojené s biokompatibilitou, biodegradáciou, minimalizáciou, trvácnosťou a napájaním senzora. Implantovateľné grafénové senzory sa zatiaľ pohybujú len na teoretickej báze vo forme simulácií.

Vďaka kombinácii mimoriadnej vodivosti, elektrochemickej stability, flexibility, mechanickej prispôsobivosti a transparentnosti je grafén takmer dokonalým riešením mnohých výziev spojených s návrhom neurónových rozhraní.

ZÁVER

Grafén predstavuje jeden z najperspektívnejších materiálov súčasnosti, najmä v oblasti flexibilnej elektroniky, kde svojimi výnimočnými vlastnosťami otvára cestu k novým technológiám a zariadeniam. Jeho využitie siaha od pokročilých senzorov a solárnych článkov, cez ohybné batérie, až po biomedicínsku diagnostiku a neurónové rozhrania.

Vývoj a výskum v tejto oblasti napredujú dynamickým tempom a neustále prinášajú nové poznatky o modifikácii grafénu a zlepšovaní jeho vodivých, mechanických a chemických vlastností. Napriek určitým výzvam, ako sú vysoké výrobné náklady a absencia zakázaného pásma, zostáva grafén kľúčovým kandidátom pre nasledujúce generácie elektroniky.

Očakáva sa, že trh s aplikáciami grafénu bude v najbližších rokoch výrazne rásť – podľa odhadov až o 45 % do roku 2030. Grafén tak naďalej zostáva v centre pozornosti výskumníkov aj priemyslu a má potenciál zásadne ovplyvniť vývoj technológií v mnohých oblastiach života.

POĎAKOVANIE

Táto práca vznikla s podporou projektu KEGA č. 011TUKE-4/2023 a projektov VEGA č. 1/0049/24 a 1/0278/24

ZDROJE

[1] T. Radadiya, A PROPERTIES OF GRAPHENE, publikované v European Centre for Research Training and Development UK, 2015

[2] A. R. URADE,I. LAHIRI,K.S. SURESH, Graphene Properties, Synthesis and Applications: A Review. 2023, roč. 75, č. 3, s. 614-630. ISSN 1047-4838

[3] M. S. A. Bhuyan,N. Uddin,M. Islam, F. A. Bipasha, S. S. Hossain, Synthesis of graphene. Online. International Nano Letters. 2016, roč. 6, č. 2, s. 65-83. ISSN 2008-9295

[4] T. H.Han, Y. Lee, M. R. Choi, S. H. Woo, S. H. Bae, B. H. Hong, J. H. Ahn, T. W. Leextremely efficient flexible organic light-emitting diodes with modified graphene anode. Online. Nature Photonics. 2012, roč. 6, č. 2, s. 105-110. ISSN 1749-4885.

[5] J. Rajesh, Flexible material in solarpanel for potential efficiency. Online. Materials Today: Proceedings. 2020, roč. 33, s. 1116-1120. ISSN 22147853.

[6] T. Q. TRUNG,H. L. LE, T. M. L. DANG, S. JU,S. Y. PARK, Fiber-Based, Wearable Temperature Sensor with Tunable Thermal Index for Healthcare Monitoring. Online. Advanced Healthcare Materials. 2018, roč. 7, č. 12. ISSN 2192-2640.

Výpočtovo efektívne metódy na identifikáciu a kompenzáciu cogging momentu

[1]*Tadeáš KMECIK,* [2]*Peter GIROVSKÝ,* [3]*Daniel GORDAN,* [4]*Tomáš BASARIK*

Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky
Technická Univerzita v Košiciach, Slovenská Republika

[1]tadeas.kmecik@tuke.sk, [2]peter.girovsky@tuke.sk, [3]daniel.gordan@tuke.sk,
[4]tomas.basarik@tuke.sk

***Abstract* — Táto štúdia sa zaoberá analýzou, identifikáciou a kompenzáciou cogging momentu v BLDC a PMSM motoroch. Riadenie je realizované pomocou vektorovej metódy Field Oriented Control (FOC). Na identifikáciu mapy cogging momentu podľa polohy rotora bola navrhnutá adaptívna schéma využívajúca q-zložku prúdu a Fast Fourier Transform (FFT). Výsledná mapa slúžila ako referencia pre návrh a porovnanie rôznych kompenzačných signálov, ktoré boli implementované do referencie q-prúdu pred vstupom do PI regulátorov. Cieľom bolo efektívne znížiť cogging moment bez zmeny riadiacej štruktúry. Navrhnuté metódy boli overené v prostredí Matlab – Simulink, pričom účinnosť sa hodnotila podľa zníženia zvlnenia uhlovej rýchlosti a presnosti kopírovania cogging momentu.**

***Keywords* — BLDC, PMSM, cogging moment, zvlnenie rýchlosti, zvlnenie momentu.**

I. Úvod

Motory s permanentnými magnetmi patria v súčasnosti medzi najrozšírenejšie typy elektrických pohonov. Vynikajú vysokým výkonom vzhľadom na objem, vysokou účinnosťou, presnosťou riadenia, dlhou životnosťou a spoľahlivou prevádzkou. Vďaka týmto vlastnostiam nachádzajú uplatnenie v širokej škále aplikácií, od domácich spotrebičov a spotrebnej elektroniky, cez elektromobilitu až po pokročilé pohony v robotike a automatizácii [1].

Rastúci dopyt po týchto motoroch zároveň zvyšuje tlak na znižovanie výrobných nákladov. To vedie výrobcov k zjednodušovaniu konštrukčných riešení, čo však môže mať za následok zvýraznenie nežiaducich parazitných momentov. Medzine patrí predovšetkým cogging moment, ale aj trecie momenty, nevyváženosť rotora či iné mechanické a elektromagnetické vplyvy. Tieto javy môžu negatívne ovplyvniť chod motorov, najmä v aplikáciách s požiadavkou na plynulosť a presnosť riadenia [1][3].

Riešenia zamerané na redukciu parazitných momentov možno rozdeliť do dvoch hlavných skupín. Prvou sú konštrukčné úpravy motorov, ktoré sú z hľadiska účinnosti potlačenia cogging momentu najefektívnejšie. Medzi tieto opatrenia patrí napríklad šikmé usporiadanie drážok statora, šikmé uloženie magnetov rotora, optimalizácia tvaru magnetov (šírka pólového oblúka, ich posunutie, zlomkový počet drážok na pól) a vytváranie zárezov v zuboch statora na potlačenie dominantných harmonických zložiek cogging momentu. Tieto riešenia však výrazne zvyšujú výrobnú náročnosť a náklady a môžu tiež obmedziť kompaktnosť a výkonovú hustotu motora [2].

Druhou možnosťou je kompenzácia cogging momentu prostredníctvom riadiacich algoritmov, ktoré injektujú vhodné kompenzačné signály do regulácie. Tento prístup je výhodný najmä tým, že ho možno použiť aj v už vyrobených motoroch bez nutnosti konštrukčných zásahov. Na druhej strane však môže mierne znížiť účinnosť motora, pričom požaduje určité požiadavky na výpočtový výkon riadiaceho systému. Kompenzácia je náročná aj vzhľadom na vysokú frekvenciu cogging momentu v porovnaní so základnou statorovou frekvenciou. Bežne používané PI regulátory vo FOC štruktúre majú obmedzenú šírku pásma, čo spôsobuje fázové oneskorenie a amplitúdový útlm pri kompenzácii vyšších harmonických. Najväčšie problémy s kompenzáciou sa preto objavujú pri stredných a vysokých rýchlostiach, kde je však cogging moment do istej miery tlmený zotrvačnosťou motora [2][3].

Najvýraznejší vplyv má cogging moment najmä pri nízkych rýchlostiach, kde zotrvačnosť už nie je dostatočná na jeho potlačenie. Prejavuje sa kmitaním uhlovej rýchlosti, zvyšovaním

vibrácií a akustického hluku. Tieto sprievodné javy môžu zvyšovať mechanické zaťaženie pripojenej mechaniky a znižovať komfort používateľa, napríklad v prípade domácich spotrebičov. Z tohto dôvodu sa táto práca zameriava práve na možnosti kompenzácie cogging momentu v oblasti nízkych rýchlostí [3][4].

V odbornej literatúre už bolo predstavených viacero pokročilých riešení založených na vysokorýchlostných regulátoroch, modelových pozorovateľoch momentu či neurónových sieťach. Ich nasadenie v nízkonákladových motoroch je však často problematické vzhľadom na obmedzené možnosti riadiaceho hardvéru. Náročné algoritmy totiž vyžadujú vysoký výpočtový výkon a ich realizácia v reálnom čase môže byť v lacnejších zariadeniach neuskutočniteľná [3][5][6].

Z uvedených dôvodov je potrebné hľadať jednoduchšie a menej výpočtovo náročné prístupy, ktoré umožnia efektívnu kompenzáciu cogging momentu bez zásadných zásahov do existujúcej riadiacej architektúry. Cieľom tejto práce je preto navrhnúť a overiť metódy identifikácie a kompenzácie cogging momentu využívajúce dostupné signály v štandardnej FOC štruktúre, ktoré budú zároveň vhodné aj pre nízko nákladové riadiace systémy.

V nízko nákladových aplikáciách je mimoriadne dôležité zohľadniť dostupné možnosti, ktoré sú k dispozícii pre identifikáciu a vytvorenie mapy cogging momentu. V praxi existuje viacero prístupov, ako túto mapu získať, pričom vo všeobecnosti ich možno rozdeliť do dvoch základných kategórií — na priame a nepriame metódy.

Priame získavanie mapy cogging momentu je založené na použití špeciálneho snímača momentu, ktorý umožňuje priamo merať parazitný cogging moment v závislosti od polohy rotora. Hoci táto metóda poskytuje veľmi presné výsledky, jej hlavnou nevýhodou sú vysoké náklady na samotný snímač. Napriek tomu, pri sériovej výrobe veľkého počtu rovnakých motorov by takéto riešenie mohlo byť pre výrobcu ekonomicky prijateľné. V takom prípade by bolo možné dodávať mapu cogging momentu priamo s motorom už od výrobcu, čím by sa používateľom uľahčila jeho kompenzácia v aplikácii [7][8].

Alternatívou k priamemu meraniu je využitie nepriamych metód, ktoré sa opierajú o dostupné snímače v bežných riadiacich systémoch, ako sú snímače polohy a prúdové senzory. V tejto publikácii je použitá práve nepriama metóda založená na vyhodnocovaní momentotvornej zložky prúdu v osi q (q-axis current). Táto zložka prúdu je úmerná elektromagnetickému momentu generovanému motorom, a preto môže byť po vynásobení momentovou konštantou motora použitá ako odhad cogging momentu [9].

Nameraný priebeh momentotvornej zložky prúdu, zodpovedajúci jednej mechanickej otáčke rotora, je následne spracovaný pomocou rýchlej Fourierovej transformácie (FFT). Výsledkom tejto analýzy sú spektrálne zložky reprezentované vektormi amplitúd a fáz najvýraznejších harmonických. Z týchto dominantných harmonických zložiek sa vytvorí referenčný kompenzačný signál, ktorý slúži ako základ pre porovnanie s rôznymi navrhnutými tvarmi kompenzačných signálov, s cieľom nájsť najvhodnejší variant pre praktickú implementáciu.

Štruktúra tejto publikácie je nasledovná. Druhá časť sa venuje podrobnej analýze cogging momentu, popisuje jeho zdroje, prejavy a faktory ovplyvňujúce jeho frekvenciu a intenzitu. Tretia časť je zameraná na proces identifikácie mapy cogging momentu vrátane popisu zvolenej metódy. Nasledujúca, štvrtá časť, predstavuje riadiacu štruktúru Field Oriented Control (FOC), použitý motor a konkrétne nastavenia riadenia. Takisto sú prezentované výsledky simulácií, kde sú jednotlivé tvary kompenzačných signálov porovnávané s referenčným signálom z hľadiska účinnosti kompenzácie a vplyvu na zvlnenie uhlovej rýchlosti motora.

V závere publikácie sú získané výsledky podrobne vyhodnotené, pričom sú identifikované nielen účinné prístupy, ale aj existujúce limity použitých metód. Zároveň sú načrtnuté možné smery ďalšieho výskumu v oblasti nízko nákladovej kompenzácie cogging momentu, najmä so zameraním na ďalšie zlepšenie účinnosti a adaptívnosti použitých algoritmov.

II. COGGING MOMENT

Cogging moment predstavuje parazitný jav vznikajúci v motoroch s permanentnými magnetmi v dôsledku interakcie medzi magnetickým poľom rotorových permanentných magnetov a železnými drážkami statora. Tento moment je obzvlášť výrazný pri nízkych rýchlostiach, kde jeho vplyv nie je dostatočne potlačený zotrvačnosťou rotora. Prejavuje sa predovšetkým kmitaním uhlovej rýchlosti motora, ako aj zvýšenou úrovňou vibrácií a akustického hluku [10].

Z fyzikálneho hľadiska možno cogging moment vyjadriť pomocou nasledujúceho vzťahu [12]:

$$T_{COGG} = -\frac{{\Phi_{PM}}^2}{2}\frac{dR(\vartheta_r)}{d\vartheta_r} \tag{1}$$

kde *R(ϑr)* predstavuje reluktanciu magnetického obvodu v závislosti od uhlovej polohy rotora ϑ_r a Φ_{PM} je magnetický tok generovaný permanentnými magnetmi [12].

Okrem fyzikálnej rovnice možno cogging moment vyjadriť pomocou Fourierovej rady, ktorá umožňuje jeho efektívnu analýzu a syntézu:

$$T_{COGG} = \sum A_{nT} \cos(nLCM\vartheta_r + \phi_n) \tag{2}$$

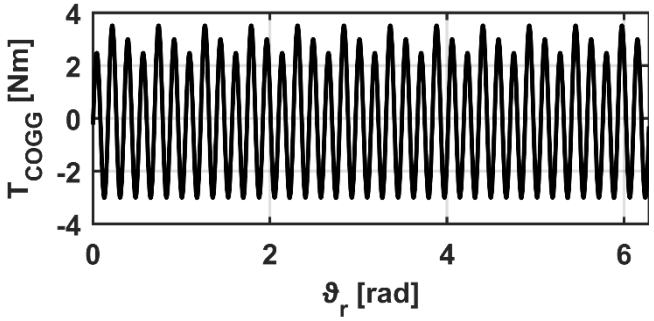
kde A_{nT} predstavuje amplitúdu n-tej harmonickej zložky, n označuje násobok základnej harmonickej frekvencie, *LCM* (najmenší spoločný násobok) vyjadruje najmenší spoločný násobok počtu drážok statora a počtu pólov rotora a ϕ_n je fáza danej n-tej harmonickej zložky. Hodnota LCM tak určuje základnú harmonickú cogging momentu, ktorá je spravidla aj najdominantnejšou zložkou. Na základe Fourierovej analýzy, konkrétne pomocou rýchlej Fourierovej transformácie (FFT), je možné identifikovať dominantné harmonické zložky cogging momentu a na ich základe vytvoriť referenčný kompenzačný signál, ktorý možno následne použiť v riadiacej štruktúre na potlačenie jeho nežiaducich účinkov. Je však potrebné zdôrazniť, že reálny priebeh cogging momentu nemusí obsahovať výlučne vyššie harmonické. V praxi to znamená, že násobok n·LCM môže byť aj menší ako samotné LCM [3].

III. Modelovanie a identifikácia cogging momentu

A. Modelovanie Cogging Momentu

V rámci realizovanej simulácie nebola k dispozícii presná mapa cogging momentu reálneho motora, ktorú by bolo možné priamo implementovať do modelu. Z tohto dôvodu bol parazitný cogging moment umelo generovaný, pričom boli zohľadnené všetky dostupné praktické poznatky o jeho typickom priebehu, amplitúde a frekvenčných vlastnostiach v závislosti od parametrov simulovaného motora.

To znamená, že pri tvorbe modelu sa vychádzalo z predpokladu, že amplitúda cogging momentu vo všeobecnosti dosahuje približne 10 % nominálneho momentu motora, čo je hodnota často uvádzaná v literatúre a zodpovedá reálnym prevádzkovým podmienkam nízko nákladových motorov. Okrem základnej harmonickej zložky cogging moment v praxi často obsahuje aj nižšie harmonické. Preto model cogging momentu je reprezentovaný ako súčet viacerých nižších harmonických, konkrétne polovičnej a tretinovej základnej harmonickej s amplitúdou opäť 10% základnej. Nasimulovaný priebeh cogging momentu je znázornený na obrázku Obr. 1 [11][13]. Takýto priebeh cogging momentu bol privedený ako záťaž na hriadeľ motora.



Obr. 1 Profil cogging momentu v závislosti od polohy

B. Identifikácia Cogging Momentu

Ako už bolo spomenuté mapu cogging momentu je možné získať na základe merania priebehu momentotvornej zložky prúdu v osi q (q-axis current). Prúd sa pri tom nevzorkuje v pravidelných časových intervaloch, ale v závislosti od presnej polohy rotora. V praktických aplikáciách môže byť táto poloha určená napríklad inkrementálnym snímačom, pričom vzorkovanie prebieha na základe uhlového natočenia medzi dvoma po sebe nasledujúcimi inkrementálnymi impulzmi.

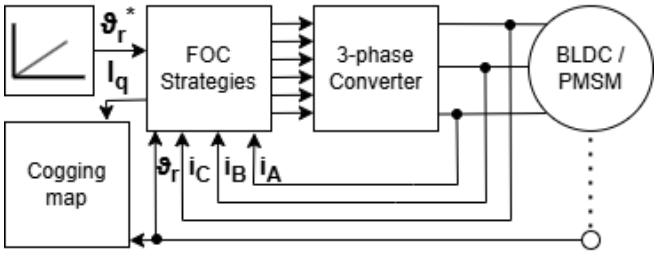
Takto získané údaje, konkrétne matica hodnôt q-prúdu a zodpovedajúcich uhlových polôh rotora, sú následne spracované pomocou FFT analýzy. Výsledkom tejto analýzy sú dva vektory obsahujúce amplitúdy a fázy jednotlivých harmonických zložiek cogging momentu odvodených z momentotvornej zložky prúdu.

Je dôležité zdôrazniť, že identifikácia mapy cogging momentu musí byť realizovaná pri nízkych polohovacích rýchlostiach, čiže pri veľmi pomalej polohovacej rampe s konštantnou uhlovou rýchlosťou. Dôvodom je minimalizácia vplyvu časových oneskorení v regulácii, ktoré

sú spôsobené dynamikou PI regulátorov, ktoré by mohli skresliť meraný priebeh prúdu. Taktiež takéto predpoklady pri identifikácií nám umožňujú zanedbať aj rozdiel medzi referenčnou a skutočnou polohou rotora.

Okrem harmonických zložiek, FFT analýza poskytuje aj jednosmernú (DC) zložku. Táto zložka v danom prípade zodpovedá prevažne trecím stratám v motore a je vhodné ju pri ďalšom spracovaní zohľadniť alebo oddeliť od samotnej mapy cogging momentu.

Obr. 2 zobrazuje blokovú schému polohového FOC riadenia používaného pri identifikácií cogging momentu. Je potrebné takisto poznamenať, že pre správnu identifikáciu je dôležité dodržať dostatočne veľkú šírku pásma prúdových regulátorov (aspoň desať násobok základnej frekvencie cogging momentu) a tiež približne (desatinu zo základnej frekvencie cogging pre polohový regulátor). Tieto predpoklady nám zabezpečia, že budeme schopný zachytiť aj vyššie harmonické cogging momentu a tiež, že polohový regulátor bude dosť pomalý na to aby ovplyvňoval identifikáciu tvaru cogging momentu. Keďže pri identifikácií sa využíva veľmi pomalá rampa polohy, z toho dôvodu bolo možné rýchlostné regulátory zanedbať.



Obr. 2 Bloková schéma identifikácie cogging momentu

Proces identifikácie začína po ustálení uhlovej rýchlosti. Pomocou buffera sa zozbierali údaje o polohe a hodnote q-zložky prúdu. Tieto dáta boli následne použité na vykonanie FFT analýzy s výsledkom uvedeným v Tab. 1. Tabuľka obsahuje potrebné amplitúdy a fázy jednotlivých harmonických z ktorých je zložený cogging moment identifikovaný prostredníctvom q-axis prúdu, vyjadrených pomocou momentovej konštanty k_t. Na základe amplitúd a fáz jednotlivých harmonických je možné s využitím vzťahu (3) získať referenčný kompenzačný signál cogging momentu. Keďže model cogging momentu bol vytvorený vopred skladaním zo známych signálov, porovnaním bola overená aj funkčnosť navrhnutej funkcie FFT analýzy. Treba podotknúť, že výstupom z FFT je aj 1 harmonická, ktorá v Matlabe predstavuje DC-zložku, čiže spomínaný napríklad trecí moment.

Tab. 1 Výsledky FFT analýzy

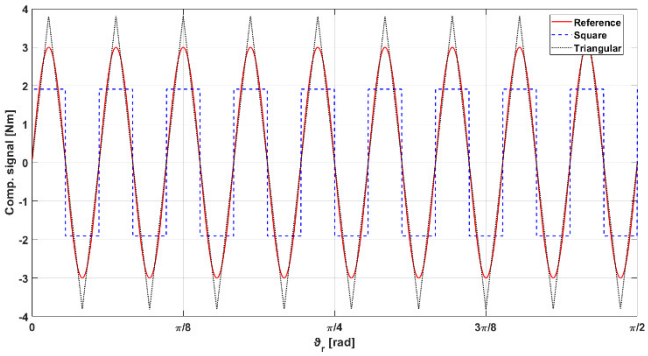
n-harmonická	$\mathbf{A_{nI}}$	$\boldsymbol{\phi_n}$
1(DC-zložka)	$-5/k_t$	-
12	$0.3/k_t$	π
24	$0.3/k_t$	$\pi/2$
36	$3/k_t$	$(3/2)\pi$

C. Modelovanie referenčného kompenzačného signálu

Ako bolo spomenuté referenčný kompenzačný signál vieme získať z výsledkov FFT analýzy uvedených v Tab. 1. Pomocou vzťahu (3) si vypočítame referenčný kompenzačný signál, ktorý sa pripočíta k žiadanej hodnote momentotvornej zložky.

$$I_{qcomp} = \sum A_{nI} \cos(n\vartheta_r + \phi_n) \tag{3}$$

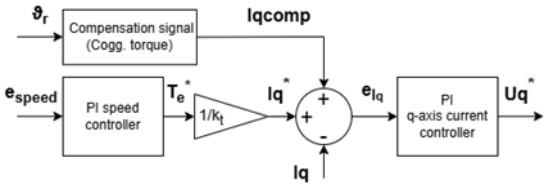
V rámci tejto práce s týmto referenčným kompenzačným signálom, budú porovnávané ďalšie výpočtovo efektívnejšie kompenzačné signály konkrétne štvorcový a trojuholníkový. Podstatou porovnania je, to že plocha ktorú tieto signály majú je dopočítavaná tak aby plochy jednotlivých signálov boli rovnaké, čo znamená aby sa ich kompenzačný účinok rovnal. Priebehy všetkých kompenzačných signálov, je možné vidieť na Obr. *3*. Ako je možné vidieť od referenčného signálu sa trojuholníkový odlišuje oveľa vyššou amplitúdou, ktorá kompenzuje rozdiel ich plochy naopak štvorcový má oveľa nižšiu amplitúdu avšak opäť plocha bola zachovaná.



Obr. 3 Alternatívne kompenzačné signály

IV. SIMULAČNÉ VÝSLEDKY

Na základe identifikovaného cogging momentu pomocou analýzy q-zložky prúdu, prezentovanej v predchádzajúcej kapitole, bol vytvorený referenčný kompenzačný signál *Iqcomp* pozostávajúci zo všetkých identifikovaných harmonických zložiek uvedených v Tab. 1. Tento signál bol následne zavedený za výstup rýchlostného PI regulátora, ktorého výstupom je žiadaný moment *Te**, resp. žiadaná hodnota q-zložky prúdu *Iq**. Schéma privedenia kompenzačného signálu do regulačnej štruktúry je znázornená na Obr. 4. Zavedením kompenzačného signálu, generovaný elektromagnetický moment pôsobí proti cogging momentu, čím dochádza k výraznému potlačeniu momentového zvlnenia. Okrem referenčného kompenzačného signálu bol v simuláciách analyzovaný aj vplyv alternatívnych, zjednodušených kompenzačných signálov uvedených vyššie, s cieľom porovnať ich účinnosť a výpočtovú náročnosť.



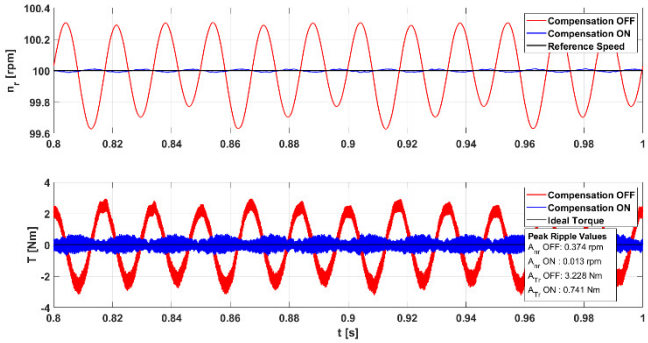
Obr. 4 Bloková schéma kompenzácie cogging momentu

Obr. 5, Obr. 6 a Obr. 7 zobrazujú priebeh rýchlosti a zvlnenia motora pred kompenzáciou (červený priebeh), po aplikácii kompenzácie (modrý priebeh) a referenčný alebo ideálny priebeh, voči ktorému sa vyhodnocuje zvlnenie (čierny priebeh). Kľúčovým parametrom pri vyhodnocovaní účinnosti kompenzácie je amplitúda zvlnenia rýchlosti A_{nr} a amplitúda rozdielu medzi momentom generovaným motorom a cogging momentom pôsobiacim na hriadeľ motora A_{Tr}.

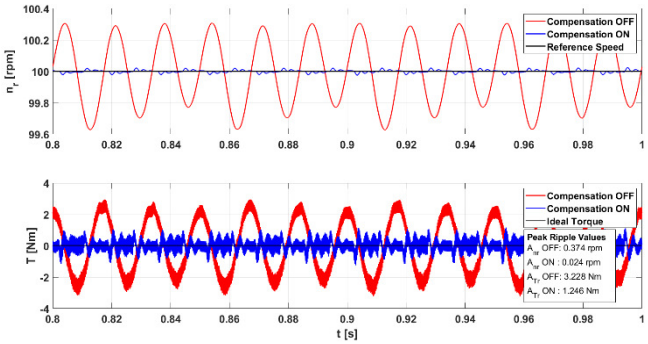
Obr. 4 amplitúdy pred kompenzáciou $A_{nr} = 0.374\ rpm$ $A_{Tr} = 3.228\ Nm$ po kompenzácií $A_{nr} = 0.013\ rpm$ $A_{Tr} = 0.741\ Nm$. Hodnoty amplitúd pred a po kompenzácii sú uvedené v legendách jednotlivých priebehov.

Obr. *5* amplitúdy pred kompenzáciou $A_{nr} = 0.374\ rpm$, $A_{Tr} = 3.228\ Nm$ po kompenzácií $A_{nr} = 0.024\ rpm$, $A_{Tr} = 1.246\ Nm$. Hodnoty amplitúd pred a po kompenzácii sú uvedené v legendách jednotlivých priebehov.

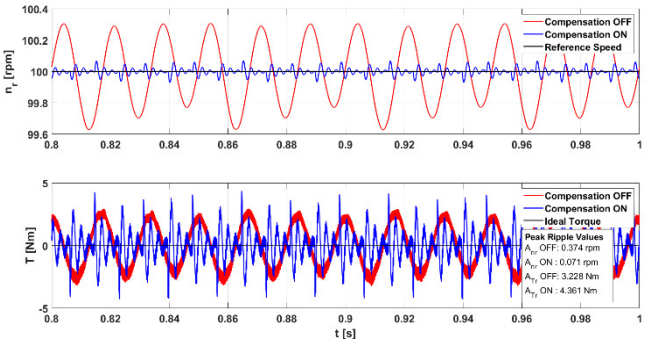
Obr. *6* amplitúdy pred kompenzáciou $A_{nr} = 0.374\ rpm$, $A_{Tr} = 3.228\ Nm$ po kompenzácií $A_{nr} = 0.071\ rpm$, $A_{Tr} = 4.361\ Nm$. Hodnoty amplitúd pred a po kompenzácii sú uvedené v legendách jednotlivých priebehov.



Obr. 4 Priebeh veličín pred a po kompenzácií referenčným signálom (žiadané otáčky n_r = 100 rpm)



Obr. 5 Priebeh veličín pred a po trojuholníkovým signálom (žiadané otáčky n_r = 100 rpm)



Obr. 6 Priebeh veličín pred a po kompenzácií štvorcovým signálom (žiadané otáčky n_r = 100 rpm)

Treba zdôrazniť, že účinnosť kompenzácie je limitovaná šírkou pásma prúdových regulátorov, čo výrazne obmedzuje použitie tejto metódy pri vyšších rýchlostiach. Z tohto dôvodu bola účinnosť kompenzácie analyzovaná pri rôznych hodnotách požadovanej rýchlosti n_r. Výsledky a zodpovedajúce účinky kompenzácie sú prehľadne zhrnuté v Tab. 2, Tab. 3 a Tab. 4.

Tab. 2 Porovnanie simulovaných hodnôt zvlnenia pre rôzne kompenzačné signály pri 50 rpm

	A_{nr} [rpm]	**A_{Tr} [Nm]**	**Redukcia zvlnenia rýchlosti [%]**	**Redukcia zvlnenia momentu [%]**
Bez kompenzácie	0,298	4,277		
Referenčný	0,009	0,532	96,98	87,56
Trojholníkový	0,015	0,718	94,97	83,21
Štvorcový	0,075	4,236	74,83	0,96

Tab. 3 Porovnanie simulovaných hodnôt zvlnenia pre rôzne kompenzačné signály pri 100 rpm

	A_{nr} [rpm]	**A_{Tr} [Nm]**	**Redukcia zvlnenia rýchlosti [%]**	**Redukcia zvlnenia momentu [%]**
Bez kompenzácie	0,374	4,228		
Referenčný	0,013	0,741	96,52	82,47
Trojholníkový	0,024	1,246	93,58	70,53
Štvorcový	0,071	4,361	81,01	-3,14

Tab. 4 Porovnanie simulovaných hodnôt zvlnenia pre rôzne kompenzačné signály pri 150 rpm

	A_{nr} [rpm]	**A_{Tr} [Nm]**	**Redukcia zvlnenia rýchlosti [%]**	**Redukcia zvlnenia momentu [%]**
Bez kompenzácie	0,285	3,778		
Referenčný	0,022	1,071	92,28	71,69
Trojholníkový	0,032	1,686	88,77	55,40
Štvorcový	0,067	4,257	76,49	-12,68

V Tab. 5 sú uvedené všetky kľúčové elektrické a mechanické parametre simulovaného motora, ako napríklad počet pólov, indukčnosť, odpor vinutí, moment zotrvačnosti. Tieto údaje zabezpečujú realistickú simuláciu správania motora a spoľahlivé vyhodnotenie účinnosti navrhnutej kompenzačnej metódy.

Tab. 5 Kľúčové parametre simulovaného PMSM motora

	Symbol	Value
Nominálne napätie [V]	U_N	48
Nominálny prúd [A]	I_N	98,8
Nominálny moment [Nm]	T_N	40,5
Počet pólov [-]	2p	12
Počet statorových drážok [-]	Q	18
Fazový odpor [mΩ]	R_S	13
Fázová indukčnsoť [mH]	L_S	0,2
Moment zotrvačnosti [kg.m^2]	J	0,02
Momentová konštanta [Nm/A]	k_t	0,27

V. ZÁVER

Táto práca sa venovala identifikácii a kompenzácii cogging momentu v BLDC a PMSM motoroch, s dôrazom na nízkonákladové aplikácie. Najvhodnejšou metódou sa ukázalo využitie údajov z bežných prúdových snímačov, ktoré sú štandardnou súčasťou pohonov. Priebeh cogging momentu bol úspešne identifikovaný na základe analýzy q-zložky prúdu za stanovených podmienok.

Identifikovaný signál obsahoval aj vyššie harmonické zložky, ktoré však nie je možné efektívne kompenzovať kvôli obmedzenej šírke pásma prúdových regulátorov. Preto bola použitá FFT analýza, pomocou ktorej boli vybrané len najdominantnejšie harmonické zložky tak, aby frekvencia kompenzačného signálu ostala v rozsahu, ktorý je regulátor schopný sledovať. Táto podmienka prirodzene obmedzuje použitie navrhovanej metódy na oblasť nízkych rýchlostí, kde sa cogging moment prejavuje najvýraznejšie.

Vzhľadom na to, že cieľové aplikácie často využívajú výpočtovo úsporné mikrokontroléry, boli preskúmané aj alternatívne signály s nižšou výpočtovou náročnosťou. V rámci simulácií boli porovnané účinky kompenzácie pomocou referenčného signálu, trojuholníkového priebehu a pravouhlého (štvorcového) signálu. Účinnosť sa posudzovala na základe amplitúdy zvlnenia uhlovej rýchlosti a rozdielu medzi generovaným elektromagnetickým momentom a cogging momentom.

Simulačné výsledky ukázali, že všetky tri signály významne znižujú zvlnenie uhlovej rýchlosti. Referenčný signál síce poskytol najlepšie výsledky, no trojuholníkový priebeh dosiahol veľmi podobnú účinnosť pri výrazne nižšej výpočtovej záťaži. Najmenej účinný bol pravouhlý signál, najmä z pohľadu zvlnenia momentu, aj keď aj ten významne znížil zvlnenie rýchlosti.

Správnosť navrhovanej metódy bola ďalej overená simuláciami pri rôznych otáčkach rotora. Potvrdilo sa, že účinnosť kompenzácie so zvyšujúcou rýchlosťou klesá, čo súvisí s pásmom priepustnosti prúdového regulátora. V prípade pravouhlého signálu sa dokonca pri vyšších otáčkach zhoršilo momentové zvlnenie, hoci zvlnenie rýchlosti ostalo znížené.

Budúci výskum by sa mohol zamerať na optimalizáciu tvaru trojuholníkového signálu, ako aj na implementáciu pokročilých metód kompenzácie aplikovaných až za prúdové regulátory, čo by mohlo umožniť efektívnu kompenzáciu cogging momentu aj pri stredných a vyšších rýchlostiach motora.

POĎAKOVANIE

Táto práca bola podporená Vedeckou grantovou agentúrou Ministerstva školstva Slovenskej republiky v rámci projektu KEGA 032TUKE-4/2024.

REFERENCIE

[1] SORESINI, Federico, et al. Noise and Vibration Modeling of Permanent Magnet Synchronous Motors: A Review. IEEE Transactions on Transportation Electrification, 2024, 10.4: 8728-8745.

[2] N. Bianchi and S. Bolognani, "Design techniques for reducing the cogging torque in surface-mounted PM motors," in IEEE Transactions on Industry Applications, vol. 38, no. 5, pp. 1259-1265, Sept.-Oct. 2002, doi: 10.1109/TIA.2002.802989.

[3] C. Xia, B. Ji and Y. Yan, "Smooth Speed Control for Low-Speed High-Torque Permanent-Magnet Synchronous Motor Using Proportional–Integral–Resonant Controller," in IEEE Transactions on Industrial Electronics, vol. 62, no. 4, pp. 2123-2134, April 2015, doi: 10.1109/TIE.2014.2354593.

[4] KRISHNAN, Ramu. Permanent magnet synchronous and brushless DC motor drives. CRC press, 2017.

[5] J. Liu, H. Li and Y. Deng, "Torque Ripple Minimization of PMSM Based on Robust ILC Via Adaptive Sliding Mode Control," in IEEE Transactions on Power Electronics, vol. 33, no. 4, pp. 3655-3671, April 2018, doi: 10.1109/TPEL.2017.2711098.

[6] A. Mora, Á. Orellana, J. Juliet and R. Cárdenas, "Model Predictive Torque Control for Torque Ripple Compensation in Variable-Speed PMSMs," in IEEE Transactions on Industrial Electronics, vol. 63, no. 7, pp. 4584-4592, July 2016, doi: 10.1109/TIE.2016.2536586.

[7] Z. Q. Zhu, "A simple method for measuring cogging torque in permanent magnet machines," 2009 IEEE Power & Energy Society General Meeting, Calgary, AB, Canada, 2009, pp. 1-4, doi: 10.1109/PES.2009.5275665.

[8] PICCOLI, Matthew; YIM, Mark. Anticogging: Torque ripple suppression, modeling, and parameter selection. The international journal of robotics research, 2016, 35.1-3: 148-160.

[9] J. Holtz and L. Springob, "Identification and compensation of torque ripple in high-precision permanent magnet motor drives," in IEEE Transactions on Industrial Electronics, vol. 43, no. 2, pp. 309-320, April 1996, doi: 10.1109/41.491355.
[10] Wei Hua and Ming Cheng, "Cogging torque reduction of flux-switching permanent magnet machines without skewing," 2008 International Conference on Electrical Machines and Systems, Wuhan, China, 2008, pp. 3020-3025.
[11] HAO, Wenjuan, et al. Methods for reducing cogging force in permanent magnet machines: A review. Energies, 2022, 16.1: 422.
[12] L. Dosiek and P. Pillay, "Cogging Torque Reduction in Permanent Magnet Machines," in IEEE Transactions on Industry Applications, vol. 43, no. 6, pp. 1565-1571, Nov.-dec. 2007, doi: 10.1109/TIA.2007.908160.
[13] N. Bianchi and S. Bolognani, "Design techniques for reducing the cogging torque in surface-mounted PM motors," in IEEE Transactions on Industry Applications, vol. 38, no. 5, pp. 1259-1265, Sept.-Oct. 2002, doi: 10.1109/TIA.2002.802989

Návrh a realizácia firemnej metalickej počítačovej siete

[1]*Ľuboš OVSENÍK,* [2]*Jakub ORAVEC,* [3]*Zuzana LIŠČINSKÁ,* [4]*Dávid ANDRIŠOV*

[1,2,3,4]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]zuzana.liscinska@tuke.sk, [4]david.andrisov@student.tuke.sk

***Abstrakt* — Príspevok sa zaoberá návrhom a realizáciou štruktúrovanej metalickej kabeláže pre potreby menšej firmy. Jej hlavným cieľom bolo na základe analýzy požiadaviek zákazníka vytvoriť návrh siete, ktorá by spĺňala aktuálne štandardy prenosových rýchlostí a zároveň zabezpečovala spoľahlivú komunikáciu v rámci firmy. Funkčnosť vybudovanej siete bola overená testovaním fyzických spojení a simuláciou prevádzky v modelových podmienkach.**

***Kľúčové slová* — metalická sieť, normy pre metalickú kabeláž, počítačová sieť, topológia sietí**

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 menšiu spoločnosť. Hlavným cieľom je vytvoriť sieťovú infraštruktúru, ktorá nielen splní súčasné technické požiadavky, ale zároveň bude pripravená aj na prípadné budúce rozšírenie [1-4].

II. Návrh metalickej kabeláže pre menšiu firmu

Návrh metalickej kabeláže pre danú firmu je dôležitým krokom pri budovaní spoľahlivej a zároveň flexibilnej sieťovej infraštruktúry. V tejto kapitole je zhrnutý postup a hlavné zásady, ktoré boli pri návrhu štruktúrovanej kabeláže zohľadnené - s dôrazom na čo najlepší výkon siete, obmedzenie rušenia a možnosť jednoduchého rozšírenia v budúcnosti.

Na schéme (Obr. 1) je znázornené priestorové rozmiestnenie sieťových prvkov, vrátane dvojzásuviek na ethernet, Wi-Fi prístupových bodov a technickej miestnosti. Návrh vychádza z analýzy potrieb firmy a je optimalizovaný pre zabezpečenie stabilného a vysoko rýchlostného pripojenia.

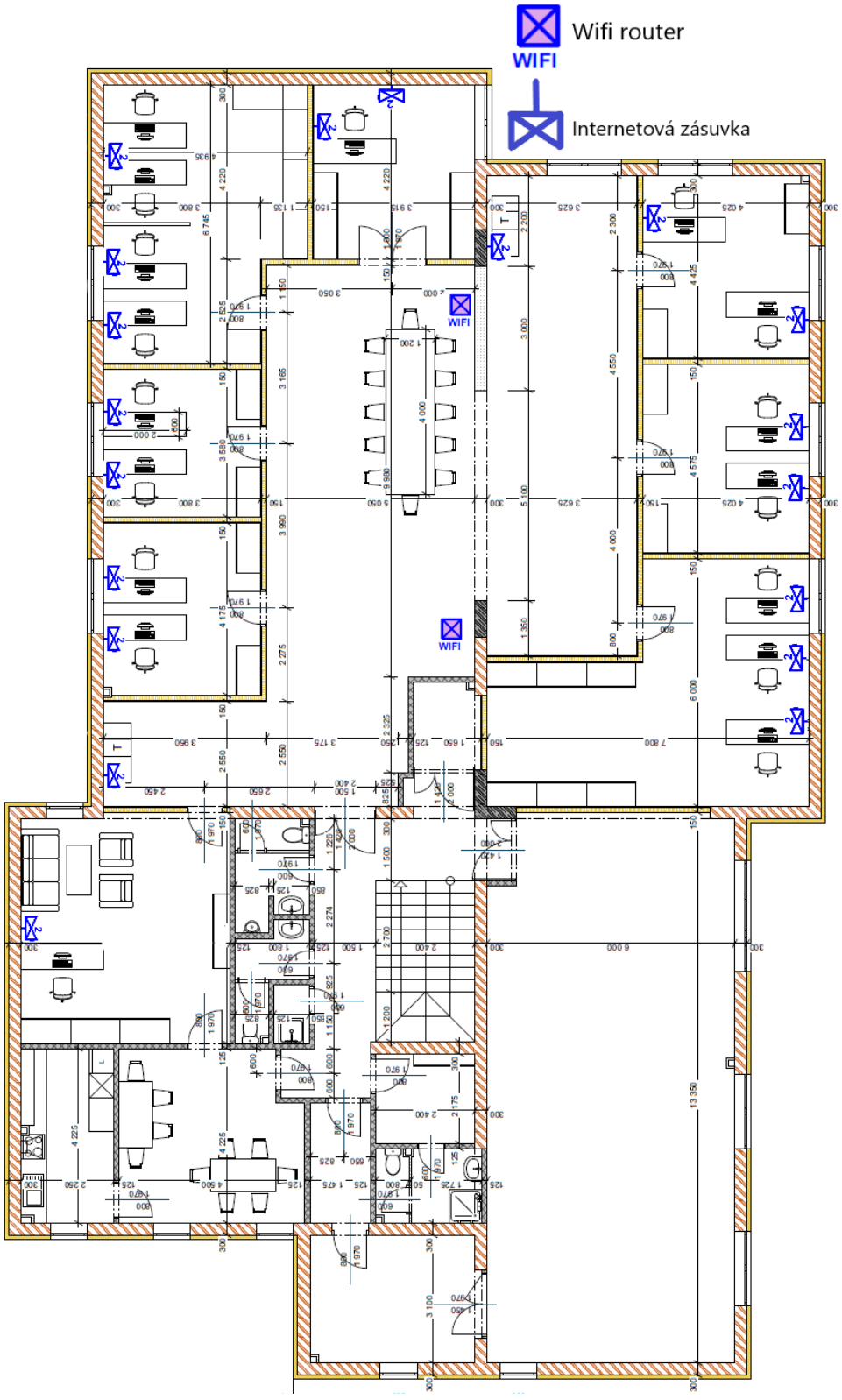
A. Analýza potrieb siete

Analýza potrieb siete je prvým krokom pri návrhu metalickej kabeláže a zahŕňa identifikáciu požiadaviek, ktoré má sieť splniť. Táto analýza sa zameriava na súčasné aj budúce potreby firmy, pričom berú do úvahy aj faktory, ako sú:

A1) Počet pracovných miest: Na základe dispozičného riešenia bolo identifikovaných 18 pracovných miest, ktoré si vyžadujú pripojenie na internet prostredníctvom pevných dvojzásuviek. Každé pracovné miesto je vybavené minimálne jednou internetovou zásuvkou, pričom v prípade potreby je možné zapojiť ďalšie zariadenia.

A2) Požiadavky na Wi-Fi pokrytie: S ohľadom na potreby mobility zamestnancov a návštevníkov sú navrhnuté dva Wi-Fi prístupové body. Tieto body sú strategicky umiestnené tak, aby zabezpečili rovnomerné pokrytie celého poschodia, vrátane kancelárskych priestorov a haly.

A3) Prenosové rýchlosti a šírka pásma: Siete musia podporovať aplikácie náročné na dáta, ako sú videokonferencie, cloudové služby a zdieľanie veľkých súborov. Na základe týchto potrieb boli vybrané káble kategórie Cat6a, ktoré poskytujú dostatočnú prenosovú kapacitu až do 10 Gbps a šírku pásma 500 MHz.



Obr. 1 Schéma budovy s rozmiestnením jednotlivých koncoviek a zariadení pre danú firmu.

A4) Typy pripojených zariadení: Sieťová infraštruktúra musí podporovať stolové počítače, notebooky, IP telefóny, tlačiarne a ďalšie zariadenia. V rámci kancelárskeho priestoru sú naplánované pevné pripojenia na internet pre zariadenia s vysokými nárokmi na stabilitu pripojenia, zatiaľ čo mobilné zariadenia budú využívať Wi-Fi.
A5) Rozšíriteľnosť a flexibilita: Návrh siete umožňuje jednoduché rozšírenie o ďalšie pracovné miesta alebo zariadenia. Pri projektovaní bol zohľadnený dostatočný počet rezervných portov v rozvádzačoch a flexibilita pri dodatočnej inštalácii káblov.

A6) Zabezpečenie siete: Sieť je navrhnutá s dôrazom na bezpečnosť, čo zahŕňa oddelenie interných a hosťovských sietí prostredníctvom VLAN, ako aj zabezpečené pripojenie Wi-Fi.

B. Požiadavky na štruktúrovanú kabeláž

Požiadavky na štruktúrovanú kabeláž určujú základné parametre a štandardy, ktoré musí sieťová infraštruktúra spĺňať. Tieto požiadavky vychádzajú z normatívnych predpisov, technologických štandardov a potrieb firmy. Pri návrhu bola kabeláž plánovaná tak, aby umožňovala vysokú stabilitu prenosu dát, škálovateľnosť a kompatibilitu s modernými technológiami:

B1) Štandardy kabeláže: Systém štruktúrovanej kabeláže je navrhnutý podľa normy TIA/EIA-568, EN 50173 a STN EN 50173 (Slovak Technical Norm) [5-10]. Použité budú káble kategórie Cat6a, ktoré spĺňajú požiadavky na šírku pásma až 500 MHz a prenosové rýchlosti až do 10 Gbps. Tieto káble poskytujú dostatočnú kapacitu na podporu dnešných aj budúcich aplikácií.
B2) Topológia siete: Hviezdicová topológia je použitá na pripojenie všetkých pracovných miest a zariadení k centrálnemu rozvádzaču (racku). Táto topológia umožňuje jednoduchú správu a údržbu siete.
B3) Dĺžka káblov: Maximálna dĺžka horizontálnych káblov (medzi dátovým rozvádzačom a koncovým zariadením) je obmedzená na 100 metrov, vrátane patch káblov. Tento limit je stanovený normami pre zabezpečenie bezstratového prenosu dát.
B4) Porty a zásuvky: Každé pracovné miesto je vybavené minimálne jednou dvojzásuvkou pre pevné internetové pripojenie. Celkovo je plánovaných 24 internetových dvojzásuviek. Zásuvky sú navrhnuté s univerzálnymi konektormi RJ-45, kompatibilnými so štandardmi Cat6a a vyššie.
B5) Napájanie cez Ethernet (Power over Ethernet, PoE): Štruktúrovaná kabeláž umožňuje napájanie zariadení, ako sú Wi-Fi prístupové body alebo IP telefóny, cez Ethernet. Tým sa minimalizuje potreba dodatočného napájania a zjednodušuje inštalácia.
B6) Bezpečnosť a odolnosť voči rušeniu: Kabeláž je navrhnutá s dôrazom na minimalizáciu EMI (Electromagnetic Interference). Použitie tienených káblov (F/UTP) zabezpečuje stabilný signál aj v prostrediach s vyššou úrovňou rušenia. Káblové trasy sú vedené oddelene od elektrických vedení, čo eliminuje presluch a rušenie.
B7) Wi-Fi prístupové body: Dva strategicky umiestnené Wi-Fi prístupové body poskytujú stabilné bezdrôtové pripojenie v celej budove. Pre každý Wi-Fi bod je pripravený dedikovaný kábel s PoE.
B8) Budúca škálovateľnosť: Kabeláž je dimenzovaná tak, aby umožňovala pripojenie ďalších zariadení alebo pracovných miest bez nutnosti zásadných zmien infraštruktúry. V centrálnom rozvádzači je rezervovaný priestor na rozšírenie patch panelov.
B9) Označovanie kabeláže: Všetky káble a zásuvky budú dôsledne označené podľa normy TIA/EIA-606, čo zjednoduší identifikáciu a údržbu kabelážneho systému.

III. Realizácia metalickej kabeláže pre menšiu firmu

Fyzická realizácia metalickej kabeláže bola vykonaná po dôkladnej projektovej príprave a výbere komponentov podľa návrhu štruktúrovanej siete. Celý proces zahŕňal inštaláciu káblových trás, zakončenie káblov do zásuviek, ich zapojenie do patch panelov v dátovom rozvádzači a následné testovanie pomocou certifikačného prístroja. Použitý bol kábel kategórie Cat6a F/UTP, ktorý spĺňa štandardy pre vysokorýchlostný prenos dát a zároveň zabezpečuje odolnosť voči rušeniu.

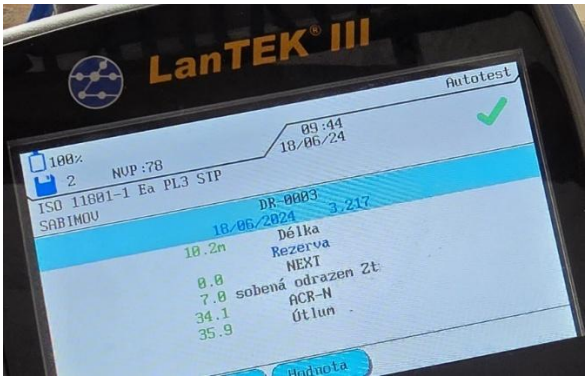
A. Inštalácia a zakončenie zásuviek

Na jednotlivé koncové body boli osadené dvojzásuvky s označením portov (napr. 9A, 9B, 10A, 10B), ktoré sú napojené na centrálne rozvádzače prostredníctvom horizontálnych trás. Každý kábel bol ručne ukončený v Keystone jack moduloch s dodržaním farebného značenia podľa normy TIA-568B. Zakončenie prebehlo pomocou nástroja pre narazenie vodičov do IDC kontaktov, pričom bola zachovaná celistvosť tienenia a uzemnenia (Obr. 2).

Obr. 2 Koncovka pre zásuvku.

B. Testovanie inštalácie

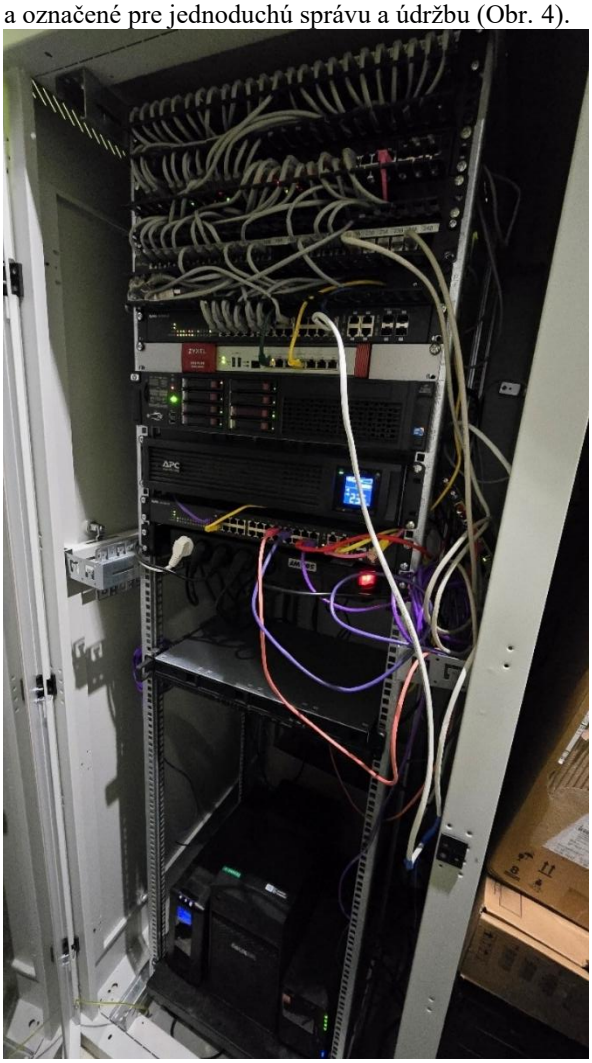
Po osadení všetkých komponentov boli jednotlivé trasy otestované pomocou LanTEK III certifikačného testera. Merania zahŕňali dĺžku vedenia, útlm, odrazenú stratu, presluch (NEXT), ACR-N a ďalšie šumové parametre. Výsledky všetkých testovaných liniek dosiahli hodnotenie PASS, čo potvrdzuje bezchybnú realizáciu a súlad s normou ISO/IEC 11801 a STN EN 50173-1 (Obr. 3).



Obr. 3 Merací pristroj LanTEK III.

C. Dátový rozvádzač

Všetky horizontálne vedenia boli ukončené v rozvádzači, kde boli osadené patch panely, prepínače, servery, záložné napájacie zdroje a ďalšie. V rozvádzači je tiež zabezpečená organizácia patch káblov, ako aj redundantné napájanie serverovej infraštruktúry. Zariadenia sú logicky usporiadané a označené pre jednoduchú správu a údržbu (Obr. 4).



Obr. 4 Dátový rozvádzač.

D. Zabezpečenie

Súčasťou návrhu siete bola aj implementácia základných bezpečnostných mechanizmov, ktoré zabezpečujú ochranu siete na úrovni dátovej vrstvy, prevenciu proti útokom a stabilitu infraštruktúry. Na prístupových portoch bola aktivovaná funkcia *Port Security* s nastaveným limitom na maximálne 1 až 2 MAC adresy, ktoré sa dynamicky ukladajú do pamäte zariadenia (známe ako *sticky* MAC adresy). V prípade porušenia pravidiel je aktivovaná akcia typu *restrict*, ktorá zabraňuje narušeniu prevádzky a zároveň zaznamenáva udalosť (Obr. 5).

```
interface FastEthernet0/X
 switchport mode access
 switchport port-security
 switchport port-security maximum 1
 switchport port-security mac-address sticky
 switchport port-security violation restrict
```

Obr. 5 Zabezpečenie portu (*Port security*).

Na ochranu proti podvrhnutým DHCP serverom bolo povolené *DHCP Snooping* pre vybrané VLANy. Porty typu *Trunk* boli explicitne označené ako *trusted,* zatiaľ čo prístupové porty na okrajových zariadeniach ostali *untrusted*, čím sa zamedzuje nežiaducemu prideľovaniu IP adries.

V nadväznosti na túto funkciu bol aktivovaný aj mechanizmus *DAI* (Dynamic ARP Inspection), ktorý zabezpečuje validáciu ARP paketov a overuje zhodu medzi MAC a IP (Internet Protocol) adresou na základe záznamov z *DHCP snoopingu*. Táto ochrana eliminuje možnosť *ARP spoofingu* v rámci siete.

Na zvýšenie stability a prevencie pred nežiaducimi slučkami boli na všetkých prístupových portoch aktivované *BPDU Guard* mechanizmy, ktoré okamžite deaktivujú port pri detekcii *BPDU rámcov*. Táto ochrana je v kombinácii s funkciou *portfast*, čím sa zabezpečí rýchlejšia inicializácia portov bez ohrozenia *STP* (Spanning Tree Protocol) topológie.

Pre vzdialenú správu bola povolená výhradne komunikácia cez *SSH* protokol, čím sa odstránila možnosť nezabezpečeného prístupu cez *Telnet*. Prístup bol obmedzený iba na špecifické IP adresy pomocou *ACL* (Access Control List), ktoré zároveň slúžili aj na filtrovanie smerovania medzi podsieťami. Všetky nevhodne definované *ACL* pravidlá boli počas návrhu odstránené z dôvodu optimalizácie a funkčnosti prevádzky.

Pre prístup do externých sietí bol nakonfigurovaný *NAT* (Network Address Translation) s využitím *ip nat inside* a *ip nat outside*, vrátane mechanizmu *overload*, ktorý umožňuje viacerým interným zariadeniam zdieľať jednu verejnú IP adresu. Z hľadiska monitoringu a správy boli zavedené centrálne služby ako *Syslog*, kde boli všetky sieťové udalosti smerované na centrálny logovací server s IP adresou **192.168.40.13**. Zároveň bol nakonfigurovaný TFTP server pre prijímanie a zálohovanie konfiguračných súborov zariadení a FTP server so zabezpečeným prístupom pre používateľov podľa definovaných práv. V rámci siete bolo zabezpečené aj interné DNS rozlíšenie pre zariadenia ako napríklad *server.local*, ktoré zjednodušuje správu siete. Pre zabezpečenie dynamického prideľovania adries bola na smerovači implementovaná funkcia *DHCP Relay* (Helper Address) smerujúca na vzdialený DHCP server. Všetky prepínače, ktoré nevyužívajú smerovanie na tretej vrstve, majú korektne nastavenú predvolenú bránu *gateway*, čo zaručuje správnu konektivitu do ostatných VLAN segmentov siete .

Tento súbor opatrení zabezpečuje vysokú úroveň ochrany a spoľahlivosti siete v súlade s odporúčaniami pre firemné infraštruktúry.

E. Výpočet nákladov na inštaláciu

Pri návrhu a realizácii metalickej siete je nevyhnutné zohľadniť nielen technické a prevádzkové aspekty, ale aj ekonomickú stránku celého riešenia. Výpočet nákladov predstavuje kľúčový krok, ktorý umožňuje presne stanoviť finančnú náročnosť projektu a pripraviť rozpočet na jeho realizáciu. Správna kalkulácia nákladov zabezpečuje nielen efektívne plánovanie zdrojov, ale zároveň minimalizuje riziko nepredvídaných výdavkov počas implementácie. Tieto údaje tvoria neoddeliteľnú súčasť projektovej dokumentácie, ktorá slúži ako podklad pre schválenie investície zo strany vedenia firmy alebo zadávateľa. Celkové náklady na inštaláciu siete boli rozdelené do dvoch hlavných kategórií: materiálové náklady a pracovné náklady. Výsledkom tohto výpočtu je komplexný rozpočet, ktorý zahŕňa všetky kľúčové položky spojené s realizáciou siete - od fyzického vybavenia až po odbornú inštaláciu a konfiguráciu zariadení.

Materiálová časť zahŕňa všetky fyzické komponenty potrebné na vybudovanie sieťovej infraštruktúry. To zahŕňa dátové rozvádzače (racky), aktívne prvky ako sú prepínače, ako aj pasívne prvky - káble, konektory, patch panely, organizéry kabeláže a doplnkové príslušenstvo. Použité komponenty zodpovedajú aktuálnym normám pre štruktúrovanú kabeláž, konkrétne pre kategóriu Cat6A s tienenou F/UTP štruktúrou, ktorá zabezpečuje spoľahlivý prenos dát a vysokú odolnosť voči rušeniu. Celkové materiálové náklady boli vyčíslené na 3912,92 € bez DPH.

Náklady na odbornú prácu zahŕňajú inštaláciu aktívnych aj pasívnych komponentov, konfiguráciu zariadení, prípravu kabeláže, testovanie spojov, zálohovanie konfigurácií a prípadné úpravy počas inštalácie. Každá z týchto činností si vyžaduje prítomnosť kvalifikovaných technikov s adekvátnymi skúsenosťami v oblasti sieťových inštalácií, ako aj dodržiavanie presne definovaných metodík. Všetky inštalačné úkony boli realizované podľa odporúčaní a metodiky vyplývajúcej zo štandardov STN EN 50173 a TIA/EIA-568. Súčasťou prác bolo aj správne zapojenie serverov, prepínačov, NAS úložiska, zabezpečovacích prvkov a elektrického napájania. Náročnejšie úkony, ako napríklad montáž patch panelov, konfigurácia portov či zapojenie prívodov 230 V, boli realizované s dôrazom na bezpečnosť a funkčnosť. Celkové náklady na inštaláciu siete boli vyčíslené na 1384,86 € bez DPH.

IV. Zhodnotenie návrhu siete

Navrhnutá sieťová infraštruktúra bola koncipovaná s dôrazom na škálovateľnosť, oddelenie prevádzky pomocou VLAN a zabezpečenie rýchlej a spoľahlivej komunikácie medzi všetkými segmentmi siete. Pri návrhu topológie sa využila kombinácia prístupových a distribučných prepínačov, ktoré sú vzájomne prepojené prostredníctvom *trunk* liniek s podporou viacerých VLAN.

Sieť je rozdelená do viacerých segmentov podľa funkcie a typu pripojených zariadení. Prístupové porty boli nakonfigurované v režime *access* a priradené k základnej VLAN pre používateľské zariadenia, ako sú pracovné stanice a tlačiarne. Medzi prepínačmi boli vytvorené *trunk* spojenia, ktoré zabezpečujú prenos viacerých VLAN, čím sa umožňuje segmentácia siete a efektívna správa prevádzky. Jeden z hlavných prepínačov je centrálne prepojený s vyššou úrovňou infraštruktúry prostredníctvom gigabitových *trunk* portov, ktoré prenášajú prevádzku VLAN 1, 20, 40 a 50. Tieto VLANy reprezentujú rôzne funkčné oblasti siete, ako napríklad kancelárska prevádzka, servery, bezdrôtové prístupové body, zálohovacie systémy a iné. Centrálny smerovač poskytuje funkciu *DHCP* servera pre automatickú konfiguráciu IP adries a zároveň slúži ako brána do iných sietí.

Každý prepínač bol nakonfigurovaný tak, aby zabezpečil správnu funkčnosť portov, pričom všetky aktívne porty sú v stave *UP*. V prípade nepoužívaných portov boli tieto ponechané ako neaktívne alebo deaktivované pre zvýšenie bezpečnosti. Z pohľadu návrhu ide o stabilnú a efektívnu topológiu, ktorá umožňuje jednoduché rozšírenie siete, pridanie ďalších VLAN alebo zariadení bez potreby zásadných zmien v architektúre. Vďaka *trunkovým* prepojeniam a centralizovanému smerovaniu je zabezpečená dobrá správa *broadcast* domén, ako aj možnosť centralizovaného dohľadu nad prevádzkou a bezpečnosťou.

Po vykonaní softvérovej simulácie siete v nástroji Cisco Packet Tracer a fyzického merania v reálnom prostredí bolo vykonané komplexné vyhodnotenie navrhnutej metalickej infraštruktúry ako je znázornené v Tabuľke 1. Cieľom porovnania bolo posúdenie, do akej miery sa návrh siete osvedčil z hľadiska funkčnosti, výkonu, spoľahlivosti a súladu s platnými technickými normami. V rámci simulácie boli testované základné sieťové parametre, ako sú latencia, priepustnosť a stabilita spojenia. Fyzické meranie následne overilo kvalitu metalickej kabeláže, a to pomocou certifikačných testov podľa normy STN EN 50173-1. Okrem toho boli realizované praktické sieťové testy (*ping, iPerf3, traceroute*), ktoré poskytli realistický obraz o správaní sa siete v reálnej prevádzke. Simulované aj reálne testy potvrdili veľmi dobrú výkonnosť a stabilitu navrhnutej a realizovanej metalickej siete. Latencia v oboch prípadoch dosahovala hodnoty medzi 1-3 ms, čo zodpovedá bežnej odozve v moderných LAN sieťach. Priepustnosť bola v reálnych podmienkach výrazne vyššia, a to na úrovni 925 – 937 Mbps, čo potvrdzuje efektívne využitie 1 Gbps linky bez výrazných strát alebo retransmisií. Z fyzického merania vyplýva, že všetky káblové trasy boli realizované kvalitne, bez chýb v tienení alebo ukončení. Všetky testované parametre ako útlm, presluch (NEXT), odrazená strata či šumové pomery (ACR) dosahovali hodnoty v súlade s požiadavkami normy. Prítomnosť rezervy a použitie Cat6a F/UTP kabeláže, zaručuje dostatočný priestor pre budúce rozšírenia a stabilitu prevádzky aj pri vyššom zaťažení siete.

Tabuľka 1
Porovnanie vybraných parametrov zo simulácie a reálneho testovania siete

PARAMETER	SIMULÁCIA (CISCO PACKET TRACER)	REÁLNE MERANIE	HODNOTENIE
Latencia	1–3 ms	1–2 ms	Zhodné / výborné
Priepustnosť	45 Mbps (FTP simulácia)	925–937 Mbps (iPerf3)	Reálna prevádzka výkonnejšia
Stabilita spojenia	1 % strata pri 100 pingoch	0 % (10 pingov), 1 % pri 100 pingoch	Porovnateľná, stabilná
Dĺžka vedenia	Simulovaná topológia	7,9 m – 31,1 m	Overené v reálnom prostredí
Útlm a presluchy (NEXT, ACR)	Nemerané	Všetky hodnoty v norme	Spĺňa požiadavky normy

Z porovnania simulácie a reálnych testov vyplýva, že navrhnutá a implementovaná sieť nielenže spĺňa technické požiadavky a normy, ale zároveň poskytuje vcelku dobré výsledky v oblasti výkonnosti a stability. Simulácia sa ukázala ako užitočný nástroj na predbežné overenie návrhu, no až fyzické merania potvrdili správnu inštaláciu kabeláže a funkčnosť siete v praxi. Táto analýza poskytuje pevný základ pre ďalšie rozšírenie siete a zvyšovanie jej kapacity.

V. 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. Sieť bola vypracovaná na základe podrobného rozboru požiadaviek konkrétnej firmy, s dôrazom na predpokladané zaťaženie siete, vnútornú organizáciu podniku a možnosti budúceho rozširovania. Praktická realizácia potvrdila, že premyslený návrh, precízna montáž a odborné ukončenie sieťových zásuviek významne ovplyvňujú stabilitu a výkon celej siete. Testovanie v programe Cisco Packet Tracer umožnilo overiť funkčnosť navrhovanej topológie ešte pred samotnou inštaláciou. Výsledky testovania potvrdili správne nastavenie a stabilné pripojenie s minimálnym oneskorením. Merania vykonané po dokončení fyzickej inštalácie siete preukázali, že všetky prípojné miesta vyhovujú požiadavkám normy STN EN 50173-1. Namerané hodnoty vykazovali nízky útlm, dobré hodnoty presluchu a vysoký odstup signálu od šumu, čo svedčí o kvalitnej inštalácii a použití vhodných komponentov. Záťažové testy potvrdili vysokú výkonnosť siete, oneskorenie sa pohybovalo v rozmedzí 1 až 2 milisekundy a prenosová rýchlosť dosahovala hodnoty nad 900 Mbps. Tieto výsledky potvrdzujú, že vybudovaná sieť zvláda bez problémov bežnú aj náročnejšiu firemnú prevádzku. Zároveň je zabezpečené oddelenie sieťovej prevádzky pomocou VLAN, stabilita prepojení medzi zariadeniami a bezpečnostné prvky ako STP, port security a ACL. Vytvorená sieť spĺňa súčasné potreby firmy a zároveň poskytuje priestor pre prípadné budúce rozšírenie. Medzi návrhy na ďalšie zlepšenie do budúcnosti možno zaradiť napríklad zavedenie optického pripojenia, rozšírenie možností monitorovania siete a vytvorenie dostatočných rezerv v kabeláži, dátovom rozvádzači aj sieťových komponentoch, čo firme uľahčí prípadné rozšírenie infraštruktúry. Z pohľadu bezpečnosti by bolo vhodné zvážiť zavedenie pokročilejších riešení, ako sú IDS/IPS systémy alebo overovanie prístupu pomocou štandardu 802.1X. Na základe vykonaných meraní a testov možno povedať, že navrhnutá sieťová infraštruktúra je spoľahlivá, výkonná a spĺňa technické požiadavky. Poskytuje stabilné zázemie pre každodennú prevádzku firmy a zároveň vytvára priestor pre jej ďalší rozvoj.

POĎAKOVANIE

Táto práca bola financovaná EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00019 – „Kľúčové digitálne technológie v oblasti bezdrôtových širokopásmových odolných prenosových sietí“ a projektu APVV-22-0400 – „Rozšírenie autonómnych aplikácií monitorovania, prostredníctvom viacpásmových UWB senzorov“.

LITERATÚRA

[1] IBM. What Is Computer Networking? [online]. [cit. 2024-10-12]. Dostupné na internete: <https://www.ibm.com/topics/networking>.
[2] Britannica. Computer Network – Definition & Types. [online]. [cit. 2024-10-19]. Dostupné na internete: <https://www.britannica.com/technology/computer-network>.

[3] CodingMinutes. Introduction to Computer Networks. [online]. [cit. 2024-10-21]. Dostupné na internete: <https://blog.codingminutes.com/introduction-to-computer-networks>.
[4] Testing and Structured Cabling Standards – Harmonization. [online]. [cit. 2024-10-22]. Dostupné na internete: <https://www.flukenetworks.com/blog/cabling-chronicles/testing-and-structured-cabling-standards-harmonization>.
[5] STN EN 50173-1 – Všeobecné požiadavky na štruktúrovanú kabeláž.
[6] STN EN 50173-3 – Kabeláž v priemyselnom prostredí.
[7] STN EN 50174-2 – Inštalácia kabeláže v budovách.
[8] STN EN 50174-3 – Inštalácia mimo budov: pokrytie vonkajších trás, kruhov vedenia, ochrany voči poveternostným podmienkam.
[9] STN EN 62012-1 – Špecifikácia viacžilových symetrických káblov pre digitalitu v drsnom prostredí.
[10] STN EN 50290-1-1 / 1-2 / 4-1 – Normy pre oznamovacie (komunikačné) káble, ktoré sa môžu uplatniť aj na metalickú dátovú kabeláž.

Optimalizácia viac-kanálového DWDM systému

[1]*Ľuboš OVSENÍK,* [2]*Jakub ORAVEC,* [3]*Zuzana LIŠČINSKÁ,* [4] *Viliam JENČA*

[1,2,3,4]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]zuzana.liscinska@tuke.sk, [4]viliam.jenca@student.tuke.sk

***Abstrakt* — Príspevok sa zaoberá návrhom viac-kanálového DWDM systému, s cieľom analyzovať vplyv rôznych modulačných formátov na kvalitu optického prenosu pri rôznych vzdialenostiach a prenosových rýchlostiach v simulačnom prostredí OptiSystem. Testované boli modulačné formáty DQPSK, OQPSK a DQPSK s CSRZ kódovaním, pri prenosových vzdialenostiach od 50 km do 230 km a rýchlostiach 10 až 15 Gbit/s. Práca poskytuje dôležité poznatky pre návrh a optimalizáciu DWDM systémov a prispieva k ďalšiemu rozvoju optických komunikačných sietí.**

***Kľúčové slová* — BER, CSRZ, DQPSK, DWDM, OQPSK, optické zosilňovače, Q-faktor**

I. Úvod

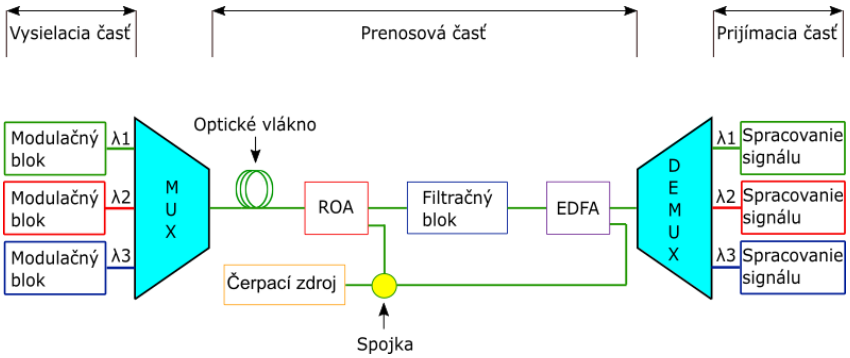
V súčasnosti sme svedkami neustále rastúcej potreby vysokorýchlostného prenosu dát, čo kladie značné nároky na kapacitu a efektivitu komunikačných systémov. S rozmachom internetu, cloudových služieb, streamovania médií a výpočtových centier je nevyhnutné neustále inovovať prenosové technológie, ktoré umožňujú rýchlu, spoľahlivú a vysokokapacitnú komunikáciu. Jednou z najefektívnejších technológií v oblasti optických sietí je DWDM (Dense Wavelength Division Multiplexing), ktorá umožňuje súčasný prenos viacerých optických signálov rôznych vlnových dĺžok cez jedno optické vlákno. Vďaka tomu sa výrazne zvyšuje celková prenosová kapacita siete bez nutnosti kladenia ďalších vlákien [1, 2].

Pri budovaní DWDM systémov však zohráva kľúčovú úlohu nielen samotná technológia multiplexovania, ale aj zosilňovanie optického signálu, ktoré je nevyhnutné najmä pri dlhých prenosových trasách. Optické zosilňovače, ako sú EDFA (Erbium-Doped Fiber Amplifiers), Ramanove zosilňovače či polovodičové zosilňovače, slúžia na obnovenie úrovne optického signálu bez potreby jeho konverzie do elektrickej formy. Ich správne umiestnenie, nastavenie a typ, má zásadný vplyv na kvalitu prenosu, dosiahnuteľnú vzdialenosť, mieru chybovosti a celkovú energetickú efektivitu systému. Cieľom tejto práce je návrh a analýza viac-kanálového DWDM systému so zameraním na vplyv rôznych modulačných formátov a na kvalitu optického prenosu [3, 4].

II. Návrh viac-kanálového DWDM systému v prostredí OptiSystem

Na Obr. 1 je znázornená bloková schéma navrhnutého optického komunikačného systému, ktorá je rozdelená na tri základné časti. V prvej z týchto troch častí je vysielacia časť, ktorá má za úlohu zabezpečiť generovanie a moduláciu optického signálu pred jeho vstupom do prenosového média. Modulačný blok predstavuje konkrétnu moduláciu pre jednotlivé použité kanály v optickom komunikačnom systéme. Každý kanál je prenášaný na odlišnej vlnovej dĺžke, pričom medzi nimi je definovaný kanálový rozostup. Signály z jednotlivých modulačných blokov sú následne vedené do optického multiplexora (MUX), ktorý ich spája do jediného optického vlákna. V druhej prenosovej časti multiplexovaný signál prechádza cez optické vlákno, kde je počas prenosu zosilňovaný a upravovaný na minimalizáciu degradácie. Najskôr je zosilnený pomocou Ramanovho zosilňovača (ROA). Tento proces je umožnený čerpacím zdrojom, ktorý dodáva energiu potrebnú na zosilnenie signálu v optickom vlákne. Následne signál prechádza cez filtračný blok, ktorý potláča nežiaduce spektrálne zložky a šum. Na ďalšie kompenzovanie útlmu je signál zosilnený erbium-dopovaným vláknovým zosilňovačom (EDFA), kde opäť energiu dodáva čerpací zdroj, čím zabezpečíme dostatočnú kvalitu pred vstupom do prijímacej časti. V prijímacej časti je optický signál najskôr demultiplexovaný (DEMUX), čím sa oddelia

jednotlivé kanály podľa ich vlnových dĺžok. Následne je prevedený na elektrický signál pomocou fotodetektora a ďalej spracovaný cez filtračné a regeneračné bloky, ktoré zabezpečujú obnovu kvality signálu. Výstupný signál je potom analyzovaný pomocou diagnostických nástrojov na hodnotenie kvality prenosu.



Obr. 1 Bloková schéma navrhnutého viac-kanálového systému.

Základne vstupné parametre viac-kanálového DWDM systému sú uvedené v Tabuľke 1. Poskytujú prehľad o nastavení systému a jeho prevádzkových podmienkach.

Tabuľka 1
Základné vstupné parametre viac-kanálového DWDM systému

Parameter	Hodnota
Bitová rýchlosť	10 Gbit/s
Dĺžka sekvencie	128 bit
Počet vzoriek na bit	64
Počet vzoriek	8192
Výkon vysielača	2 mW
Počiatočná frekvencia	193,1 THz
Kanálový rozstup	100 GHz
Počet kanálov	16 - 32
Dĺžka vlákna	140 – 175 km

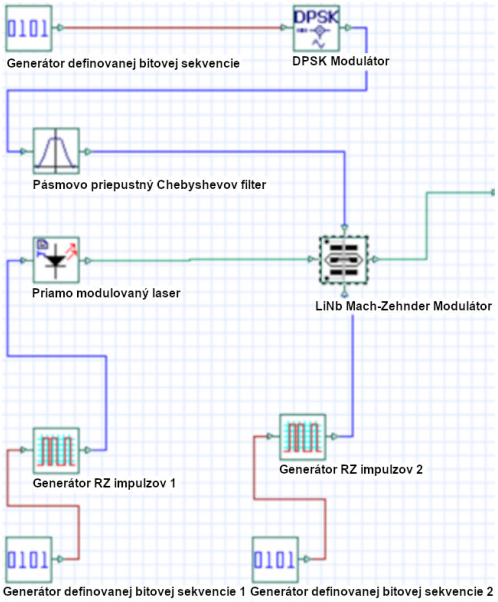
A. Návrh modulačných formátov pre vysielaciu časť

Výber vhodnej modulácie zohráva dôležitú úlohu pre optický prenosový systém. Cieľom je vyhodnotiť a vybrať vhodné modulačné techniky, ktoré zabezpečia spoľahlivý a efektívny prenos signálu. Pre náš systém sme sa zaoberali týmito modulačnými technikami: DQPSK (Differential Quadrature Phase Shift Keying), OQPSK (Offset Quadrature Phase Shift Keying) a DQPSK s CSRZ (Differential Quadrature Phase Shift Keying with Carrier-Suppressed Return-to-Zero) [2, 5].

A1) DQPSK modulácia: je fázová modulácia, ktorá kóduje dva bity na symbol posunom fázy nosného signálu, to znamená, že modulácia pracuje so štyrmi fázovými stavmi (0°, 90°, 180°, 270°), pričom každý symbol reprezentuje 2 bity. Namiesto prenosu absolútnych hodnôt fázy využíva fázové rozdiely medzi po sebe nasledujúcimi symbolmi, čo zjednodušuje detekciu na prijímači. Na Obr. 2 môžeme vidieť návrh simulácie DQPSK modulácie v programe OptiSystem.

Tabuľka 2
Parametre DQPSK modulácie

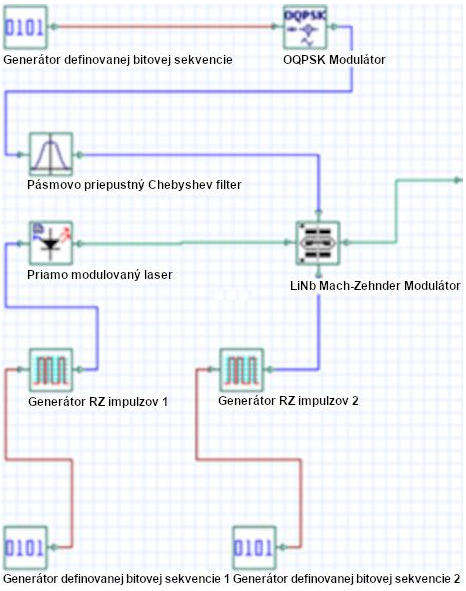
Parameter	Hodnota
Výkon vysielača	2 mW
Frekvenčný rozsah kanálov	193,1 - 194,6 THz
Kanálový rozstup	100 GHz
Počet kanálov	16
Dĺžka vlákna	175 km



Obr. 2 Návrh DQPSK modulácie v programe OptiSystem.

Návrh pozostáva z viacerých blokov, ktoré zabezpečujú generovanie signálu, jeho moduláciu a následnú konverziu do optickej podoby. Proces začína generátorom definovanej bitovej sekvencie, ktorý vytvára digitálny dátový tok. Dátový signál vstupuje do elektrického DPSK modulátora, ktorý vykonáva diferenciálnu fázovú moduláciu a mení fázu signálu podľa vstupnej bitovej sekvencie. Je nastavený na 2 bity pre symbol, takže každý symbol kóduje dva bity informácie. Modulovaný signál následne prechádza cez pásmovo priepustný Chebyshevov filter, ktorý odstráni nežiadúce frekvenčné zložky, šum a zabezpečí správnu šírku pásma. Optický priamo modulovaný laser generuje optický signál na počiatočnej vlnovej dĺžke 193,1 THz s výkonom 2 mW a tento nosný signál putuje do elektrooptického modulátora LiNb MZM (Mach-Zehnder Modulátor), kde sa ďalej moduluje. Ďalšími vstupmi LiNb MZM sú DPSK modulovaný signál a dva RZ (Return-to-Zero) impulzy. Tieto RZ impulzy sú generované na základe užívateľom definovaných bitových sekvencií. Základné nastavené parametre DQPSK modulácie sú zobrazené v Tabuľke 2.

A2) OQPSK modulácia: je variant QPSK modulácie, pri ktorej sa minimalizujú prudké zmeny amplitúdy signálu, čím sa znižuje pravdepodobnosť fázových skokov o 180° a zlepšuje sa odolnosť voči šumu a rozptylu. Znižuje náhle zmeny amplitúdy signálu tým, že oneskoruje Q-kanál o polovicu symbolového intervalu. Na Obr. 3 môžeme vidieť návrh OQPSK modulácie pre náš prenosový systém v programe OptiSystem.



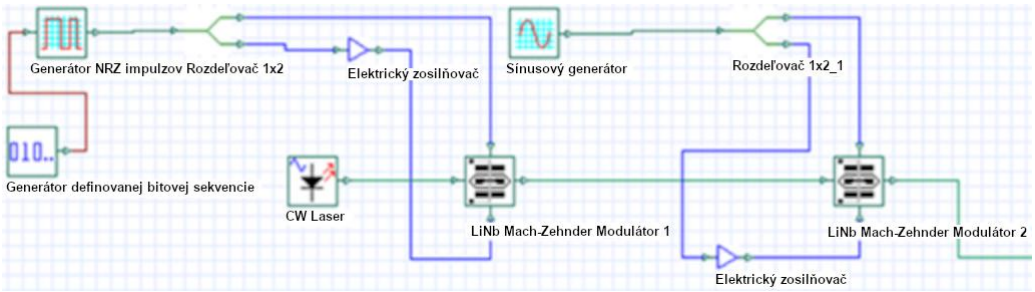
Obr. 3 Návrh OQPSK modulácie v programe OptiSystem.

Tabuľka 3
Parametre OQPSK modulácie

Parameter	Hodnota
Výkon vysielača	2 mW
Frekvenčný rozsah kanálov	193,1 - 194,6 THz
Kanálový rozstup	100 GHz
Počet kanálov	16
Dĺžka vlákna	160 km

Na začiatku je generátor bitovej sekvencie, ktorý vytvára digitálny dátový tok na prenos. Tento dátový tok vstupuje do OQPSK modulátora, ktorý rozdeľuje sekvenciu na dva kanály – I-kanál (In-phase) a Q-kanál (Quadrature-phase). Nepárne bity sú priradené I-kanálu, zatiaľ čo párne bity sú priradené Q-kanálu. Následne je Q-kanál oneskorený o polovicu symbolového intervalu. Po modulácii OQPSK signálu prechádza výstup cez pásmový priepustný Chebyshevov filter, ktorý odstraňuje nežiadúce frekvenčné zložky a šum. Dôležitý blok je priamo modulovaný laser, ktorý generuje optický nosný signál s nosnou frekvenciou 193,1 THz a s výkonom 2 mW. Nasleduje LiNb Mach-Zehnder modulátor, ktorý vykonáva elektro-optickú moduláciu vstupného optického signálu na základe OQPSK signálu a pomocných pulzov. Do MZM vstupuje modulovaný signál z OQPSK modulátora spolu s dvoma RZ impulzmi, ktoré sú generované na základe užívateľom definovaných bitových sekvencií. Tieto RZ impulzy slúžia na časovanie a formovanie signálu. V Tabuľke 3 sú zobrazené nastavené parametre OQPSK modulácie.

A3) DQPSK modulácia s kódovaním CSRZ: Pre zlepšenie odolnosti voči disperzii a zvýšenie spektrálnej efektivity sa k DQPSK pridáva CSRZ kódovanie. CSRZ sa používa na potlačenie nosnej vlny a vrátenie signálu na nulu na začiatku každého symbolu. DQPSK s CSRZ je technika prenosu dát, ktorá kombinuje výhody DQPSK a CSRZ, tým že sa v DQPSK informácie kódujú pomocou rozdielu fáz medzi po sebe idúcimi symbolmi, čo zvyšuje odolnosť proti šumu. V CSRZ sa nosná vlna potláča a signál sa vracia na nulu, čím sa zjednodušuje detekcia. Na Obr. 4 môžeme vidieť návrh DQPSK modulácie s kódovaním CSRZ.



Obr. 4 Návrh DQPSK modulácie s kódovaním CSRZ v programe OptiSystem.

V tomto systéme sa najskôr generuje binárna sekvencia pomocou generátora bitovej sekvencie, ktorá sa konvertuje na NRZ (Non-Return-to-Zero) signál. Generátor NRZ impulzov zabezpečuje správnu časovú reprezentáciu bitov pred ich fázovou moduláciou. Optický signál poskytuje CW laser (Continuous Wave) so spojitou vlnou pracujúci na počiatočnej frekvencii 193,1 THz s výkonom 2 mW. Tento signál vstupuje do MZM, ktorý vykonáva fázovú moduláciu na základe NRZ signálu. Diferenciálne kódovanie zabezpečuje, že informácia je zakódovaná ako rozdiel medzi fázami susedných symbolov, čím odstraňuje potrebu referenčného bodu na prijímacej strane. Po modulácii sa signál rozdelí pomocou rozdeľovača 1x2 na dve vetvy – I (In-phase) a Q (Quadrature). Sínusový generátor s frekvenciou 32 GHz a fázovým posunom 90° vytvára dva ortogonálne signály pre kvadratúrnu moduláciu.

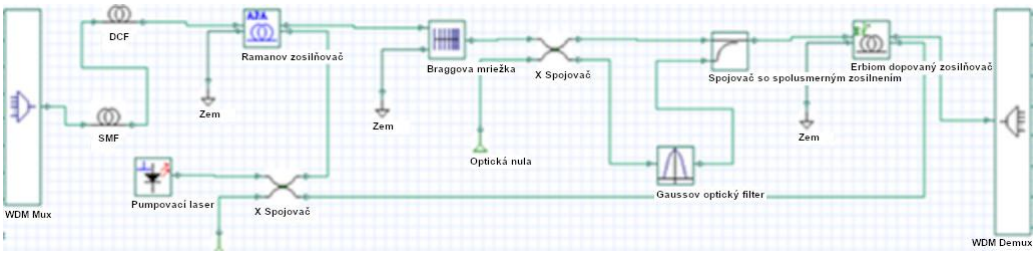
Elektrické zosilňovače upravujú úroveň signálu, aby zabezpečili optimálnu moduláciu v ďalšom stupni. Druhý MZM vykonáva CSRZ kódovanie, pri ktorom optická intenzita v strede symbolovej periódy klesá na nulu, čím sa potláča nosná zložka spektra a zlepšuje spektrálna efektivita. V Tabuľke 4 sú zobrazené nastavené parametre DQPSK modulácie s CSRZ kódovaním.

Tabuľka 4
Parametre DQPSK modulácie s CSRZ kódovaním

Parameter	Hodnota
Výkon vysielača	2 mW
Frekvenčný rozsah kanálov	193,1 - 196,2 THz
Kanálový rozstup	100 GHz
Počet kanálov	32
Dĺžka vlákna	140 km

B. Návrh prenosovej časti prenosového systému

Prenosová časť optického systému je navrhnutá na zosilnenie a kompenzáciu disperzie pri prenose signálu na dlhé vzdialenosti. Na zosilnenie signálu používame hybridný optický zosilňovač spojením Ramanovho zosilňovača a erbiom dopovaného zosilňovača. Na Obr. 5 môžeme vidieť návrh prenosovej časti prenosové systému v programe OptiSystem.



Obr. 5 Návrh prenosovej časti systému v programe OptiSystem.

Po multiplexovaní signálu cez WDM multiplexor sa optické signály prenášajú cez SMF (Single Mode Fiber) jednomódové optické vlákno s prenosovou dĺžkou od 140 – 175 km podľa použitého typu modulácie. Keďže dlhé optické vlákno spôsobuje chromatickú disperziu, tak pomocou disperzne kompenzujúceho vlákna DCF (Dispersion Compensating Fiber) s dĺžkou 13,6 km túto disperziu kompenzujeme. Následne signál vstupuje do Ramanovho zosilňovača, ktorý dosahuje zosilnenie čerpaním energie z čerpacieho zdroja s vlnovou dĺžkou 1490 nm a výkonom 900 mW. V Tabuľke 5 môžeme vidieť nastavené parametre Ramanovho zosilňovača.

Ďalej nasleduje Braggova mriežka, ktorá sa používa na čiastočnú disperznú kompenzáciu. Signál ďalej rozdeľuje cez optický krížový spojovač, ktorý umožňuje delenie signálov na rôzne cesty. Na jednej z ciest sa nachádza Gaussov optický filter, ktorý potláča nežiadúce šumové zložky signálu a optimalizuje jeho spektrálnu charakteristiku. Na druhej ceste spojovač so spolu smerným zosilnením, ktorý smeruje zosilnený dátový signál do EDFA zosilňovača. Následne sa signál zosilní pomocou EDFA zosilňovača pomocou energie z čerpacieho zdroja a nakoniec zosilnený signál prechádza do WDM demultiplexora, kde sa jednotlivé kanály rozdelia na svoje špecifické vlnové dĺžky pre ďalšie spracovanie. V Tabuľke 6 vidíme základné nastavené parametre EDFA zosilňovača.

Tabuľka 5
Parametre Ramanovho zosilňovača

Parameter	Hodnota
Dĺžka vlákna	10 km
Útlm vlákna	0,2 dB/km
Čerpací zdroj – vlnová dĺžka	1490 nm
Čerpací zdroj – výkon	900 mW

Tabuľka 6
Parametre EDFA zosilňovača

Parameter	Hodnota
Dĺžka vlákna	5 m
Doba života Er^{3+}	10 ms
Numerická apertúra	0,24
Čerpací zdroj – vlnová dĺžka	1490 nm
Čerpací zdroj – výkon	900 mW

C. Návrh prenosovej časti prenosového systému

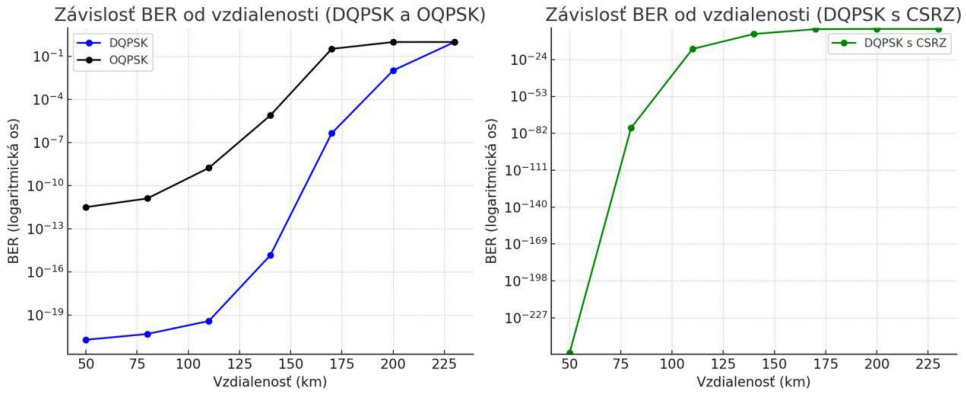
Prijímacia časť prenosového systému spracúva optický signál a konvertuje ho na elektrický signál. Neskôr si vieme tento spracovaný a prekonvertovaný signál vyhodnotiť a zistiť tak kvalitu prenosu pomocou rôznych analyzátorov ako napríklad sú diagram oka alebo BER analyzátor.

C1) Prijímacia časť DQPSK a OQPSK modulácie: je v oboch prípadoch rovnaká a využíva 4-stupňovú kaskádovú sústavu Braggových mriežok, ktorá efektívne kompenzuje chromatickú

disperziu. Môžeme si to predstaviť ako sériu zrkadiel, ktorá upravuje signál odrazom v každej mriežke a tým minimalizuje BER a tiež aj Q-faktor, ktorý predstavuje kvalitu signálu. Po prechode tejto sústavy nasleduje Besselov optický filter, ktorý redukuje spektrálne zvlnenia. Fotodetektor konvertuje optický signál na elektrický a následne dolnopriepustný Besselov filter odstraňuje šum. Potom 3R regenerátor (reamplifikácia, pretvarovanie, prečasovanie) obnovuje signál. Nakoniec už hodnotíme kvalitu prenosu pomocou rôznych analyzátorov.
C2) Prijímacia časť DQPSK modulácie s CSRZ kódovaním: obsahuje optický prijímač, ktorý premieňa prijatý optický signál na elektrický. Prijímač obsahuje aj dolnopriepustný Besselov filter s orezávacou frekvenciou 0,75-násobku prenosovej rýchlosti, ktorý redukuje šum a nežiaduce vysokofrekvenčné zložky signálu. Následne je signál analyzovaný nástrojmi diagramu oka a BER analyzérom.

III. Výsledky a experimentálne testovanie viac-kanálového DWDM systému

Testovanie navrhnutého systému prebiehalo s cieľom overiť jeho funkčnosť a efektivitu pre rôzne nastavené podmienky a dosiahnuť nízku chybovosť a vysokú kvalitu DWDM systému. Pri testovaní boli analyzované dôležité parametre, ako je bitová chybovosť BER a kvalita signálu Q-faktor. Testovanie zahŕňa porovnanie rôznych modulačných formátov, ako aj overenie správania systému pri rôznych prenosových rýchlostiach a vzdialenostiach. Výsledky testovania poskytujú dôležité údaje o výkonnosti a spoľahlivosti systému.



Obr. 6 Porovnanie závislostí BER od prenosovej vzdialenosti pre rôzne modulácie.

A. Porovnanie modulácií pri rýchlosti 10 Gbit/s

Z hľadiska počtu prenesených kanálov dokázala DQPSK modulácia s CSRZ kódovaním preniesť najviac použiteľných kvalitných signálov a to až 32 kanálov. Pre porovnanie pri DQPSK a OQPSK modulácii sa podarilo preniesť 16 kvalitných signálov, keďže pri navýšení kanálov pri týchto moduláciách systém už viac kanálov nedokázal spracovať. Z hľadiska chybovosti BER a kvality Q-faktoru dosiahla najlepšie výsledky DQPSK modulácia, ktorá mala najnižšie hodnoty BER a zároveň vysoký Q-faktor v jednotlivých kanálov. Aj keď OQPSK modulácia dosiahla taktiež veľmi kvalitné výsledky, no jej BER nebol taký nízky ako pri DQPSK, navyše dosiahla kratšiu vzdialenosť optického vlákna o 15 km ako DQPSK, ale z hľadiska konzistentnosti bola práve OQPSK modulácia najstabilnejšia.

Ak sa pozrieme na kvalitu signálu pri predĺžení optického vlákna o 30 km, tak najlepšie výsledky dosiahla OQPSK modulácia, ktorá ešte dosahovala relatívne kvalitne a konzistentné hodnoty BER a Q-faktoru. DQPSK modulácia s CSRZ už značne postihla degradácia signálu, teda mala najhoršie výsledky v tejto štatistike. Ako sa menil BER pri zvyšovaní prenosovej vzdialenosti od 50 km do 230 km je graficky znázornené na Obr. 6.

B. Porovnanie modulácií pri rýchlosti 15 Gbit/s

Vplyv zvýšenej prenosovej rýchlosti na 15 Gbit/s najlepšie zvládla DQPSK modulácia s CSRZ kódovaním, aj keď iba niektoré kanály dosiahli ešte relatívne dobré výsledky. Najhoršie si v tejto štatistike počínala DQPSK modulácia, ktorá v niektorých kanáloch dosiahla BER = 1 a Q-faktor = 0, čo znamená, že každý bit bol prijatý nesprávne a signál bol úplne nečitateľný.

Ak porovnáme všetky modulácie na základe spomenutých štatistík, nedá sa presne povedať, ktorá bola najvhodnejšia, keďže každá z nich mala výhody a nevýhody. Ak sa pozrieme na kvalitu signálu v pôvodných parametroch, tak jasne vyhrala DQPSK modulácia, no nepreniesla toľko kanálov ako DQPSK modulácia s CSRZ a nebola ani tak konzistentná v jednotlivých kanáloch ako OQPSK modulácia, ale preniesla signál na najdlhšiu vzdialenosť 175 km.

IV. ZÁVER

V tejto práci sme sa venovali návrhu a optimalizácii viac-kanálového DWDM systému s cieľom zabezpečiť kvalitný a spoľahlivý prenos optického signálu na rôzne vzdialenosti a prenosové rýchlosti. Primárnym cieľom bolo analyzovať vplyv rôznych modulačných formátov na kvalitu prenosu signálu v DWDM systémoch.

Pri použití DQPSK modulácie sa dosiahli veľmi dobré výsledky pri vzdialenosti 175 km a rýchlosti 10 Gbit/s bola dosiahnutá chybovosť BER na úrovni $9{,}07 \cdot 10^{-14}$ a vysoký Q-faktor až 7,33. S rastúcou vzdialenosťou nad 175 km a prenosovou rýchlosťou až na 15 Gbit/s, však táto modulácia strácala na kvalite, čím sa prejavila jej zvýšená citlivosť na degradačné javy a bola použiteľná len v obmedzenom počte 16 kanálov. OQPSK modulácia vynikala predovšetkým svojou stabilitou kvality signálu naprieč všetkými kanálmi. Pri vzdialenosti 160 km a rýchlosti 10 Gbit/s boli dosiahnuté hodnoty BER rádovo na úrovni 10^{-10} a Q-faktor približne 6, čo predstavuje spoľahlivý prenos. Táto modulácia však výrazne strácala na kvalite pri zvýšení prenosovej rýchlosti na 15 Gbit/s, kde sa signál stával už nepoužiteľným, avšak pri zvyšovaní prenosovej vzdialenosti nad 160 km ešte dokázala preniesť stabilné, ale menej kvalitné signály. Modulačný formát DQPSK s CSRZ kódovaním sa ukázal ako optimálny z hľadiska kapacity systému, keď umožňoval spoľahlivo preniesť až 32 kanálov pri vzdialenosti 140 km a prenosovej rýchlosti 10 Gbit/s, pričom hodnoty BER boli stabilne na úrovni 10^{-7} až 10^{-15} a Q-faktor v rozmedzí od 5 až 7,8. Táto modulácia však mala výrazný pokles kvality pri predĺžení trasy, kde niektoré kanály vykazovali vysoké hodnoty BER. Na druhej strane, práve táto modulácia ako jediná zachovávala dobré parametre prenosu aj pri zvyšovaní rýchlosti až na 15 Gbit/s, na kratších vzdialenostiach do 140 km. Z dosiahnutých výsledkov simulácií vyplýva, že každý zo sledovaných modulačných formátov dosahoval špecifické výhody a súčasne obmedzenia v rôznych scenároch nasadenia. Výsledky tejto práce sú teda príspevkom k návrhu efektívnych a spoľahlivých optických prenosových systémov s potenciálom uplatnenia v reálnych komunikačných sieťach a môžu slúžiť ako základ pre ďalší výskum a optimalizácie v tejto oblasti.

POĎAKOVANIE

Táto práca bola financovaná EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00019 – „Kľúčové digitálne technológie v oblasti bezdrôtových širokopásmových odolných prenosových sietí“ a projektu VEGA 1/0260/23 – „Umelá inteligencia, robustné MANET a viacpreskokové D2D siete integrované do siete 6.generácie“.

LITERATÚRA

[1] Chadha, D. Optical WDM Networks: From Static to Elastic Networks. s.l. : IEEE Press, 2019. 9781119393269.
[2] Rajeev, & Kumar, Chakresh. Performance Investigation of NRZ-DQPSK, CSRZ-DQPSK and MDRZ-DQPSK Modulation Techniques for 450 Channel UD-WDM System for Long Haul Communication. Journal of Scientific & Industrial Research, Vol. 82, August 2023, pp. 870-884. DOI: 10.56042/jsir.v82i08.1053.
[3] Attaouia, B. and Malika, K. Performance of EDFA Amplifier Position (pre-, post- and In-line Amplification) for Reach Extendibility of GPON-FTTH. International Journal of Electronics Mechanical and Mechatronics Engineering, 2017, Vol. 7, no. 1, pp. 1327–1334.
[4] Honde, V., Mhatre, A., Tonde, S., Barkul, S. and Pund, P.. Performance Analysis of WDM Network based on EDFA Amplifier with Different Pumping Techniques. International Journal on Recent and Innovation Trends in Computing and Communication, 2016, Vol. 4, no. 4, pp. 480–485. ISSN 2321-8169.
[5] Faruque, S. Phase shift keying (PSK). Radio Frequency Modulation Made Easy. SpringerBriefs in Electrical and Computer Engineering. Cham: Springer, 2017. DOI: https://doi.org/10.1007/978-3-319-41202-3_6.

Komunikácia FSO systémov v UV spektre

[1]*Ľuboš OVSENÍK,* [2]*Jakub ORAVEC,* [3]*Zuzana LIŠČINSKÁ,* [4]*Branislav KAČMARIK*

[1,2,3,4]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]zuzana.liscinska@tuke.sk, [4]branislav.kacmarik@student.tuke.sk

***Abstrakt* — Príspevok sa zaoberá návrhom a simuláciou bezdrôtových optických systémov šíriacich signál voľným prostredím (FSO) s využitím ultrafialového spektra v pásmach UV-A, UV-B a UV-C. Cieľom práce je analyzovať vplyv vlnovej dĺžky žiarenia, počet odrazov a odrazivosť materiálov na kvalitu prenosu, pričom hlavný hodnotiaci parameter je chybovosť prenášaných bitov. Simulácie boli realizované v prostredí OptiSystem, ktorý umožňuje modelovať FSO kanály s ohľadom na rôzne podmienky šírenia, ako aj charakteristiky vysielača, prijímača a odrazového materiálu. Výsledky ukazujú, že pri UV-C prenose sú použiteľné len vysoko reflektívne materiály, ako striebro alebo titán, zatiaľ čo UV-A a UV-B prenos je možný aj s bežnými materiálmi ako drevo či betón. Priemerný počet použiteľných odrazov sa pohyboval v rozmedzí od 1 do 5 v závislosti od spektrálneho pásma a vlastností povrchov. Zistenia poukazujú na potenciál UV FSO systémov v špecifických aplikáciách a na možnosti ich ďalšieho výskumu.**

***Kľúčové slová* —bitová chybovosť, FSO, NLOS prenos, odrazivosť materiálu, UV žiarenie**

I. Úvod

V súčasnosti rastie dopyt po vysokorýchlostnej a bezpečnej bezdrôtovej komunikácii, ktorá by bola nezávislá od konvenčného preťaženého rádiového spektra. Jednou z perspektívnych oblastí výskumu sú optické bezdrôtové komunikačné systémy (FSO), ktoré využívajú optický signál na prenos dát. Zatiaľ čo infračervené FSO systémy sú celkom dobre preskúmané, komunikácia v ultrafialovom spektre v pásmach UV-A, UV-B a UV-C, predstavuje nový a málo prebádaný smer, ktorý ponúka výhody najmä pri nepriamočiarom režime šírenia signálu (NLOS) [1].

Táto práca sa zaoberá návrhom a simuláciou UV FSO prenosových systémov v rôznych prostrediach od satelitnej komunikácie vo vákuu na nízkej obežnej dráhe Zeme, až po interiérové aplikácie v bežných podmienkach na povrchu Zeme. V rámci simulácií bol skúmaný vplyv viacerých parametrov na kvalitu prenosu, predovšetkým odrazivosť rôznych materiálov, počet viacnásobných odrazov a použitá vlnová dĺžka ultrafialového žiarenia. Hlavným cieľom práce bolo vyhodnotenie použiteľnosti UV FSO systémov na základe dosiahnutej chybovosti prenosu pri rôznych prenosových podmienkach.

II. Ultrafialové žiarenie a jeho využitie v FSO komunikačných systémoch

Ultrafialové UV (Ultraviolet) žiarenie je súčasťou elektromagnetického spektra s kratšími vlnovými dĺžkami ako viditeľné svetlo v rozsahu od približne 100 nm do 400 nm. Jeho zdrojom je primárne Slnko, avšak môže byť generované aj umelo pomocou špeciálnych výbojok alebo LED diód. UV spektrum sa delí na tri hlavné časti: blízke ultrafialové UV-A s rozsahom 315 – 400 nm, stredné ultrafialové UV-B v rozsahu 280 – 315 nm a vzdialené ultrafialové UV-C s vlnovými dĺžkami od 100 do 280 nm. Každý z týchto rozsahov má špecifické vlastnosti a využitie v rôznych oblastiach, od komunikačných systémov, medicíny až po sterilizáciu [2, 3].

Ultrafialové žiarenie je využívané vo vedeckých výskumoch, pri ktorých sa skúma jeho potenciál pri zabezpečených komunikačných kanáloch napr. NLOS FSO, pretože UV signál sa ťažšie odpočúva v porovnaní s rádiovými alebo infračervenými vlnami, čím získava dobrú úroveň zabezpečenosti pred vonkajšími útokmi.

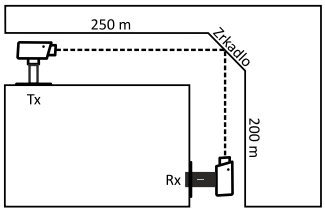
Využitie ultrafialového spektra v FSO systémoch predstavuje veľkú výzvu, či už pri využití v interiéri alebo exteriéri. Aj napriek svojim nevýhodám, ako je rušenie slnečným žiarením pri

využívaní za denného svetla v exteriéri alebo bezpečnostným rizikám pre zrak a pokožku človeka pri využití v interiéri, si dokázala táto komunikácia nájsť uplatnenie napr. pri prenose dát na krátke vzdialenosti, pri komunikácií medzi družicami na nízkej obežnej dráhe, či pri vojenských a bezpečnostných operáciách v mestskom prostredí.

Aj keď UV komunikácia ponúka výhody v oblasti flexibility prenosu, dosah signálu je stále limitovaný absorpciou aerosólmi a molekulami v atmosfére. Tento vysoký útlm zároveň znamená, že signály za určitým bodom už nie je možné zachytiť, čo môže byť výhodné pre bezpečné a šifrované taktické aplikácie [4].

A. Optická komunikácia vo voľnom prostredí pri nepriamej viditeľnosti

Jednou z dôležitých aplikácií ultrafialového žiarenia je jeho využitie v bezdrôtovej komunikácii, najmä v NLOS FSO systémoch. UV FSO dokáže prenášať signál aj v prostredí s prekážkami, pretože UV lúče sa rozptyľujú v atmosfére a umožňujú tak nepriamu komunikáciu. Táto vlastnosť je výhodná v prostredí, kde nie je možné vytvoriť priame optické spojenie, napríklad v husto zastavaných mestských oblastiach alebo v interiéroch (Obr. 1).



Obr. 1 NLOS FSO komunikácia.

UV žiarenie má špecifické vlastnosti, vďaka ktorým je vhodné na NLOS komunikáciu. Keďže krátkovlnné UV svetlo sa silnejšie rozptyľuje na malých časticiach a molekulách v atmosfére, jeho signál sa môže prenášať aj v nepriaznivých podmienkach, napríklad medzi prekážkami, v prašnom prostredí alebo v hustom mestskom prostredí. Tento efekt je spôsobený Rayleighovým rozptylom, ktorý je výraznejší pri kratších vlnových dĺžkach. Rozptyl signálu vedie k zníženiu jeho intenzity a účinnosti, čím sa obmedzuje maximálna dosiahnuteľná prenosová rýchlosť. Preto je dôležité optimalizovať výkon vysielačov a citlivosť prijímačov, aby bol zabezpečený stabilný prenos informácií [4, 5].

B. Odrazové prvky pre UV komunikáciu

Odrazivosť materiálov v UV spektre zohráva dôležitú úlohu v mnohých technických a vedeckých aplikáciách. Niektoré materiály dokážu UV svetlo efektívne odrážať, zatiaľ čo iné ho skôr absorbujú alebo prepúšťajú. V nasledujúcej tabuľke si podrobnejšie opíšeme jednotlivé materiály, ktoré sú alebo by mohli byť využívané pri FSO UV NLOS komunikácii. Materiály sú zoradené podľa ich schopnosti odrážať UV žiarenie, od najefektívnejších po najmenej efektívne (Tabuľka 1).

Tabuľka 1
Odrazivosť UV od jednotlivých materiálov

Materiál	Odrazivosť v UV spektre (% UV odrazeného žiarenia)	Poznámka
Striebro	90 – 98 %	Vysoká odrazivosť, ale oxiduje
Titánová fólia (TiO_2 povrch)	85 – 95 %	Silný UV reflektor, stabilný
Biely oxid titaničitý (TiO_2 náter)	85 – 95 %	Používaný v reflexných náteroch
Leštený hliník	80 – 95 %	Jeden z najlepších UV reflektorov
Mylar s UV ochrannou vrstvou	75 – 90 %	Používa sa na ochranu proti UV žiareniu
Zlatá fólia	60 – 80 %	Dobrá odrazivosť, používa sa v niektorých vesmírnych aplikáciách
Nerezová oceľ (leštená)	50 – 70 %	Nižšia UV odrazivosť, ale stále efektívna
Betón, tehla, drevo	10 – 30 %	Nízka odrazivosť, vysoká absorpcia
Akrylové a plastové povrchy	5 – 25 %	Väčšina plastov UV absorbuje
Bežné sklo	4 – 20 %	Väčšina UV prechádza alebo sa absorbuje

B1) Striebro (90 – 98 % odrazivosť v UV spektre): Striebro je najlepším známym reflektorom UV žiarenia, dosahujúcim odrazivosť až 98 %. Vďaka svojej vysokej odrazivosti sa využíva v

optických systémoch a pri výrobe zrkadiel. Nevýhodou je jeho náchylnosť na oxidáciu, ktorá môže znížiť jeho efektivitu. Preto sa často potiahne ochrannými vrstvami, ktoré zabraňujú jeho degradácii.
B2) Titánová fólia s povrchom TiO_2 (85 – 95 % odrazivosť v UV spektre): Tento materiál patrí medzi veľmi efektívne UV reflektory, pričom je stabilný a odolný voči vonkajším vplyvom. Povrchová vrstva oxidu titaničitého (TiO_2) zvyšuje jeho odrazivosť, čo ho robí ideálnym pre použitie v reflexných systémoch.
B3) Biely oxid titaničitý (TiO_2 náter) (85 – 95 % odrazivosť v UV spektre): Reflexné nátery na báze oxidu titaničitého sú často využívané v stavebníctve a priemyselných aplikáciách na odrážanie UV žiarenia. Tento materiál poskytuje veľmi dobrú ochranu pred UV degradáciou a znižuje tepelné účinky slnečného žiarenia.
B4) Leštený hliník (80 – 95 % odrazivosť v UV spektre): Leštený hliník je široko používaný v UV reflektoroch v priemyselných a vedeckých aplikáciách. Má veľmi dobrú odrazivosť a zároveň je lacnejší a menej náchylný na oxidáciu ako striebro. Jeho výhodou je tiež vysoká dostupnosť a možnosť využitia v rôznych povrchových úpravách.
B5) Mylar s UV ochrannou vrstvou (75 – 90 % odrazivosť v UV spektre): Mylar je tenká plastová fólia, ktorá sa často používa v optických a ochranných aplikáciách. Keďže je vybavená špeciálnou UV ochrannou vrstvou, dokáže efektívne odrážať UV žiarenie, čo z nej robí populárnu voľbu pre izoláciu a ochranu citlivých povrchov.
B6) Zlatá fólia (60 – 80 % odrazivosť v UV spektre): Zlato nie je tak efektívnym UV reflektorom ako striebro či hliník, ale stále poskytuje dostatočne vysokú odrazivosť. Používa sa predovšetkým v špecifických aplikáciách, ako je vesmírny priemysel, kde chráni satelity a citlivé zariadenia pred UV a infračerveným žiarením.
B7) Leštená nerezová oceľ (50 – 70 % odrazivosť v UV spektre): Hoci nerezová oceľ nie je primárne známa ako UV reflektor, leštené povrchy môžu dosiahnuť relatívne dobrú odrazivosť. V porovnaní s inými kovmi má však nižšiu účinnosť a je častejšie využívaná pre svoju odolnosť voči korózii než pre optické vlastnosti.
B8) Betón, tehla, drevo (10 – 30 % odrazivosť v UV spektre): Tieto materiály majú najnižšiu odrazivosť v UV spektre, pretože ich štruktúra a zloženie umožňujú vysokú absorpciu žiarenia. Sú to primárne stavebné materiály, ktoré sa zvyčajne nevyužívajú na optické aplikácie.
B9) Akrylové a plastové povrchy (5 – 25 % odrazivosť v UV spektre): Väčšina plastov má nízku UV odrazivosť a skôr absorbuje UV žiarenie, čo môže viesť k ich degradácii. Existujú však špeciálne upravené polyméry s pridanými UV blokátormi, ktoré môžu zvýšiť odolnosť voči UV žiareniu.
B10) Bežné sklo (4 – 20 % odrazivosť v UV spektre): Bežné sklo má veľmi nízku odrazivosť v UV spektre, pričom väčšina UV žiarenia ním prechádza alebo je absorbovaná. Preto sa na ochranu proti UV žiareniu často používajú špeciálne vrstvy alebo UV blokujúce sklá.

Odrazivosť materiálov v UV spektre sa výrazne líši v závislosti od ich chemického zloženia a povrchovej úpravy. Striebro a leštený hliník sú najefektívnejšími reflektormi, zatiaľ čo betón, tehla a drevo majú jednu z najnižších schopností odrážať UV žiarenie. Pri výbere vhodného materiálu je potrebné zohľadniť nielen jeho optické vlastnosti, ale aj stabilitu, odolnosť a aplikáciu, v ktorej bude použitý. Efektívne využitie správnych materiálov môže výrazne zlepšiť výkonnosť optických systémov a ochranných povrchov v prostredí vystavenom UV žiareniu [4, 6].

III. Simulácia UV FSO systémov v programe OptiSystem

Pri simuláciách ultrafialových bezdrôtových optických prenosových systémov v prostredí programu OptiSystem, sme analyzovali tri rôzne vlnové pásma: UV-A, UV-B a UV-C, pričom každé z nich bolo využité v rozdielnych podmienkach vzhľadom na ich fyzikálne vlastnosti a interakciu s prostredím.

UV-A a UV-B (použitie v interiéri): pásma UV-A (315–400 nm) a UV-B (280–315 nm) boli v simuláciách využívané na interiérové aplikácie, kde nehrozí výrazný vplyv slnečného žiarenia. V exteriéri by bolo využitie týchto pásiem problematické, kvôli silnému prirodzenému UV žiareniu na pozadí, ktoré by spôsobovalo výrazné rušenie signálu. V interiéri však tieto pásma umožňujú efektívny prenos bez významného šumu z vonkajšieho prostredia.

UV-C (použitie v exteriéri (LEO)): pásmo UV-C (100–280 nm) bolo simulované pre exteriérové použitie, konkrétne na nízkej obežnej dráhe Zeme (LEO) medzi satelitmi. UV-C žiarenie sa bežne absorbuje v atmosfére, a preto je jeho použitie bezpečné len mimo atmosféry, napríklad vo vákuu, kde zároveň umožňuje čistý FSO prenos bez atmosférického rušenia a zároveň bez ohrozenia ľudského zdravia.

A. Simulačné scenáre a materiál odrazových prvkov

Vo všetkých prípadoch boli do simulácie zahrnuté odrazové prvky. Počet odrazových prvkov sa menil, tak isto ako odrazový materiál daného prvku. Pri výbere odrazových materiálov do jednotlivých simulácií sme sa riadili snahou o čo najrealistickejšie podmienky pre dané prostredie. Pre simulácie v interiéri na Zemi (UV-A a UV-B pásmo) boli zvolené bežné stavebné a interiérové materiály (drevo, betón, plast a sklo), ktoré sa často vyskytujú v miestnostiach, skladoch alebo kanceláriách. Ich odrazivosť voči UV žiareniu je síce nízka, ale realistická pre testovanie prenosu signálu. Naopak, pre simulácie UV-C prenosu na nízkej obežnej dráhe LEO, kde sa signál môže odrážať od konštrukčných prvkov satelitov alebo reflektorov, sme zvolili vysoko reflexné materiály, ako striebro, titánová fólia, leštený hliník a zlatá fólia. Tieto materiály sa používajú v reálnych kozmických aplikáciách kvôli svojej schopnosti odrážať UV-C žiarenie a odolnosti voči extrémnym podmienkam vesmíru (Tabuľka 1).

Cieľom simulácie bolo preskúmať, ako jednotlivé UV pásma reagujú na rôzne podmienky s viacerými odrazmi a ako sa mení úroveň signálu na prijímači v závislosti od konfigurácie systému.

B. Parametre FSO systémov využitých pri jednotlivých simuláciách v programe OptiSystem

V nasledujúcej tabuľke (Tabuľka 2) sú definované parametre parametre FSO systémov, ktoré sme pri simuláciách využívali. Vysielač IndoorComm 370 využíva ako optický prijímač kremíkovú fotodiódu s úpravou pre možnosť vysielania na vlnovej dĺžke UV-A spektra (UV-enhanced Si photodiode). Z dôvodu obmedzení knižnice sme pri simuláciách použili ako optický prijímač lavínovú fotodiódu APD (Avalanche photodiode). Komunikačný systém SecureLink 295 využíva ako optický prijímač fotomultiplikátorovú trubicu necitlivú na slnečné svetlo (PMT Solar-blind), ktorú sme nahradili PIN fotodiódou, na ktorú sme aplikovali zosilnenie pre kompenzáciu nižšej citlivosti. Dôvodom tejto zmeny je obmedzená knižnica optických prijímačov, ktoré voľne dostupná verzia programu OptiSystem neponúka.

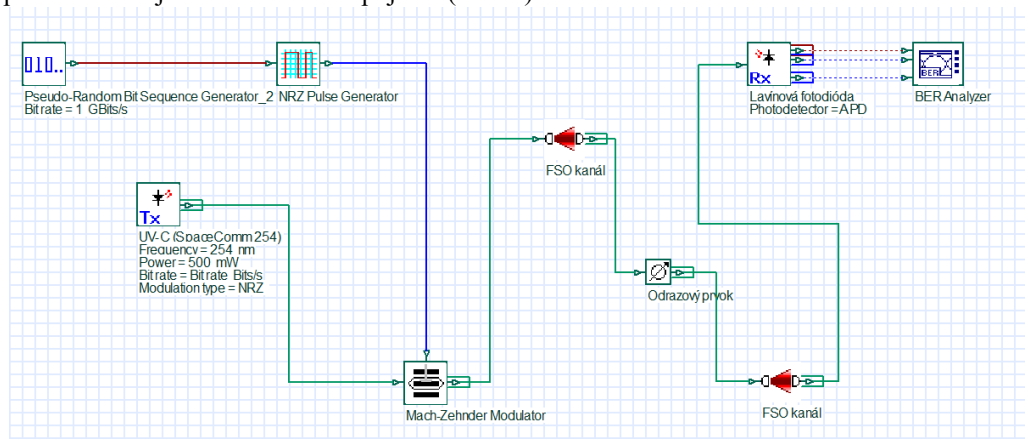
Tabuľka 2
Parametre FSO systémov využívané v simuláciách

Spektrum	UV-A	UV-B	UV-C
Vlnová dĺžka [nm]	370	295	254
Názov hlavice	IndoorComm 370	SecureLink 295	SpaceComm 254
Typ prostredia	Interiér (vzduch)	Interiér (špeciálne aplikácie)	Exteriér (vákuum, vesmír)
Vysielací výkon	1 mW	0,5 mW	500 mW
Dosah	20 m	40 m	50 000 km
Priemer vysielacej šošovky [mm]	20	30	250
Divergencia lúča	2 mrad	1,5 mrad	0,0001 mrad
Priemer detekčnej šošovky [mm]	25	35	300
Optický prijímač	UV-enhanced Si photodiode	PMT Solar-blind	APD avalanche photodiode
Norma EN 60825-1 (bezpečnostná trieda)	Trieda 1M	Trieda 1	Trieda 3B
Príklady použitia	Prenos dát v budovách	Špeciálne laboratórne aplikácie	Medzi satelitná komunikácia

C. Návrh simulačnej schémy UV FSO systémov

Začiatok tvorí Pseudo-Random Bit Sequence Generator, ktorý vytvára náhodnú postupnosť bitov a predstavuje zdroj informácií, ktoré sme prenášali cez UV-FSO kanál. Následne sa tento digitálny signál upraví pomocou Pulzného Generátora, kde sa preklopí do formátu NRZ (Non-Return-to-Zero), čo je štandardný spôsob, ako kódovať digitálny signál pre optickú moduláciu. Súbežne s tým, je aktívny aj UV laserový zdroj. V tomto prípade UV-C SpaceComm 254, ktorý vysiela svetlo s vlnovou dĺžkou 254 nm a výkonom 500 mW. Tento laser zatiaľ nenesie žiadne dáta, až Mach-Zehnder modulátor spojí tieto dva signály tak, že modulovaním intenzity svetla zakóduje do lúča pôvodné dáta. Modulovaný signál potom putuje do prvého FSO kanála, ktorý v simulácii predstavuje voľný priestor. Po prekonaní danej vzdialenosti narazí svetlo na odrazový prvok. Tento prvok simuluje reálny odraz od danej plochy, môže to byť napríklad stena, povrch satelitu alebo iný objekt. Odrazový prvok znižuje výkon signálu o konkrétnu hodnotu, čo reprezentujú straty pri odraze. Oslabený signál ďalej pokračuje cez druhý FSO kanál smerom k

prijímaču. Na strane prijímača je použitý fotodetektor, v tomto prípade lavínová fotodióda APD. APD je tu zvolená preto, že má vnútorné zosilnenie a dokáže detegovať aj veľmi slabý signál, čo je pri UV-FSO prenosoch po odraze kľúčové. Nakoniec prijatý elektrický signál vstupuje do analyzátora miery chybovosti BER (Bit Error Rate), ktorý zistí, koľko chýb vzniklo počas prenosu a aká je celková kvalita spojenia (Obr. 2).



Obr. 2 Návrh schémy pre prenos optických signálu v UV spektre s využitím FSO kanálov.

Útlm na trase sa bude meniť v závislosti od výkonu vysielača, počtu odrazových prvkov, útlmu odrazového materiálu, vzdialenosti medzi jednotlivými odrazovými prvkami, útlmom samotného FSO kanála, rozptylu lúča, priemerom apertúry vysielača a prijímača. Výsledkom je BER, ktorý ukáže, či by bol daný prenos v reálnych podmienkach spoľahlivý.

IV. Zhrnutie výsledkov simulácií

Na základe vykonaných simulácií v prostredí programu OptiSystem bolo možné analyzovať správanie FSO prenosov v troch rôznych ultrafialových spektrálnych oblastiach – UV-A (370 nm), UV-B (295 nm) a UV-C (254 nm).

A. Pôsobenie materiálov

Z pozorovania výsledkov (Tabuľka 3) môžeme konštatovať, že maximálna využiteľná odrazivosť v jednotlivých UV spektrách výrazne závisí od fyzikálnych vlastností materiálov a vlnovej dĺžky žiarenia. Pre UV-A pásmo sú vďaka dlhšej vlnovej dĺžke a nižšej absorpcii, aj bežné interiérové materiály, ako drevo, či betón schopné zabezpečiť spoľahlivý prenos, až po 4., či 5. odraz. V UV-B pásme sa účinnosť odrazu znižuje a len kvalitnejšie materiály s vysokou efektívnosťou odrazu, ako drevo alebo betón zvládnu 3–4 odrazy, zatiaľ čo slabšie reflektory, ako sklo či plast, rýchlo strácajú schopnosť prenosu a degradácia signálu je výrazná. V UV-C pásme pri simulácií vo vákuu sme nemali možnosť sledovať vplyv absorpcie na FSO kanál z dôvodu nulového útlmu. V UV-C spektre sú použiteľné len bežné materiály, teda vysoko odrazivé kovy, ako striebro alebo titán, ktoré umožňujú maximálne 3–4 úspešné odrazy bez výrazného zhoršenia kvality signálu. Táto závislosť podčiarkuje potrebu prispôsobiť výber odrazového materiálu konkrétnemu UV spektru a prevádzkovému prostrediu.

Tabuľka 3
Parametre FSO systémov využívané v simuláciách

Materiál	Odrazivosť	UV-A Max. odraz/BER	UV-B Max. odraz/BER	UV-C Max. odraz/BER
Drevo	30%	$5/1{,}9.10^{-91}$	$4/4{,}5.10^{-19}$	-
Betón	20%	$5/2{,}7.10^{-8}$	$3/7{,}8.10^{-16}$	-
Plast	10%	$3/3{,}4.10^{-14}$	$2/7{,}1.10^{-22}$	-
Sklo	4%	$2/9{,}7.10^{-9}$	$1/1{,}2.10^{-61}$	-
Striebro	98%	-	-	$4/4{,}8.10^{-11}$
Titán	85%	-	-	$4/4{,}4.10^{-6}$
Hliník	80%	-	-	$4/8{,}3.10^{-5}$
Zlato	60%	-	-	$3/1{,}9.10^{-10}$

B. Pôsobenie rôznych vlnových dĺžok

Skracovaním vlnovej dĺžky sa zvyšuje citlivosť systému na kvalitu odrazových materiálov, čo súvisí so silnejším rozptylom, nižšou priepustnosťou a vyššou absorpciou pri kratších UV vlnových dĺžkach. FSO systémy vo vákuu vykazujú vyššiu účinnosť prenosu aj na väčšie vzdialenosti, no sú extrémne citlivé na útlm po odrazoch. FSO systémy, ktoré pôsobia v interiéri sú ovplyvnené prachom, difúziou a Rayleighovým rozptylom, avšak pri správnom návrhu sú schopné zachovať použiteľný signál aj po viacerých odrazoch.

Správne nastavenie jednotlivých parametrov na komponentoch v simuláciách je kľúčové najmä pri simuláciách s viacerými odrazmi, kde sa optický výkon výrazne znižuje s každým ďalším odrazom. V takýchto prípadoch sa ukazuje, že príliš vysoký útlm materiálu alebo príliš veľký rozptyl lúča môže spôsobiť úplný výpadok prenosu. Vo všeobecnosti teda možno konštatovať, že výsledky každej simulácie sú úzko previazané s technickými parametrami jednotlivých komponentov systému a ich správna voľba je nevyhnutná pre dosiahnutie realistických a použiteľných výsledkov.

V. Záver

Cieľom práce bolo preskúmanie možnosti prenosu signálu prostredníctvom UV FSO systémov a zhodnotenie vplyvu vybraných parametrov, predovšetkým odrazových materiálov, na kvalitu prenosu v rôznych UV pásmach. Pomocou simulačného prostredia OptiSystem boli vytvorené modely, ktoré realisticky reprezentujú prenos v UV-C spektre medzi satelitmi vo vákuu, ako aj UV-A a UV-B komunikáciu v interiéri s prítomnosťou prekážok.

Výsledky simulácií preukázali, že schopnosť materiálu odraziť UV žiarenie zohráva kľúčovú úlohu v zabezpečení prijateľnej úrovne BER pri viacnásobných odrazoch. Z materiálov použitých pre UV-C sa najlepšie osvedčilo striebro, ktoré dokázalo zabezpečiť stabilný signál aj po štyroch odrazoch. Titánová fólia, leštený hliník a zlatá fólia boli spoľahlivé do troch odrazov (BER lepší ako 10^{-9}). V interiérovej simulácii s UV-B a UV-A žiarením, poskytlo najlepšie výsledky drevo, nasledované betónom. Materiály ako plast a sklo vykazovali vysokú stratu už pri prvých odrazoch a boli nevhodné pre viacnásobné odrazy.

Simulácie zároveň ukázali, že vo vákuu možno FSO systém považovať za kvázi LOS prenos, zatiaľ čo v interiéri sa signál šíri predovšetkým v NLOS režime pomocou difúznych odrazov. Výber vhodného fotodetektora, optimalizácia jeho citlivosti, ako aj správne nastavenie parametrov FSO kanála boli kľúčové pre dosiahnutie realistických výsledkov.

Prínosom práce bolo vytvorenie simulačného modelu pre testovanie FSO prenosov v UV spektre, ktorý možno ďalej rozširovať o pokročilejšie modely rozptylu, odrazové materiály, či rôzne druhy modulácií. Do budúcna by bolo vhodné preskúmať aj hybridné systémy kombinujúce UV a RF/FSO, ako aj praktické merania v reálnych podmienkach na overenie simulačných výstupov.

Poďakovanie

Táto práca bola financovaná EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00019 – „Kľúčové digitálne technológie v oblasti bezdrôtových širokopásmových odolných prenosových sietí“ a projektu č. 09I03-03-V05-00015 - „Využitie AI v oblasti bezdrôtových širokopásmových odolných prenosových sietí“.

Literatúra

[1] F. Salahdine, T. Han and N. Zhang, „5G, 6G, and Beyond: Recent advances and future challenges. Annals of Telecommunications“, 2023, Vol. 78, pp. 525–549. Dostupné z: https://doi.org/10.1007/s12243-022-00938-3.

[2] S. A. H. Mohsan, A. Mazinani and H. B. Sadiq, „A survey of optical wireless technologies: practical considerations, impairments, security issues and future research directions“, Optical and Quantum Electronics, 2022, Vol. 54, no. 187. Dostupné z: https://doi.org/10.1007/s11082-021-03442-5.

[3] R. Saiyyed, M. Sindhwani, S. Sachdeva and al., „Enabling high-speed and large-capacity data transmission in optical inter-satellite communication links under various conditions“, Journal of Optics, 2024. Dostupné z: https://doi.org/10.1007/s12596-024-02227-9.

[4] A, Zinecker, B. Spivey, G. Jayne, M. Hasegawa Jacobs, „Solar Absorptivity Degradation of Spacecraft Materials Due to UV and Charged Particles in the Gateway Environment“, Jacobs NASA, [online] 2022 [cit. 2025-03-02]. Dostupné z: https://ntrs.nasa.gov/citations/20220011070.

[5] A. Majumdar, „Non-Line-Of-Sight (NLOS) Ultraviolet And Indoor Free-Space Optical (FSO) Communications“, Advanced Free Space Optics (FSO), Springer Series in Optical Sciences, Vol. 186, New York: Springer, 2015. Dostupné z DOI: 10.1007/978-1-4939-0918-6_5.

[6] P. de Sternberg Stojalowski and J. Fairfoull, „Comparison of Reflective Properties of Materials Exposed to Ultraviolet-C Radiation“, Journal of Research of the National Institute of Standards and Technology, 2021, 126:126017. Dostupné z DOI: 10.6028/jres.126.017.

Inteligentný psí obojok s monitorovaním jeho vitálnych funkcií

[1]*Ľuboš OVSENÍK,* [2]*Jakub ORAVEC,* [3]*Zuzana LIŠČINSKÁ,* [4] *Daniel León LABANC*

[1,2,3,4]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]zuzana.liscinska@tuke.sk, [4]daniel.leon.labanc@student.tuke.sk

***Abstrakt* — Príspevok sa zaoberá návrhom a vývojom inteligentného obojku pre psy, ktorý by umožňoval monitorovanie ich základných vitálnych funkcií – konkrétne srdcovej frekvencie a telesnej teploty. Hlavným cieľom projektu bolo vytvoriť kompaktné a nositeľné zariadenie, ktoré by poskytovalo presné údaje o zdravotnom stave zvieraťa v reálnom čase a mohlo tak pomôcť majiteľom a veterinárom efektívne sledovať a vyhodnocovať jeho zdravotné zmeny.**

***Kľúčové slová* — inteligentný obojok, IoT, monitorovanie, pes, senzor, vitálne funkcie**

I. Úvod

Rozvoj internetu vecí (Internet of Things, IoT) viedol v posledných rokoch k výraznému nárastu inteligentných riešení určených na monitorovanie živých organizmov vrátane ľudí a zvierat. V oblasti starostlivosti o domácich miláčikov, najmä psov, sa čoraz častejšie uplatňujú technológie, ktoré umožňujú sledovanie ich zdravotného stavu, pohybu a správania v reálnom čase. Technológie pôvodne vyvinuté pre ľudí, ako sú nositeľné senzory, neinvazívne meranie vitálnych funkcií alebo vzdialený prenos dát, sa začali adaptovať aj pre potreby veterinárnej praxe a chovateľov. Vzhľadom na to, že komerčne dostupné riešenia monitorovania psov majú často obmedzenia ako je vysoká cena, obmedzená otvorenosť systému či nemožnosť úprav, cieľom tejto práce bolo navrhnúť a vytvoriť vlastný funkčný prototyp inteligentného obojka, ktorý by umožnil spoľahlivo a efektívne monitorovať vybrané vitálne parametre psa. Systém bol navrhnutý tak, aby umožňoval meranie srdcovej frekvencie a telesnej teploty, sledovanie polohy pomocou GPS a bezdrôtový prenos údajov do mobilného zariadenia cez Bluetooth Low Energy (BLE). Obojok bol realizovaný s dôrazom na kompaktnosť, energetickú efektívnosť, spoľahlivosť a vhodnosť pre každodenné použitie. Pri návrhu sa prihliadalo aj na komfort zvieraťa a bezpečné osadenie senzorov tak, aby neprekážali počas jeho pohybu. Zariadenie sme testovali v interiéri aj v exteriéri a jeho vývoj bol podporený vlastným návrhom puzdra, ktoré bolo vyrobené pomocou práškovej 3D tlače. Význam projektu spočíva najmä v možnosti včasného rozpoznania potenciálneho zdravotného problému zvieraťa, čím sa môže predísť jeho zhoršeniu alebo fatálnym následkom. Sledovanie polohy zároveň poskytuje možnosť rýchlej lokalizácie psa v prípade jeho straty, čím sa zvyšuje celková bezpečnosť.

II. Internet vecí (IoT)

Internet vecí predstavuje moderný technologický rámec, ktorý umožňuje digitálne prepojenie fyzických objektov s cieľom zdieľania údajov, ovládania zariadení a automatizovaného rozhodovania. IoT umožňuje objektom vnímať svoje okolie, komunikovať medzi sebou alebo s infraštruktúrou a vykonávať akcie bez zásahu človeka. Zariadenia v rámci IoT sú navrhnuté tak, aby boli schopné snímať údaje z prostredia, komunikovať medzi sebou navzájom, prípadne s cloudom, a na základe zozbieraných dát samostatne alebo riadene reagovať [1].

Rozmanitosť IoT zariadení odráža ich prispôsobivosť a univerzálnosť v rôznych oblastiach života. IoT zariadenia môžu byť veľmi rôznorodé: od jednoduchých senzorov teploty, cez inteligentné hodinky a domáce spotrebiče, až po zložité systémy na sledovanie dopravy alebo správy výrobných závodov. Vzhľadom na rastúce nároky na spoľahlivosť a rýchlosť reakcie sa IoT čoraz častejšie spolieha na lokálne spracovanie údajov (edge computing), ktoré znižuje záťaž na cloudové servery a zároveň umožňuje okamžité reakcie v prostredí, kde každá sekunda

rozhoduje – napríklad v autonómnych vozidlách alebo v zdravotníckych systémoch [1].

S rastúcim množstvom pripojených zariadení sa však zvyšuje aj riziko ich zneužitia. Rozšírenie IoT otvára cestu aj novým výzvam. Medzi najväčšie patrí bezpečnosť a súkromie používateľov. IoT zariadenia sú často zraniteľné voči útokom, keďže bežia na jednoduchom hardvéri bez silných šifrovacích schopností. Neautorizovaný prístup k dátam alebo ovládanie zariadení môže viesť k strate súkromia alebo poškodeniu majetku. Preto sa vyvíjajú špecializované bezpečnostné rámce, ako aj regulačné opatrenia (napr. GDPR) na ochranu používateľov [1].

V súčasnosti sa do popredia dostávajú technológie, ktoré posúvajú IoT na vyššiu úroveň autonómie a inteligencie. Dôležitým trendom v oblasti IoT je aj využitie umelej inteligencie, strojového učenia a edge computingu. Tieto technológie umožňujú lokálne spracovanie dát, autonómne rozhodovanie a znižujú latenciu aj zaťaženie sietí. Vďaka tomu môže byť reakcia systému okamžitá aj bez potreby cloudovej analytiky, čo je zásadné napríklad v zdravotníctve, autonómnej doprave alebo priemyselnom riadení [1].

IoT tak predstavuje nielen spojenie fyzického a digitálneho sveta, ale aj dynamickú platformu pre nové inovatívne služby. Očakáva sa, že do roku 2030 bude pripojených viac než 50 miliárd zariadení, ktoré budú generovať exabajty dát denne. Úspešná implementácia IoT bude preto závisieť od štandardizácie, interoperabilných platforiem, zabezpečenia a škálovateľnej infraštruktúry [2].

S týmto vývojom úzko súvisí aj rastúci význam tzv. inteligentných hrán (smart edges). S nástupom umelých neurónových sietí a rozšírením edge AI sa v IoT rozširuje pojem smart edges. Tieto zariadenia sú schopné samostatne rozhodovať, klasifikovať situácie, predikovať vývoj udalostí a minimalizovať potrebu prenosu dát do vzdialených serverov. Vďaka tomu sa znižuje latencia, zvyšuje bezpečnosť a zlepšuje odolnosť systémov voči výpadkom internetového pripojenia. Takéto systémy sa uplatňujú najmä v autonómnych vozidlách, robotike a zdravotníckych prístrojoch, kde oneskorenie v reakcii môže byť kritické [3].

Práve v kontexte urbanizácie a moderného mestského plánovania sa IoT ukazuje ako kľúčový nástroj pre rozvoj smart cities. Systémy pre inteligentné osvetlenie, adaptívne riadenie semaforov, senzory kvality ovzdušia či prediktívne systémy zberu odpadu, znižujú prevádzkové náklady miest a zároveň minimalizujú dopad na životné prostredie. IoT sa zároveň využíva aj na monitorovanie statiky budov, stavu kanalizácií, detekciu požiarov či zemetrasení, čím sa zvyšuje odolnosť infraštruktúry voči havarijným stavom [3].

Záverom možno povedať, že koncept digitálneho dvojčaťa (digital twin) predstavuje ďalší krok k integrácii fyzického a virtuálneho sveta. Tento prístup umožňuje vytvárať digitálne repliky fyzických objektov, procesov či dokonca celých miest. Tieto digitálne modely, napájané reálnymi dátami z IoT senzorov, slúžia na simulácie, optimalizácie a predikcie správania systémov. V priemysle sa digitálne dvojčatá používajú na optimalizáciu výroby, údržbu zariadení alebo školenie personálu. V kombinácii AI s IoT sa tak vytvárajú systémy schopné samostatného učenia, adaptácie a autonómneho riadenia [3].

III. Návrh riešenia systému

Architektúra navrhovaného systému bola koncipovaná ako modulárna a kompaktná, s dôrazom na zníženú spotrebu energie, jednoduchú údržbu a možnosť ďalšieho rozširovania. Celý systém je postavený na jednom centrálne riadiacom prvku (ESP32), ku ktorému sú pripojené jednotlivé periférne komponenty.

A. Funkčné požiadavky a ciele návrhu

Hlavným cieľom návrhu bolo vytvoriť inteligentné a nositeľné zariadenie v podobe obojku pre psa, ktoré bude schopné monitorovať jeho základné vitálne funkcie v reálnom čase. Primárnymi meranými údajmi sú srdcová frekvencia a telesná teplota. Na meranie pulzu bol zvolený optický senzor typu PPG, konkrétne model MAX30102, ktorý využíva infračervené a červené LED diódy spolu s fotodiódou na zaznamenanie zmien objemu krvi v kapilárach. Tento senzor poskytuje vysokú presnosť merania a vďaka kompaktnému integrovanému puzdru je vhodný na zabudovanie do nositeľných riešení [4].

Druhou sledovanou veličinou je telesná teplota, ktorá sa zisťuje pomocou bezkontaktného infračerveného senzora MLX90614. Tento senzor umožňuje meranie objektovej teploty bez nutnosti priameho kontaktu, čo je ideálne pri aplikácii na psa, kde sa zohľadňuje srsť a pohyb zvieraťa. Meranie prebieha v krátkych intervaloch, pričom výstupy sú presné a stabilné aj pri meniacich sa podmienkach prostredia [5].

Dôležitým aspektom návrhu bola komunikácia medzi obojkom a externým zariadením, napr. smartfónom. Na tento účel bol zvolený BLE (Bluetooth Low Energy) modul, ktorý je integrovaný v mikrokontroléri ESP32. BLE zabezpečuje obojsmerný prenos dát pri nízkej spotrebe energie a

umožňuje jednoduché napojenie na bežné mobilné aplikácie ako napríklad LightBlue. Prenášané údaje sú vizualizované vo forme notifikácií, čím je zabezpečený rýchly prístup k informáciám bez potreby trvalého internetového pripojenia. Zariadenie je teda plne funkčné aj v offline režime, čo je výhodné najmä pri vonkajšej aktivite alebo na miestach bez sieťového pokrytia [6].

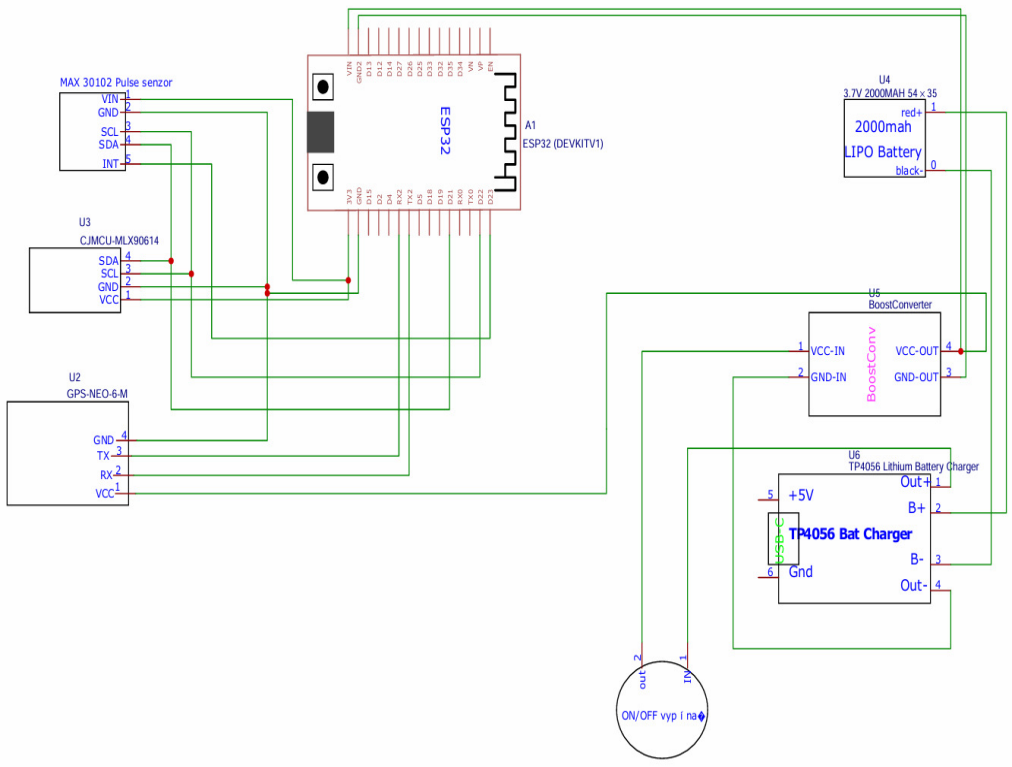
Z pohľadu ergonómie a komfortu bolo dôležité, aby celé zariadenie bolo ľahké, kompaktné a jednoducho nositeľné. Komponenty sú preto integrované do malého puzdra, ktoré je možné uchytiť priamo na bežný psí obojok. Dizajn zohľadňuje aj pohybovú aktivitu zvieraťa, t.j. zariadenie nesmie obmedzovať pohyb, spôsobiť odreniny alebo byť citlivé na nárazy. Materiály musia byť zdravotne nezávadné a odolné voči vonkajším vplyvom [6].

Napájanie zariadenia je riešené pomocou Li-Ion batérie s kapacitou 3000 mAh, ktorá zabezpečuje dlhú výdrž pri zachovaní kompaktných rozmerov. Batéria je dobíjaná pomocou modulu TP4056 s ochranou proti prebitiu a pod-vybitiu. Cieľom bolo navrhnúť systém, ktorý vydrží niekoľko dní v prevádzke bez nutnosti častého nabíjania a ktorý umožní jednoduchú obsluhu používateľom, vrátane nabíjania cez bežný USB-C konektor [6].

Pri návrhu bola taktiež kladená pozornosť na rozšíriteľnosť systému. Modulárna architektúra umožňuje budúce doplnenie ďalších senzorov, napr. na monitorovanie aktivity, vlhkosti alebo GPS sledovanie. Dôležité je, že zariadenie má fungovať spoľahlivo aj pri absencii dátového pripojenia a súčasne zabezpečuje nízku spotrebu energie vďaka režimom spánku mikrokontroléra. Tým sa dosahuje predĺženie životnosti batérie a zároveň sa zabezpečuje spoľahlivá činnosť systému v reálnych podmienkach [4].

B. Schéma systému

Schéma systému (Obr. 1) jasne vyjadruje smer toku údajov: od senzorov (MAX30102 a MLX90614), ktoré sú pripojené prostredníctvom I2C zbernice, do centrálneho mikrokontroléra ESP32. Ten zabezpečuje zber údajov v pravidelných intervaloch, aplikuje jednoduché filtre na odstránenie šumu a vykonáva výpočty potrebné na stabilizáciu výstupu. Po spracovaní sú získané hodnoty prenášané pomocou BLE profilu ako notifikácie do mobilného zariadenia používateľa. Táto komunikácia prebieha cez štandardný GATT profil, pričom každá hodnota je sprístupnená ako charakteristika so zapnutou funkciou notifikácie. Používateľ tak v aplikácii LightBlue Explorer okamžite vidí aktuálne hodnoty tepu a teploty. Napájanie systému zabezpečuje batéria Li-Ion s kapacitou 3000 mAh, ktorá je chránená a dobíjaná cez modul TP4056. Výstupné napätie z batérie je zvýšené pomocou meniča MT3608 na potrebných 5 V, ktoré napájajú ESP32 aj senzory. Do napájacieho obvodu bol zaradený aj vypínač, ktorý umožňuje fyzické odpojenie systému a LED dióda ako signalizačný prvok o stave napájania. Takto navrhnuté logické prepojenie zabezpečuje stabilnú a spoľahlivú prevádzku všetkých komponentov.



Obr. 1 Návrh schémy zapojenia inteligentného obojka.

C. Komunikačný model s BLE

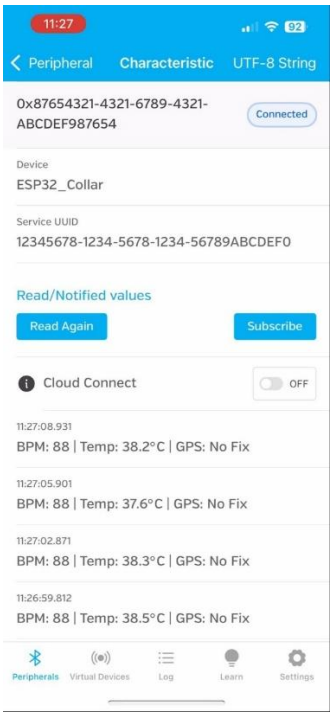
Komunikácia medzi zariadením a mobilným telefónom prebieha pomocou technológie Bluetooth Low Energy (BLE). Táto technológia bola zvolená najmä z dôvodu nízkej spotreby energie, čo je kľúčové pre zariadenia napájané batériou, a tiež pre jej jednoduchú integráciu do mobilných aplikácií bez nutnosti internetu [7].

GATT profil

Zariadenie sa po zapnutí správa ako BLE periféria, ktorá vysiela *advertising* pakety. Po pripojení mobilného zariadenia (centrála) sa aktivuje *GATT* (Generic Attribute Profile), v ktorom sú definované [*20*]:

- Služba (Service): napríklad *Health Monitoring Service*,
- Charakteristiky (Characteristics):
  - *Heart Rate* – obsahuje poslednú hodnotu pulzu,
  - *Body Temperature* – obsahuje poslednú hodnotu teploty.

Obidve charakteristiky sú nastavené na *Notify*, čo znamená, že pri každej zmene hodnoty je odoslaný automaticky nový údaj do aplikácie.

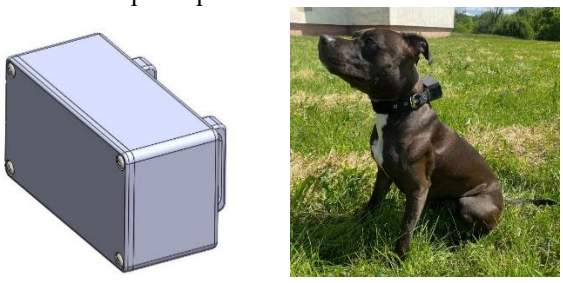


Obr. 2 Aplikácia Light Blue.

Týmto spôsobom sa dosiahlo jednoduché a efektívne prepojenie medzi hardvérom a mobilným používateľom bez potreby zložitej softvérovej infraštruktúry (Obr. 2). Komunikačný model zároveň umožňuje budúce rozšírenie napríklad o možnosť zapisovania parametrov alebo konfigurácie priamo z aplikácie [7].

D. Konštrukcia puzdra a uchytenie na obojok

Na základe rozmerov jednotlivých komponentov bol v programe SolidWorks vytvorený 3D model puzdra (Obr. 3), ktorý bol následne vytlačený pomocou práškovej 3D tlače. Puzdro bolo navrhnuté na mieru pre tento projekt, pričom rešpektovalo potrebu kompaktnosti, ochrany súčiastok a zároveň jednoduchého prístupu k senzorom a konektorom.



Obr. 3 Návrh puzdra a uchytenie na obojku.

Charakteristika puzdra:

- *Materiál*: technický polymér vhodný na práškovú SLS tlač.
- *Upevnenie*: spodná strana puzdra obsahuje úchyty na prevlečenie obojku, čím je zabezpečená stabilita a pohodlné nosenie.
- *Umiestnenie senzorov*: káble zo senzorov tepu a teploty prechádzajú cez výrezy v spodnej časti priamo do priestoru medzi obojok a telo psa.
- *Tesnenie*: medzi vrchným a spodným dielom je miesto na gumové tesnenie, ktoré znižuje prenikanie prachu a vlhkosti do puzdra.
- *Krytie*: puzdro je uzavreté pomocou 4 skrutiek, čím sa zabezpečuje pevné uchytenie a možnosť budúcej údržby alebo výmeny komponentov.
- *Pripojenie a ovládanie*: na vonkajšej strane puzdra sa nachádza USB-C port na nabíjanie a mechanický prepínač (On/Off).

IV. Zhrnutie testovania inteligentného obojku v reálnych podmienkach

Po dokončení návrhu a hardvérovej implementácii nasledovala fáza testovania, ktorej cieľom bolo overiť praktickú funkčnosť celého riešenia, spoľahlivosť merania fyziologických parametrov psa a stabilitu bezdrôtového prenosu dát.

A. Meranie teploty a srdcovej frekvencie

Počas testovania sa podarilo úspešne otestovať meranie telesnej teploty aj srdcovej frekvencie v reálnych podmienkach. Senzor MAX30102 na meranie tepu bol umiestnený tak, aby mal priamy kontakt s pokožkou psa v oblasti spodného krku. MLX90614 bol umiestnený s miernym sklonom smerom na telo, vo vzdialenosti niekoľkých milimetrov až centimetrov.

Merania prebiehali v intervaloch každé 3 sekundy a boli prenášané prostredníctvom Bluetooth do mobilnej aplikácie LightBlue (Obr. 2). Pokiaľ bol obojok stabilne nasadený a senzory držali na mieste, zariadenie poskytovalo konzistentné a reálne hodnoty, ktoré korešpondovali s očakávaným fyziologickým rozsahom.

Pri pohybe psa alebo pri zosunutí obojka však senzory stratili kontakt s vhodným meracím miestom, čo spôsobovalo nepresné alebo nulové výstupy. Tento jav je pri optických senzoroch očakávaný, keďže presnosť merania závisí od kvality kontaktu so snímanou plochou. Aj napriek týmto obmedzeniam možno konštatovať, že v dobrej polohe boli oba senzory funkčné a použiteľné na reálne monitorovanie vitálnych funkcií.

B. Komunikačná stabilita a výdrž

Bluetooth Low Energy (BLE) spojenie medzi ESP32 a mobilnou aplikáciou LightBlue bolo počas testov stabilné a bez výpadkov, pokiaľ sa používateľ nachádzal v štandardnom dosahu do 10 metrov. Údaje sa odosielali každé 3 sekundy a zobrazovali sa v reálnom čase.

Napájanie systému zabezpečovala Li-Ion batéria s kapacitou 3000 mAh, ktorá bola pripojená cez nabíjací modul TP4056 s ochranou proti prebitiu a pod-vybitiu. Nabíjanie cez USB-C port prebehlo bez problémov a systém sa správne aktivoval po pripojení alebo zapnutí cez mechanický vypínač, ktorý tiež pracoval spoľahlivo.

Výdrž batérie na základe meraní a výpočtov odhadujeme na približne 1,5 dňa pri odosielaní údajov každé 3 sekundy. Pre dlhodobejšie nasadenie by bolo vhodné zvoliť batériu s vyššou kapacitou (napr. 5000 mAh).

Počas vývoja bola zistená chyba pri napájaní GPS modulu NEO-6M, ktorý bol pôvodne napojený na 3.3 V vetvu, čo nestačilo na stabilný príjem signálu. Modul bol následne prepojený na 5 V, no jeho funkčnosť sa už nepodarilo overiť, pretože po výmene microUSB za USB-C port sa mikrokontrolér začal hlásiť ako úložisko namiesto sériového portu, čo znemožnilo ďalšie programovanie a testovanie.

C. Zhrnutie testovania

Testovanie prototypu preukázalo, že systém je schopný reálne monitorovať základné vitálne funkcie psa, pokiaľ sú splnené určité podmienky, najmä stabilné nasadenie, dostatočný kontakt senzorov s telom a správna konfigurácia komponentov:

- meranie telesnej teploty a tepu fungovalo spoľahlivo pri správnom umiestnení senzorov,
- bluetooth komunikácia prebiehala bez problémov v rámci bežného dosahu,
- nabíjanie batérie a systém vypínania boli funkčné a spoľahlivé,
- puzdro preukázalo mechanickú odolnosť a stabilitu.

GPS modul sa nepodarilo otestovať vo finálnom zariadení kvôli chybe s USB-C portom, ktorá znemožnila jeho naprogramovanie. Tento nedostatok predstavuje najväčšie obmedzenie projektu,

no zároveň vytvára priestor pre jeho ďalšie zlepšenie.

Výsledky testovania potvrdili, že navrhnutý obojok je plne funkčný prototyp s potenciálom pre ďalší vývoj. Zariadenie umožňuje základné fyziologické monitorovanie v reálnom čase a predstavuje prínosný nástroj pre sledovanie zdravotného stavu zvieraťa.

V. Záver

Táto práca sa zaoberala návrhom, vývojom a testovaním inteligentného obojku pre psa, ktorý umožňuje monitorovanie vybraných vitálnych funkcií, t.j. srdcovej frekvencie a telesnej teploty v reálnom čase. Riešenie je postavené na platforme ESP32 s bezdrôtovým prenosom údajov cez Bluetooth Low Energy a integruje senzory MAX30102 a MLX90614. Cieľom bolo navrhnúť zariadenie, ktoré je kompaktné, prenosné, dostatočne presné a zároveň jednoducho rozšíriteľné do budúcnosti.

Na základe teoretického štúdia oblasti IoT, bezdrôtových senzorových sietí a nositeľnej elektroniky pre zvieratá bol navrhnutý systém, ktorý zohľadňuje požiadavky na energetickú efektívnosť, použiteľnosť v praxi a komfort pre zviera. V rámci vývoja bola navrhnutá kompletná architektúra systému, vybrané komponenty, vytvorený funkčný firmvér a vyrobené vlastné puzdro pomocou 3D tlače.

Počas testovania v reálnych podmienkach sa potvrdila funkčnosť systému, meranie tepu a teploty prebiehalo korektne s aktualizáciou každé 3 sekundy a prenosom údajov do mobilnej aplikácie. Ukázalo sa, že zariadenie funguje spoľahlivo, pokiaľ sú splnené podmienky pre stabilné osadenie senzorov. Komunikačný modul BLE, napájací systém aj mechanická konštrukcia boli funkčné a preukázali svoju praktickú použiteľnosť.

Hlavným obmedzením prototypu bola nefunkčnosť GPS modulu vo finálnej verzii zariadenia v dôsledku chyby pri výmene konektora, ktorá znemožnila ďalšie programovanie mikrokontroléra. Táto skutočnosť neumožnila úplné otestovanie lokalizačnej funkcionality, ktorá bola pôvodne súčasťou návrhu.

Medzi hlavné prínosy práce patrí návrh a čiastočne overená implementácia systému pre základné monitorovanie zdravotného stavu psa, ktorý predstavuje funkčný základ pre ďalší vývoj. Systém preukázal technickú životaschopnosť riešenia a zároveň identifikoval konkrétne miesta pre optimalizáciu, zmenšenie rozmerov, stabilnejšie uchytenie na krk, možnosť rozšírenia o SIM kartu a ďalšie senzory, ako aj vývoj vlastnej aplikácie.

Ďalší vývoj by sa mal zamerať najmä na:

- vyriešenie problémov s napájaním a prenosom dát z GPS modulu,
- návrh vlastnej DPS a miniaturizáciu systému,
- spoľahlivú lokalizáciu nezávislú od BLE dosahu (napr. pomocou GSM/GPS),
- zlepšenie ergonómie a protišmykovosti obojku,
- rozšírenie o ďalšie vitálne funkcie, napríklad monitorovanie stresu, aktivity, respiračnej frekvencie či nasadenie algoritmov pre predikciu a detekciu zdravotných rizík.

Výsledky a skúsenosti získané pri realizácii tejto práce poskytujú pevný základ pre ďalší výskum a vývoj v oblasti veterinárneho IoT a ukazujú, že aj jednoduchý prototyp dokáže priniesť reálne využiteľné informácie pre zdravie zvieraťa.

Poďakovanie

Táto práca bola financovaná EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00019 – „Kľúčové digitálne technológie v oblasti bezdrôtových širokopásmových odolných prenosových sietí“ a KEGA grantu 041TUKE-4/2025 – „Laboratórium na výskum a vývoj pokročilých IoT senzorových sietí a širokopásmových senzorov“.

Literatúra

[1] M. Miller, „Internet of Things“, 2015, 1st Edition, Kindle Edition. Dostupné z: ISBN 978-0134021331.
[2] A. Minteer, „Analytics for the Internet of Things (IoT)“, 2017, 1st Edition, Packt Edition. Dostupné z: ISBN 978-1787120730.
[3] A. Tamboli, „Build Your Own IoT Platform“, 2019, 1st Edition, Apress Edition. Dostupné z: ISBN 978-1484244974.
[4] Analog Devices. MAX30102: High-Sensitivity Pulse Oximeter and Heart-Rate Sensor. [cit. 2025-03-02]. Dostupné z: https://www.analog.com/media/en/technical-documentation/data-sheets/max30102.pdf.
[5] MLX90614: Infrared Thermometer for Non-Contact Temperature Measurements. [cit. 2025-02-18]. Dostupné z: https://www.melexis.com/-/media/files/documents/datasheets/mlx90614-datasheet-melexis.pdf
[6] Adafruit Inustries Powering wearables with li-Ion Batteries. [cit. 2024-12-10]. Dostupné z: https://www.adafruit.com/category/574
[7] R. Heydon, „Bluetooth Low Energy: The Developer's Handbook“, 2012, Prentice Hall. Dostupné z: ISBN: 978-0132888368.

Využitie dronov a umelej inteligencie v bezpečnostných systémoch

[1]*Ján BAČÍK*

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

[1]jan.bacik.2@tuke.sk

***Abstract* — Súčasná bezpečnostná situácia je čoraz viac ohrozovaná hybridnými hrozbami, ktoré si vyžadujú detekciu nebezpečenstiev v reálnom čase. V tomto článku sme predstavili integrovaný monitorovací rámec, ktorý kombinuje BVLOS letové operácie dronov s objektovou detekciou využívajúcou modely YOLO. Takto umožňujeme spoľahlivú identifikáciu ozbrojených osôb v okolí kritickej infraštruktúry. Na trénovanie detekčného modelu sme vytvorili experimentálny dataset prispôsobený špecifikám sledovaného prostredia.**

***Keywords* — dataset, BVLOS, drony, YOLO, AI**

I. ÚVOD

Od začiatku ozbrojeného konfliktu na Ukrajine sa bezpečnostná situácia v Európe zmenila z tradičných vojenských stretov na sofistikované hybridné operácie, v ktorých sa kybernetické útoky, dezinformačné kampane a bezpilotné lietajúce systémy spájajú, aby podkopávali stabilitu spoločnosti aj mimo územia konfliktu. Najmä komerčne dostupné aj špecializované drony s FPV (First-Person View) prenosom videa sa rozšírili nielen na fronte, ale vyvolali aj značné obavy v oblastiach vzdialených od bojových zón.

Namiesto priameho vedenia vojenských akcií sa protivníci čoraz častejšie zameriavajú na civilné oblasti života prostredníctvom koordinovaných hrozieb, od anonymných bombových poplachov zasielaných e-mailom až po falošné hlásenia o ozbrojených vniknutiach v blízkosti verejných inštitúcií. Školy a univerzity sa stávajú častým terčom nútených evakuácií, uzatváraní a nasadzovania zásahových tímov, čo so sebou prináša vysoké finančné nároky a zároveň zaneprázdňuje bezpečnostné zložky, ktoré môžu byť v daný moment potrebné na inom mieste.

V súčasnosti proaktívnu obranu pomáhajú zabezpečiť moderné metódy umelej inteligencie. Real-time analýza videozáznamu využívajúca pokročilé siete YOLO (You Only Look Once) spája rýchlosť spracovania s vysokou presnosťou detekcie. Najnovšia séria modelov YOLO11 (varianty n, s, m, l, x) umožňuje flexibilné prispôsobenie na špecifické scenáre. Nasadením týchto modelov na drony v režime BVLOS a použitím doménovo špecifického datasetu môže náš systém autonómne identifikovať ozbrojené osoby, čo výrazne skráti reakčný čas na potencionálne riziká a umožní lepšiu koordináciu bezpečnostných zložiek.

II. SÚVISIACE PRÁCE

Detekcia objektov je kľúčová disciplína s univerzálnym uplatnením v zdravotníctve, bezpečnosti, priemysle či ďalších odvetviach. Vďaka moderným GPU (Graphics Processing Unit) a CPU (Central Processing Unit) je možné nasadiť náročné modely strojového učenia v reálnom čase bez výrazných obmedzení výpočtového výkonu.

Zhang et al. [1] demonštrovali, že integrácia YOLO do autonómnej inšpekcie lietadiel umožňuje dronu oblietať automaticky okolo trupu lietadla a zároveň prenášať video na pozemnú stanicu na následné spracovanie. Tým sa eliminuje potreba výkonného onboard hardvéru priamo umiestneného na drone. Absencia kvalifikovanej obsluhy na druhej strane znižuje náklady na samotnú inšpekciu. Di et al. [2] využili drony na detekciu poškodení elektrických vedení, čím výrazne znížili riziká a náklady spojené s tradičnými manuálnymi inšpekciami. Do et al. [3]

ponúkli metódu detekcie osôb z obrazu zhotoveného z výšky pomocou dronu, dosahujúc približne 90% mAP@0,5 na datasete Human Drones, čo svedčí o robustnosti tohto prístupu aj pri malých objektoch.

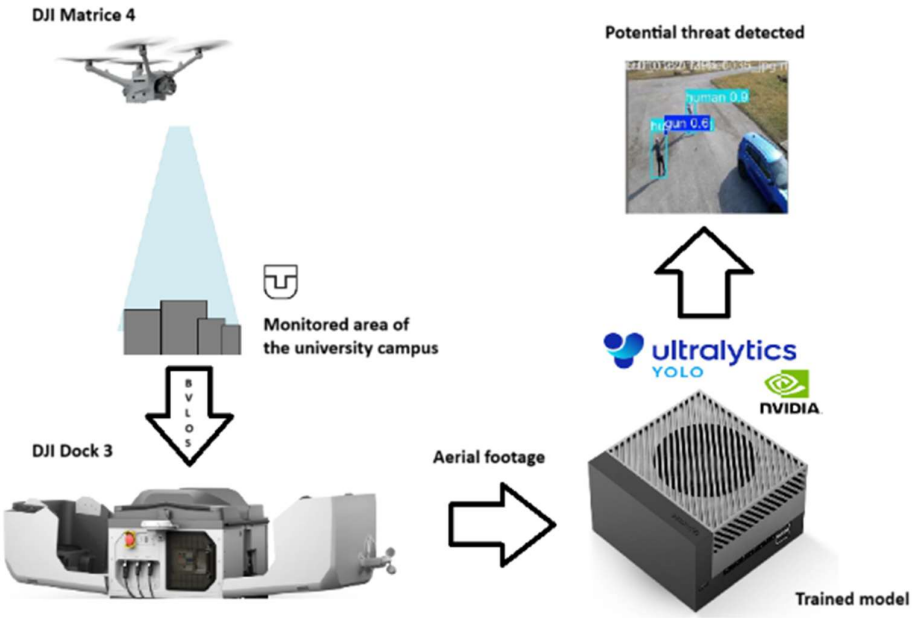
V prostredí letísk a odbavovania cestujúcich, prinášajú automatizované systémy bezpečnostnej kontroly batožiny dramatické zrýchlenie oproti ručnému vyhodnocovaniu röntgenových snímok batožiny. Kundilokovit et al. [4] porovnali architektúry CNN, R-CNN, Detectron, RetinaNet a YOLO pri detekcii nebezpečných predmetov v batožine. Mimoriadnu presnosť YOLO pri rozpoznávaní zbraní potvrdili aj Sumi et al. [5] a Chitravanshi et al. [6].

III. Koncept

Kľúčovým prvkom nášho konceptu sú operácie Beyond Visual Line-Of-Sight (BVLOS). Na rozdiel od letu v rámci pravidiel VLOS, pri ktorom operátor neustále udržiava vizuálny kontakt s dronom, BVLOS umožňuje rozšíriť operačný priestor za hranice priamej viditeľnosti operátora. Vďaka špecializovaným komunikačným protokolom, redundantným GNSS modulom a automatizovaným systémom detekcie prekážok dokáže autonómny systém monitorovať oblasti, v ktorých sa môže pohybovať potencionálna hrozba, ozbrojená osoba, ktorá by mohla pri operácii VLOS ohroziť operátora drona.

V roku 2025 spoločnosť DJI uviedlo na trh dron Matrice 4 spolu s pozemnou stanicou DJI Dock 3, ktoré predstavujú komplexné riešenie navrhnuté priamo pre bezpečné BVLOS misie. Matrice 4 sa vyznačuje letovou obálkou s dĺžkou desiatok kilometrov, pričom Dock 3 poskytuje plne automatizované odletové a pristávacie procedúry, stabilné pripojenie cez 4G/5G sieť a šifrované video-streamovanie do riadiacej stanice.

Náš bezpečnostný koncept (obr. 1) kombinuje platformu DJI Matrice 4 s pozemnou stanicou DJI Dock 3 na zabezpečenie autonómnych BVLOS operácií a prenosu videostreamu. Z dronu je real-time video odosielané na pozemnú stanicu, kde je prijaté výpočtovým modulom NVIDIA Jetson AGX Orin 64 GB. Na tomto module, ktorý pri spotrebe len 60 W dosahuje až 275 TOPS AI výkonu, prebieha offboard detekcia ozbrojených osôb pomocou vlastných modelov vytrénovaných s využitím knižníc Ultralytics YOLO. Táto metóda spája hotové priemyselné komponenty s vlastnými AI riešeniami, čo umožňuje rapid prototyping, nízku latenciu a vysokú spoľahlivosť detekcie v reálnom čase.



Obr. 1 Schéma bezpečnostného konceptu s DJI Matrice 4, Dock 3 a NVidia Jetson AGX

Naša metodika tréningu modelu pozostáva z troch hlavných fáz:

Zostavenie dátovej množiny

- Zhromaždenie doménovo špecifických záberov z rôznych letových výšok, uhlov a poveternostných podmienok. Údaje zahŕňajú scény s civilnou infraštruktúrou, lesnatými úsekmi aj mestskými oblasťami, aby bol model dostatočne robustný pri nasadení v rozmanitom prostredí.

Anotačný proces

- Použitie špecializovaného anotátorského nástroja na označenie tried ako „osoba“, „ozbrojená osoba“ či „potenciálna hrozba“. Každý obrázok prechádza dvojitou kontrolou kvality: najprv ručným označením, potom automatizovanou validáciou na základe predtrénovaného modelu na podobnej doméne.

Tréning a doladenie modelu

- Fine-tuning vybraného variantu YOLO11 (m) na GPU klastroch s batch size 16 a adaptívnou lr schedulou po dobu 200 epôch. Počiatočné váhy modelu pochádzajú z pretrénovania na veľkej verejnej množine, čo skráti konvergenčný čas. V každej fáze sa sledujú metriky presnosti (precision), citlivosti (recall) a priemerná presnosť ($mAP_{50–95}$), aby bol zaistený optimálny kompromis medzi rýchlosťou inferencie a spoľahlivosťou detekcie.

IV. TRÉNOVANIE MODELU

Hodnotili sme výkonnosť troch variantov YOLO 11 (n, m a x) na našom vlastnom datasete. Všetky modely boli trénované 200 epoch s veľkosťou batchu 16 na GPU NVIDIA A100-SXM4-40 GB v prostredí Google Colab, čo zabezpečilo rýchlu konvergenciu a vysokú výpočtovú efektivitu. Po trénovaní sme porovnali ich výsledky, aby sme identifikovali architektúru s najlepšími výsledkami pre ďalšiu optimalizáciu a nasadenie.

Pre komplexné hodnotenie detekčnej schopnosti každého modelu sme vypočítali tri kľúčové metriky:

Presnosť (Precision)

- udáva pomer správne predpovedaných pozitívnych detekcií, pričom penalizuje falošné poplachy:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

– *TP* (True Positives) = správne detegované objekty
– *FP* (False Positives) = nesprávne detegované objekty

Citlivosť (Recall)

- zachytáva schopnosť modelu nájsť všetky relevantné objekty, pričom penalizuje neodhalené detekcie:

$$Recall = \frac{TP}{TP+F} \tag{2}$$

– *FN* (False Negatives) = objekty prítomné, ale nezachytené

Poslednou hodnotenou metrikou v tejto štúdii je $mAP_{50–95}$, ktorá poskytuje komplexné zhodnotenie modelu na základe analýzy jeho presnosti aj citlivosti pri rôznych prahoch IoU (Intersection over Union) v rozsahu od 50 % do 95 %. Analýzou týchto metrík pre každú variantu YOLO 11 sme určili, ktorá konfigurácia najlepšie vyvažuje rýchlosť inferencie a presnosť detekcie pre real-time BVLOS operácie.

V. VÝSLEDKY

Ako už bolo uvedené, trénovali sme tri varianty modelov YOLO: n, m a x. Ako je zrejmé z Tabuľky 1, najrýchlejší je variant n, dosahujúci inferenčný čas 0,7 ms s $mAP_{50–95}$ na úrovni 0,603. Tento model je optimálnou voľbou pre aplikácie, kde je prioritou rýchlosť. Naopak variant m dosiahol najvyššiu hodnotu $mAP_{50–95}$, konkrétne 0,675, čím prekonal variant x z hľadiska presnosti. Na základe týchto výsledkov sa variant m javí ako najvhodnejší pre aplikácie uprednostňujúce presnosť, pri zachovaní relatívne rýchlej inferencie.

Tabuľka 1
Porovnanie výkonu modelov s rôznymi variantmi YOLO11

Variant	All classes			
	Inference speed	Precision	Recall	mAP50-95
n	0.7 ms	0.864	0.756	0.603
m	2.4 ms	0.858	0.806	0.675
x	4.9 ms	0.879	0.799	0.645

Napriek celkovo dobrým výsledkom je stále priestor na zlepšenie. Obrázok 2 ukazuje, že model pri detekcii osôb dosahuje vysoké dôvery hodnostné skóre, no pri rozpoznávaní strelných zbraní sú tieto skóre podstatne nižšie. Tento rozdiel pravdepodobne pramení z nevyváženosti tried v trénovacej množine (menej príkladov zbraní) a nedostatočného pokrytia reálnych variantov zbraní vo videách. Ďalej malá veľkosť zbraní na dronových snímkach zvyšuje náročnosť ich detekcie, keďže neurónové siete majú tendenciu horšie identifikovať menšie objekty.

Obr. 2 Inferencia modelu: (a) vonkajšie prostredie, (b) vnútorné prostredie.F9P

VI. ZÁVER

V posledných rokoch sa drony v spojení s počítačovým videním založeným na sieťach YOLO ukázali ako silný nástroj proaktívnej bezpečnosti, najmä v operáciách Beyond Visual Line-Of-Sight (BVLOS). Naša štúdia potvrdzuje, že pri podpore robustnej dátovej pipeline — ktorá obsahuje vlastný dataset pre vnútorné i vonkajšie prostredie, precíznu anotáciu a cielenú augmentáciu — môžu modely YOLO excelovať v leteckej detekcii objektov. Variant YOLO-m dosiahol presnosť 0,858, citlivosť 0,806 a $mAP_{50–95}$ na úrovni 0,675 pri inferencii len 2,4 ms, čo zdôrazňuje jeho vhodnosť pre monitorovanie v reálnom čase. Ďalšia práca by mala skúmať pokročilé techniky augmentácie a optimalizované protokoly tréningu s cieľom zlepšiť detekciu malých či čiastočne zakrytých objektov v real-time dronových záberoch.

POĎAKOVANIE

Článok bol publikovaný s podporou projektu APVV-19-0210.

Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00018.

REFERENCIE

[1] D. Zhang et al., Research on Aircraft Patrol Inspection Method Using UAV Based on YOLO Algorithm, proceedings of the 4th International Conference on Electronic Information Engineering and Computer (EIECT), Shenzhen, China, 2024.
[2] T. Di et al., Research on Real-Time Power Line Damage Detection Method Based on YOLO Algorithm, proceedings of the 3rd IEEE International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 2023.
[3] M.-T. Do et al., Human Detection Based Yolo Backbones-Transformer in UAVs, proceedings of the International Conference on System Science and Engineering (ICSSE), Ho Chi Minh, Vietnam, 2023.
[4] P. Kundilokovit et al., Weapon Detection in X-ray Image of Baggages, proceedings of the 23rd International Symposium on Communications and Information Technologies (ISCIT), Bangkok, Thailand, 2024.
[5] L. Sumi et al., A Deep Learning Approach to Gun Detection in Surveillance System Using YOLOv9, proceedings of the Control Instrumentation System Conference (CISCON), Manipal, India, 2024.
[6] D. Chitravanshi et al., Weapon Detection from Images Using YOLO and OpenCV, proceedings of the 1st International Conference on Technological Innovations and Advance Computing (TIACOMP), Bali, Indonesia, 2024.

Simulácia vplyvu defektov na distribúciu tepla vo veľkoplošných spojoch výkonových modulov

[1]*Peter NEMERGUT,* [2]*Alena PIETRIKOVÁ*

[1,2] Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovensko

[1]peter.nemergut@tuke.sk, [2]alena.pietrikova@tuke.sk

***Abstrakt* — Tento článok sa zaoberá simuláciou vplyvu defektov vo veľkoplošných spojoch výkonových modulov na tepelný manažment. Simulácie boli vyhotovené v softvéri Solid Edge 2024 Student Edition. Simulačné modely obsahovali spájkovaciu zliatinu 96,5Sn3,5Ag s rôznym množstvom a polohou defektov. Cieľom článku je analyzovať vplyv defektov medzi DBC substrátom a základovou doskou na tepelný odpor štruktúry z pohľadu ich množstva a polohy v rámci plochy spoja. Počas simulácií boli zaznamenávané teploty na silikónovom čipe a spodnej strane základovej dosky priamo pod čipom. Z týchto teplôt bol následne vypočítaný tepelný odpor štruktúry. Výsledky ukazujú, že pri zachovaní rovnakých simulačných podmienok narastá tepelný odpor štruktúry výkonového modulu s narastajúcim množstvom defektov v spoji. Kritická situácia nastáva najmä vtedy, keď sú defekty sústredené priamo pod čipom, ktorý je hlavným zdrojom tepla.**

***Kľúčové slová* — výkonový modul, veľkoplošné spoje, tepelný manažment, Solid Edge, simulácie**

I. Úvod

Problémy spojené s tepelným manažmentom (obzvlášť v oblasti výkonovej elektroniky) sú kritické najmä v oblasti veľkoplošných spojov [1], [2]. Tepelný manažment je kľúčovým faktorom, pretože efektívne riadenie tepla zohráva dôležitú úlohu pri zvyšovaní spoľahlivosti a výkonu zariadení [3], [4], [5]. Vo výkonových aplikáciách, ako sú napájacie zdroje, meniče alebo elektromotory, sú komponenty vystavené vysokým tepelným zaťaženiam, ktoré vznikajú pri prevádzke. Efektívny odvod tepla patrí medzi kľúčové aspekty tepelného manažmentu, pretože jeho nedostatočnosť môže viesť k defektom ako sú delaminácia, prehnutia vplyvom rozdielnych koeficientov tepelnej rozťažnosti [6], [7], vzniku voidov (prázdnych vzduchových miest vo vnútri spájkovaného spoja) alebo zhoršenie kvality spojov medzi komponentmi. Tieto problémy sú pri veľkoplošných spojoch obzvlášť nebezpečné, pretože aj malé defekty v spoji môžu počas prevádzky viesť k lokálnemu prehrievaniu spoja alebo čipu, čo urýchľuje degradáciu materiálov a znižovaniu spoľahlivosť výkonového modulu.

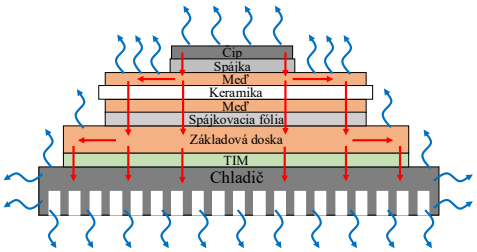
Veľkoplošné spoje [8] zabezpečujú kvalitný odvod tepla a podporujú mechanickú robustnosť zariadenia. Veľkoplošné spoje, napr. medzi čipmi a substrátom alebo medzi substrátom a základovou doskou majú zabezpečiť rovnomerný prenos tepla, čím sa zníži riziko teplotného namáhania súčiastok a zlepší celková stabilitu zariadenia. Kvalita týchto spojov je kľúčová nielen pre dlhodobú funkčnosť a spoľahlivosť, ale aj pre efektívny odvod tepla počas prevádzky. Nevhodný spôsob výroby môže viesť k poškodeniam spoja, hlavne vplyvom tepelného cyklovania. Spoj ovplyvňuje nielen mechanickú pevnosť modulu, ale aj efektivitu a rovnomernosť odvodu tepla. Preto je dôležité sa zaoberať simuláciami teplotného rozloženia, ktoré umožňujú optimalizovať návrh a výrobu veľkoplošných spojov.

Vo výkonových moduloch sú hlavným zdrojom tepla výkonové súčiastky ako tranzistory, diódy a iné aktívne komponenty. Pre optimalizáciu ich výkonu a životnosti je nutné porozumieť tepelnému manažmentu a kvality spojov [5], [9]. Realizácia simulácií teplotného rozloženia je nevyhnutná pre efektívny návrh a optimalizáciu tepelného manažmentu [4], [10]. Tieto simulácie umožňujú analyzovať rozloženie tepla v rôznych častiach zariadenia, identifikovať potenciálne rizikové miesta a predchádzať problémom, ktoré by mohli vzniknúť počas výroby alebo prevádzky. Pomocou simulačných nástrojov je možné optimalizovať návrh, zlepšiť kvalitu spojov a predchádzať nežiaducim defektom. To má za následok zvyšovanie spoľahlivosti a životnosti výkonových moduloch, ale aj znižovanie nákladov na testovanie.

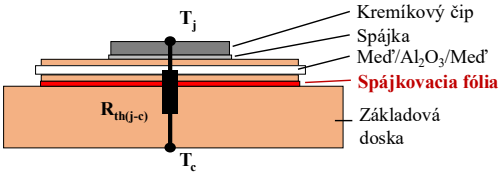
II. Rozloženie Tepla vo Výkonových Moduloch

A. *Základná štruktúra výkonového modulu*

Vo všeobecnosti štruktúra výkonových modulov [2], [11], [12] pozostáva z výkonových komponentov (napr. tranzistory alebo diódy) vo forme čipu (Obr. 1). Tie môžu byť vyrobené z kremíka Si, karbidu kremíka SiC alebo gálium nitridu GaN. Čip je prispájkovaný k DBC (Direct Bonded Copper) substrátu pomocou spájkovanej pasty alebo spájkovacej fólie (preformy), ktorá zabezpečuje elektrické, ale aj tepelné vlastnosti spoja. DBC substrát pozostáva z izolačnej keramickej vrstvy (AlN, Al_2O_3, Si_3N_4) a dvoch medených vrstiev na obidvoch stranách keramiky. Elektrické prepojenie jednotlivých čipov medzi sebou a medzi čipom a substrátom je realizované pomocou ultrazvukovo zváraných drôtových spojov.



Obr. 1 Štruktúra výkonového modulu a tepelný tok (prenos tepla vedením – červené šípky, prenos tepla prúdením – modré šípky).



Obr. 2 Štruktúra základného simulačného modelu v programe Solid Edge (T_j – teplota čipu, T_c – teplota na spodnej strane základovej dosky, $R_{th(j-c)}$ – tepelný odpor štruktúry).

DBC substrát je následne prispájkovaný k základovej doske pomocou spájkovacej fólie. Základová doska slúži ako mechanická podpora pre výkonový modul a zároveň zabezpečuje odvod tepla z výkonového modulu na externý chladič, ktorý je k základovej doske pripevnený pomocou skrutiek. Medzi základovou doskou a externým chladičom sa používa tepelne vodivý materiál (TIM - Thermal Interface Material). Všetky komponenty výkonového modulu sú uložené v puzdre, ktoré môže byť zaliate silikónovým gélom alebo epoxidom. Na povrchu puzdra sú vyvedené výkonové konektory a riadiace piny (v závislosti od typu výkonového modulu) [13].

B. *Spôsoby prenosu a odvodu tepla vo výkonovom module*

Typickou prevádzkovou charakteristikou výkonového modulu (napr. IGBT) je, že pracuje rádovo pri vysokých napätiach (600 – 6500 V) a prúdoch (6 – 3600 A) [5], [14]. Pracovný cyklus výkonového modulu pozostáva z pravidelného zapínania a vypínania v určitých časových intervaloch (t.j. spínanie). Počas prevádzky dochádza k ohrievaniu a ochladzovaniu súčiastok, pričom je dôležité zobrať do úvahy aj tepelný tok v závislosti od veľkosti plochy substrátu, na ktorej sú súčiastky prispájkované [15]. Teplo, ktoré sa vytvorí na súčiastke počas prevádzky, môže vznikať dvoma základnými spôsobmi a to ako prechodové (spínacie) straty, ktoré vznikajú pri spínaní a ohmické (odporové) straty, ktoré vznikajú pri samotnom prechode prúdu súčiastkou. Teplota čipu (T_j – Junction Temperature) nesmie presiahnuť výrobcom stanovenú hodnotu teploty, aby nedošlo k jej trvalému poškodeniu. Je preto potrebné vzniknuté teplo odvádzať z danej súčiastky, aby bola zabezpečená dlhodobá funkčnosť a spoľahlivosť modulu. Pre komplexné porozumenie priebehu tepelných tokov vo výkonovom module je dôležité definovať zdroje tepla, spôsoby prestup tepla a ich smerovanie. Vo všeobecnosti rozlišujeme tri spôsoby prestupu tepla a to vedením, prúdením a sálaním [16], [17].

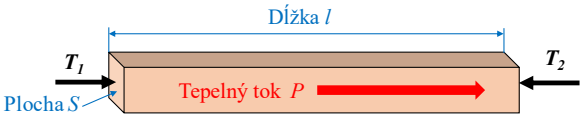
Hlavným zdrojom tepla vo výkonových moduloch sú výkonové polovodičové súčiastky (čipy), ktoré sú prispájkované na substráte. V nečinnom stave sa komponenty nachádzajú v termodynamickej rovnováhe, kedy ich teplota je rovnaká. Počas prevádzky dochádza k ich intenzívnemu zohrievaniu. Z dôvodu rozdielnych teplôt (teplotného gradientu) medzi čipom a externým chladičom prebieha prenos tepla vedením z oblasti s vyššou teplotou (čip) do oblasti s nižšou teplotou (externý chladič) prostredníctvom tuhých materiálov. Smer tepelného toku je orientovaný od čipu smerom k externému chladiču [18]. Efektivita odvodu tepla je výrazne ovplyvnená pomerom plochy osadených čipov k celkovej ploche DBC substrátu.

Vo výkonových moduloch sa okrem prenosu tepla vedením vyskytuje aj prenos tepla prúdením, ktorý sa najviac uplatňuje v oblasti externého chladiča. Ten je navrhnutý tak, aby zabezpečoval prenos tepla zo základovej dosky do okolia pomocou prúdenia vzduchu. To môže byť realizované prirodzeným prúdením vzduchu alebo nútene pomocou ventilátorov. Vzduch, ktorý prúdi cez chladič odvádza teplo z jeho povrchu do okolia, čím zabezpečuje jeho chladenie. V uzavretých výkonových moduloch, kde nie je prúdenie vzduchu, môže dochádzať okrem vedenia a prúdenia aj k sálaniu tepla medzi vnútornými komponentmi a stenami obalu.

C. *Termofyzikálne vlastnosti materiálov*

Štruktúra výkonového modulu (Obr. 1) obsahuje niekoľko komponentov, pričom každý komponent je vyrobený z odlišného materiálu (Tab. 1). Medzi dôležité parametre materiálov z pohľadu tepelného manažmentu môžeme zaradiť tepelnú vodivosť a tepelný odpor. Významnú úlohu zohráva aj tepelná rozťažnosť, ktorá ovplyvňuje mechanické namáhanie spojov (pôsobí na mechanické pnutie a únavu spoja pri tepelnom cyklovaní). Dôležitá je aj tepelná kapacita materiálov, ktorá ovplyvňuje dynamiku zahrievania (chladenia) a kvalita kontaktných plôch, ktorá ovplyvňuje tepelný odpor na rozhraní. Tepelný odpor môže byť ovplyvnený aj javmi, ako sú elektromigrácia a termomigrácia v spájkach.

Tepelná vodivosť λ (W/mK) je schopnosť materiálu viesť teplo a patrí medzi kľúčové parametre pri návrhu výkonových modulov. Vysoká tepelná vodivosť materiálov (napr. Cu alebo AlN) zabezpečuje rovnomernejšie rozloženie teploty, nižšie teplotné gradienty, efektívnejší odvod tepla z čipov a zníženie tepelného odporu, čím sa zvyšuje spoľahlivosť a životnosť spoja. Vďaka tomu má výrazný vplyv na účinnosť chladenia aj celkovú životnosť modulu. Táto vodivosť je daná hlavne molekulárnou štruktúrou jednotlivých materiálov. Závisí však aj od teploty, tlaku a stavu materiálu [19]. Tepelný odpor R_{th} (K/W) vyjadruje schopnosť materiálu brániť prestupu tepla. Je definovaný ako teplotný rozdiel medzi dvoma ľubovoľnými bodmi (T_1, T_2) vydelený množstvom tepla (tepelný tok resp. výkon P – množstvo tepla prenesené za jednotku času), ktoré prúdi medzi nimi (1). Tepelný odpor [20] súvisí aj s tepelnou vodivosťou materiálu λ, jeho dĺžkou l a plochou prestupu tepla S (Obr. 3).



Obr. 3 Ilustrácia prenosu tepla vedením.

$$R_{th} = \frac{T_1 - T_2}{P} \tag{1}$$

$$R_{th} = \frac{l}{S \cdot \lambda} \tag{2}$$

Vzťah (2) platí za ideálnych podmienok, čo znamená, že daný materiál je obklopený tepelnou izoláciou. Tepelný odpor bude väčší, ak bude väčšia dĺžka materiálu (horší prestup tepla) alebo menšia plocha (vyšší tepelný odpor), cez ktorú bude teplo prechádzať. Vyššia tepelná vodivosť znamená nižší tepelný odpor a teda lepší prestup tepla. V oblasti výkonových modulov je preto dôležité zabezpečiť čo najnižší tepelný odpor (napr. použitím materiálov s vyššou tepelnou vodivosťou), aby sa zabezpečil efektívny odvod tepla zo súčiastok, ktoré ho generujú.

III. SIMULÁCIE

Simulácie teplotného rozloženia vo výkonových moduloch boli realizované pomocou programu Solid Edge 2024 Student Edition. Výhodou softvéru je bezplatnosť pre aktívnych študentov a je určený na akademickú prácu pre jednotlivca alebo skupinu študentov. Obsahuje pomerne jednoduché a intuitívne používateľské rozhranie, ktoré umožňuje vytváranie 3D modelov, rôznych animácií a niekoľko druhov simulácií (napr. mechanické a tepelné simulácie). Nevýhodou študentskej verzie je nekompatibilita s komerčnou verziou softvéru a čiastočne obmedzená funkcionalita [21].

A. *Modelovanie jednotlivých komponentov*

Pre simuláciu teplotného rozloženia vo veľkoplošných spojoch bol vytvorený základný simulačný model (Obr. 2), ktorý pozostáva zo 7 komponentov. Tieto komponenty sú usporiadané nad sebou a vzájomne sa ovplyvňujú z pohľadu prenosu tepla. Simulácia sa sústreďuje na oblasť od čipu po základovú dosku. Kľúčovým prvkom pri hodnotení vplyvu defektov (v našom prípade voidov) na rozloženie teploty je spájkovacia fólia medzi DBC substrátom a základovou doskou. Cieľom analýzy je posúdiť vplyv voidov v spájkovacej fólii (na Obr. 2 označené červenou farbou) na zmenu tepelného odporu. Čip je umiestnený v strede vrchnej strany DBC substrátu, čo umožňuje presnejšiu analýzu tepelného správania výkonového modulu. DBC substrát je zložený z troch komponentov (vrstiev): Meď / Al_2O_3 / Meď. Prvým krokom bolo vytvorenie 3D komponentov na základe technických dokumentácií a dostupných údajov (Tab. 1), [3], [13], [22], [23], [24]. Tie boli následne spojené do zostavy, čím vznikol základný simulačný model (Obr. 2), ktorý umožnil simulovať rozloženie teploty vo veľkoplošnom spoji. Plochy spojov boli

konštantné pre všetky realizované simulácie. Hlavný dôraz bol kladený na spoj medzi DBC substrátom a základovou doskou, pričom simulácie teplotného rozloženia zohľadňujú aj vplyv ostatných komponentov výkonového modulu.

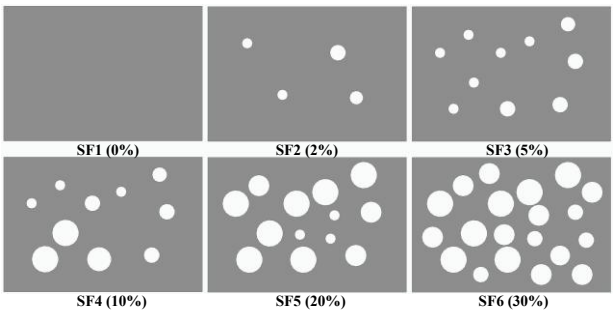
Tab. 1
Technické parametre komponentov výkonového modulu

Komponent	Materiál	Rozmery (mm)	Hrúbka (mm)
Výkonový čip	Si	11,9 x 11,9	0,44
Spájka	96,5Sn3,5Ag	12 x 12	0,08
	Cu	37 x 25	0,3
DBC substrát	Al_2O_3	38 x 26	0,38
	Cu	37 x 25	0,3
Spájkovacia fólia	96,5Sn3,5Ag	37 x 25	0,08
Základová doska	Cu	104 x 56	3

Pri tvorbe simulácií bolo potrebné určenie spôsobu, akým by bolo možné simulovať voidy v spájkovanom spoji, ktoré môžu byť vyplnené vzduchom alebo tavivom. Voidy sa typicky vyskytujú vo forme kruhových tvarov (avšak môžu mať aj rôzne nepravidelné tvary), ktoré môžu prechádzať celou hrúbkou spájky, čím sa ovplyvňujú mechanické a tepelné vlastnosti spoja. Zjednodušene môžeme o voidoch uvažovať ako o valcových útvaroch rôzneho priemeru, ktoré na rôznych miestach prechádzajú celou hrúbkou spájky [25]. Na tomto mieste je vhodné poukázať na to, že výskyt voidov súvisí s výrobou modulov a problematika ich výskytu v prípade veľkoplošných spojov môže predstavovať kritický problém počas prevádzky. Na základe úvahy spomenutej vyššie bolo vytvorených osem variantov komponentu spájkovacej fólie (SF1 – SF6), (Tab. 2), ktoré sa odlišujú rozdielnym množstvom voidov (%), resp. ich polohou (Obr. 5). Voidy boli simulované náhodným rozmiestnením valcových otvorov prechádzajúcich celou hrúbkou komponentu, ktoré predstavujú poruchy v spájkovacej fólií po výrobe (Obr. 4).

Tab. 2
Parametre jednotlivých variantov komponentu spájkovacej fólie

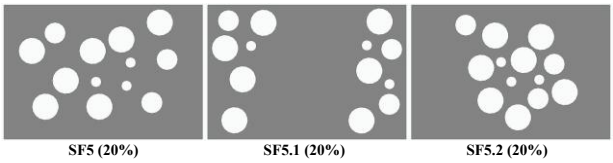
Variant	Objem (mm^3)	Množstvo voidov (%)	Počet voidov	Priemer voidov (mm)
SF1	74	0	0	0
SF2	72,5	2	4	2 - 3
SF3	70,3	5	10	2 - 3
SF4	66,6	10	10	2 - 5
SF5	59,2	20	13	2 - 5
SF5.1	59,2	20	13	2 - 5
SF5.2	59,2	20	13	2 - 5
SF6	51,8	30	20	2 - 5



Obr. 4 Vizuálne zobrazenie rozloženia voidov vo variantoch komponentu spájkovacej fólie.

Množstvo voidov bolo definované ako percentuálny podiel objemu voidov k objemu spájky. Tento prístup umožňuje validáciu výsledkov aj v prípade, ak by voidy neboli ideálne valcovité a neprechádzali celou hrúbkou spájky, ale iba jej časťou. Vyhotovené varianty spájkovacej fólie (SF1, SF2, SF3, SF4, SF5 a SF6) boli integrované do základného simulačného modelu, čím bolo vytvorených šesť variantov základného simulačného modelu s cieľom analýzy vplyvu množstva voidov na zmenu tepelného odporu. Zároveň boli vytvorené dve varianty základného simulačného modelu, ktoré simulujú rozdielnu polohu voidov (SF5.1 a SF5.2) v spájkovacej fólii (Obr. 5) pri rovnakom množstve (20 %). Každý variant simulačného modelu obsahoval jeden variant komponentu reprezentujúci spájkovaciu fóliu. Základný simulačný model nie je funkčným elektrickým modelom, ale zachováva kľúčové štrukturálne a materiálové parametre výkonového modulu pre účely simulácií teplotného rozloženia. Hodnoty percentuálneho

množstva voidov (%) boli určené pomocou softvéru, pričom ich počet a priemer boli zvolené náhodne. Parametre materiálov použité pri simuláciách sú uvedené v Tab. 3.



Obr. 5 Vizuálne porovnanie variantov spájkovacej fólie s 20 % množstvom voidov.

Tab. 3
Parametre materiálov

Materiál	Si	96,5Sn3,5Ag	Cu	Al_2O_3	Vzduch
Hustota (kg/m^3)	2329	7360	8940	3720	1,204
Koef. tepel. rozťažnosti (10^{-6}/K)	4,1	30,2	18	8,3	3400
Tepelná vodivosť (W/mK)	148	33	391	24	0,026
Špecifické teplo (J/kg°C)	700	233	385	880	1005
Modul elasticity (GPa)	98	41	117,21	331	0
Poissonov pomer (-)	0,25	0,38	0,31	0,25	-
Medza pevnosti (MPa)	7000	38,75	310	400	0

B. Simulačné nastavenia v softvéri Solid Edge

Pred začatím simulácie bolo nevyhnutné definovať jej typ ako aj ďalšie kľúčové parametre potrebné na získanie presných výsledkov. V softvéri Solid Edge 2024 Student Edition bola zvolená metóda časovej teplotnej analýzy (Transient Heat Transfer Study), ktorá umožňuje sledovať vývoj teplotného rozloženia v priebehu času počas tepelného zaťaženia. Táto analýza bola aplikovaná na celú zostavu základného modelu, ktorý obsahuje komponenty s odlišnými tepelnými vlastnosťami, poskytuje pohľad na dynamiku teplotných zmien v rôznych časových okamihoch, čo umožňuje presnejšie hodnotenie tepelných procesov v simulovanom modeli. Po výbere simulačnej metódy boli nastavené základné parametre: celkový simulačný čas na 60 s, časový krok výpočtu 2,5 s, typ intervalu na konštantný a počiatočná teplota pre simuláciu 20 °C.

Ďalším krokom bolo definovanie tepelných zaťažení (Loads), ktoré ovplyvňujú teplotné správanie modelu. Hlavným zdrojom tepla je kremíkový čip, ktorý sa v čase prevádzky správa ako generátor tepla. Keďže čip vyžaruje teplo do všetkých smerov, na tento komponent bola aplikovaná tepelná záťaž (Heat Generation) s hodnotou 100 W. Táto hodnota závisí od konkrétnych parametrov čipu a spôsobu jeho prevádzky (v našom prípade volené ako univerzálna hodnota). Vzhľadom na to, že čip generuje teplo, je potrebné definovať spôsob jeho odvodu zo simulačného modelu, na čo slúži základová doska. Na tento komponent bola aplikovaná tepelná záťaž vo forme prúdenia, pričom koeficient prestupu tepla (Film Coefficient) bol nastavený na hodnotu 20 W/m²K, čo odpovedá prirodzenému tepelnému prestupu do okolia. Teplota okolia bola nastavená na 20 °C, čím sa simuluje normálne prostredie. Tepelná záťaž vo forme prúdenia s rovnakými parametrami bola aplikovaná aj na hornú stranu DBC substrátu, kde taktiež dochádza k prenosu tepla do okolia. Posledným tepelným zaťažením, ktoré bolo potrebné zadefinovať, bola počiatočná teplota modelu na začiatku simulácie (Body Temperature), ktorá bola nastavená na 20 °C, čo predstavuje bežnú teplotu na začiatku tepelného cyklu.

Po nastavení tepelných zaťažení bolo potrebné definovať mriežku (Mesh), ktorá je základným nástrojom pre numerickú analýzu tepelného prestupu v simuláciách. Mriežka slúži na diskretizáciu geometrie modelu, čo umožňuje výpočty na základe malých geometrických prvkov. Pre náš model bol zvolený typ mriežky Tetrahedral, ktorý je vhodný pre komplexné geometrie a nerovnomerné tvary. Zároveň bola aktivovaná možnosť Body Mesh, ktorá zabezpečuje optimalizáciu a vyššiu kvalitu vytvorenej štruktúry mriežky.

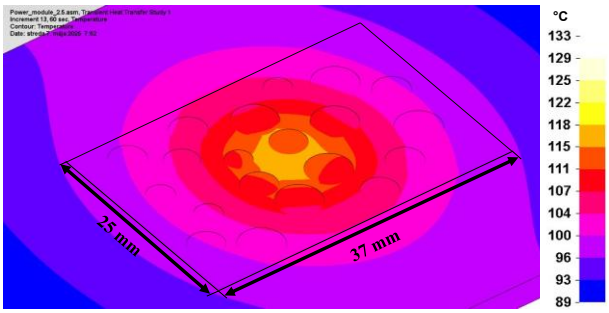
Na zabezpečenie správnej interakcie medzi vrstvami bolo potrebné definovať typ prepojení medzi jednotlivými komponentami zostavy. V tomto prípade sme zvolili typ prepojení Thermal Glue (Penalty Value: 100), ktorý umožňuje simuláciu ideálneho tepelného prepojenia medzi komponentmi, pričom sa neberie do úvahy kontaktný tepelný odpor. Tento typ prepojenia je vhodný pre modelovanie tepelných spojov, najmä pri simuláciách s materiálmi ako spájka, kde je potrebné sledovať vplyv voidov v spájke na tepelný prestup. Každé prepojenie musí byť správne definované z hľadiska smeru prestupu tepla: od plochy, ktorá predstavuje zdroja tepla (Source) ku cieľovej ploche (Target), čím sa zabezpečí realistické zobrazenie tepelného toku medzi vrstvami materiálov. Tieto parametre boli použité na všetkých ôsmich simulačných modeloch, pričom každý z modelov obsahoval komponent s odlišným podielom voidov v spájke.

IV. VÝSLEDKY A DISKUSIA

Teplotné simulácie sa zameriavali na analýzu teplotného rozloženia vo výkonových moduloch pri rôznom množstve voidov v spájkovanom spoji medzi DBC substrátom a základovou doskou. Simulácie boli realizované na šiestich základných simulačných modeloch s rôznym množstvom voidov (0 - 30 %), pričom cieľom bolo vyhodnotiť ich vplyv na tepelný odpor. Zároveň boli vykonané dve doplňujúce simulácie (5.1 a 5.2) s cieľom vyhodnotiť vplyv polohy voidov pri rovnakom množstve voidov v spájke (20 %).

Všetky použité parametre, vrátane tepelných zaťažení, mriežky, a prepojení medzi komponentami, boli aplikované na základný simulačný model, ktorý simuloval reálne podmienky tepelného prestupu vo výkonných moduloch so špecifickými vlastnosťami materiálov. Zároveň bola simulácia pripravená tak, aby zahrnula vplyv voľného prúdenia vzduchu, čo ovplyvňuje účinnosť tepelného prestupu, a tým aj teplotný profil výkonového modulu počas prevádzky.

Vizualizácia výsledkov poskytla detailný pohľad na distribúciu teploty v rôznych vrstvách výkonového modulu. Horizontálne rezy modelu (Obr. 6) ukazujú teplotné rozloženie v oblasti medzi DBC substrátom a základovou doskou. Vo vizualizáciách sú dobre viditeľné oblasti s prítomnosťou voidov. Tieto oblasti vykazujú nižšiu teplotu (menej ako 120 °C), čo poukazuje na zhoršený prenos tepla. Tento jav je spôsobený nižšou tepelnou vodivosťou vzduchových medzier v porovnaní so spájkovacím materiálom, čo negatívne ovplyvňuje efektívny tepelný tok v štruktúre výkonového modulu.



Obr. 6 Príklad vizualizácie teplotného rozloženia v spájkovacej fólii s 30 % množstvom voidov (horizontálny rez).

A. *Vyhodnocovanie simulácií*

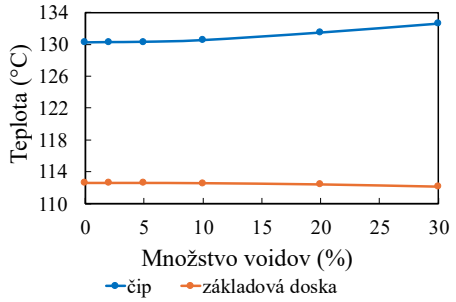
Vyhodnocovanie prebiehalo po skončení simulácie pomocou nástroja PROBE, ktorý umožňoval odčítať teploty z dvoch miest a to na povrchu komponentu, ktorý predstavuje čip (T_j), kde bola teplota simulačného modelu najvyššia a na spodnej strane základovej dosky (T_c) v mieste pod čipom. Týmto spôsobom bol simulovaný proces, pri ktorom bola výkonová súčiastka (čip) v pracovnom režime a generovala teplo 100 W, ktorý predstavoval spínacie straty.

B. *Interpretácia výsledkov*

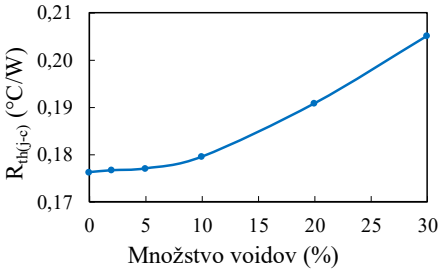
Získané výsledky (Tab. 4) poukazujú na jednoznačnú koreláciu medzi množstvom voidov v spájkovanom spoji a teplotnými charakteristikami modulu. Pri 30 % množstve voidov došlo k nárastu teploty na čipe o 2,44 °C (oproti 0 % množstvu voidov), čo indikuje zhoršenie tepelného odporu a obmedzený prenos tepla cez spájkovaný spoj. Výsledky taktiež ukazujú, že pri menšom množstve voidov v spájke bol vplyv na teplotu čipu (T_j) menej výrazný. S narastajúcim množstvom voidov však dochádza nielen k nárastu teploty čipu, ale zároveň k poklesu teploty na spodnej strane základovej dosky (Obr. 7). Tento efekt je dôsledkom zhoršenia prenosu tepla medzi čipom a základovou doskou, čo vedie k zvýšeniu tepelného odporu celej štruktúry. Na základe nameraných hodnôt bol pomocou vzorca (1) vypočítaný tepelný odpor $R_{th(j-c)}$ medzi čipom a základovou doskou. Výsledky potvrdzujú (Obr. 8), že s rastúcim podielom voidov sa tepelný odpor štruktúry zvyšuje, čo koreluje so zhoršenou schopnosťou odvádzať teplo z čipu. To v reálnej prevádzke vedie k zvýšenému teplotnému namáhaniu čipu, čo negatívne ovplyvňuje jeho spoľahlivosť a životnosť. Taktiež je vidieť, že závislosť tepelného odporu štruktúry od množstva voidov v spájke nie je lineárna. Simulácie teplotného rozloženia preukázali, že poloha voidov v spájke pri 20 % množstve voidov (SF5.1 a SF5.2) má vplyv na tepelný odpor medzi čipom (umiestneným v strede štruktúry) a základovou doskou. V prípade koncentrácie voidov v strede spájky (SF5.2) je tepelný prestup menej efektívny, čo vedie k vyššej teplote čipu v porovnaní s variantom SF5 o 1,21 °C. Na druhej strane, ak sú voidy koncentrované na okraji spájky (SF5.1), prestup tepla z čipu je účinnejší, čo vedie k nižšej teplote čipu v porovnaní s variantom SF5 o 1,25 °C. Výpočty tepelného odporu štruktúry potvrdzujú, že tepelný odpor sa zvyšuje so zvyšujúcim sa množstvom voidov koncentrovaných v oblasti priamo pod čipom.

Tab. 4
Simulačné výsledky

Simulácia	Množstvo voidov (%)	Variant spájkovacej fólie	Tj (°C)	T_c (°C)	Vypočítaný Rth (°C/W)
1	0	SF1	130,22	112,59	0,1763
2	2	SF2	130,26	112,59	0,1767
3	5	SF3	130,3	112,59	0,1771
4	10	SF4	130,52	112,56	0,1796
5	20	SF5	131,5	112,41	0,1909
5.1	20	SF5.1	130,25	112,63	0,1762
5.2	20	SF5.2	132,71	112,18	0,2053
6	30	SF6	132,66	112,14	0,2052



Obr. 7 Teplotná závislosť kremíkového čipu a spodnej strany základovej dosky na množstve voidov v spájkovacej fólií.



Obr. 8 Závislosť tepelného odporu štruktúry na množstve voidov v spájkovacej fólií.

V. ZÁVER

V tomto článku boli realizované simulácie teplotného rozloženia vo veľkoplošných spojoch, ktoré sa nachádzajú vo výkonových moduloch, prostredníctvom softvéru Solid Edge 2024 Student Edition. Závislosť tepelného odporu vo výkonových moduloch medzi čipom a základovou doskou je výrazne ovplyvnená prítomnosťou voidov v spájkovanom veľkoplošnom spoji (37 x 25 mm) medzi DBC substrátom a základovou doskou. Množstvo, veľkosť, poloha a distribúcia týchto voidov ovplyvňujú celkový tepelný odpor štruktúry výkonového modulu. Väčšie množstvo voidov zvyšuje tepelný odpor nelineárne. V dôsledku zvýšeného tepelného odporu, ako aj v dôsledku toho, že vzduch prítomný vo voidoch pôsobí ako izolant, sa znižuje efektívnosť prestupu tepla z čipu na základovú dosku. Výsledkom simulácií je, že pri tepelnej záťaži 100 W na čipe počas 60 s, a pri 30 % množstve voidov v spájke došlo k nárastu teploty na čipe o 2,44 °C oproti 0 % množstvu voidov. Rovnako aj veľkosť voidov a ich poloha (náhodné rozloženie, rozloženie na okrajoch a priamo pod čipom) majú vplyv na tepelný tok a tiež na homogenitu rozloženia teploty. Dochádza k zvýšenému riziku prehriatia a negatívnemu vplyvu na celkový výkon modulu. Pre dosiahnutie optimálneho tepelného manažmentu je nevyhnutné minimalizovať množstvo voidov v spájkovaných spojoch, čím sa zabezpečí efektívnejší odvod tepla a predĺži životnosť výkonových súčiastok.

Softvér Solid Edge 2024 Student Edition je vhodný nástroj pre analýzu teplotného rozloženia vo veľkoplošných spojoch pre orientačné simulácie v akademickom a výskumnom prostredí. Tento softvér umožňuje presné simulácie, avšak v našich experimentoch bol aplikovaný zjednodušený model a podmienky. Zo simulácií vyplýva, že vplyv voidov v spájke nie je až tak výrazný, avšak v simuláciách boli uvažované ideálne podmienky a zanedbanie vplyvov, ako napr. tepelný odpor spojov, mechanická deformácia materiálov v dôsledku rozdielnych koeficientov tepelnej rozťažnosti a niektoré ďalšie tepelné zaťaženia ako prúdenie tepla z čipu, sálanie tepla medzi komponentami, sálanie tepla medzi vnútornými komponentmi a stenami obalu a ďalšie. Tento softvér zároveň umožňuje realizovať aj iné modely výkonových modulov, ktorých súčasťou je väčšie množstvo veľkoplošných spojov a komponentov na základovej doske. Nevýhodou tohto softvéru je, že neumožňuje simulovať komplexnejšie simulácie na predikciu defektov ako napríklad delaminácie, prasknutí alebo iných poškodení počas prevádzky.

POĎAKOVANIE

Táto práca vznikla s podporou projektov KEGA 011TUKE-4/2023, VEGA 1/0049/24, VEGA 1/0278/24, APVV 22-0289.

ZDROJE

[1] A. George, J. Zipprich, M. Breitenbach, M. Klingler, a M. Nowottnick, "Reliability investigation of large area solder joints in power electronics modules and its simulative representation", *Microelectronics Reliability*, roč. 88–90, s. 762–767, sep. 2018, doi: 10.1016/j.microrel.2018.06.049.

[2] J. Lutz, H. Schlangenotto, U. Scheuermann, a R. De Doncker, *Semiconductor Power Devices*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-70917-8.

[3] S. E. Nebo, E. H. Amalu, a D. J. Hughes, "Critical solder joint in insulated gate bipolar transistors (IGBT) power module for improved mechanical reliability", *Microelectronic Engineering*, roč. 291, s. 112200, aug. 2024, doi: 10.1016/j.mee.2024.112200.

[4] G. Lee, J. Kim, S. Choi, a B. Ma, "Comparative Thermal Analysis of Wide Band Gap Power Module", v *2023 22nd IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, máj. 2023, s. 1–6. doi: 10.1109/ITherm55368.2023.10177652.

[5] N. Iwamuro a T. Laska, "IGBT History, State-of-the-Art, and Future Prospects", *IEEE Transactions on Electron Devices*, roč. 64, č. 3, s. 741–752, mar. 2017, doi: 10.1109/TED.2017.2654599.

[6] S. Gao, R. Wang, H. Wang, a R. Kang, "Warping model of high-power IGBT modules subjected to reflow soldering process", *International Journal of Mechanical Sciences*, roč. 251, s. 108350, aug. 2023, doi: 10.1016/j.ijmecsci.2023.108350.

[7] A. Sitta, M. Renna, A. A. Messina, G. Mirone, G. D'Arrigo, a M. Calabretta, "Power Module Ceramic Substrates: mechanical characterization and modeling", v *2020 21st International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, júl. 2020, s. 1–5. doi: 10.1109/EuroSimE48426.2020.9152670.

[8] A. J. George, J. Zipprich, M. Klingler, M. Breitenbach, a M. Nowottnick, "Study on the interdependence of soldering profile, ageing conditions and intermetallic layer thickness in large area solder joints and its influence on reliability", v *2017 40th International Spring Seminar on Electronics Technology (ISSE)*, máj. 2017, s. 1–6. doi: 10.1109/ISSE.2017.8000907.

[9] Y. Jiang, L. Che, a G. Lei, "Design of a High-power Density Power Module with Large-area (10 mm × 10 mm) SiC MOSFETs for Traction Inverter Application", v *2024 25th International Conference on Electronic Packaging Technology (ICEPT)*, aug. 2024, s. 01–06. doi: 10.1109/ICEPT63120.2024.10668675.

[10] T. Girašek a E. Pietriková, "Visualization of temperature fields distribution on power module within web interface", v *2016 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, nov. 2016, s. 69–72. doi: 10.1109/ICETA.2016.7802062.

[11] I. R. Swan, A. T. Bryant, N.-A. Parker-Allotey, a P. A. Mawby, "3-D thermal simulation of power module packaging", v *2009 IEEE Energy Conversion Congress and Exposition*, sep. 2009, s. 1247–1254. doi: 10.1109/ECCE.2009.5316235.

[12] L. Wang *et al.*, "Enhanced Thermal Performance of Power Module with Integrated AlN Ceramic Heatsink", v *2023 IEEE 2nd International Power Electronics and Application Symposium (PEAS)*, nov. 2023, s. 1695–1700. doi: 10.1109/PEAS58692.2023.10394736.

[13] "SKM600GB07E3", Semikron Danfoss. Cit: 07. apríl 2025. [Online]. Available at: https://www.semikron-danfoss.com/products/p/skm600gb07e3-22895570

[14] "IGBT modules", Infineon Technologies AG. Cit: 07. apríl 2025. [Online]. Available at: https://www.infineon.com/cms/en/product/power/igbt/igbt-modules/

[15] IBE, "Power Module: Working Principle, Structural Features, and Process", IBE Electronics. Cit: 29. marec 2025. [Online]. Available at: https://www.pcbaaa.com/power-module/

[16] C. H. Forsberg, "Chapter 1 - Introduction to heat transfer", v *Heat Transfer Principles and Applications*, C. H. Forsberg, Ed., Academic Press, 2021, s. 1–21. doi: 10.1016/B978-0-12-802296-2.00001-9.

[17] "What Is Heat Transfer (Heat Flow)? Complete Guide", SimScale. Cit: 29. marec 2025. [Online]. Available at: https://www.simscale.com/docs/simwiki/heat-transfer-thermal-analysis/what-is-heat-transfer/

[18] M. Knecht, R. Tschirbs, a R. Ott, "Superior Assembly Technology", Power Systems Design. Cit: 02. október 2024. [Online]. Available at: https://www.powersystemsdesign.com/articles/superior-assembly-technology/22/5855

[19] K. Stewart, "Thermal conductivity | Definition, Science, & Applications", Britannica. Cit: 29. marec 2025. [Online]. Available at: https://www.britannica.com/science/thermal-conductivity

[20] "Fundamentals of Thermal Resistance and Heat Dissipation: Thermal Resistance in Conduction", TechWeb Powered by ROHM. Cit: 29. marec 2025. [Online]. Available at: https://techweb.rohm.com/product/thermal-design/9551/

[21] "Solid Edge Student Edition", Siemens Digital Industries Software. Cit: 29. marec 2025. [Online]. Available at: https://resources.sw.siemens.com/en-US/download-solid-edge-student-edition/

[22] "curamik® Power". Cit: 11. november 2024. [Online]. Available at: https://www.rogerscorp.com/advanced-electronics-solutions/curamik-ceramic-substrates/curamik-power

[23] "SKT11 PDF", SEMIKRON [Semikron International]. Cit: 07. apríl 2025. [Online]. Available at: http://www.alldatasheet.com/datasheet-pdf/view/418168/SEMIKRON/SKT11.html

[24] "Solder Ribbon | Solder Foil", Indium Corporation. Cit: 23. marec 2025. [Online]. Available at: https://www.indium.com/products/solders/solder-ribbon/

[25] S. Y. Jo *et al.*, "Effect of solder void on mechanical and thermal properties of flip-chip light-emitting diode: Statistical analysis based on finite element modeling", *Heliyon*, roč. 10, č. 12, s. e33242, jún. 2024, doi: 10.1016/j.heliyon.2024.e33242.

Real-time detekcia obrazcov pri obmedzených datasetoch pomocou YOLO11

[1]*Ladislav HRIC,* [2]*Daniela PERDUKOVÁ*

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

[1]ladislav.hric@tuke.sk, [2] daniela.perdukova@tuke.sk

***Abstract* — Detekcia objektov je kľúčovou úlohou počítačového videnia s využitím v rôznych odvetviach. Jednou zo špecifických aplikácií je rozpoznávanie obrazcov, ktorá často využíva konvolučné neurónové siete. Na vytvorenie spoľahlivého modelu je zvyčajne potrebný rozsiahly dataset, no v praxi to nie je vždy možné. Táto práca sa zaoberá výzvou trénovania modelu na reálnu detekciu obrazcov pomocou obmedzeného datasetu. Ako základ bol použitý moderný model YOLO11, známy vyváženosťou medzi rýchlosťou a presnosťou. Trénované boli rôzne varianty YOLO11 a ich výkon bol porovnaný. Výsledky ukazujú, že aj pri malom množstve dát možno dosiahnuť spoľahlivú detekciu obrazcov v reálnom čase.**

***Keywords* — Detekcia objektov, Deep learning, Obmedzený dataset, YOLO 11**

I. ÚVOD

Detekcia objektov je široko diskutovaná téma vzhľadom na jej rozsiahle využitie v kľúčových odvetviach ako zdravotníctvo, bezpečnosť, automatizácia a ďalšie. Dopyt po týchto algoritmoch existuje už dlhodobo, no ich praktické využitie sa do popredia dostalo až v posledných rokoch. Konvolučné neurónové siete sa intenzívne využívajú pri úlohách počítačového videnia, a to vďaka svojej efektivite pri učení sa vzorov z obrázkov. Takéto algoritmy dokážu rozpoznať všetko od jednoduchých čiar a kriviek až po zložité objekty reálneho sveta ako sú ľudia, psy a mačky – v podstate akékoľvek vzory, na ktoré bol model natrénovaný [1]. Kvalita a efektívnosť neurónových sietí závisí od ich štruktúry. Existuje mnoho modelov, ktoré môžu slúžiť ako východiskový bod [2] pre ďalšie aplikácie a vylepšenia. V oblasti detekcie objektov v reálnom čase sú modely YOLO (You Only Look Once) od Ultralytics považované za špičku, známe pre vyváženie medzi rýchlosťou a presnosťou [3]. Najnovšia vydaná verzia je YOLO11, ktorá je vhodná pre mnohé úlohy. Táto práca sa zameriava na detekčné úlohy s využitím rôznych variant YOLO: n, s, m, l a x. Každý variant ponúka jedinečnú rovnováhu medzi rýchlosťou a presnosťou, aby pokryl širokú škálu aplikácií. Variant *n* má najjednoduchšiu architektúru, zatiaľ čo *x* má najzložitejšiu.

Na vytvorenie robustného modelu je kľúčové zamerať sa na zber veľkého množstva tréningových dát. Ako je uvedené v práci [4], mnohé reálne scenáre vyžadujú prácu s obmedzenými dátovými zdrojmi. Jedným z takýchto príkladov je detekcia zriedkavých ochorení, kde nedostatok dát vyžaduje použitie techník na ich predspracovanie a augmentáciu na zlepšenie tréningového procesu. Podobné výzvy boli riešené aj v práci [5], ktorá skúmala detekciu špirál nakreslených na papieri za účelom identifikácie a klasifikácie štádií Parkinsonovej choroby. Detekcia obrazcov má široké spektrum využitia, pričom štúdia [6], ktorá používa neurónové siete na rozpoznávanie špecifických elektrických symbolov, je len jedným z mnohých príkladov.

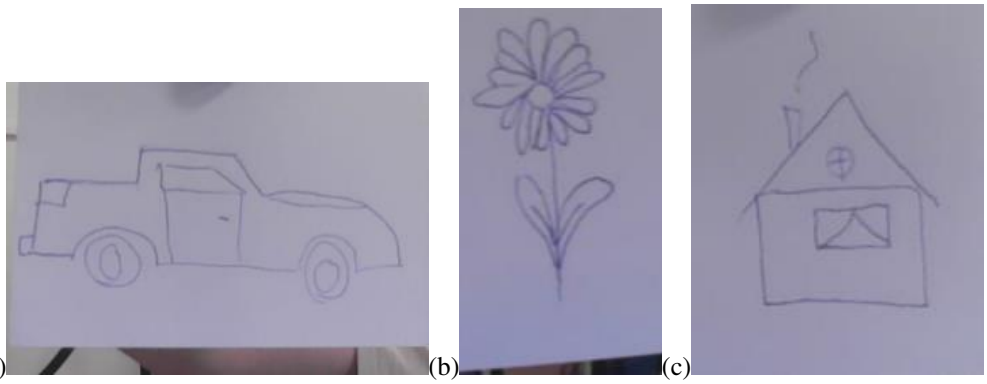
V tejto štúdii napodobňujeme takúto situáciu a vyvíjame model na detekciu v reálnom čase, trénovaný na malom datasete. Dataset obsahuje tri základné kresby: autá, kvety a domy. Hoci sú tieto kresby pre ľudí jednoduché na rozpoznanie, pre počítačové videnie predstavujú náročnú úlohu kvôli ich abstraktnosti a kombinácii rôznych kriviek. Modely schopné detekcie základných kresieb majú potenciál v mnohých aplikáciách.

Hlavným prínosom tejto práce je vyhodnotenie modelov založených na YOLO11 pre aplikáciu detekcie obrazcov v reálnom čase trénovanú na malom datasete.

II. ZBER DÁT

V tejto práci používame vlastný dataset zložený z jednoduchých obrazcov rozdelených do troch kategórií: autá, kvety a domy (Obr. 1). Tieto kresby predstavujú kombinácie kriviek, ktoré musí model rozpoznať, aby ich vedel presne detekovať a klasifikovať. Pre každú kategóriu bolo zozbieraných 45 trénovacích obrázkov, 25 validačných obrázkov a 30 testovacích obrázkov, čím vznikol základ malého datasetu. Na tieto obrázky boli aplikované rôzne techniky predspracovania a rozšírenia dát s cieľom dataset rozšíriť.

(a) (b) (c)



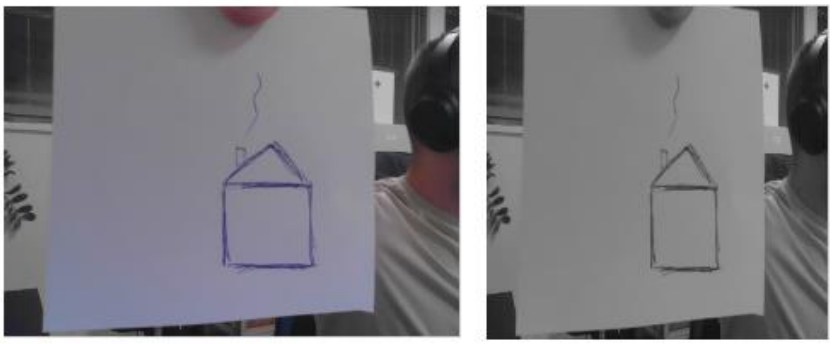
Obr. 1 Kategórie datasetu: (a) autá, (b) kvety, (c) domy

III. ÚPRAVA DÁT

V tejto práci bola ako prvá aplikovaná technika predspracovania zmena veľkosti obrázkov. Ide o kľúčový krok v oblasti počítačového videnia, ktorý pomáha modelom učiť sa rýchlejšie na menších obrázkoch. Navyše, mnohé architektúry hlbokého učenia vyžadujú, aby všetky obrázky mali rovnakú veľkosť. Tento krok je teda často nevyhnutný, pretože zozbierané obrázky sa veľkosťou líšia. V tejto práci boli všetky obrázky upravené na 640x640.

Detekcia hrán je kľúčová pre detekciu objektov. Úprava kontrastu ako krok predspracovania zvýrazňuje hrany, čím uľahčuje neurónovým sieťam ich rozpoznávanie. Preto bola zvolená ako ďalší krok predspracovania.

Následne boli všetky obrázky v datasete konvertované do odtieňov sivej (grayscale). Tým sa zjednodušuje trénovací proces, pretože model sa môže sústrediť na tvary a ignorovať závislosť od farieb. Keďže náš dataset obsahuje iba kresby a cieľom je rozpoznať vzory, farebné informácie sú v tomto prípade zbytočné. Konverzia do odtieňov sivej predstavuje záverečný krok predspracovania. Obr. 2 znázorňuje obrázok pred a po aplikácii techník predspracovania dát.



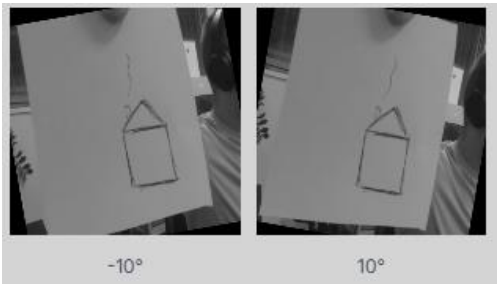
Obr. 2 Predspracovanie obrázkov: vľavo je pôvodný obrázok; vpravo je obrázok po aplikovaní zmeny veľkosti, úpravy kontrastu a prevodu do odtieňov sivej

IV. AUGMENTÁCIA DÁT

Augmentácia obrázkov je manipulácia s existujúcimi obrázkami za účelom rozšírenia pôvodného datasetu. Realizuje sa vytváraním rôznych verzií pôvodných obrázkov, čo pomáha modelu lepšie generalizovať a predchádzať preučeniu (overfittingu). Je to výhodné v reálnych podmienkach, kde môže byť ťažké získať obrázky objektu z rôznych uhlov, vzdialeností, pri rôznych svetelných podmienkach či v iných scenároch. Tento krok je v našom prípade kľúčový,

pretože dataset pozostáva z malého počtu obrázkov. V tejto práci bolo zvolených niekoľko techník augmentácie.

Rotácia obrázkov je dôležitý krok augmentácie v prípadoch, keď vstupný obrázok nemá pevnú orientáciu a môže sa otáčať. V našom prípade môžu byť vstupné obrázky otočené, preto je tento krok dôležitý. V tejto práci boli generované nové obrázky s rotáciami v rozsahu od -10° do +10° (Obr. 3).

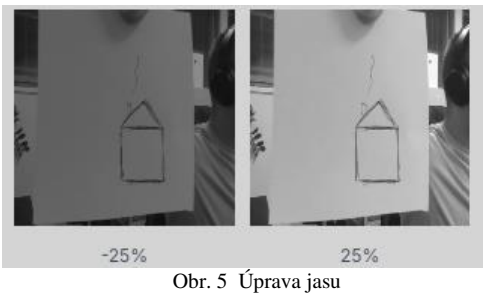


Obr. 3 Rotácia obrázkov

Úprava sklonu bola zvolená kvôli zmene perspektívy obrázka (Obr. 4). Sklon bol menený v rozsahu ±15° horizontálne a ±15° vertikálne.

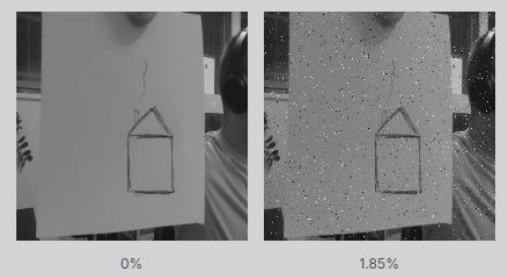
Obr. 4 Úprava sklonu

V mnohých prípadoch, aj keď je možné získať obrázky cieľového objektu z rôznych uhlov, môže byť stále náročné meniť svetelné podmienky. Preto rozšírenie datasetu o obrázky s upraveným jasom pomáha modelu získať odolnosť voči rôznym svetelným podmienkam. V našom prípade bola úprava jasu nastavená v rozsahu od -25 % do +25 % (Obr. 5).



Obr. 5 Úprava jasu

Pridanie šumu do obrázkov je prospešné z hľadiska zrobustnenia modelu. V tomto kroku sú náhodné pixely zmenené čierne alebo biele (Obr. 6). V našom datasete sme túto augmentáciu využili s úpravou až 1,85 % pixelov.



Obr. 6 Pridanie šumu

V. PRÍPRAVA MODELU

Séria modelov YOLO11 ponúka modely špecializované pre rôzne úlohy počítačového videnia. V tomto článku sme najskôr pripravili všetky varianty n, s, m, l a x modelu YOLO11 a porovnali ich efektivitu. Následne sme najlepší model trénovali s rôznymi hodnotami batch size.

Batch size je kľúčový pojem v strojovom učení. Určuje počet trénovacích dát použitých v jednej iterácii trénovania modelu. Má vplyv na efektivitu, rýchlosť trénovania a výkonnosť modelu.

VI. VÝSLEDKY

Výsledky výstupných modelov založených na rôznych variantoch modelu YOLO11 sú uvedené v tabuľke 1. Všetky modely boli trénované s batch size (veľkosťou dávky) 16 počas 100 epôch, čo je predvolená konfigurácia tréningového procesu v rámci frameworku YOLO. Modely boli testované na testovacej množine z datasetu. Na základe definovaných metrík sa ako najvýkonnejší model pri malom datasete ukázal variant n.

Tabuľka 1
Porovnanie výkonnosti modelov pri rôznych variantoch YOLO11

Variant	Precision	Recall	mAP50-95
n	**0.997**	**1**	**0.842**
s	0.997	1	0.833
m	0.986	1	0.834
l	0.996	1	0.816
x	0.978	1	0.817

Tabuľka 2 ďalej zobrazuje modely založené na variante n modelu YOLO11, ale s modifikovanými tréningovými nastaveniami. V tomto prípade sme upravili parameter batch size. Začali sme s nasteveným batch size = 16 a následne sme parameter zmenšovali a zväčšovali s cieľom zvýšiť kvalitu modelu. Ako je vidieť v tabuľke 2, najlepšie výsledky dosiahol model s nastaveným batch size16. Tento model dosiahol presnosť 0.986, recall 0.989 a hodnotu 0.836 pre mAP50–95 na testovacom datasete.

Tabuľka 2
Porovnanie výkonu modelov variant n

Batch size	Precision	Recall	mAP50-95
14	0.997	1	0.829
15	0.994	1	0.816
16	**0.997**	**1**	**0.842**
17	0.997	0.99	0.835
18	0.997	1	0.836
19	0.997	1	0.825
20	0.997	1	0.825

VII. ZÁVER

Modely založené na špičkovej architektúre YOLO11, využívajúce konvolučné neurónové siete, preukázali svoju efektivitu pri detekcii v reálnom čase, ako napríklad pri rozpoznávaní

jednoduchých kresieb s vysokou presnosťou. Tieto modely je možné využiť v rôznych oblastiach, vrátane autonómnych vozidiel na detekciu dopravných značiek, rozšírenej reality na rozpoznávanie náčrtov v interaktívnych zážitkoch, interaktívnych tabúľ pre vzdelávanie či digitálnych umeleckých platforiem.

Takéto modely je možné efektívne trénovať aj na obmedzených datasetoch za použitia vhodných techník predspracovania a augmentácie. V tejto práci boli navrhnuté rôzne modelové varianty trénované na vlastnom datasete, ktorý bol vytvorený s cieľom nasimulovať náročné situácie, v ktorých je potrebné detegovať zriedkavé alebo obmedzené kresby. Prvá skupina modelov sa líšila architektúrou a najlepší z nich bol zvolený ako kandidát na ďalšie zlepšovanie. Tento model bol následne trénovaný v rôznych konfiguráciách s meniacou sa veľkosťou dávky.

Najlepší model dosiahol presnosť 0.986, recall 0.989 a hodnotu 0.836 pre mAP50–95. Tieto výsledky nám dávajú dôveru, že modely založené na YOLO11 môžu spoľahlivo fungovať aj v takýchto typoch aplikácií.

Budúce zlepšenia by sa mali zamerať na rozšírenie datasetu, preskúmanie rôznych techník predspracovania alebo augmentácie a vylepšenie architektúry neurónovej siete pre tento typ úlohy, čo by mohlo viesť k zaujímavým výsledkom.

POĎAKOVANIE

Článok bol publikovaný s podporou projektu APVV-19-0210.

Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00018.

REFERENCES

[1] Y. Ji et al., “Lightweight road damage detection algorithm based on the improved YOLO model,” in Proc. 2024 5th Int. Conf. on Artif. Intelligence and Electromech. Automation (AIEA), 2024, pp. 832–835. DOI: 10.1109/AIEA62095.2024.10692408.

[2] R. Hamzah et al., “Comparing modified YOLO v5 and Faster Regional Convolutional Neural Network performance for recycle waste classification,” in Proc. 2024 IEEE Int. Conf. on Automatic Control and Intelligent Systems (I2CACIS), 2024, pp. 415–419. DOI: 10.1109/I2CACIS61270.2024.10649835.

[3] A. Liu, Y. Liu, and S. Kifah, “Deep convolutional neural network for enhancing traffic sign recognition developed on YOLO v5,” in Proc. 2024 Int. Conf. on Advancements in Smart, Secure and Intelligent Computing (ASSIC), 2024, pp. 1–6. DOI: 10.1109/ASSIC60049.2024.10508025.

[4] K. Sanchez et al., “Subspace-based domain adaptation using similarity constraints for pneumonia diagnosis within a small chest X-ray image dataset,” in Proc. 2021 IEEE 18th Int. Symp. on Biomedical Imaging (ISBI), 2021, pp. 1232–1235. DOI: 10.1109/ISBI48211.2021.9434173.

[5] B. Karthikeyan, J. J. Futina, P. P. Abinaya, and R. S. Priya, “Parkinson’s disease detection and classification of stages from drawing patterns using deep learning,” in Proc. 2023 Int. Conf. on Bio Signals, Images, and Instrumentation (ICBSII), 2023, pp. 1–7. DOI: 10.1109/ICBSII58188.2023.10181055.

[6] Y. Luo, T. Yu, J. Zheng, and Y. Ding, “Design of engineering drawing recognition system based on YOLO v4,” in Proc. 2022 IEEE 6th Information Technology and Mechatronics Engineering Conf. (ITOEC), vol. 6, 2022, pp. 1221–1225. DOI: 10.1109/ITOEC53115.2022.9734453.

Využitie papiera ako substrátu v elektronike

Filip Mahdal, Peter Lukács, Miroslav Jusko a Tomáš Lenger

Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Košice, Slovenská republika

peter.lukacs@tuke.sk

***Abstrakt* — Článok analyzuje možnosti využitia substrátov na báze papiera vo flexibilnej a nositeľnej elektronike ako ekologickej alternatívy ku konvenčným polymérnym fóliám. Popisujú sa základné typy papierových substrátov vrátane celulózových a aramidových vlákien, ich elektrické a mechanické vlastnosti, ako aj technologické obmedzenia a výhody. Zvlášť sú rozobrané prístupy na povrchovú úpravu papiera, ktoré zlepšujú jeho kompatibilitu s funkčnými vrstvami. Článok taktiež sumarizuje dostupné technológie nanášania elektricky vodivých materiálov, ako je sieťotlač, InkJet alebo netradičné metódy ako „pencil-on-paper". V druhej časti sú prezentované konkrétne aplikácie ako antény, OLED diódy či biosenzory integrované do papierového substrátu. Výsledky poukazujú na veľký potenciál papiera ako udržateľnej platformy pre nízkonákladovú elektroniku budúcnosti.**

***Kľúčové slova* — flexibilná elektronika, papierový substrát, aramid, ekologická elektronika**

I. Úvod

Dopyt po flexibilnej elektronike sa v posledných rokoch prudko zvyšuje. Avšak, v súčasnosti vzniká snaha znižovania emisií a ekologickej stopy pri výrobe elektroniky. Taktiež sa zvyšujú požiadavky na jednoduchú recykláciu elektroniky. Medzi konvenčné využívané materiály používajúce sa ako substráty využívané vo flexibilnej elektronike patria polyetyléntereftalát (PET), polyetylénnaftalát (PEN), polyimid (PI) a ďalšie. Niektoré z týchto materiálov je možné jednoducho recyklovať. Avšak doba potrebná na kompletné rozloženie polymérnych flexibilných substrátov sa pohybuje v stovkách rokov. Kvôli jednoduchej recyklovateľnosti a rozložiteľnosti sa javia substráty na báze papiera ako vhodná substitúcia za konvenčne využívané substráty [1], [2].

Myšlienku používania papiera ako substrátu uviedli v šesťdesiatych rokoch minulého storočia vedci na čele s Brodym. Avšak ich metóda pre finančnú náročnosť pri výrobe nebola konvenčne zavedená [1].

Medzi vlastnosti konvenčne využívaného papiera patria vysoká miera nasiakavosti, vysoká drsnosť povrchu. Avšak tieto vlastnosti nie sú žiadúce a ani vhodné pre využitie papiera ako substrátu v oblasti flexibilnej elektroniky. Napriek tomu je však možné potlačiť tieto nežiadúce charakteristiky papiera pomocou nanášania ochranných vrstiev polyetylénu, polypropylénu, polyetyléntereftalátu, vosku alebo hliníka. Tieto vrstvy taktiež vykonávajú úlohu ochrany pred environmentálnymi vplyvmi prostredia ako je vlhkosť. Taktiež zamedzujú prílišnému absorbovaniu funkčných atramentov do vnútorných štruktúr papiera, a tým zlepšujú tlačiteľnosť na papierové substráty. Avšak príliš hladký povrch spôsobuje rozptyl atramentu do strán čo má za následok zníženie rozlíšenia tlače [1], [2].

Okrem iného konvenčný papier obsahuje vo svojej štruktúre póry, ktoré za bežných podmienok obsahujú vzduch ktorý závažnou mierou ovplyvňuje dielektrické vlastnosti. Tento jav je spôsobený nízkou relatívnou permitivitou vzduchu ktorá dosahuje hodnotu približne 1. Preto sa relatívna permitivita konvenčného papiera pohybuje v intervale od 1,3 do 4. To je však možné ovplyvniť vyplnením pórov materiálom s vyššou relatívnou permitivitou. Napriek možnostiam jednoduchého ovplyvňovania a zlepšovania vlastností papiera sa na papieroch zvyčajne očakáva vyšší elektrický odpor ako na plastových fóliách. Priblížený povrch papiera je možné vidieť na Obr. 1 pričom si je možné všimnúť pórov v jeho štruktúre [2].

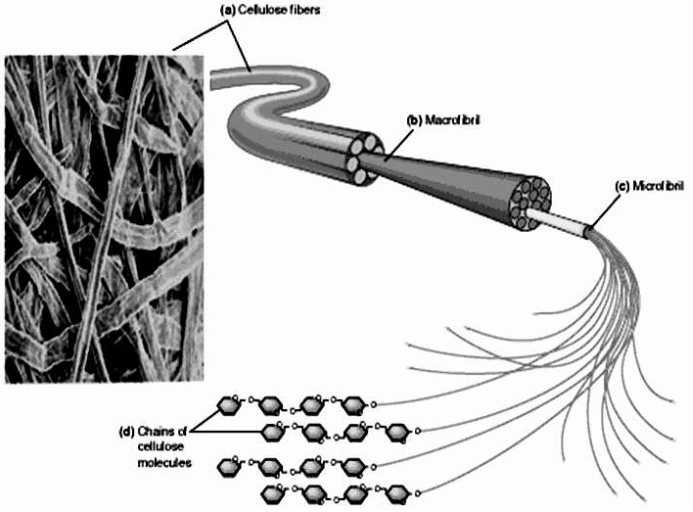
Obr. 1 Povrch papiera zobrazený pomocou elektrónového mikroskopu [3]

Elektronika na báze papiera je široko skúmanou oblasťou. Papier sa hodnotil ako substrát pre termochromické a elektrochromické displeje, odporové pamäťové zariadenia, tranzistory, kondenzátory, jednorazové štítky na rádiofrekvenčnú identifikáciu (RFID), batérie, fotovoltické články a senzory a aktuátory. V posledných rokoch sa rozšírilo používanie substrátov na báze papiera v diagnostických, farmaceutických, energetických a nositeľných aplikáciách vďaka vysokej priedušnosti papiera, ktorá vyplýva z jeho pórovitosti, flexibility a udržateľnosti. Príkladom týchto zariadení sú nositeľné snímače tlaku a vlhkosti, termoelektrické generátory, ako aj inteligentné obväzy. Elektronika na báze papiera bola demonštrovaná aj pomocou veľkoobjemových techník hromadnej výroby [4].

II. Papiere využívané ako substráty v elektronike

A. Papiere založené na báze celulózy

Papier je vo všeobecnosti obsahuje vo svojej štruktúre biopolymér s názvom celulóza. Avšak mechanické a elektrické vlastnosti jednotlivých druhov papiera sa od seba líšia v dôsledku rozdielov v zložení materiálu a výrobných procesov. Je zložité generalizovať súbor vlastností papiera, pretože obsah a typ zložiek papiera sa medzi jednotlivými typmi líšia. Každý typ papiera slúži na iný špecifický účel, funkciu na základe jeho štruktúry a zloženia. Tlačový papier sa skladá z celulózových vlákien s veľkým množstvom výplňového materiálu. Výplňovým materiálom môžu byť buď prírodné materiály (vápenec, íl a mastenec) alebo syntetické alternatívy (uhličitan vápenatý, oxid titaničitý a sadra). Výplňové materiály určujú výrobné náklady a vlastnosti papiera. Papier je nehomogénny materiál, pretože je to lisovaná zmes celulózy a plnidiel, jedná sa o štruktúru v ktorej dominujú vlákna celulózy a výplňového materiálu, táto štruktúra zostáva celkovo nerovnomerná a náhodná vo všetkých typoch papiera. Celulóza je taktiež náchylná na degradáciu po vystavení teplotám vyšším ako 100°C, čo vytvára prekážku pri výrobných procesoch elektroniky ktoré túto teplotu prekračujú bežne. Na Obr. 2 je možné vidieť štruktúru celulózového vlákna [5].



Obr. 2 Štruktúra celulózového vlákna [6]

Existujú taktiež bezdrevné papiere, avšak, ich názov je zavadzajúci pretože taktiež sú vyrábané z celulózy ktorej jedným z primárnych zdrojov je drevo. Rozdiel od drevnatých papierov spočíva v rozdielom obsahu a inej výrobnej technológie na získavanie celulózy.

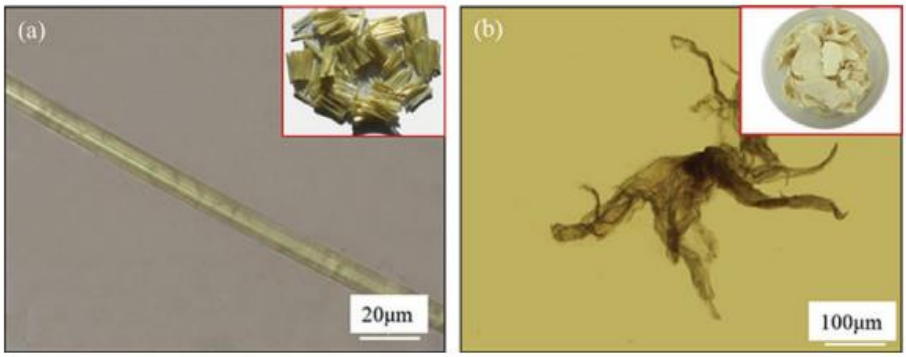
Celulóza potrebná na výrobu papiera sa môže získavať dvoma rôznymi spôsobmi. Prvým možným prístupom je získavanie celulózy z chemickej buničiny, pričom sa jednotlivé vlákna celulózy oddeľujú varením. Tento proces má za následok vznik dlhých celulózových vlákien a jednoduché odstránenie lignínu. Lignín spôsobuje žltnutie papieru a oslabovanie mechanickej pevnosti. Papier vyrobený touto metódou sa označuje ako bezdrevný papier. Celulózové vlákna oddelené danou technológiou sú hygroskopické, môžu však po exponovaní vlhkosti napučať až o 20% v smere hrúbky ale v pozdĺžnom smere napučiavajú len o 1%.

Druhým možným prístupom získavania celulózových vlákien z buničiny je mechanický výrobný proces. V referencii s chemickým procesom je tento proces lacnejší. Avšak vo výslednom papieri ostávajú zvyšky lignínu a hemiceulózy, ktoré spôsobujú krehnutie a žltnutie papiera pri vystavení UV žiareniu. Takto vyrobený papier sa označuje ako drevnatý [5], [6].

B. Papiere založené na báze aramidových vlákien

Ďalšou možnosťou ako sa vysporiadať s nepriaznivými vlastnosťami celulózy v papieri je čiastočná alebo úplná náhrada za aromatický polyamid, známy pod názvom aramid alebo kevlar. Tento materiál sa vyznačuje odolnosťou voči teplu, nehorľavosťou, vysokou mechanickou odolnosťou voči ťahu. Medzi najznámejšie papiere využívajúce vo svojej štruktúre aramidové vlákna patria produkty firmy DuPont pod menom Nomex [7], [8], [9].

Štruktúrnymi zložkami aramidového papiera sú vlákna, ktoré je možné rozdeliť do dvoch subkategórii a to flocs a fibrids. Aramidové flocky sú krátke úseky dlhých aramidových vlákien, ktoré majú dĺžku približne 6 mm, a priemer dosahuje hodnoty 10 µm. Aramidové fibridy nadobúdajú nepravidelný tvar, majú dĺžku 0,2 až 1 mm a hrúbku menej ako 1 µm. Fibridy pôsobia ako spojivo, ktoré udržujú aramidové vlákna pohromade, a tak tiež plnia úlohu plnivá v aramidovom papieri. Adhézna sila medzi aramidovými flokmi a aramidovými fibridmi je päťkrát slabšia ako sila pôsobiaca vzájomne medzi fibridmi. Adhézna sila medzi fibridmi je väčšia pretože ich kontaktná plocha je väčšia ako pri vláknach. Na Obr. 3 je možne vidieť aramidové vlákna [7], [8], [9].



Obr. 3 Aramidové vlákna (a) flocs (b) fibrids [8]

Papiere na báze aramidových vlákien sú primárne vyžívané ako izolačné vrstvy v transformátoroch. Medzi ich vlastnosti patrí vysoká dielektrická pevnosť, mechanická odolnosť, pružnosť, flexibilnosť, tepelná odolnosť. Taktiež je možné ho využívať aj ako substrát. Nomex 410 obsahuje vo svojej štruktúre výlučne aramid, pri výrobe je kalandrovaný pre zlepšenie jeho vlastnosti. Naopak Nomex 411 nie je kalandrovaný čo spôsobuje výrazné zhoršenie mechanickej odolnosti, tlačiteľnosti a dielektrických vlastnosti. Na Obr. 4 je možné vidieť Nomex 410 [10], [11], [12].

Obr. 4 Nomex 410 [13]

V Tabuľke. 1 je možné vidieť porovnanie vlastnosti kalandrovaného aramidového papiera Nomex 410 a nekalandrovaného Nomex 411. Pričom si je možné všimnúť horšie vlastnosti nekalandrovaného papiera [11], [12].

Tabuľka. 1 Porovnanie vybraných vlastností Nomex 410 a Nomex 411 [11], [12]

Vlastnosti	Nomex 410	Nomex 411
Hrúbka [mm]	0,25	0,25
Hustota [g/cm^3]	0,96	0,31
Pevnosť v ťahu [N/cm] MD	296	35
Pevnosť v ťahu [N/cm] XD	161	20
Dielektrická konštanta (60 Hz)	1,2	2,7

MD- pozdĺžny smer, XD- priečny smer

.

III. MOŽNOSTI NANÁŠANIA FUNKČNÝCH VRSTIEV

Pre nanášanie funkčných vrstiev na povrchu substrátu sú využívané rôzne aditívne technologické metódy. Medzi tieto technológie patrí sieťotlač, Inkjet Printing, rotačná hĺbkotlač, ofsetová tlač, flexografia.

Sieťotlač je vhodná na vytváranie hrubých vrstiev približne do hrúbky 100 µm avšak rozlíšenie tlače nedosahuje vysokej úrovne a to maximálne 50 µm čo je v referencii s ostatnými výrobnými technológiami horší výsledok [2].

Technológia Injekt Printing dokáže tlačiť vrstvy o hrúbke 0,3 µm, pri rozlíšení 20 µm. Medzi výhody tejto technológie patrí jednoduchá zmena tlačového motívu, malá spotreba funkčného materiálu. Využívaná je na vytváranie antén, tenko vrstvových tranzistorov [2].

Rotačná hĺbkotlač, ofsetová tlač a flexografia využívajú na nanášanie motívu tlačový valec na ktorom povrchu je vygravírovaný tlačový motív alebo upevnený reliéf vyrobený z mäkkého materiálu ako napríklad guma. Principiálne sa tieto technológie líšia iba počtom valcov a ich usporiadaním. Využívané sú hlavne na výrobu veľkých sérii pretože zmena tlačového motívu je nákladná a zložitá, avšak rýchlosť tlače je násobne vyššia ako v porovnaní so sieťotlačou a Inkjet printingom. Rozlíšenie tlače dosahujú obdobné hodnoty ako pri Inkjet printingu [2].

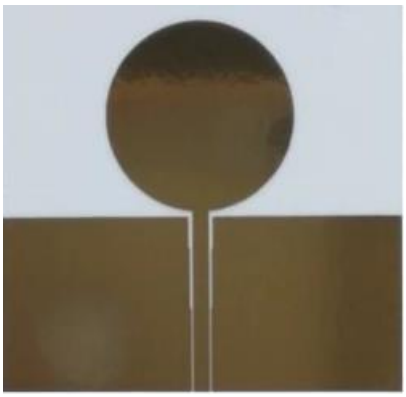
Jednou z ďalších možných metód nanášania funkčných vrstiev na povrch papierového substrátu je experimentálna metóda Pencil on paper. Táto metóda využíva grafit v ceruzke na nanesenie motívu na papier. Adhézia grafitu je zabezpečená drsnosťou papiera [14].

IV. VYBRANÉ APLIKÁCIE PAPIERA V ELEKTRONIKE

V nasledujúcich podkapitolách budú priblížené problematiky a princípy fungovania niektorých vybraných aplikácii využívajúcich vo svojej štruktúre papier.

A. Antény

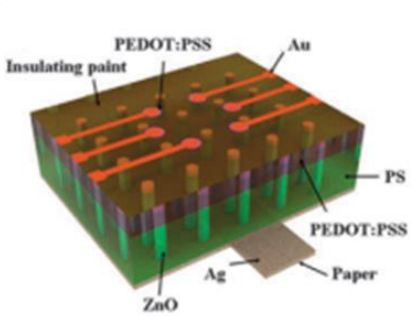
Medzi najviac rozšírené aplikácie papiera vo flexibilnej elektronike patria antény integrované v papierových substrátoch. Využívané sú napríklad ako zabezpečovacie zariadenie proti krádežiam. Pre korektné fungovanie antén je potrebné vybrať správny materiál prípadne ho upraviť spôsobom aby dokázal vydržať proces sintrovania, vplyv vlhkosti a teplôt, mal vhodné elektrické vlastnosti. Tlač antén na povrch papiera si vyžaduje vysporiadať sa s nepravidelným, drsným povrchom substrátu. Jedným z riešení je vytlačenie antény na povrch hladkého polyméreneho substrátu a následne integrovanie do štruktúry papiera. Ďalším z riešení je pokrytie povrchu ochrannou vrstvou (Coating), alebo kalandrovanie. Na vytvorenie antény je možné zvoliť ľubovoľnú výrobnú technológiu. Mechanické namáhanie antény môže spôsobiť zmenu rezonančnej frekvencie a elektrickej vodivosti. Na Obr. 5 je možné vedieť anténu vytlačenú za pomoci Inkjet printingu na povrchu papierového substrátu [15].



Obr. 5 UWB anténa vytlačená na papierovom substráte [15]

B. Flexibilné organické svetlo emitujúce diódy

Organické svetlo emitujúce diódy (OLED) konvertujú elektrickú energiu na svetelnú energiu. Svetlo je emitované do prostredia vďaka rekombinácii elektrónov v organických častiach OLED. Tým vedcov na čele so Zhangom vyvinuli svetlo emitujúcu diódu s použitím papierového substrátu. Ako substrát použili chromatografický papier značky Whatman, pre vysporiadanie sa s drsnosťou a nasiakavosťou zvolili vedci metódu ošetrenia povrchu. Následne bola nanesená na takto upravený papier s pomocou sieťotlače vodivá strieborná vrstva. Po nanesení ďalších vrstiev vznikla dióda ktorá emitovala žlto-zelené svetlo. Túto svetlo emitujúcu diódu je možné vidieť na Obr. 6 [16].

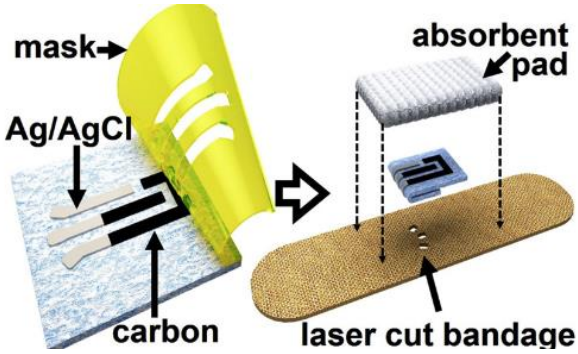


Obr. 6 Štruktúra svetlo emitujúcej diódy na papierovom substráte [16]

C. Nositeľný senzor monitorujúci stav rany

Chronické rany ktoré sa ťažko hoja spôsobujú ťažkosti ľudom a zdravotnému systému. Preto existuje potreba tieto rany lacno a jednoducho monitorovať. Pre túto potrebu vznikol nositeľný senzor v podobe náplasti ktorá sleduje hodnoty pH a kyseliny močovej. Na základe nameraných koncentrácii sa vie identifikovať bakteriálna infekcia v rane. Sieťotlačou bola nanesená senzorová časť na papier, pri tejto aplikácii bolo dôležité zachovanie priedušnosti papiera a preto

bol na povrch nastriekaný omnifóbny roztok. Na Obr. 7 je možné vidieť náplasť monitorujúcu biomarkery [17].



Obr. 7 Nositeľný senzor monitorujúci stav rany integrovaný v náplasti [17]

V. ZÁVER

Ekologická transformácia elektronického priemyslu vyžaduje nové materiálové riešenia, ktoré sú lacné, dostupné, recyklovateľné a biologicky rozložiteľné. Papierové substráty, napriek svojim prirodzeným nevýhodám ako je pórovitosť, nasiakavosť či drsnosť povrchu, preukazujú po vhodnej úprave veľmi sľubné vlastnosti pre použitie vo flexibilnej elektronike. Aplikácie ako tlačené RFID a UWB antény, OLED diódy alebo biosenzory ukazujú reálny potenciál papiera ako nosnej platformy. V kombinácii s nízkoteplotnými technológiami nanášania funkčných vrstiev a vývojom nových hybridných papierov (napr. aramidových) sa papier čoraz viac stáva atraktívnou voľbou pre vývoj udržateľných elektronických systémov. Praktické aplikácie prezentované v článku – vrátane papierových antén, biosenzorov, OLED zariadení či RFID tagov – potvrdzujú, že papier ako substrát nie je len laboratórnou zaujímavosťou, ale má reálny potenciál nájsť uplatnenie v komerčnej výrobe, najmä v oblastiach jednorazovej alebo nositeľnej elektroniky. Dôležitým benefitom je aj možnosť integrácie biodegradovateľných komponentov, čím vznikajú elektronické systémy s minimálnym dopadom na životné prostredie

POĎAKOVANIE

Článok bol podporený projektmi KEGA č. 011TUKE-4/2023, APVV 22-0289 a VEGA č. 1/0049/24 a č. 1/0278/24 a financovaný Európskou úniou z programu NextGenerationEU prostredníctvom Plánu obnovy a odolnosti Slovenskej republiky v rámci projektu č. 09I03-03-V04-00640.

POUŽITÁ LITERATÚRA

[1] J. Liu *et al.*, "Future paper based printed circuit boards for green electronics: fabrication and life cycle assessment", *Energy Env. Sci*, roč. 7, č. 11, s. 3674–3682, 2014, doi: 10.1039/C4EE01995D.

[2] D. Tobjörk a R. Österbacka, "Paper Electronics", *Adv. Mater.*, roč. 23, č. 17, s. 1935–1961, máj. 2011, doi: 10.1002/adma.201004692.

[3] R.-J. Yang, H.-H. Pu, a H.-L. Wang, "Ion concentration polarization on paper-based microfluidic devices and its application to preconcentrate dilute sample solutions", *Biomicrofluidics*, roč. 9, č. 1, s. 014122, jan. 2015, doi: 10.1063/1.4913366.

[4] E. Jansson *et al.*, "Suitability of Paper-Based Substrates for Printed Electronics", *Materials*, roč. 15, č. 3, s. 957, jan. 2022, doi: 10.3390/ma15030957.

[5] S. M. Khan, J. M. Nassar, a M. M. Hussain, "Paper as a Substrate and an Active Material in Paper Electronics", *ACS Appl. Electron. Mater.*, roč. 3, č. 1, s. 30–52, jan. 2021, doi: 10.1021/acsaelm.0c00484.

[6] D. A. Bocchini Martins *et al.*, "Agroindustrial Wastes as Substrates for Microbial Enzymes Production and Source of Sugar for Bioethanol Production", v *Integrated Waste Management - Volume II*, S. Kumar, Ed., InTech, 2011. doi: 10.5772/23377.

[7] O. H. Arroyo-Fernández, J. Jalbert, E. M. Rodriguez-Celis, S. Duchesne, B. Morin, a I. Fofana, "Changes in mechanical properties of impregnated Nomex papers 410 and 910 during accelerated aging", *Polym. Test.*, roč. 83, s. 106358, mar. 2020, doi: 10.1016/j.polymertesting.2020.106358.

[8] Y. Zhao, W. Dang, Z. Lu, L. Wang, L. Si, a M. Zhang, "A novel mica-based composite with hybrid aramid fibers for electrical insulating applications: largely improved mechanical properties and moisture resistance", *Polym. Int.*, roč. 67, č. 2, s. 204–211, feb. 2018, doi: 10.1002/pi.5498.

[9] J. K. Fink, "Aramids", v *High Performance Polymers*, Elsevier, 2014, s. 301–320. doi: 10.1016/B978-0-323-31222-6.00013-3.

[10] P. Lukacs, A. Pietrikova, I. Vehec, a T. Lenger, "Long-Term Stability Analysis of the Paper-Based Nomex 410 Substrate's Dielectric Properties", v *2024 International Conference on Diagnostics in Electrical Engineering (Diagnostika)*, Pilsen, Czech Republic: IEEE, sep. 2024, s. 1–4. doi: 10.1109/Diagnostika61830.2024.10693901.

[11] "DuPont, 'DuPontTM Nomex® Paper Type 410'". [Online]. Available at: https://www.dupont.com/content/dam/dupont/amer/us/en/safety/public/documents/en/DPT16_21668_Nomex_410_Tech_Data_Sheet_me03_REFERENCE.pdf

[12] "DuPont, 'DuPontTM Nomex® Paper Type 411". [Online]. Available at: https://www.dupont.com/content/dam/dupont/amer/us/en/safety/public/documents/en/Nomex-Type-411.pdf

[13] J. Zhang, Y. Guo, Q. Chen, a C. Luo, "Kinetic Analysis, Solid Residues and Volatile Products of Pyrolyzing Nomex Insulation Paper in Nitrogen and Air Atmosphere", *Fire Technol.*, roč. 60, č. 2, s. 1119–1141, mar. 2024, doi: 10.1007/s10694-022-01266-7.
[14] N. Kurra a G. U. Kulkarni, "Pencil-on-paper: electronic devices", *Lab. Chip*, roč. 13, č. 15, s. 2866, 2013, doi: 10.1039/c3lc50406a.
[15] T. K. Saha, T. N. Knaus, A. Khosla, a P. K. Sekhar, "A CPW-fed flexible UWB antenna for IoT applications", *Microsyst. Technol.*, roč. 28, č. 1, s. 5–11, jan. 2022, doi: 10.1007/s00542-018-4260-0.
[16] Y. Zhang *et al.*, "Flexible paper-based ZnO nanorod light-emitting diodes induced multiplexed photoelectrochemical immunoassay", *Chem Commun*, roč. 50, č. 12, s. 1417–1419, 2014, doi: 10.1039/C3CC48421A.
[17] R. V. Martinez, "Wearables, E-textiles, and Soft Robotics for Personalized Medicine", v *Springer Handbook of Automation*, S. Y. Nof, Ed., v Springer Handbooks. , Cham: Springer International Publishing, 2023, s. 1265–1287. doi: 10.1007/978-3-030-96729-1_59.

Flexibilné mikrozberače energie z ľudského tela

Jerguš Ševec, Peter Lukács, Miroslav Jusko a Tomáš Lenger

Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Košice, Slovenská republika

peter.lukacs@tuke.sk

***Abstrakt* — Článok spracováva problematiku mikrozberačov energie z ľudského tela v zábere flexibilnej elektroniky. Súčasný vývoj kladie čoraz väčšie nároky na flexibilitu, účinnosť a nízke rozmery využívaných zdrojov energie pre potreby nositeľných a implantovateľných biomedicínskych zariadení, nositeľnej elektroniky a rôznych zariadení IoT. Voľa takýchto systémov umožňuje eliminovať potrebu pravidelnej výmeny konvenčných elektrochemických článkov. V práci je spracovaný prehľad potenciálnych zdrojov neelektrickej energie využiteľných k premene na energiu elektrickú. Taktiež sú v práci vysvetlené fundamentálne princípy funkčnosti jednoltivých využívaných prístupov, ako aj praktické príklady súčasných aplikácií vo vybraných oboroch.**

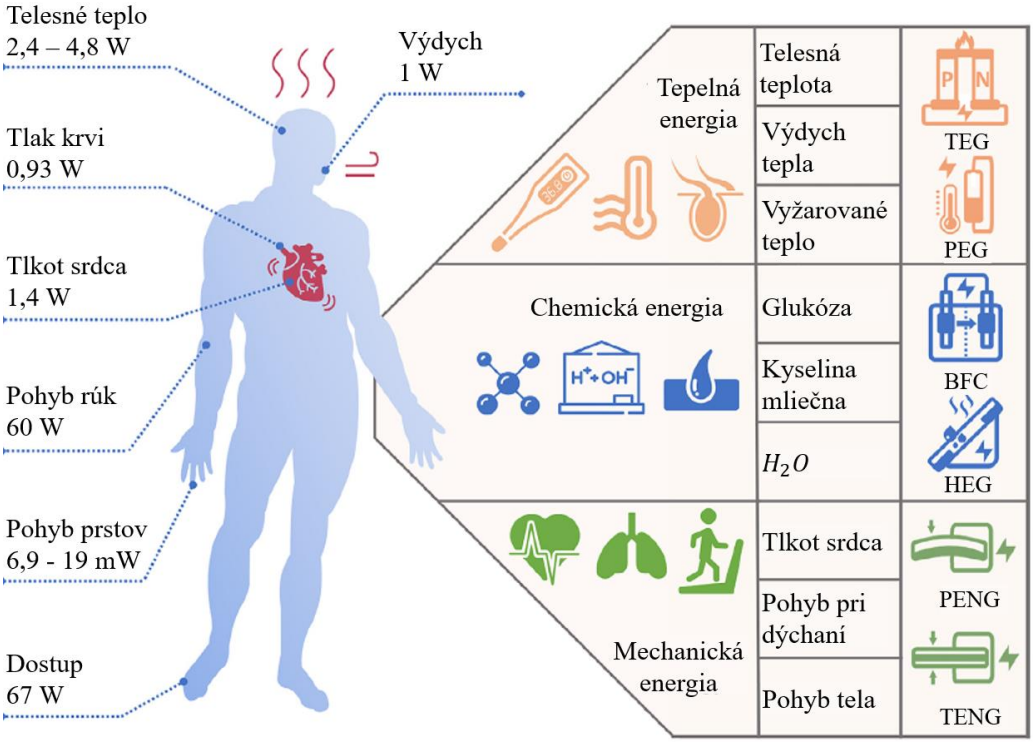
***Kľúčové slová* — flexibilné mikrozberače energie, flexibilná elektronika, biomedicínske zariadenia**

I. Úvod

Mikrozberače energie predstavujú štruktúry o rozmeroch rádovo v mikrometroch, poprípade nanometroch, ktorých účel spočíva v transformovaní neelektrickej energie produkovanej ľudským telom na elektrický potenciál schopný napájať rôzne elektronické zariadenia, ako sú napríklad nositeľná elektronika, rôzne implantovateľné a nositeľné biomedicínske prístroje, zariadenia IoT, elektronická koža a rôzne ďalšie. Účelom týchto zariadení je monitorovanie telesných funkcií pacientov, podávanie liečiv, komplexné vyhodnocovanie pohybu pri športe, poprípade rôzne ďalšie špecifickejšie účely. V dnešnej dobe je vo vedeckých kruhoch vynakladané čoraz väčšie úsilie pri vývoji a kompatibilite týchto systémov s ľudským telom s cieľom zvyšovať účinnosť premeny zbieranej energie a nahradiť tak konvenčné riešenia založené na elektrochemických článkoch, ktoré neposkytujú dostatočné kvality pre použitie vo flexibilnej elektronike. Tá vyžaduje vysokú mieru flexibility komponentov, prispôsobivosť nerovnostiam ľudského tela, pomerne vysokú účinnosť premeny energie spojenú s vysokou mierou integrovateľnosti a odstránenie potreby pravidelnej výmeny napájacích článkov.

II. Dostupné zdroje energie v ponímaní ľudského tela

Ľudské telo ako také predstavuje prirodzenú zásobáreň rozličných druhov energie s potenciálom ich konverzie na energiu elektrickú. V bežnej praxi sú využívané rôzne mechanizmy založené na zbere energie z chemických reakcií súvisiacich s metabolickými procesmi, ako napríklad chemické komponenty prítomné vo vylučovanom pote, slzách, slinách, poprípade iných chemických látok. Zber tepelnej energie súvisí s využívaním vyžarovaného tepla z organizmu do prostredia, v ktorom sa nachádza, či už v ponímaní telesnej teploty, samotného tepelného žiarenia, alebo vydychovanej tepelnej energie. V neposlednom rade zdroje mechanickej energie predstavujú rôzne vibrácie, nárazy, alebo ohyb končatín. Energiu tak možno zbierať z každodenného vedomího pohybu, ako je gestikulácia tváre, rúk, chôdza, dostup chodidiel. Potenciálny zdroj predstavuje aj pohyb nevedomý. Ide o pohyb hrudného koša pri dýchaní, pohyb respiračnej sústavy, tlkot srdca, poprípade pulzovanie ciev. Tieto tri základné zásobárne neelektrických druhov energie predstavujú základné delenie princípov činnosti dostupných technológií mikrozberačov energie z ľudského tela. Prehľad potenciálnych zdrojov energie a využívané technológie zberu spolu s aktuálne dostupným výkonom môže čitateľ pozorovať na Obr. 1.

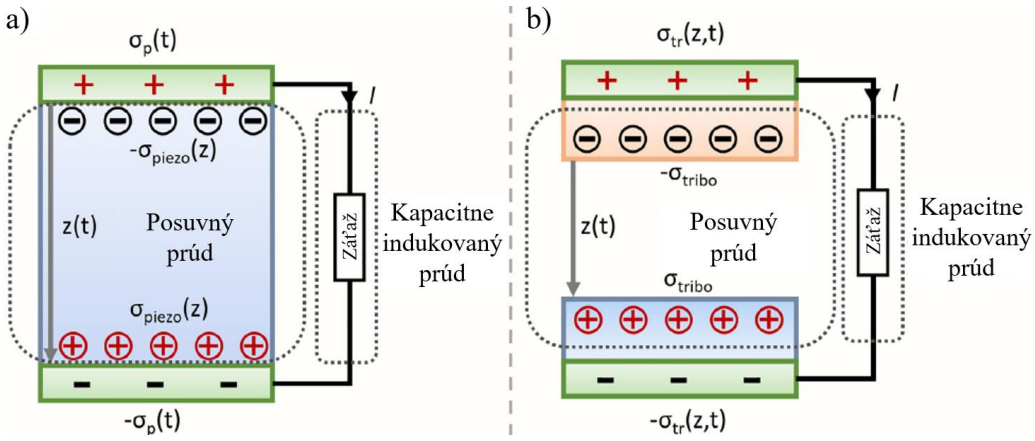


Obr. 1 Distribúcia a charakter potenciálnych zdrojov neelektrickej energie pre konverziu na energiu elektrickú a využívané princípy na tento účel [1]

III. PREMENA MECHANICKEJ ENERGIE

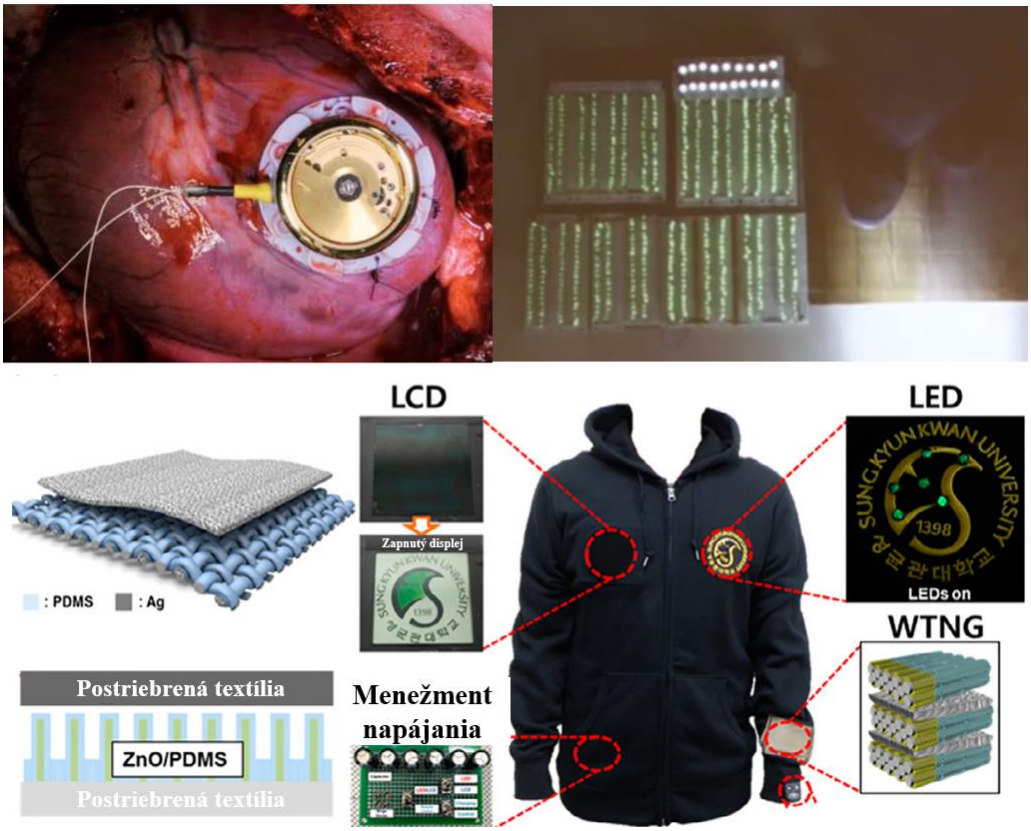
Hlavným problémom v prípade konverzie mechanickej energie z pohybu človeka predstavuje fakt, že tá má dominantne nízkofrekvenčný a zároveň nepravidelný charakter. Bežne dostupné technológie založené na generovaní elektromagnetického poľa však operujú pri rádovo vyšších frekvenciách, čo vytvára precedens potreby nových spôsobov zberu energie pracujúcich na odlišných fyzikálnych princípoch. Dominantne využívané technológie na konverziu mechanickej energie preto predstavujú piezoelektrické (PENG – Piezoelectric nanogenerator) a triboelektrické (TENG – Triboelectric nanogenerator) nanogenerátory. Tie možno považovať za protikladnú technológiu, pretože pracujú s prevádkovými frekvenciami vhodnými pre použitie na zber energie práve z pohybov ľudského tela, pričom ďalší rozdiel predstavuje tvorba vysokých napätí miesto vysokých prúdov. Je tak možné povedať, že nadobúdajú napäťový charakter [2].

Triboelektrické nanogenerátory využívajú triboelektrický jav, teda vznik napätia vplyvom kombinácie elektrifikácie materiálov spojenej s ich vzájomným trením a elektrostatickej indukcie. Vzniká tak kombinácia posuvného prúdu vo vnútri materiálu a jeho kapacitne indukovaného proťaljšku na povrchu. V praxi sú využívané štyri pracovné módy týchto zariadení. Ide o opakovanú separáciu kontaktov, trenie vzniknuté planárnym posunom, jednoelektródový princíp a princíp voľne ležiaceho telesa. Oba spomínané javy sú graficky znázornené na Obr. 2.



Obr. 2 Princíp činnosti PENG (vľavo) a TENG (vpravo) [1]

Piezoelektrické nanogenerátory pracujú na báze piezoelektrického javu. Ten je prítomný v polovodičových látkach, ktorých vnútorná štruktúra oplýva stredovou nesúmernosťou, pričom sú prítomné iónové väzby medzi molekulami. Vplyvom mechanického namáhania materiálu a jeho následnej deformácie dochádza ku predkupeniu vnútornej mriežky takým spôsobom, pri ktorom vznikajú dipólové momenty, čo ma za následok vznik rozdielu v potenciáli medzi elektródami umiestnenými na protiľahlých stranách deformovaného materiálu. Nutné je však spomenúť, že tieto piezoelektrické vlastnosti sú prítomné iba po takzvanú Curieho teplotu, po prekročení ktorej dochádza ku usporiadaniu vnútornej štruktúry materiálu a strate stredovej nesúmernosti [1].



Obr. 3 Praktické aplikácie technológie PENG a TENG. Kardiostimulátor napájaný PENG (vľavo vyššie) [3], Demonštrácia TENG ako snímača prítomnosti (vpravo vyššie) [2] a Textília so zabudovaným TENG pre napájanie nositeľnej elektroniky (nižšie) [4]

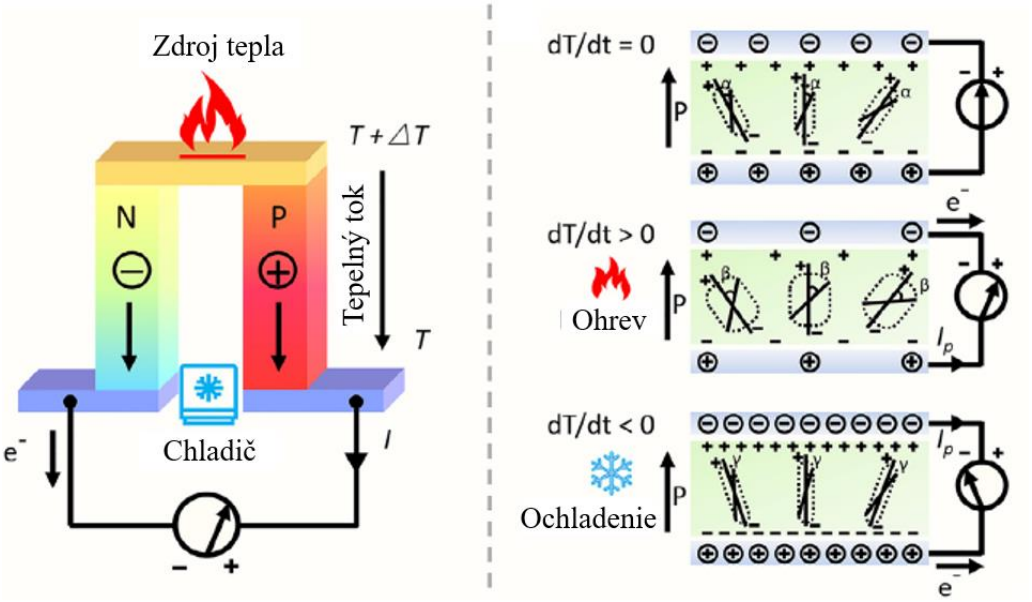
Z hľadiska konkrétneho použitia v praxi je možné spomenúť niekoľko vyvíjaných technológií. Na obrázku Obr. 3 (vľavo vyššie) demonštruje zabudovanie PENG do implantovateľných medicínskych zariadení. Takýto prístup ku konštrukcii odstraňuje potrebu výmeny batérií, čím sú znižované riziká spojené s opakovanými operáciami. Jednu z možností integrácie TENG predstavuje zabudovanie do podlahového systému (Obr. 3 vpravo vyššie), čím je možné vytvoriť napríklad inteligentné osvetlenie. Poprípade detekciu prítomnosti osôb. Ďalší príklad opisuje zabudovanie TENG textílií do oblečenia. Takýmto spôsobom je možné napájať rôznu nositeľnú elektroniku, technológie IoT, poprípade v textílii zabudované výstražné prvky.

IV. VYUŽITIE ODPADOVÉHO TEPLA

V bežnej praxi sú zastúpené dva prístupy k zberu energie z prebytočného tepla. Ide o termoelektrický mechanizmus založený na Seebeckovom jave (TEG) a Pyroelektrický princíp (PEG). Seebeckov jav predstavuje usmernený tok elektrónov prítomný v kovoch a polovodičoch spôsobený difúziou nosičov náboja z miesta s vyššou teplotou do miesta s teplotou nižšou. Z hľadiska generácie energie ide o zdroje jednosmerného napätia, kde sú jednotlivé články zapájané do sérioparalelných štruktúr obdobne ako v prípade elektrochemických článkov.

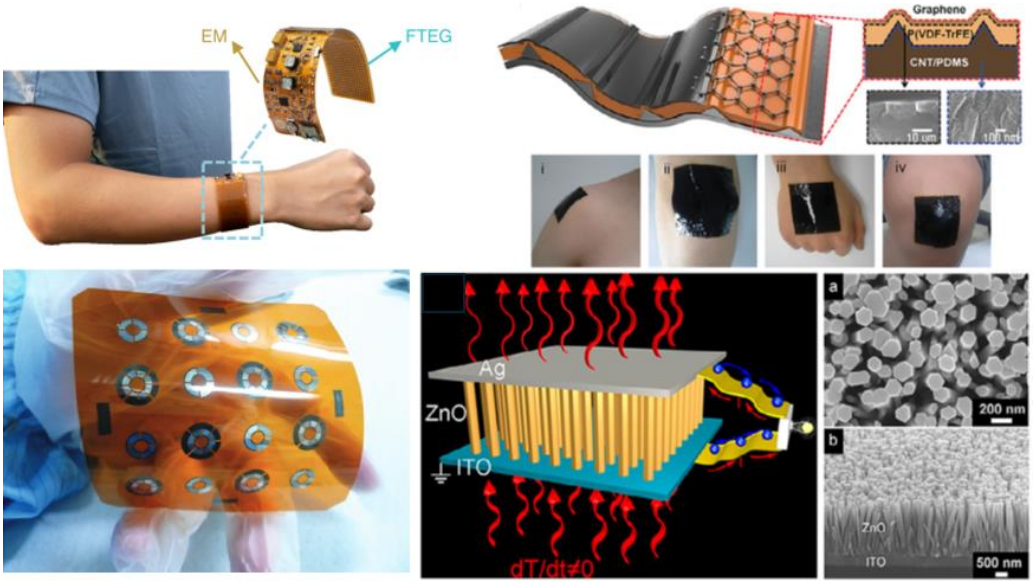
Pyroelektrizmus predstavuje jav, pri ktorom dochádza k povrchovej polarizácii u látok oplývajúcich vlastnosťami spontánnej polarizácie vnútornej štruktúry vplyvom zmeny teploty v čase. Sekundárne je tento jav spôsobený aj vplyvom tvorby vnútorných pnutí a rozťažnosti materiálu, ktorý psôsobuje dodatočnú tvorbu náboja. Zásadný rozdiel medzi jednotlivými

technológiami teda spočíva v skutočnosti, že zatiaľ čo TEG zbiera energiu z priestorového teplotného gradientu, PEG je schopný napájať zariadenia iba v prípade zmeny časovej [1], [2].



Obr. 4 Princíp činnosti TEG (vľavo) a PEG (vpravo) [1]

TEG je vzhľadom na svoj charakter zvyčajne integrovaný na zber energie z teplotného rozdielu medzi telesnou teplotou a vonkajším prostredím. Na Obr. 5 (vľavo vyššie) je znázornená flexibilná štruktúra (FTEG) s prítomnou správou napájania (PM) vo forme nositeľného náramku. Obr. 5 (vľavo nižšie) predstavuje vyvinutý flexibilný TEG na Národnej Taiwanskej univerzite v Taipei založený na báze zliatin s využitím Bi a Te. Štruktúra bola vytvorená pomocou deponovania z plynnej fázy. Príklady uvedené na pravej strane Obr. 5 predstavujú implementáciu PEG systémov. V hornej časti je vyobrazená kompozitná štruktúra vysokoelastického hybridného nanogenerátora využívajúceho kombináciu princípov PEG s PENG. Trojvrstvá štruktúra kombinuje vrchnú elektródu tvorenú grafénom, spodnú elektródu na báze PDMS-CNT a funkčnú vrstvu tvorenú P(VDF-TrFE). V spodnej časti sa nachádza experimentálna implementácia PEG vo forme nanodrôtov založených na zliatine ZnO. V tomto prípade je vrchná elektróda tvorená striebornou mikrovrstvou, zatiaľ čo spodná je tvorena oxidom cínu india (ITO). Podľa uvedených údajov je daná štruktúra schopná generovať 50 – 80 mVm^2/W [2], [5], [6], [7].



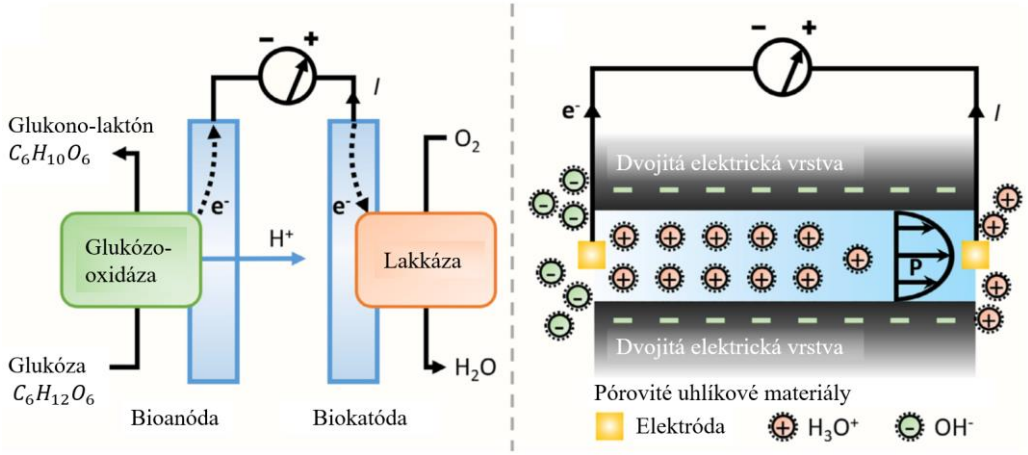
Obr. 5 Modelové aplikácie technológií TEG a PEG konvertujúcich tepelnú energiu na energiu elektrickú. Flexibilný termoelektrický generátor (FTEG) s manažmentom napájania (EM) (vľavo vyššie) [2], FTEG na báze zliatin využívajúcich Bi a Te (vľavo nižšie) [5], vysokoelastický hybridný nanogenerátor (vpravo vyššie) [6] a kompozitná štruktúra využívajúca nanodrôty na báze ZnO (vpravo nižšie) [7]

V. CHEMICKÉ ZDROJE ĽUDSKÉHO TELA

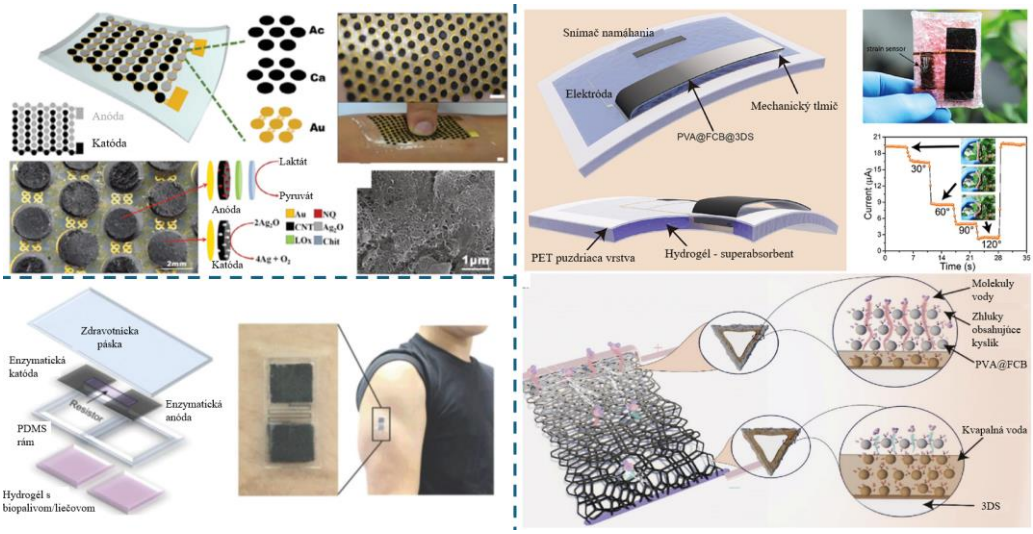
Zber eneregie z chemických zdrojov prítomných v ľudskom tele je založený na elektrónovej výmene medzi prítomnými chemickými zlúčeninami. Využité môžu byť telesné tekutiny, ako napríklad pot, slzy, sliny a iné. Takáto technológia je označovaná ako biopalivový článok (BFC). Druhá metóda, hydrovoltické generátory využíva reakcie molekúl vody s nanomateriálmi na zber energie z vyparovania, poprípade zmeny vlhkosti prostredia (HEG).

BFC operujú na báze redoxnýxh reakcií palivových látok obsiahnutých v živých organizmoch (príkladom môže byť glukóza, poprípade laktát). U organickej zlúčeniny tak dochádza súčasne k oxidácii (strate elektrónov) a redukcii (zisku elektrónov). Vo všeobecnosti sú tieto systémy rozdeľované na enzymatické a mikrobiálne. Mikrobiálne BFC oplývajú lepšou životnosťou a stabilitou z dôvodu využívania živých organizmov, avšak značnú nevýhodu predstavuje prítomné riziko infekcií a znížené difúzne podmienky vplyvom prítomnosti biofilmu [2].

K vzniku hydrovoltického javu dochádza pri interakcii molekúl vody s nanomateriálmi, ako sú napríklad karbónové nanotrubice (CNT), poprípade grafén. V zásade ide o súbor rôznych elektrifikačných procesov na fázovom rozhraní kvapaliny a tuhej látky, ktoré zahŕňajú elektrokinetické javy a javy súvisiace s odparovaním kvapalín [1].



Obr. 6 Princíp činnosti BFC na báze enzymatického štepenia glukózy (vľavo) a HEG na báze uhlíkových nanomateriálov (vpravo) [1]



Obr. 7 Praktické ukážky využitia enzymatických BFC a HEG. Enzymatický BFC systém pre zber energie z ľudského potu (vľavo hore) [8], Kožná náplasť na podávanie liečiv na báze BFC (vľavo dole) [2] a HEG určený na napájanie nositeľnej elektroniky (vpravo) [9]

Enzymatické BFC je možné aplikovať na zber energie z ľudského potu pre účely nositeľnej elektroniky (viď Obr. 7 vľavo hore). Zvolená bola štruktúra na báze zlata tvorená ostrovčekmi s meandrovitými prepojeniami z dôvodu zvýšenia flexibility štruktúry. Katóda je tvorená nanotrubicami na báze uhlíka a Ag_2O, pričom katóda je tvorená hustou spleťou CNT [8].

Pomocou technológie BFC je taktiež možné stimulovať transdermálne podávanie liečiv. Ukážka takejto kožnej náplasti je vyobrazená na Obr. 7 (vľavo dole). Kožná náplasť využíva

princíp takzvanej osmotickému transdermálnemu presunu do podkožného priestoru. Vyobrazená ilustrácia využíva uhlíkový výplet s cieľom zvýšiť pohyblivosť náboja (zvyšuje sa tak účinnosť zariadenia) [2].

Na Obr. 7 (vpravo) je vyobrazená štruktúra HEG pre nositeľnú snímaciu elektroniku. Zariadenie pozostáva z hydrogélu určeného pre potreby absorbcie, PED/PDMS puzdra, dvoch medených elektród a PVA@FCB@3D filmu. Elektrická energia je zbieraná počas vyparovania sa vody z povrchu PVA@FCB@3D filmu, pričom voda za pomoci kapilárnych síl preteká funkčnými PVA@FCB nanočasticami [9].

VI. SÚČASNÝ STAV A PERSPEKTÍVA

Technológia flexibilných zberačov energie prestavuje obor s rastúcim potenciálom. Napriek neustálym pokrokom, tieto systémy ešte niesú schopné úplne nahraadiť konvenčné riešenia v plnom rozmedzí. Ďalší vývoj kladie dôraz na zvyšovanie účinnosti zberu energie PEN a TENG (v súčasnosti najvyššie dosahované výkony). Termoelektrické generátory taktiež narážajú na technické problémy vyplývajúce z nízkej účinnosti zberu energie pri malých teplotných gradientoch. Technológie využívané pre zber energie z chemických zdrojov ľudského tela čelia striktným požiadavkám na vysokú biokompatibilitu a zdravotnú nezávadnosť. Súčasné trendy v oblasti výskumu odpovedajú na zvyšujúce sa nároky na vývoj nových materiálov s cieľom zlepšovať flexibilitu, odolnosť a starnutie materiálov. Jedným z takýchto smerov pracuje so zahrnutím tekutých kovov do súčasných štruktúr. Implementáciou rôznych eutektických zlúčenín, ako napríklad eutektika gália a india je možné preklenúť nevýhody využívaných vodivých vrstiev súvisiacich hlavne s relatívne vysokým vnútorných odporom týchto látok, s cieľom ďalšie navyšovania účinnosti zberu energie. Navyše takýmto spôsobom je možné zabezpečiť ďalšie kvality, ako napríklad prirodzenú regeneráciu materiálu [2].

VII. ZÁVER

Jedným z dôležitých v súčasnosti skúmaných faktorov implementácie pokročilých flexibilných zariadení v oblasti implantovateľných a nositeľných medicínskych zariadení, nositelnej elektroniky a zariadení IoT predstavuje prístup k napájaniu týchto systémov. Konvenčné metódy implementujúce elektrochemické články a batérie neposkytujú požadovanú mieru flexibility štruktúr, prispôsobivosti k povrchu ľudského tela, účinnosti, spoľahlivosti a oplývajú potrebou pravidelnej výmeny vybitých degradovaných článkov, pričom tie následne vytvárajú riziká pri potrebách opetovných operácií pri implantovateľných biomedicínskych zariadeniach, ako napríklad kardiostimulátoroch, nehovoriac o dopade na životné prostredie. Vyvíjané je preto čoraz väčšie úsilie na premenu neelektrickej energie produkovanej ľudským telom na energiu schopnú napájania týchto systémov. Sľubnú náhradu konvenčných riešení napájania tak poskytujú flexibilné mikrozberače energie premieňajúce mechanickú, chemickú, poprípade tepelnú odpadovú energiu z ľudského tela.

POĎAKOVANIE

Článok bol podporený projektmi KEGA č. 011TUKE-4/2023, APVV 22-0289 a VEGA č. 1/0049/24 a č. 1/0278/24 a financovaný Európskou úniou z programu NextGenerationEU prostredníctvom Plánu obnovy a odolnosti Slovenskej republiky v rámci projektu č. 09I03-03-V04-00640.

POUŽITÁ LITERATÚRA

[1] Zou, Y., Bo, L., & Li, Z. (2021). Recent progress in human body energy harvesting for smart bioelectronic system. Fundamental Research, 1(3), 364–385. doi:10.1016/j.fmre.2021.05.002

[2] Biswas, Swarup & Lee, Sang Won & Lee, Yongju & Choi, Hyo-Jeong & Chen, Jianjun & Yang, Xiao & Du, Yuxuan & Falcone, Natashya & Barros, Natan & Lee, Sung-Min & Kim, Hyeok & Khademhosseini, Ali & Yangzhi, Zhu. (2024). Emerging Energy Harvesters in Flexible Bioelectronics: From Wearable Devices to Biomedical Innovations. Small Science. 4. 10.1002/smsc.202300148.

[3] Chuck Seegert, Ph.D. Motion-Powered Pacemaker Runs Like Clockwork. Med Device Online. [Online] Dostupné na internete: https://www.meddeviceonline.com/doc/motion-powered-pacemaker-runs-like-clockwork-0001

[4] Ö. Faruk Ünsal a A. Çelik Bedeloğlu, "Recent Trends in Flexible Nanogenerators: A review", Material Science Research India, roč. 15, č. 2. Oriental Scientific Publishing Company, s. 114–130, júl. 30, 2018. doi: 10.13005/msri/150202. Y

[5] Hsiao, Chun-Ching & Wu, Yi-Syuan. (2011). Fabrication of flexible thin-film thermoelectric generators. Journal of the Chinese Institute of Engineers. 34. 809-816. 10.1080/02533839.2011.591552.

[6] Lee, J.-H., Lee, K. Y., Gupta, M. K., Kim, T. Y., Lee, D.-Y., Oh, J., … Kim, S.-W. (2013). Highly Stretchable Piezoelectric-Pyroelectric Hybrid Nanogenerator. Advanced Materials, 26(5), 765–769. doi:10.1002/adma.201303570

[7] Yang, Y., Guo, W., Pradel, K. C., Zhu, G., Zhou, Y., Zhang, Y., … Wang, Z. L. (2012). Pyroelectric Nanogenerators for Harvesting Thermoelectric Energy. Nano Letters, 12(6), 2833–2838. doi:10.1021/nl3003039
[8] Bandodkar, A. J., You, J.-M., Kim, N.-H., Gu, Y., Kumar, R., Mohan, A. M. V., … Wang, J. (2017). Soft, stretchable, high power density electronic skin-based biofuel cells for scavenging energy from human sweat. Energy & Environmental Science, 10(7), 1581–1589. doi:10.1039/c7ee00865a
[9] Li, L., Hao, M., Yang, X., Sun, F., Bai, Y., Ding, H., … Zhang, T. (2020). Sustainable and Flexible Hydrovoltaic Power Generator for Wearable Sensing Electronics. Nano Energy, 104663. doi:10.1016/j.nanoen.2020.104663

Nové trendy vo vývoji a využití elektrických strojov s permanentnými magnetmi

[1]*Ján KAŇUCH,* [2]*Peter GIROVSKÝ,* [3]*Norbert REŠOVSKÝ*
[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]norbert.resovsky@student.tuke.sk

***Abstrakt* — V tomto príspevku sú prezentované nové trendy vo vývoji a využití elektrických strojov s permanentnými magnetmi v súčasných priemyselných podmienkach, pričom sa zameriava na ich výhody a nevýhody v kontexte moderných aplikácií. V posledných niekoľkých desaťročiach elektrické stroje a pohony zohrali kľúčovú úlohu v elektrifikácii v širokej škále aplikácií. Bolo zavedených niekoľko iniciatív na nahradenie mechanických a hydraulických systémov elektrickými systémami a v prípade potreby na prechod na pokročilé pohony s premenlivou rýchlosťou. Článok poskytuje prehľad kľúčových faktorov elektrifikácie, rôznych aplikačných oblastí a niekoľko príkladov pokročilých elektrických strojov a pohonov, ktoré boli buď vyvinuté, alebo sa v súčasnosti vyvíjajú.**

***Kľúčové slová* — permanentný magnet, elektrické stroje s permanentnými magnetmi, vývoj elektrických strojov**

I. Úvod

Elektrické stroje a pohony v posledných desaťročiach zohrávali kľúčovú úlohu v elektrifikácii v širokej škále aplikácií. Veľa iniciatív sa usiluje o nahradenie mechanických a hydraulických systémov elektrickými systémami, čo je teda v prípade potreby prechod na pokročilé pohony s premenlivou rýchlosťou. Pokiaľ ide o výkon, spoľahlivosť, účinnosť a robustnosť, tak elektrické systémy a pohony ponúkajú veľmi významné výhody, pričom sa elektrické systémy stali trendom pre mnohé priemyselné odvetvia, ako je napr. trakcia, letectvo, ovládanie, ťažba ropy a zemného plynu a ďalšie iné priemyselné aplikácie. Široká téma obnoviteľných zdrojov a ich integrácie do siete súvisí s nasadením inteligentných sietí, tzv. mikrosietí a distribuovanej výroby. Tento posun smerom k väčšej elektrifikácii predstavuje významné výzvy, najmä pokiaľ ide o požiadavky na výkon elektrických strojov a pohonov. Teraz je jasné, že existuje kritická potreba najmodernejších elektrických strojov a pohonov, aby spĺňali požiadavky rôznych aplikácií vrátane drsných prostredí, vysokej hustoty výkonu, vysokej účinnosti a odolnosti voči poruchám v aplikáciách kritických aj z hľadiska bezpečnosti [1].

II. Topológie rotorov s permanentnými magnetmi

Rotor môže byť vnútorný alebo vonkajší vzhľadom na polohu statora. Vnútorná topológia (Obr. 1a) zaručuje dosiahnutie vyššej hodnoty amplitúdy hustoty toku vo vzduchovej medzere blízkej hustote magnetického toku permanentného magnetu (PM), čo umožňuje prevádzku v širšom rozsahu otáčok a vedie aj k zníženiu toku prúdu. Vyššie uvedené vlastnosti sa zhoršujú, keď sa zvoli konštrukcia s vonkajším rotorom (Obr. 1b).

a) b)

Obr. 1 Motor s PM s vnútorným a vonkajším rotorom.

Pre takýto motor je chladenie pomerne komplikované, pretože cievky vinutia statora (v ktorých Joulové straty sú zvyčajne hlavným zdrojom tepla) sú umiestnené vo vnútornej časti stroja. Kvôli horšiemu odvodu tepla sa pozornosť zamerala na vývoj pokročilých chladiacich systémov [2]. Naopak, montáž cievok vinutia je oveľa jednoduchšia, pretože zuby statora smerujú von. Okrem toho môže byť axiálna dĺžka motora menšia. Motory majú vysokú hodnotu pomeru vonkajšieho priemeru a axiálnej dĺžky, čo podstatne prospieva k zvýšeniu krútiaceho momentu. Čistá hmotnosť vonkajšieho rotora môže byť až o 15 % menšia ako zodpovedajúca hmotnosť vnútorného rotora [3]. Motory s vonkajším rotorom sú primárnou voľbou v aplikáciách, kde je potrebné zníženie mechanických strát pri prenose počas pohonu a tiež tam , kde je tiež možnosť rekuperačného brzdenia. Vyžadujú si však vhodnú a presnú konštrukciu a musia splniť nie len vysoké požiadavky na krútiaci moment, ale aj dosahovať potrebnú rýchlosť. Vyšší moment zotrvačnosti zároveň prispieva k zmierneniu vibrácií a zníženiu harmonických krútiaceho momentu. Z vyššie uvedených dôvodov sú motory s PM a s vonkajším rotorom dominantné pre použitie napr. v kolesách elektrických vozidiel. Vysokorýchlostná prevádzka nie je pre tento stroj veľmi vhodná, pretože tu pôsobia veľké odstredivé sily.

Pokiaľ ide o umiestnenie magnetov, existuje ďalšia klasifikácia pre rotory s PM. Z tohto hľadiska je uloženie magnetov:
a) povrchové (Obr.2a),
b) vnútorné (Obr.2b).

a) b)

Obr. 2 Rotor s povrchovo a vnútorne uloženými PM.

V prvom prípade sú PM pripevnené na vnútornom alebo vonkajšom obvode rotora a sú radiálne magnetizované. Efektívna dĺžka vzduchovej medzery motora sa rovná súčtu výšky PM a skutočnej dĺžky vzduchovej medzery, pretože magnetická permeabilita PM je porovnateľná s magnetickou permeabilitou vzduchu. Relatívne veľká efektívna dĺžka vzduchovej medzery znižuje vlastnú indukčnosť motora. PM sú však priamo vystavené reakčnému magnetickému poľu kotvy motora a teda existuje vysoké riziko, že budú náchylné na čiastočnú nezvratnú demagnetizáciu [4].

Ako už bolo spomenuté, v druhou prípade sú permanentné magnety buď vložené alebo zabudované do rotora a môžu byť magnetizované buď tangenciálne alebo radiálne. Najdôležitejšími výhodami sú väčšia mechanická odolnosť magnetov, ako aj lepšie tienenie proti demagnetizácii. Sú to v podstate stroje s vyniklými pólmi a preto indukčnosť v osi d je nižšia ako indukčnosť v osi q. Výsledný reluktančný moment motora s vnútornými magnetmi ale veľmi závisí nie len od rozmerov, polohy a konfigurácie PM, ale aj od umiestnenia a geometrie bariér, ktoré usmerňujú dráhu magnetického toku v rotore [5].

III. NOVÉ TRENDY VO VÝVOJI ELEKTRICKÝCH STROJOV S PERMANENTNÝMI MAGNETMI

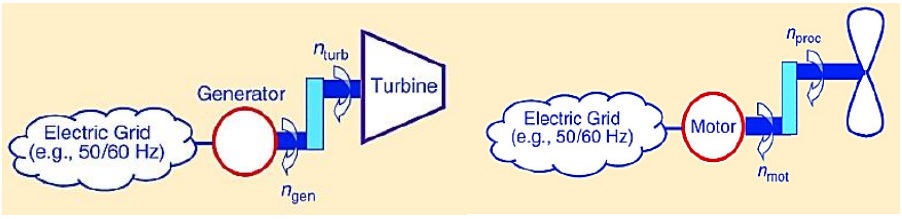
Pojem „elektrifikácia“ sa vo veľkej miere používa v širokej škále aplikácií. Elektrifikáciu možno definovať rôznymi spôsobmi:

- rozšírenie používania elektrických systémov na nahradenie mechanických a hydraulických systémov,
- vyššia penetrácia a integrácia obnoviteľných zdrojov energie,
- výmena starých elektrických pohonov s pevnou rýchlosťou za pokročilé výkonnejšie pohony s premenlivou rýchlosťou.

Stredobodom elektrifikácie je použitie nových elektrických strojov a pohonov. Existuje niekoľko mega trendov, ktoré prispeli k rastúcej vlne elektrifikácie:

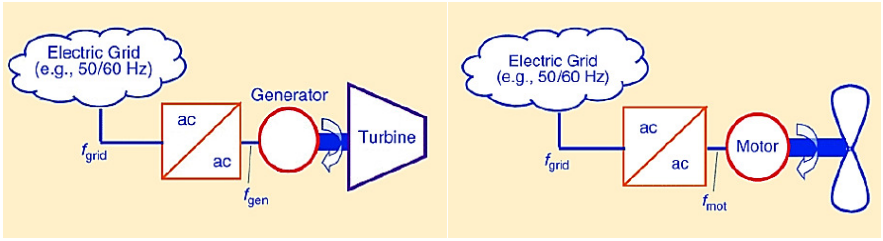
- elektrifikácia dopravy,
- väčšia penetrácia a integrácia obnoviteľných zdrojov do siete,
- potreba elektrických strojov a pohonov, ktoré môžu fungovať v nepriaznivom prostredí (najmä v baníctve, či v ropnom alebo plynárenskom priemysle),
- rastúci dopyt na rozvodnú sieť a rastúca úloha distribuovanej výroby a mikrosietí [6].

Evolúcia výkonovej elektroniky a prechod na pokročilé pohony s premenlivou rýchlosťou, boli významným posunom paradigmy v technológii. Ako je vidieť na Obr. 3, tradične systémy pozostávali z elektrického stroja (generátora alebo motora), ktorý bol priamo pripojený k elektrickej sieti a pracoval na frekvencii siete (t.j. 50/60 Hz v závislosti od krajiny).



Obr. 3 Konvenčný generátor/motor pripojený k hnaciemu stroju/záťaži cez prevodovku [1].

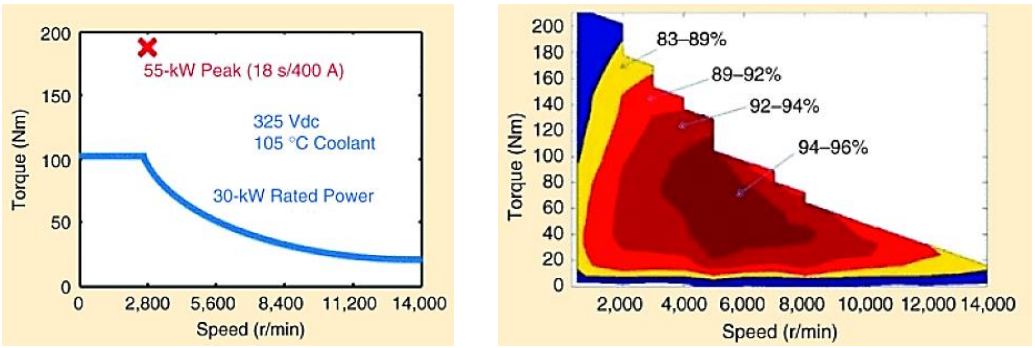
Pohonný stroj/záťaž boli zvyčajne navrhnuté/optimalizované na základe určitej prevádzkovej rýchlosti, ktorá sa zvyčajne líšila od optimálnej prevádzkovej rýchlosti generátora/motora. Na riešenie tohto nesúladu rýchlosti sa tradične používali prevodovky, ktoré predstavovali značné problémy so spoľahlivosťou a údržbou systému. Vývoj nových frekvenčných meničov umožnil frekvenčnému meniču efektívne fungovať ako „elektronická prevodovka“, ako je znázornené na Obr. 4. To umožnilo optimalizovať systémy tak, aby pracovali s vyššou účinnosťou, lepším riadením, vysokou spoľahlivosťou a zníženou údržbou.



Obr. 4 Konvenčný generátor/motor pripojený k hnaciemu stroju/záťaži cez frekvenčný menič [1].

A. Stroje pre ľahké hybridné/elektrické vozidlá

Jedným z kľúčových trendov, ktoré vyvolali súčasnú elektrifikačnú vlnu, je elektrifikácia dopravy. Podstatný dôraz sa kladie na ľahké úžitkové vozidlá, pretože predstavujú najväčší trh v sektore dopravy z hľadiska objemu. Za posledných niekoľko desaťročí sa celosvetovo vyvíjalo trvalé úsilie o pokračovanie v rozvoji najmodernejších elektrických a ľahkých hybridných-elektrických vozidiel (EV/HEV). Stredobodom tohto úsilia bol vývoj pokročilých elektrických strojov, meničov energie a akumulátorov. Americké ministerstvo energetiky zohralo dôležitú úlohu stanovením náročných cieľov pre výkonové štandardy hybridných a elektrických pohonných jednotiek. Tieto ciele, resp. požiadavky sa spravidla označujú ako špecifikácie „FreedomCar 2020“ a sú zobrazené na Obr. 5.



a) Krivka krútiaceho momentu b) Mapa účinnosti

Obr. 5 Špecifikácie „FreedomCar 2020“ [1]..

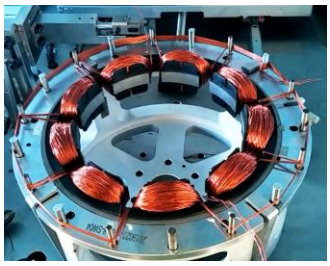
Začal sa vývoj nových koncepcií pohonov, ktoré spĺňajú tento súbor požiadaviek. Bolo vyvinutých niekoľko nových topológií strojov v spojení s pokročilým tepelným manažmentom. Konštrukcia stroja obsahovala segmentovanú štruktúru statora s vinutím a bolo aplikované aj

pokročilé chladenie rotora s vnútorným vŕtaním. Tento návrh využíval neodymové magnety (NdFeB) zo vzácnych zemín a bol najlepším z hľadiska účinnosti a špecifickej hustoty výkonu.

Pred niekoľkými rokmi sa toto zameranie posunulo smerom k vývoju vysokovýkonných motorov, ktoré dokážu splniť požiadavky „FreedomCar2020“ a znížiť alebo odstrániť materiály vzácnych zemín. Bolo to spôsobené najmä takmer 70% nárastom cien magnetov zo vzácnych zemín. Aj keď cena magnetov zo vzácnych zemín neskôr klesla na úroveň pred zvýšením, tak pretrvávajú veľké obavy o jej dlhodobú udržateľnosť a stabilitu. Z tohto dôvodu existuje značný záujem o zníženie alebo odstránenie magnetov zo vzácnych zemín v trakčných motoroch novej generácie [7].

B. Návrh redukujúci magnety zo vzácnych zemín

Permanentné magnety používané v trakčných motoroch sú typicky magnety NdFeB, ktoré zahŕňajú ľahké materiály vzácnych zemín, ako sú neodým a bór. Čo sa týka ťažkých materiálov vzácnych zemín, ako je dysprózium (Dy), tak ono je tiež zahrnuté v týchto magnetoch (v malom množstve) Dysprózium má významný vplyv nie len na koercitivitu magnetov a tým aj na ich náchylnosť k demagnetizácii, ale podstatne vplýva aj na zvyšenie ceny magnetu. Bolo vyvinutých niekoľko strojov využívajúcich magnety bez dysprózia, vrátane viacvrstvových strojov s vonkajším rotorom a aj strojov s prepínaním magnetického toku (FSM). Jednou z topológií, ktoré sa ukázali ako príťažlivý kandidát v tejto kategórii, je Dy-free FSM. Jednou z jeho hlavných výhod je, že vzhľadom na povahu magnetického obvodu a skutočnosť, že permanentné magnety sú umiestnené v statore, sú magnety menej náchylné na demagnetizáciu, vďaka čomu sú magnety bez dysprózia vhodné pre túto topológiu. Prototyp FSM bez dysprózia, bol postavený a plne otestovaný a je zobrazený na Obr. 6 [8].

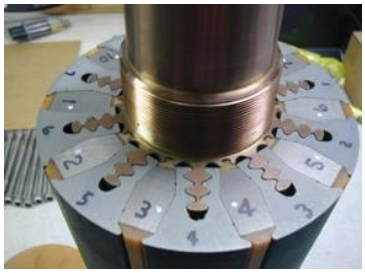


Obr. 6 FSM motor s magnetmi bez Dysprózia.

C. Návrhy motorov s permanentnými magnetmi z iných než vzácnych zemín

Medzi permanentné magnety, ktoré neobsahujú materiály zo vzácnych zemín, patria okrem iného aj feritové Alnico magnety. Okrem nižšej energetickej hustoty v porovnaní s magnetmi zo vzácnych zemín, tak feritové magnety majú aj negatívny tepelný koeficient koercitivity. V prípade Alnico magnetov, hoci majú pomerne vysokú remanenciu, je tendencia k demagnetizácii nižšia, (menej ako polovica) ako u magnetov zo vzácnych zemín. Bolo vytvorených niekoľko topológií strojov, ktoré používajú ferity a Alnico, vrátane viacvrstvových IPM strojov a radiálnych IPM strojov. Kvôli obmedzeniam, pokiaľ ide o magnetické vlastnosti feritov a Alnico, všetky stroje v tejto kategórii boli väčšie v porovnaní so strojmi, ktoré používajú magnety zo vzácnych zemín. Konštrukcia radiálneho statora využívajúca feritové magnety sa ukázala ako najlepšia možnosť z hľadiska hustoty výkonu kvôli účinku koncentrácie toku v lúčovej konfigurácii, ako aj eliminácii mostíkov. Majú tiež pomerne vysokú účinnosť, najmä pri nízkej záťaži, ako aj pri vyšších rýchlostiach. Prototyp bol postavený a plne otestovaný a je znázornený na Obr. 7 [9].

a) stator



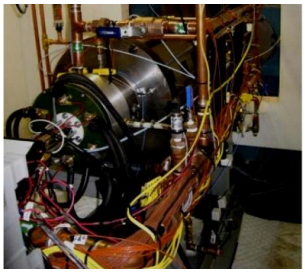
b) rotor

Obr. 7 Konštrukcia radiálneho motora s feritovými magnetmi v rotore.

D. Elektrické stroje s PM pre použitie v letectve a kozmonautike

S neustálym nárastom počtu elektrických pohonov v lietadlách sa oživil záujem o stroje s permanentnými magnetmi (kvôli ich veľkosti a účinnosti) pre letecké aplikácie. V strojoch s permanentnými magnetmi sa v posledných rokoch veľa pracovalo na riešení odolnosti týchto strojov voči poruchám. V priebehu rokov sa vyvinulo veľa riešení pre dobré pochopenie toho, čo je potrebné z hľadiska konštrukcie a riadenia stroja zrealizovať, aby sa dosiahla vysoká odolnosť týchto strojov voči poruchám, ktoré by mali za následok nevyčísliteľné škody a v neposlednej miere dokonca aj straty na ľudských životoch.

Na Obr. 8 je zobrazený kvapalinou chladený generátor s permanentnými magnetmi, ktorý je v maximálnej miere odolný voči poruchám. Tento stroj bol navrhnutý špeciálne ako primárny generátor elektrickej energie pre lietadlo. S využitím permanentných magnetov na báze vzácnych zemín Samarium–Kobalt je funkčný aj pri maximálnej prevádzkovej teplote 200°C [10].



Obr. 8 Zostavený generátor s PM nainštalovaný na skúšobnom stojane [10].

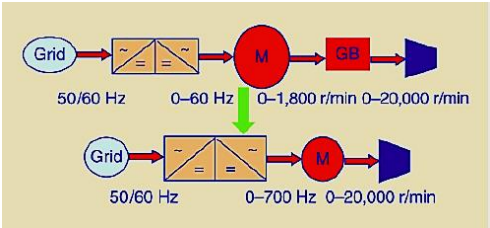
Základné prevádzkové parametre tohto stroja sú:

- trojfázový elektrický výkon – 170 kW,
- prevádzková rýchlosť - 3000 – 12000 ot.min^{-1},
- striedavé regulované napätie na 230 V (ph RMS), 500 – 2000 Hz,
- jednosmerné napätie zbernice – 700 V.

V poslednom období sa výskumné a vývojové aktivity pokrývajúce široké spektrum aplikácií zameriavajú od čisto elektrických pohonných systémov pre bezpilotné a malé lietadlá až po hybridné systémy určené pre veľké komerčné lietadlá. Toto je zásadne odlišné využitie elektrických systémov, čo predstavuje odklon od tradičných elektrických systémov, ktoré sa zameriavali na distribúciu energie. Táto oblasť je nevyhnutná pre vývoj pokročilých elektrických strojov a riadi súčasný stav techniky, najmä pokiaľ ide o špecifický výkon a účinnosť, pretože veľmi vysoký merný výkon a účinnosť sú potrebné na dosiahnutie zníženej spotreby paliva pri minimalizovaní straty dodatočnej hmotnosti elektrických komponentov. Stroje s permanentnými magnetmi a elektrické pohony využívajúce tieto motory sú najlepšími kandidátmi v tejto oblasti.

E. Elektrické stroje s PM pre ropný a plynárenský priemysel

Ropný a plynárenský priemysel je dobrým príkladom aplikačného priestoru, ktorý tradične nehral kľúčovú úlohu pri využívaní elektrických strojov a pohonov. To sa za posledných niekoľko rokov zmenilo a elektrifikácia ropného a plynárenského priemyslu sa stala dôležitou iniciatívou. Dôležitou aplikáciou v ropnom a plynárenskom priemysle je použitie hnacieho ústrojenstva pre vysokorýchlostné kompresory. Tradičná základná architektúra systému, ktorá poháňa vysokorýchlostné kompresory, zahŕňa menič výkonu, nízko rýchlostný motor, prevodovku a kompresor, ako je znázornené na Obr. 9 hore.



Obr. 9 Transformácia elektrického pohonu vysokorýchlostných kompresorov [11].

Táto architektúra systému prešla na pokročilejšiu, ktorá zahŕňa vysokofrekvenčný výkonový menič a vysokorýchlostný motor priamo spojený s kompresorom. Táto transformácia hnacieho

ústrojenstva v spojení s použitím strojov s permanentnými magnetmi vedie k viacerým výhodám, vrátane výrazného zlepšenia špecifického výkonu a okrem toho sú stroje s permanentnými magnetmi vhodnejšie na zapuzdrenie (ako napr. klasické indukčné stroje), čo je dôležité pri aplikáciách s kyselinami a kyslými plynmi [11].

Boli vyvinuté pohony s výkonom v rozsahu od 5 MW (pri 20 000 ot.min^{-1}) do 20 MW (pri 10000 ot.min^{-1}), pričom boli dosiahnuté špecifické výkony približne 2 – 3 kW/kg. Okrem ropy a zemného plynu možno rovnakú technológiu a architektúru hnacieho ústrojenstva využiť aj v iných synergických oblastiach, vrátane výroby energie v námornej doprave.

F. Pohonné systémy ťažkých terénnych a banských vozidiel

Pokiaľ ide o elektrifikáciu dopravy, hoci sa hlavný dôraz kládol na ľahké úžitkové vozidlá, elektrifikáciou prechádzajú aj stredne ťažké a ťažké vozidlá. Jednou kategóriou ťažkých úžitkových vozidiel (Obr. 10a), o ktoré je záujem, sú terénne vozidlá (OHV), ktoré sa zvyčajne používajú v stavebnom a ťažobnom priemysle. Niektoré z týchto vozidiel majú diesel – elektrické pohonné systémy, kde dieselový motor poháňa alternátor, ktorý napája trakčný motor resp. motory. Tieto trakčné motory sú v mnohých prípadoch uložené priamo v kolesách, ako je znázornené na Obr. 10b) [12].

a) Ťažké úžitkové vozidlo b) Trakčný kolesový motor

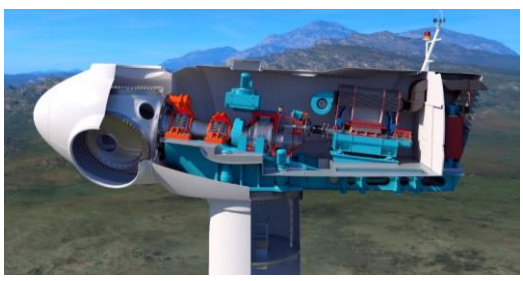
Obr. 10 Ťažké úžitkové vozidlo OHV General Electric s pohonným motorom v kolese.

V poslednom období vzrástol aj záujem o podzemné banské vozidlá. Ako je znázornené na Obr. 11, niektoré z týchto vozidiel sú už napájané z batérie. Snahou je nahradiť diesel – elektrické pohonné systémy čisto elektrickým, aby sa znížili emisie z nafty. To viedlo k potrebe vyvinúť spoľahlivé, kompaktnejšie bezporuchové elektrické pohonné systémy. V tejto oblasti doteraz síce prevládajú indukčné stroje, ale vývoj sa zameral aj na stroje s permanentnými magnetmi z dôvodu obmedzeného priestoru, ktorý je k dispozícii na uloženie batérie a samotného pohonu.

Obr. 11 Podzemné elektrické naberače uhlia.

G. Veterné generátory

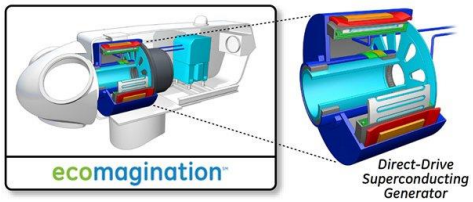
Využívanie obnoviteľných zdrojov a potreba ich integrácie do siete je jednou z oblastí využitia elektrických strojov ako veterných generátorov. Pre veternú energiu na pevnine sa najčastejšie používa riešenie s prevodom, v ktorom je prevodovka umiestnená medzi veternú turbínu a elektrický generátor (Obr. 12).



Obr. 12 Konštrukcia veternej elektrárne s prevodovkou [13].

Vykonalo sa veľa práce na optimalizácii celkového hnacieho ústrojenstva, (t.j. počtu prevodových stupňov, prevodového pomeru, a teda aj rýchlosti generátora) pre dosiahnutie čo najvyššej účinnosti. Pre pobrežné veterné elektrárne, najmä tie s nižším výkonom, sa využívajú indukčné generátory. Generátory s permanentnými magnetmi sa používajú nie len na pevnine, ale stále viac sa využívajú aj na mori pre ich vyššie výkony a kvôli ich vyššej energetickej využiteľnosti a účinnosti. Z toho dôvodu pred niekoľkými rokmi sa priemysel začal výrazne posúvať smerom k generátorom s permanentnými magnetmi, avšak tento posun sa spomalil z dôvodu prudkého nárastu cien magnetov zo vzácnych zemín. V súčasnosti ceny týchto magnetov išli opäť dole, ale aj tak stále pretrvávajú dlhodobé obavy o ich udržateľnosť a cenovú stabilitu. Veterné generátory spravidla používajú jednu metrickú tonu magnetov zo vzácnych zemín na 1 MW výkonu. Z tohto dôvodu môže mať akékoľvek výrazné kolísanie ceny magnetov vzácnych zemín podstatný vplyv na cenu veterných generátorov [14].

Vzhľadom na problémy s prístupnosťou k niektorým veterným elektrárňam (napr. na mori), by sa požiadavky na údržbu mali minimalizovať. Preto sa používa hnacie ústrojenstvo s priamym pohonom, čím sa eliminuje prevodovka a s tým súvisiaca aj všetka potrebná údržba, ktorá je pre bezporuchový chod potrebná. Konfigurácia s priamym pohonom, okrem vysokého výkonu veterných turbín na mori a veľkosti generátora, predstavuje výzvu, najmä preto, že v budúcnosti sa očakáva výkon nad 10 MW. To je dôvod, prečo stroje s permanentnými magnetmi, majú jasné výhody pre veterné pohony s priamym pohonom na mori, pokiaľ ide o hustotu krútiaceho momentu a účinnosť. Keďže výroba elektrickej energie na mori sa neustále posúva smerom k vyšším hodnotám výkonu, možno tu zvážiť aj použitie supravodivých generátorov (Obr. 13).

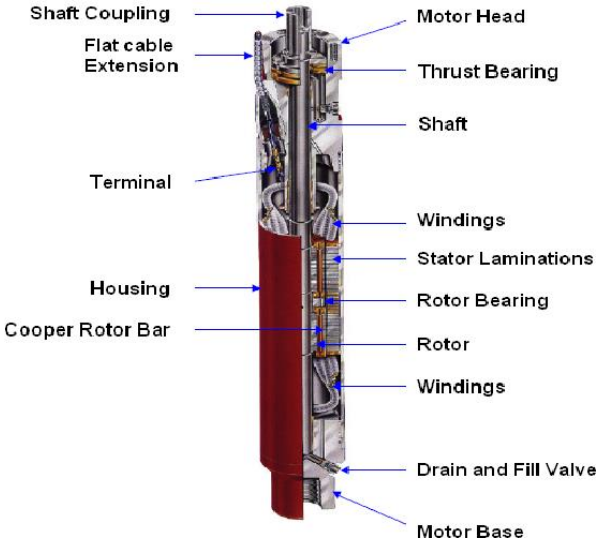


Obr. 13 Supravodivý generátor pre veternú elektráreň [15].

H. Geotermálna energia - Motor pre geotermálne čerpadlo

Hoci sa pozornosť sústredila na solárnu a veternú energiu, existuje mnoho ďalších obnoviteľných zdrojov energie, ktoré zohrávajú významnú úlohu. Niektoré z týchto iných zdrojov predstavujú zaujímavé výzvy z dôvodu ich spôsobu prevádzky a systémových požiadaviek. V jednej takejto aplikácii, ako je geotermálne čerpadlo, je teplota do 300°C a tlaky do 300 barov. Okrem toho je tiež aj priestor pomerne obmedzený. Motor má viacero stupňov, pričom každý z rotorových stupňov je podopretý nezávislou sadou ložísk (Obr. 14).

Tento motor má extrémne malý pomer priemeru a dĺžky stroja. Kvôli vyššie uvedeným požiadavkám sú preto potrebné elektrické stroje s veľmi vysokou hustotou krútiaceho momentu. Tiež musia byť splnené vysoké požiadavky na spoľahlivosť a životnosť a sú potrebné aj pokročilé izolačné systémy, ktoré dokážu odolať aj veľmi vysokým teplotám [16].



Obr. 14 Motor pre geotermálne čerpadlo [20].

IV. ZÁVER

Tento článok poskytuje prehľad širokej škály aplikácií, v ktorých zohrávajú kľúčovú úlohu pokročilé pohony, kde je možné využiť elektrické stroje s permanentnými magnetmi. Existuje mnoho aplikácií, ktoré si vyžadujú prudkú zmenu vo výkone elektrických strojov a pohonov. Okrem nových topológií a optimalizácie návrhu bude pri plnení neustále rastúcich požiadaviek vyššie uvedených aplikácií zohrávať kľúčovú úlohu široká škála podporných technológií, vrátane pokročilých materiálov a tepelného manažmentu. Pre zaujímavosť, ak porovnáme 15 kW indukčný motor (IM) s motorom toho istého výkonu, ale je to motor s permanentnými magnetmi (SPM), tak motor s PM má váhu 54,4 kg a IM váži až 113,5 kg, pričom objem SPM je len 0,0178 m^3 a objem IM je 0,336 m^3.

Súdiac podľa šírky aplikácií a veľmi náročných požiadaviek z hľadiska efektívnosti, nákladov a spoľahlivosti sa očakáva, že v budúcnosti sa výrazne zvýši úsilie vo výskume a vývoji v oblasti pokročilých elektrických strojov a pohonov s permanentnými magnetmi aj vzhľadom na ich vyššiu cenu. Prechod na pokročilé pohony s premenlivou rýchlosťou s vyššou účinnosťou bude mať významný vplyv aj na množstvo globálnej energie spotrebovanej v priemyselných odvetviach..

POĎAKOVANIE

Táto práca bola podporená projektom vedeckej agentúry KEGA podľa zmluvy číslo: 032TUKE-4/2024.

LITERATÚRA

[1] El-Refaie. Ayman M., „Toward a Sustainable More Electrified Future: The Role of Electrical Machines and Drivers“, IEEE Electrification Magazine, Vol. 7, No. 1, 2023), pp. 49-59.
[2] Yu, Z.; Li, Y.; Jing, Y.; Wang, J., „Cooling System of Outer Rotor SPMSM for a Two-Seater All-Electric Aircraft Based on Heat Pipe Technology“, IEEE Trans. Transp. Elektrif. 2022, 8, pp. 1656–1664.
[3] Beniakar, M.E.; Kakosimos, PE; Krasopoulos, CT; Sarigiannidis, A.G.; Kladas, A.G., „Comparison of in-wheel permanent magnet motors for electric traction“, In International Conference on Electrical Machines (ICEM), Berlín, Nemecko, 2. – 5. sept. 2014; pp. 2472–2478.
[4] Fang, H.; Qu, R.; Li, J.; Zheng, P.; Fan, X., „Rotor design for high-speed permanent magnet synchronous motor.“, IEEE Trans. Ind. Appl. 2017, 53, pp.3411–3419.
[5] Chung SU, Kim JW, Chun YD, Woo BC, Hong DK., „Fractional Slot Concentrated Winding PMSM With Consequent Pole Ro-tor for a Low-Speed Direct Drive: Reduction of Rare Earth Permanent Magnet.“ IEEE Transactions on Energy Conversion. 2015;30(1), pp. 103-109.
[6] Badahdah, Hesham S., El-Refaie. Ayman M., „Investigation of Different Rotor Configurations of a Hybrid-PM Variable Flux Machine“, IECON 2024 - 50th Annual Conference of the IEEE Industrial Electronics Society, IL, USA, 2024, pp. 1-6, doi: 10.1109/IECON55916.2024.10905085.
[7] El-Refaie. Ayman M. „Motors/generators for traction/propulsion applications", IEEE Veh. Technol. Mag., vol. 8, no. 1, Mar. 2013, pp. 90-99.
[8] Lovelace. E. C., Jahns. T. M., Keim, T. A. and J. H. Lang, „Mechanical design considerations for conventionally laminated, high-speed, interior PM synchronous machine rotors,” IEEE Transactions on Industry Applications, vol. 40, no. 3, May-June 2024, pp. 806-812.
[9] Weeber, K. R. et al., „Advanced permanent magnet machines for a wide range of industrial applications", Proc. IEEE Power & Energy Society (PES) General Meeting, 2010, pp. 1-6.
[10] NASA technology roadmaps: TA 2: In-space propulsion technologies, [online]. [Dátum: 12. Okt. 2024]. Dostupné na internete: <https://docslib.org/doc/10174227/nasa-technology-roadmaps-ta-2-in-space-propulsion-technologies>.
[11] El-Refaie, A. M., Shah, M. R., Huh, K. -K., „High power-density fault-tolerant PM generator for safety critical applications,“ International Electric Machines & Drives Conference, Chicago, IL, USA, 2013, pp. 30-39.
[12] GE Transportation, „Transforming mining together.“ [Online]. [Dátum: 29. November. 2024]. Dostupné na internete: <http://www.getransportation.com >.
[13] Geared Wind Turbine Drivetrain Working Principle Animation, [online]. [Dátum: 10. Januar 2025]. Dostupné na internete: <https://siamagazin.com/geared-wind-turbine-drivetrain-working-principle-animation/ >.
[14] Jahns, T. M. et al., „Design and experimental verification of a 50-kW interior permanent magnet synchronous machine“, Proc. Conf. Rec. IEEE Industry Applications Conf. 41st IAS Annu. Meeting, 2006, pp. 1941-1948.
[15] Windpower monthly, [online]. [Dátum: 15. Januar 2025]. Dostupné na internete: <https:// www. windpowermonthly.com/article/1212397/advancing-wind-power-superconductivity >.
[16] Lobianco, L. F.,Wardani, W., „Electrical submersible pumps for geothermal applications“, [online]. [Dátum: 26. Januar. 2025]. Dostupné na internete: < https://www.semanticscholar.org/paper/Electrical-Submersible-Pumps-for-Geothermal-Lobianco-Wardani/d7104ddaad70b14e199e06422bfde487a5531147>.

Návrh a vyhotovenie pracovného nástroja pre vykonávanie dotykov na dotykových obrazovkách

[1]*Peter NEMERGUT,* [2]*Peter GIROVSKÝ*

[1] Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovensko
[2] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovensko

[1]peter.nemergut@tuke.sk, [2]peter.girovsky@tuke.sk

***Abstrakt* — Tento článok sa zaoberá návrhom a vyhotovením pracovného nástroja určeného na vykonávanie dotykov na dotykových obrazovkách. V rámci výskumu bolo navrhnutých, vyhotovených a otestovaných sedem verzií pracovného nástroja, ktoré boli vyrobené technológiou 3D tlače. Jednotlivé verzie sa odlišujú v konštrukcii a použitých mechanizmoch pre lineárne vedenie a pohyb. Na základe praktického testovania boli identifikované nedostatky jednotlivých verzií, čo umožnilo optimalizovať návrh a vyhotoviť finálnu verziu pracovného nástroja s dostatočnou spoľahlivosťou a funkčnosťou.**

***Kľúčové slová* — 3D tlač, dotyk, dotyková obrazovka, konštrukcia, pracovný nástroj, testovanie**

I. Úvod

V súčasnosti môžeme pozorovať dynamický rozvoj digitálnych technológií obsahujúcich dotykové obrazovky [1] ako sú smartfóny, tablety, informačné systémy, priemyselné panely a pod. So zvyšujúcou sa komplexnosťou softvérových riešení, ktoré sú integrované v týchto zariadeniach narastá aj potreba ich dôsledného a spoľahlivého testovania [2]. Automatizované testovanie dotykových rozhraní (napr. Android aplikácie) je dôležitou súčasťou kontroly kvality softvéru nie len počas jeho vývoja, ale aj pri finálnom testovaní funkčnosti [3], [4]. Oproti manuálnemu testovaniu ponúkajú automatizované testy časové úspory, čo umožňuje rýchlejšie testovanie rôznych testovacích scenárov. Zároveň sa znižuje chybovosť, ktorá môže byť spôsobená ľudským faktorom, čo prispieva k zabezpečeniu vyššej konzistentnosti výsledkov počas testovania. Vzhľadom na to, že mnohé aplikácie musia správne reagovať na rôzne typy dotykov (napr. krátky, dlhý a pod.), je nevyhnutné vyvinúť spoľahlivý a opakovateľný spôsob aj ich fyzickej simulácie. Z pohľadu samotného spôsobu testovania aplikácie spustenej na zariadení, ktoré obsahuje dotykovú obrazovku, je potrebné pre zabezpečenie správnej funkčnosti testovať rôzne scenáre automatizovaným alebo manuálnym spôsobom [5].

Tento článok sa zameriava na návrh pracovného nástroja určeného na vykonávanie dotykov na dotykových obrazovkách, ktorý by umožňoval zautomatizovať fyzické stláčanie obrazovky. Cieľom je vytvoriť pracovný nástroj, ktorý bude schopný presne, opakovateľne a bez poškodenia dotykovej obrazovky simulovať ľudský dotyk.

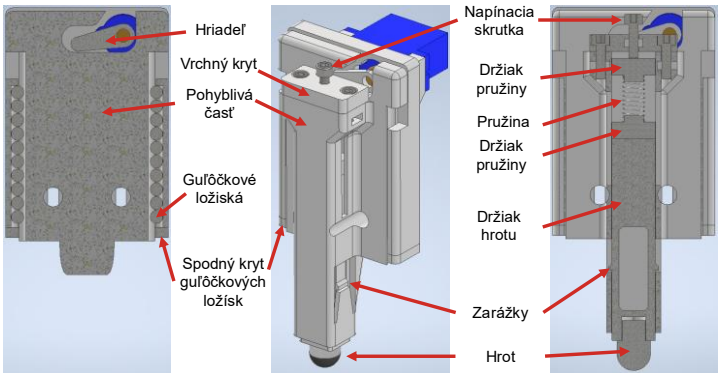
II. Návrh a Vyhotovenie Pracovného Nástroja

Pri realizácií pracovného nástroja sa vychádzalo z požiadavky vykonávať dotyky na dotykovej obrazovke a merať aktuálne vytváraný tlak pomocou snímača. Aby bolo možné vykonávať dotyky na obrazovke, je potrebné, aby pracovný nástroj umožňoval pohyb vo vertikálnom smere (t.j. v smere osi Z, teda kolmo na dotykovú obrazovku) [6], [7], [8]. Všetky navrhnuté pracovné nástroje pozostávajú z pevnej a pohyblivej časti, ktoré boli vyhotovené technológiou 3D tlače. Pevná časť umožňuje prichytenie pomocou skrutiek na pracovnú hlavu testovacieho zariadenia, ktoré realizuje pohyb pracovnej hlavy v osiach X a Y. Pohyblivá časť pracovného nástroja zabezpečuje fyzické vykonávanie dotykov na dotykovej obrazovke vo vertikálnom smere.

V prvej fáze návrhu bol pre realizáciu pohybu vo verziách pracovného nástroja 1.0 a 2.0 zvolený servopohon s impulzovou šírkovou moduláciou DFRobot DMS-MG90-A (180°), ktorý vďaka jednoduchému ovládaniu polohy, kompaktným rozmerom, nízkej hmotnosti a možnosti priameho napájania predstavoval jednoduché riešenie pre rýchly vývoj.

A. Pracovný nástroj vo verzii 1.0

Pevná časť pracovného nástroja pozostáva z jedného komponentu, do ktorého je pomocou skrutiek pripevnený servopohon (Obr. 1). Z dolnej strany je upevnený spodný kryt guľôčkových ložísk, ktorý vymedzuje uloženie guľôčkových ložísk s priemerom 3 mm. Tie zabezpečujú lineárne vedenie pre pohyb pohyblivej časti vo vertikálnom smere. Samotný pohyb je realizovaný pohybom hriadele, ktorý je upevnený na servopohone. Pohyblivá časť pozostáva z hrotu a jeho držiaka, ktorý je zasunutý do pohyblivej časti. Zarážky zabraňujú vypadnutiu držiaka hrotu smerom dole. Dva držiaky pružiny spolu s pružinou zabezpečujú odtláčanie držiaka hrotu od vrchného krytu, pričom napínacou skrutkou je možné nastaviť požadovanú silu odtláčania.



Obr. 1 Pracovný nástroj vo verzii 1.0.

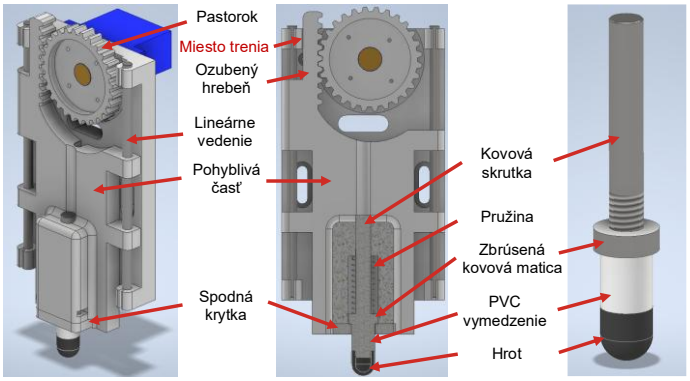
Výhodou riešenia je, že ak by pri nesprávnom zatlačení hrotu na dotykovú obrazovku bola vytvorená väčšia sila, ako je sila pružiny, dôjde k jej stlačeniu, čím sa držiak hrotu spolu s hrotom zatlačia do pohyblivej časti a zníži sa riziko poškodenia dotykovej obrazovky. Po 3D tlači a montáži komponentov bolo zistených niekoľko nedostatkov. Vyhotovenie komponentov technológiou 3D tlače je realizované nanášaním vrstiev materiálu na seba. Tieto jemné vrstvy však spôsobujú zasekávanie guľôčkových ložísk pri pohybe pohyblivej časti. To má negatívny vplyv aj na komponenty vo vnútri pohyblivej časti, kde taktiež dochádza k zasekávaniu komponentov a nedochádza k správnemu stláčaniu pružiny. Ďalšou nevýhodou je nepriame spojenie pohyblivej časti s hriadeľom servopohonu, čo sa prejavuje nepresným polohovaním hrotu z dôvodu vôlí. Zároveň použitý hrot z dotykového pera nebol schopný dostatočne absorbovať elektrický náboj a dochádzalo iba k náhodnému aktivovaniu dotykovej obrazovky.

B. Pracovný nástroj vo verzii 2.0

Pevná časť pozostáva z jedného komponentu, do ktorého je pomocou skrutiek upevnený servopohon (Obr. 2). Lineárne vedenie pre pohyb pohyblivej časti vo vertikálnom smere je realizované pomocou nerezových tyčí namiesto použitia guľôčkových ložísk. Lineárny pohyb pohyblivej časti je realizovaný pomocou pastorka umiestneného na servopohone a ozubeného hrebeňa, ktorý je súčasťou pohyblivej časti. Základom držiaka hrotu je zrezaná kovová skrutka, na ktorej je zbrúsená kovová matica do kruhového tvaru, podlepená pevnostným lepidlom. Na nej je nasadený dotykový hrot s priemerom 7 mm. PVC (Polyvinylchlorid) vymedzenie slúži na vymedzenie priestoru medzi hrotom a zbrúsenou kovovou maticou. Pružina zabezpečuje odtláčanie kovovej skrutky spolu s maticou smerom dole vo vnútri pohyblivej časti.

Podobne ako pri verzii 1.0, aj tu sa nachádza pružinový mechanizmus, ktorý zabezpečuje zníženie rizika poškodenia dotykovej obrazovky. Po vyhotovení pracovného nástroja bolo zistené výrazné zlepšenie pohybu vo vertikálnom smere, ktoré funguje spoľahlivo s minimálnou vôľou a trením. Funkčnosť lineárneho pohybu sa ukázala lepšia v porovnaní s verziou 1.0, pretože sa zmenšili vôle pri pohybe. Taktiež sa výrazne zlepšila funkčnosť pružinového mechanizmu. Zistené nedostatky po realizácií pracovného nástroja sa ukázali až pri prevádzke. V krajných polohách pohyblivej časti (hore a dole) nemá servopohon dostatočný moment pre pohyb pohyblivej časti a dochádza k preťažovaniu servopohonu, čo sa prejavuje jeho prehrievaním. Ďalšou nevýhodou je slabá citlivosť polohovania vo vertikálnom smere. Servopohon totiž nedokáže vyvinúť dostatočnú silu pri malej požadovanej zmene polohy a dochádza k skokovitým

zmenám polohy. Taktiež bolo pozorované miesto trenia ozubeného hrebeňa voči pevnej časti, čo negatívne ovplyvňovalo plynulý pohyb pohyblivej časti. Poslednou zistenou nevýhodou je nedostatočné pohlcovanie elektrického náboja hrotom a jeho držiakom, čo zapríčiňuje iba náhodné aktivovanie dotykovej obrazovky. Tento nedostatok bol čiastočne odstránený pridaním tenkého elektrického vodiča upevneného z jednej strany medzi zbrúsenú kovovú maticu a pružinu. Druhá strana vodiča bola vyvedená von z pohyblivej časti pracovného nástroja.

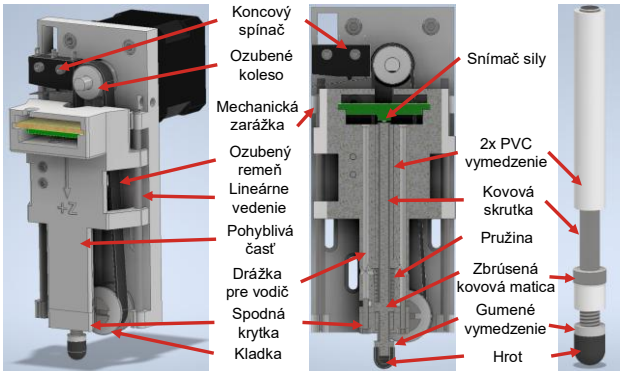


Obr. 2 Pracovný nástroj vo verzii 2.0.

Na základe analýzy predchádzajúcich verzií a identifikovaných konštrukčných nedostatkov bol v druhej fáze vývoja implementovaný presnejší a robustnejší krokový motor NEMA14 SY35ST36-1004A od spoločnosti Pololu Corporation. Návrh pracovného nástroja vo verziách 3.0 až 3.4 zahŕňa integráciu vývojovej dosky Force 5 Click [9] pre snímanie aktuálne pôsobiaceho tlaku na dotykovú obrazovku, ktorá bola z dôvodu zníženia rizika deformácie pripevnená na samostatnú dosku plošného spoja. Výsledkom je technicky komplexnejší pracovný nástroj obsahujúci nielen pohon, ale aj koncový spínač AVT3212AT od výrobcu Panasonic a snímač sily, čím došlo k výraznejšiu zlepšeniu jeho funkčnosti a spoľahlivosti.

C. Pracovný nástroj vo verziách 3.0 – 3.3

Pevná časť pozostáva z úchytov pre uchytenie krokového motora, koncového spínača, kladky a lineárneho vedenia (vo forme nerezových tyčí rovnako ako vo verzii 2.0). Lineárny pohyb je realizovaný krokovým motor s ozubeným kolesom, ktorý poháňa ozubený remeň okolo kladky (Obr. 3). Začiatok a koniec ozubeného remeňa je spojený v napínacom mechanizme, ktorý sa priamo pripája zo zadnej strany k pohyblivej časti pomocou dvoch skrutiek (Obr. 4). Pri pohybe smerom nahor dôjde tesne pred mechanickou zarážkou k zopnutiu koncového spínača. Držiak hrotu pozostáva zo zrezanej kovovej skrutky, na ktorej je zbrúsená kovová matica do kruhového tvaru, podlepená pevnostným lepidlom. Na nej je nasadený dotykový hrot s priemerom 7 mm. Držiak hrotu sa pohybuje vo vnútri PVC vymedzenia, čím sa minimalizuje jeho trenie. Pri vytvorení tlaku na hrot dôjde k pohybu držiaka hrotu, ktorý priamo pôsobí na snímač sily. Po uvoľnení tlaku pružina spôsobí odtlačenie držiaka hrotu od snímača sily.



Obr. 3 Pracovný nástroj vo verzii 3.1.

Hlavnou výhodou verzií 3.0 až 3.3 oproti verzii 2.0 je presnejšia a spoľahlivejšia funkčnosť polohovania hrotu vďaka použitiu krokového motora. Tieto verzie zároveň umožňujú merať silu, ktorou hrot pôsobí na dotykovú obrazovku. Konkrétne rozdiely vo verziách sú nasledovné:

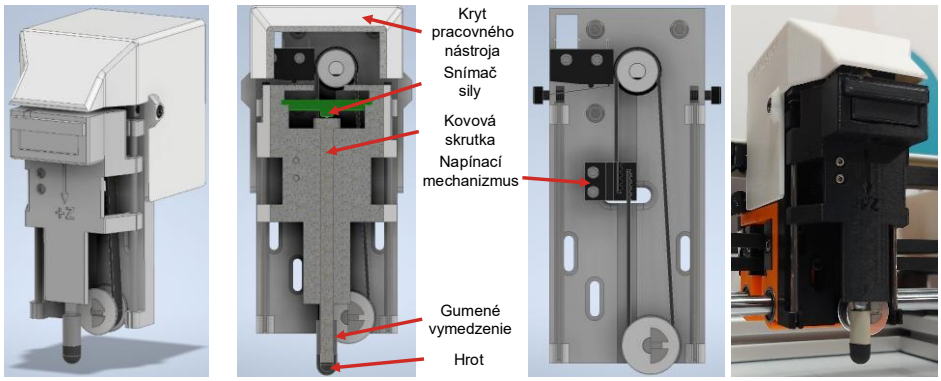
- 3.0 – Kovová matica držiaka hrotu má šesťhranný tvar. Problémom je však zasekávanie matice vo vnútri pohyblivej časti.

- 3.1 – Kovová matica držiaka hrotu je zbrúsená do kruhového tvaru, avšak dochádza k rovnakému problému ako vo verzii 3.0.
- 3.2 – Riešením problémov z verzie 3.0 a 3.1 je minimalizovanie množstva komponentov v pohyblivej časti. Verzia 3.2 neobsahuje spodný kryt, pružinu, zbrúsenú maticu a PVC vymedzenia. Držiak hrotu pozostáva iba z nezbrúsenej skrutky zo šesťhrannou hlavou, gumeným vymedzením a hrotom. Pri tlaku na hrot dochádza k pôsobeniu tlaku na snímač sily. Po uvoľnení tlaku gravitácia spôsobí samovoľné odtlačenie držiaka hrotu od snímača.

Tieto verzie majú v úzkej drážke umiestnený tenký vodič, ktorého jeden koniec je ovinutý okolo kovovej skrutky (držiaka hrotu). Druhá strana vodiča je vyvedená z pohyblivej časti smerom von. Pri používaní bolo zistené, že časom ovinutý vodič stratil kontakt s kovovou skrutkou. Tento nedostatok bol vyriešený zmenou vodivého spojenia tenkého vodiča s kovovým držiakom hrotu použitím pružného prítlačného mechanizmu vo verzii 3.3. Všetky uvedené zistenia z týchto verzií boli aplikované pri návrhu pracovného nástroja vo verzii 3.4 (Obr. 4).

D. Pracovný nástroj vo verzii 3.4

V tejto verzii bolo implementovaných viacero úprav, ktoré významne prispeli k zlepšeniu funkčnosti. Hlavnou zmenou je implementovanie kovového prítlačného mechanizmu umiestneného na zadnej strane pohyblivej časti, ktorý zabezpečuje spoľahlivý elektrický kontakt vodiča s kovovou skrutkou. Súčasťou tejto verzie je taktiež kryt pracovného nástroja v bielom vyhotovení, ktorý mechanicky chráni ozubeného koleso, koncový spínač a pripojené vodiče. Spoľahlivosť navrhnutého riešenia sa prejavila v dosiahnutí vysokej úspešnosti pri detekcii dotyku na dotykovej obrazovke počas používania. Kľúčovou zmenou bolo uzemnenie konca vyvedeného tenkého vodiča, čo zabezpečilo dostatočné absorbovanie elektrického náboja. Porovnanie technických parametrov a vlastností vyhotovených pracovných nástrojov je v Tab. 1.



Obr. 4 Pracovný nástroj vo verzii 3.4.

III. Výsledky a Odporúčania

Počas návrhu a vyhotovenia pracovného nástroja určeného na fyzické vykonávanie dotykov na dotykových obrazovkách bolo vytvorených a otestovaných sedem rôznych verzií. Každá z týchto verzií predstavovala určité úpravy v konštrukcii, pričom boli postupne odstraňované zistené nedostatky a optimalizované technické riešenia. Prvé dve verzie (1.0 a 2.0) využívali servopohon, ktorý síce umožňoval jednoduché ovládanie a rýchlu implementáciu, avšak jeho použitie prinieslo viacero problémov ako nepresné polohovanie v dôsledku mechanických vôlí, nedostatočný moment a nespoľahlivá detekcia dotykov spôsobená slabým elektrickým kontaktom medzi držiakom hrotu a tenkým vodičom.

Významný posun bol dosiahnutý zavedením krokového motora vo verziách 3.0 až 3.4. Tento pohon poskytol vyššiu presnosť polohovania a umožnil plynulý a opakovateľný vertikálny pohyb pohyblivej časti. Pridaním snímača sily sa podarilo zabezpečiť meranie pôsobiaceho tlaku, čo rozšírilo možnosti použitia nástroja pre komplexnejšie testovanie. Ako vhodné sa ukázalo použitie lineárneho vedenia z nerezových tyčí v kombinácii s ozubeným remeňom a kladkou, ktoré zabezpečili hladký a spoľahlivý chod pohyblivej časti. Napriek zlepšeniam vo verziách 3.0 až 3.3 sa vyskytli problémy v oblasti elektrického spojenia medzi kovovými časťami hrotu a uzemnením, ktoré boli realizované tenkým vodičom. Ukázalo sa však, že tento spôsob nie je dostatočne spoľahlivý z dôvodu nedostatočného elektrického spojenia vodiča s kovovým držiakom hrotu. Preto bol vo finálnej verzii 3.4 implementovaný konštrukčne zložitejší, ale účinnejší prítlačný mechanizmus umiestnený na zadnej strane pohyblivej časti, ktorý zabezpečil stabilný elektrický kontakt. Toto riešenie odstránilo problém s náhodným aktivovaním dotykovej obrazovky a pomohlo dosiahnuť vysokú úspešnosť detekcie dotyku počas testovania. Finálna

verzia pracovného nástroja neobsahuje pružinový mechanizmus, ktorý v prvotných verziách plnil ochrannú funkciu pred poškodením dotykovej obrazovky. Implementácia takéhoto mechanizmu by mohla byť súčasťou budúcich verzií pracovného nástroja.

Tab. 1
Porovnanie technických parametrov a vlastností pracovných nástrojov

Verzia	1.0	2.0	3.0, 3.1, 3.2, 3.3	3.4
Typ pohonu	PWM servopohon	PWM servopohon	Krokový motor	Krokový motor
Presnosť polohovania	Nízka	Stredná	Vysoká	Vysoká
Lineárne vedenie	Guľôčkové ložiská	Nerezové tyče	Nerezové tyče	Nerezové tyče
Lineárny pohyb	Nepriamy	Pastorok a ozub. hrebeň	Remeňový systém	Remeňový systém
Snímač sily	Nie	Nie	Áno	Áno
Spoľahlivosť dotykov	Malá	Stredná	Stredná	Vysoká
Lineárne trenie	Veľké	Stredné	Malé	Malé
Mechanická zarážka	Nie	Áno	Áno	Áno
Pružinový mechanizmus	Áno	Áno	3.0,3.1 -Áno 3.2,3.3 - Nie	Nie
Celková funkčnosť	Slabá	Dobrá	Dobrá	Výborná

Na základe získaných poznatkov je dôležité poukázať na použitie krokových motorov ako lepšej alternatívy v aplikáciách, kde je vyžadovaná vysoká presnosť a spoľahlivosť pohybu. Zároveň je potrebné venovať pozornosť dostatočnému uzemneniu a dobrému elektrickému spojeniu medzi hrotom a konštrukciou, keďže ide o dôležitý faktor ovplyvňujúci spoľahlivosť dotykov na dotykovú obrazovku. Po mechanickej stránke sa ako vhodné ukázalo zníženie počtu komponentov a minimalizácia trenia pomocou jednoduchšieho konštrukčného usporiadania.

IV. ZÁVER

Cieľom tohto článku bolo navrhnúť a zrealizovať funkčný pracovný nástroj pre vykonávanie dotykov na dotykových obrazovkách, ktorý je možné upevniť na testovacie zariadenie. V priebehu vývoja bolo navrhnutých sedem verzií nástroja vyrobených technológiou 3D tlače, ktoré boli postupne testované a optimalizované. Prvé verzie obsahovali servopohon, ktorý sa ukázal ako nedostatočný pre túto aplikáciu. Pri ďalších verziách s krokovým motorom a snímačom sily došlo k zlepšeniu presnosti, spoľahlivosti a celkovej funkčnosti nástroja.

Finálna verzia nástroja úspešne splnila požiadavky na presné polohovanie a spoľahlivú simuláciu ľudského dotyku na dotykovej obrazovke. Meranie sily pôsobiacej na dotykovú obrazovku priamo v pracovnom nástroji umožňuje testovanie softvérových aplikácií, ktoré reagujú aj na rôznu silu dotyku. Výsledky potvrdzujú, že správne navrhnutý pracovný nástroj dokáže výrazne zefektívniť a spresniť testovanie softvéru, a tým prispieť k zvyšovaniu jeho kvality v rôznych typoch elektronických zariadení, ktoré obsahujú dotykovú obrazovku.

POĎAKOVANIE

Táto práca bola podporená projektom KEGA 032TUKE-4/2024.

ZDROJE

[1] The InTouch Team, “5 Types of Touch Screen Monitor Technologies”, InTouch Screens. Cit: 22. máj 2025. [Online]. Available at: https://intouchscreens.com.au/5-types-of-touch-screen-monitor-technologies/
[2] A. Krysik, “21 Types of Software Testing Every Engineer Should Know”, Stratoflow. Cit: 22. máj 2025. [Online]. Available at: https://stratoflow.com/types-of-software-testing/
[3] “Čo je testovanie softvéru?”, skillmea.sk. Cit: 22. máj 2025. [Online]. Available at: https://skillmea.sk/blog/co-je-testovanie-softveru
[4] Y. Chernak, “Understanding the Logic of System Testing”, StickyMinds. Cit: 22. máj 2025. [Online]. Available at: https://www.stickyminds.com/article/understanding-logic-system-testing
[5] M. Adil, “Android Automated Testing and Manual Testing : Which to Follow?”, Testsigma Technologies Inc. Cit: 22. máj 2025. [Online]. Available at: https://testsigma.com/blog/android-application-testing-comparison-of-the-two-approaches-manual-and-automated/
[6] “Robotic Touch Panel Tester - TakTouch 1000”, Tactile Automation, Inc. Cit: 22. máj 2025. [Online]. Available at: https://tactileautomation.com/shop/touch-panel-test-systems/robotic-touch-panel-tester-taktouch-1000/
[7] “T100”, SPEA. Cit: 22. máj 2025. [Online]. Available at: https://www.spea.com/en/products/t100-flying-actuator-optomechatronics-tester/
[8] “TOUCH: Automated touch screen testing”. Cit: 22. máj 2025. [Online]. Available at: https://www.optofidelity.com/touch
[9] “Force 5 Click”, MIKROE. Cit: 22. máj 2025. [Online]. Available at: http://www.mikroe.com/force-5-click

Scalable Infrastructure Monitoring System for Corporate Environments

[1*]*Christian Gucik,* [2*]*Ondrej Kainz,* [3*]*Miroslav Murin,*
[4*]*Ivana Novakova*

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

[1]christian.gucik@student.tuke.sk, [2]ondrej.kainz@tuke.sk, [3]miroslav.murin@tuke.sk, [4]ivana.novakova@tuke.sk

Abstract **– This paper introduces a novel infrastructure monitoring system designed for corporate environments, addressing critical needs for robust IT management and improved system reliability. The proposed solution efficiently collects and securely stores infrastructure metrics, providing intuitive real-time visualizations. Emphasizing flexibility, scalability, and seamless integration, the implementation leverages modern technologies and interactive interfaces, significantly enhancing monitoring capabilities and enabling faster issue resolution. Compared to established monitoring stacks, this solution prioritizes simplicity, minimal overhead, and ease of deployment, making it suitable for constrained or educational environments.**

Keywords **– alerting, infrastructure monitoring, metrics visualization, software dashboard**

I. Introduction

Modern organizations heavily rely on IT infrastructure, which continues to grow in scale and complexity. As digital operations expand, ensuring system reliability and availability becomes a critical business objective. Article [1] describes how insufficient infrastructure monitoring can lead to limited system visibility, delayed incident detection, and an increased risk of outages. These issues can result in financial losses and significant reputational damage for companies.

On-premises infrastructure offers advantages such as enhanced control, customization, and potential cost savings over time. However, article [1] also highlights that monitoring such environments poses unique challenges. Many companies lack centralized monitoring solutions, leading to fragmented oversight and operational inefficiencies. While commercial tools may provide robust features, they often come with high licensing costs. Conversely, open-source alternatives demand significant setup effort, technical knowledge, and ongoing maintenance, as discussed in article [2].

To address these gaps, this paper introduces a lightweight, cost-effective monitoring system specifically designed for on-premises environments. The system prioritizes simplicity, operational efficiency, and security. It incorporates automated alerts and proactive issue detection to enhance service reliability, reduce administrative overhead, and streamline infrastructure management.

II. Monitoring Principles and Metrics Collection Systems

This section introduces core monitoring principles, outlines their significance, and explores current technologies used for gathering and analyzing metrics. Understanding these foundational concepts enables informed decisions on selecting appropriate monitoring tools tailored to organizational needs.

A. What is Monitoring?

According to IBM [3], infrastructure monitoring is the process of observing, analyzing, and managing the performance, availability, and overall health of all backend components within an organization's technology stack. These components ranging from memory chips and processors to operating systems and application servers are crucial for delivering applications and services to end users, and may exist across cloud, on-premise, or hybrid environments.

Monitoring is essential to avoid service degradation, application outages, and potential revenue loss or reputational damage. It relies on specialized tools that automatically collect, aggregate, and analyze metrics and data from servers, virtual machines, containers, databases, and other backend systems. These tools cover key metrics such as CPU and memory usage, network traffic, disk space, response time, and error rates. Alerts are generated when thresholds are breached or anomalies detected, allowing IT teams to act proactively.

As described by the author in [4], infrastructure monitoring can be categorized into two primary types:

- **Agent-Based Monitoring:** This approach involves installing software agents on monitored devices. These agents collect performance data from infrastructure components such as servers and applications, then transmit this information to the monitoring system for analysis.
- **Agentless Monitoring:** A more lightweight method that relies on APIs and network protocols such as SNMP, HTTP, WMI, ICMP, and IPMI to monitor servers and applications without the need to install additional software on the monitored devices.

A visual comparison of agent-based and agentless monitoring approaches is shown in Figure 1.

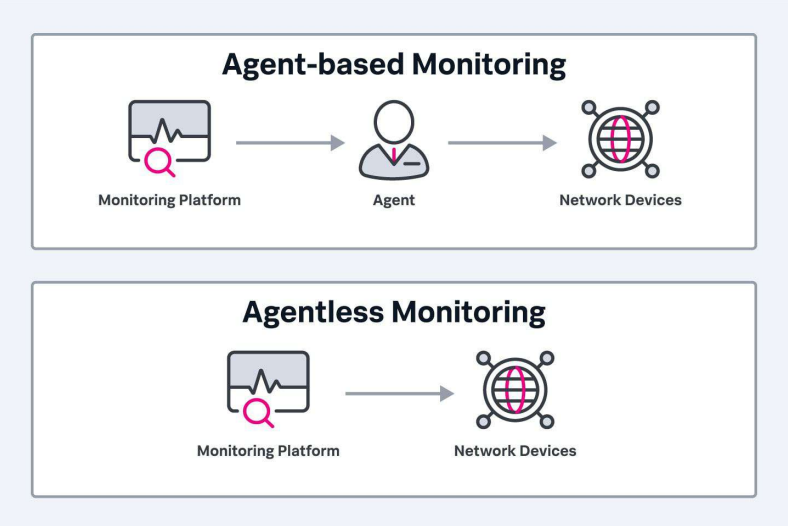


Fig. 1: Infrastructure Monitoring Types
Source Splunk: [5]

B. Analysis of current technologies for metrics collection

As the article [6] explains, metrics collection systems play a crucial role in infrastructure monitoring by gathering, storing, and analyzing data. Prometheus, as described in [7], is an open-source system designed for monitoring and alerting, utilizing a pull-based approach to gather metrics from nodes discovered through static configuration or built-in service discovery methods. Its core strengths include a powerful query language, PromQL, for sophisticated data querying and real-time alerts. Alerts can be set using fixed thresholds, as illustrated by the latency alert example in Listing 1.

```
alert: high_latency
expr: node_latency_seconds{job="some-node"} > 1.0
for: 10m
```

Listing 1: Example of alert rule using Prometheus [7]

The architecture detailed in articles [8] and [9] comprises several components: the Prometheus server, exporters for metric collection, Pushgateway, Alertmanager for alerting, PromQL, and a WebUI for data visualization, as shown in Figure 2. This section introduces core monitoring principles, outlines their significance, and explores current technologies used for gathering and analyzing metrics. Understanding these foundational concepts enables informed decisions on selecting appropriate monitoring tools tailored to organizational needs.

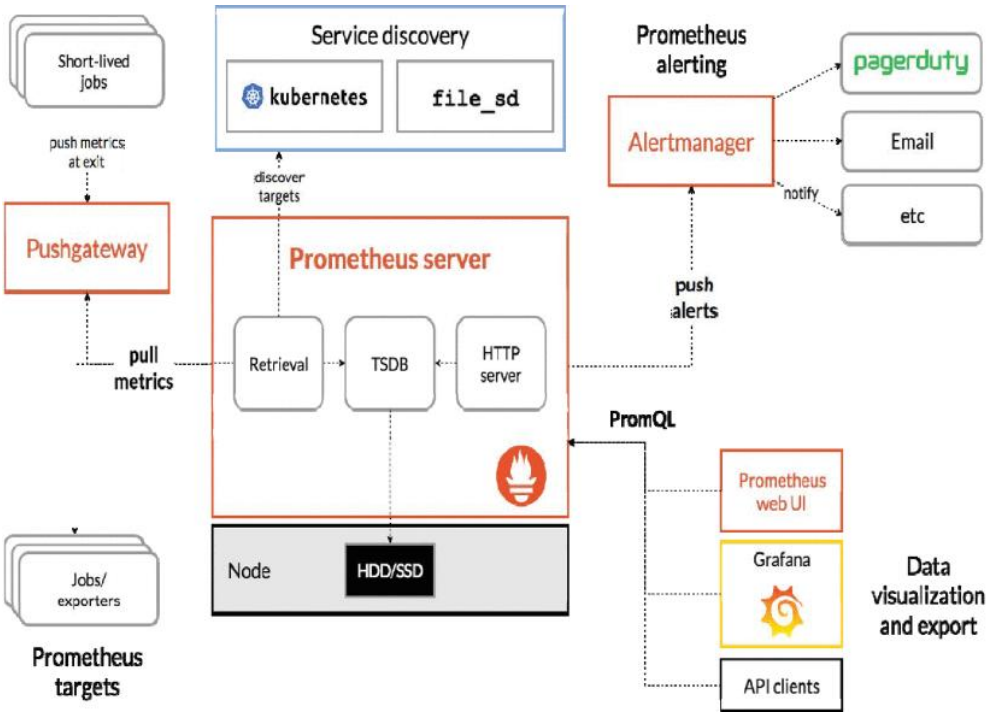


Fig. 2: Overview of Prometheus core architecture.
Source: [8]

Articles [10] and [11] explains that Prometheus provides a comprehensive HTTP API for data retrieval, configuration, and integration with visualization tools like Grafana.

Fig. 3: Graphite system architecture.
Source: [12]

Graphite, on the other hand, is another prominent open-source time-series monitoring system introduced in 2008, as outlined in [13]. It follows a push-based model, relying on applications to send metrics directly to its backend. Graphite's modular architecture, shown in Figure 3, includes components such as Carbon for data reception, Whisper for storage, and Graphite-Web for rendering visualizations [12]. Data ingestion into Graphite typically occurs via auxiliary tools

such as StatsD or Collectd. Furthermore, Graphite supports both TCP and UDP protocols for metric data transmission [14]. TCP provides reliable data transfer at the cost of increased latency due to synchronization overhead, whereas UDP is favored in environments tolerant to some data loss but requiring rapid, high-volume metric processing.

A direct comparison between Prometheus and Graphite is presented in [15], which outlines their differences in architecture, metric collection models, and scalability.

Table 1: Comparison of Existing Data Collection Systems

Tool	**Supported Protocols**	**API Support / Open Source**
Prometheus	TCP	Yes / Yes
Graphite	TCP and UDP	Yes / Yes

III. Conceptual Design of the Application

This section outlines the architectural concept of a system focused on metrics collection and visualization. A web application based on Node.js is proposed to provide user interaction and visualization functionality.

To collect data from devices, a metrics collection service with a standardized API is recommended. This API facilitates automated, efficient data transfer and seamless integration between the collection and visualization layers. It also ensures reliable, scheduled updates of current metrics.

The proposed system architecture is illustrated in Conceptual diagram 4, showing how individual components are interconnected and how data flows from source devices to the user interface via the metrics service.

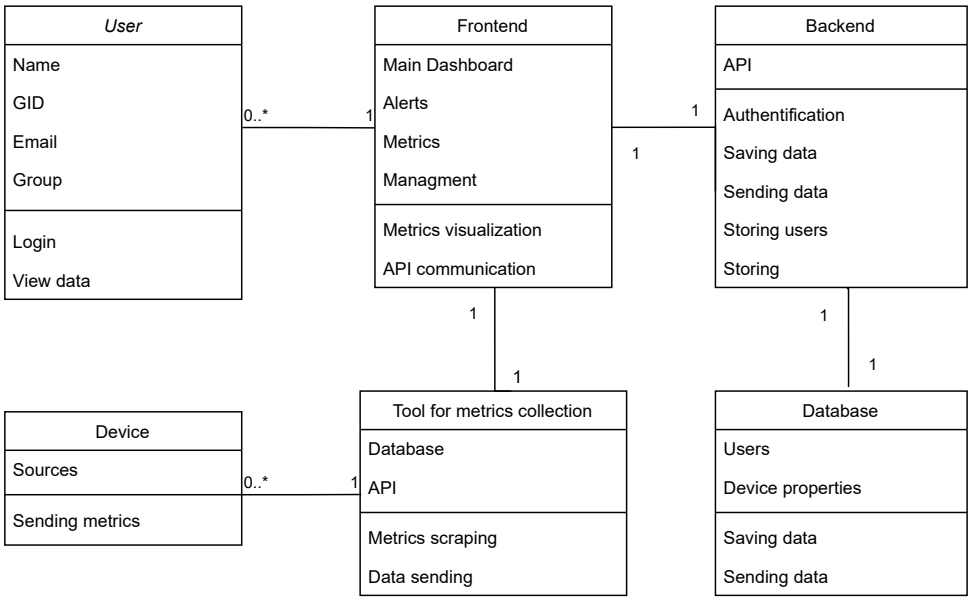


Fig. 4: Conceptual system architecture

The system is designed in a modular way to ensure high flexibility, scalability, and data processing efficiency. It includes the following components:

- **Web Application** — Serves as the main user interface for interacting with system data. It includes:
  - *Dashboard* – Displays system status and key indicators.
  - *Alerts* – Notifies users of anomalies or important events.
  - *Visualization* – Provides graphical representation of metrics.
  - *API Communication* – Facilitates integration with the metrics collection layer.

  The web application communicates with the metrics collection service and the backend server.
- **Metrics Collection Service** — Acts as the central node for acquiring, processing, and storing data from devices. Its functions include:
  - *Data Management* – Acquiring and aggregating metrics from devices.
  - *API Interface* – Exposing data in a standardized format.

  - *Data Storage* – Long-term retention of device-collected metrics.
- **Backend Server** — Operates as a middleware between the web application and the database. Responsibilities include:
  - User authentication and authorization.
  - Interaction with the local database.
- **Database** — Provides structured, persistent storage for collected metrics and user-related data.
- **Devices** — Serve as data sources via sensors or measurement units. The relationship between devices and the Prometheus-based collector is many-to-one (**0..***), allowing the system to scale with an arbitrary number of input sources.

IV. Implementation

This section describes the implementation of the proposed system, focusing on technologies used for metrics collection, backend communication, database storage, and data visualization.

A. Implementation Tools

Figure 5 illustrates the major components of the implementation stack and their interdependencies.

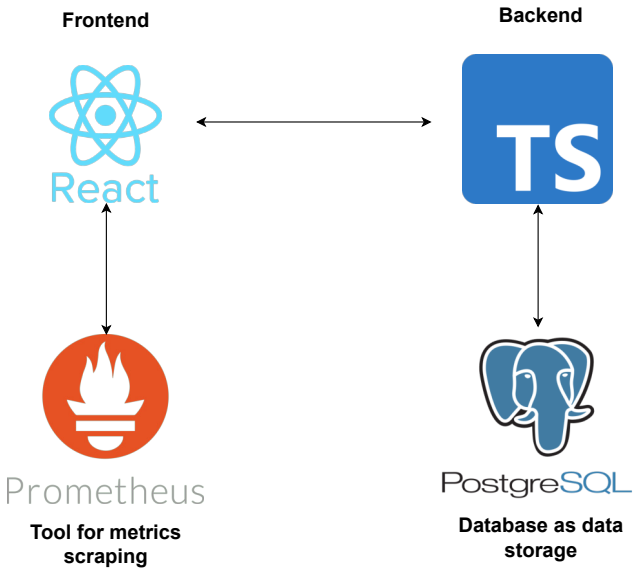


Fig. 5: Overview of implementation technologies and their roles in the system.

B. Web Application

The frontend application is implemented using the *React* framework, providing the graphical interface and communicating with the backend via REST API endpoints. Dependency management is handled by `Node.js`, based on definitions in the `package.json` file.

C. Backend Application

The backend service, found in the `backend` directory, is written in *TypeScript* and provides a RESTful API interface. It serves as a bridge between the frontend and the database, handling data retrieval, transformation, and user authentication logic.

D. PostgreSQL Database

The PostgreSQL database stores key system data, including user information, device metadata, and exporter configurations. The server-side code uses the `pg` library for interacting with PostgreSQL, offering an efficient and straightforward API for executing SQL queries and managing connections. PostgreSQL was chosen due to its robust support for relational data, active development community, and seamless integration with Node.js-based backend services. In contrast to document-oriented databases such as MongoDB, it provides transactional integrity and a mature ecosystem for access control and structured querying, which are beneficial for multi-user monitoring environments.

E. Metrics Collection Tool

Prometheus serves as the core tool for collecting and storing metrics from devices and services. It is deployed along with supporting exporters like `Node Exporter` and `cAdvisor` using containerized environments to ensure modularity and ease of deployment. Prometheus was selected over alternatives like ELK or Graphite for its efficient pull-based architecture, native alerting support, and tight integration with exporter tools such as Node Exporter and cAdvisor. Its simple deployment model and compatibility with containerized environments made it particularly suited for modular on-premises installations.

F. Chart Rendering

The graph visualization is implemented using the **Chart.js** library, which supports multiple chart types including line, bar, pie, doughnut, and text-based outputs. Each chart is rendered based on the type of metric and the format of data received from the exporter.

The data visualized in the charts is obtained from the Prometheus HTTP API. During runtime, each chart is instantiated via a reference to a **Chart.js** object, which is updated asynchronously. Every 60 seconds, an asynchronous request fetches the latest data from Prometheus through the backend API. These values are appended to the dataset, and the chart is re-rendered using internal methods of the `Chart.js` library. This ensures that users interact with live data without needing to refresh the page. The results of the chart rendering can be seen in Figure 6, which shows live metrics visualization within the application.

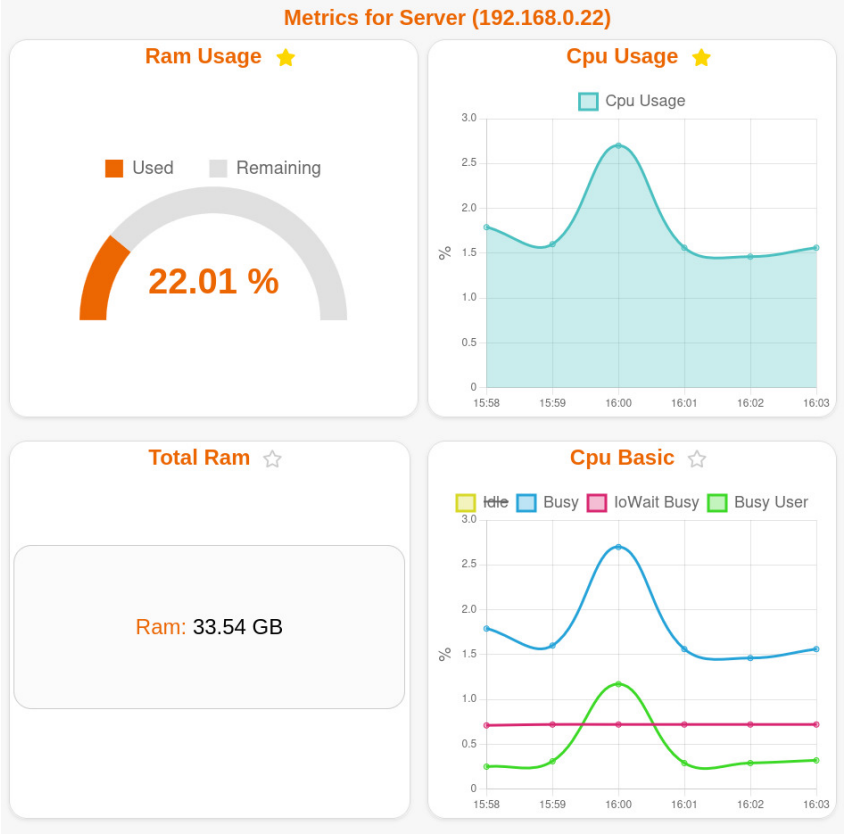


Fig. 6: Live metrics visualization using Chart.js, showing real-time system parameters such as CPU and memory usage. Data is refreshed asynchronously every 60 seconds.

V. Evaluation

This section evaluates the final system deployment, emphasizing usability, functionality, and overall effectiveness. Evaluation was conducted in two phases with target users, employing the User Experience Questionnaire (UEQ) to systematically assess various usability aspects such as efficiency, ease of use, and interface attractiveness. Additional qualitative feedback was collected through open-ended questions, providing insights into user experiences and suggestions for system improvements.

Table 2: Average UEQ Scores for the Deployed Application

UEQ Scale	Mean	Variance
Attractiveness	2.18	0.20
Perspicuity	2.08	0.49
Efficiency	2.23	0.46
Dependability	2.30	0.22
Stimulation	1.70	0.59
Novelty	1.70	0.61

A. *User Feedback*

UEQ results summarized in Table 2 indicate positive user perceptions across all categories.

Highest scores were recorded for Dependability (2.30), Efficiency (2.23), and Attractiveness (2.18), indicating high reliability, functionality, and visual appeal. Lower, yet positive, scores for Stimulation and Novelty suggest room for enhancing user engagement and innovative aspects.

Benchmark comparison with UEQ standards in Table 3 confirmed outstanding results, placing most categories within the top 10% of reference samples.

Table 3: Application Results Compared to UEQ Benchmark

UEQ Scale	Score	Category	Interpretation
Attractiveness	2.18	Excellent	Top 10%
Perspicuity	2.08	Excellent	Top 10%
Efficiency	2.23	Excellent	Top 10%
Dependability	2.30	Excellent	Top 10%
Stimulation	1.70	Good	Top 25%
Novelty	1.70	Excellent	Top 10%

Qualitative feedback highlighted the application's intuitiveness, speed, and ease of use, with minor suggestions regarding navigational improvements.

B. *System Functionality Evaluation*

The deployed system, based on Prometheus and various specialized exporters, actively monitors 20 devices, demonstrating robust scalability and reliability. Continuous monitoring confirmed the system's capability to manage increasing data volumes without significant performance degradation or downtime.

Performance and stability were consistently high throughout the testing period, effectively handling both routine operations and fault scenarios. Identified limitations relate primarily to storage requirements, highlighting the necessity for planned infrastructure scaling and performance monitoring to ensure ongoing system efficiency and reliability.

The implemented system was conceptually and functionally compared against two prevalent open-source monitoring stacks: the ELK stack (Elasticsearch, Logstash, and Kibana) and the Prometheus + Grafana combination.

The ELK stack, while powerful for log indexing and search, introduces significant complexity in deployment and operation. It is less suited for lightweight, metric-focused monitoring, and typically requires higher system resources.

Prometheus + Grafana is architecturally closer to our solution. However, Grafana, though feature-rich, brings additional dependencies and configuration requirements. By contrast, our use of Chart.js enables seamless integration within a single-page application and avoids the need for an external visualization layer, providing a leaner and more controlled frontend experience.

Overall, our architecture offers a balance between customizability and simplicity, targeting scenarios where rapid deployment, ease of maintenance, and tight backend–frontend coupling are prioritized over advanced data aggregation features.

VI. Conclusion

This paper presented a lightweight, cost-effective system designed to address monitoring challenges within on-premises infrastructure environments. The implemented solution distinguishes itself from existing dashboards by integrating a modular, scalable frontend-backend design tailored for custom deployment environments. It successfully achieves its objectives by efficiently collecting, processing, and visualizing infrastructure metrics, demonstrating robust scalability and operational reliability.

User evaluations using the User Experience Questionnaire (UEQ) highlighted the system's strong points, such as dependability, efficiency, and overall attractiveness. Qualitative feedback further reinforced these results, with users appreciating the intuitive interface, rapid response times, and ease of use. Benchmark comparisons confirmed the system's exceptional user experience, ranking among the top 10% in most evaluated categories.

Operational evaluations affirmed the system's stability and capability to manage continuous data collection and real-time monitoring tasks. Despite positive outcomes, limitations regarding storage capacity and long-term data retention were identified, suggesting careful infrastructure planning and scalability considerations for future expansions.

Future improvements could include expanding the system to manage multiple geographic locations, integrating local artificial intelligence for advanced analytics, and pursuing full automation of infrastructure optimization. Overall, the proposed system effectively enhances infrastructure visibility, reduces issue detection times, and facilitates proactive management in a cost-effective manner.

VII. Abbreviations and Acronyms

The following abbreviations and acronyms are used throughout the article. Each is defined upon first mention in the text:

- **API** – Application Programming Interface
- **UEQ** – User Experience Questionnaire
- **HTTP** – Hypertext Transfer Protocol
- **SNMP** – Simple Network Management Protocol
- **WMI** – Windows Management Instrumentation
- **ICMP** – Internet Control Message Protocol
- **IPMI** – Intelligent Platform Management Interface
- **SQL** – Structured Query Language
- **UI** – User Interface
- **UX** – User Experience
- **CPU** – Central Processing Unit

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. 004 ŽU-4/2024.

References

[1] Lumos, "Mastering it infrastructure monitoring: A comprehensive guide for it operations," https://www.lumos.com/topic/it-operations-it-infrastructure-monitoring, 2025, accessed: 2025-05-02.
[2] Obkio. (2024) Open-source network monitoring tool vs. paid: Pros & cons. Accessed: 2025-05-02. [Online]. Available: https://obkio.com/blog/open-source-network-monitoring-tools-vs-paid/
[3] IBM, "What is infrastructure monitoring?" https://www.ibm.com/topics/infrastructure-monitoring, 2023, accessed: 2025-05-02.
[4] R. LaFlamme, "Agent vs agentless monitoring: Which is best?" https://www.auvik.com/franklyit/blog/agent-vs-agentless-monitoring/, 2023, accessed: 2025-05-02.
[5] Splunk Inc., "What is infrastructure monitoring?" https://www.splunk.com/en_us/blog/learn/infrastructure-monitoring.html, 2023, accessed: 2025-05-02.
[6] Dynatrace. (2025) What is infrastructure monitoring? Accessed: 2025-05-02. [Online]. Available: https://www.dynatrace.com/news/blog/what-is-infrastructure-monitoring-2/
[7] O. Mart, C. Negru, F. Pop, and A. Castiglione, "Observability in kubernetes cluster: Automatic anomalies detection using prometheus," in *Proc. IEEE Int. Conf. High Performance Computing and Communications (HPCC); Smart City; Data Science and Systems (DSS)*, 2020, pp. 565–570.
[8] L. Chen, M. Xian, and J. Liu, "Monitoring system of openstack cloud platform based on prometheus," in *Proc. Int. Conf. Computer Vision, Image and Deep Learning (CVIDL)*, 2020, pp. 206–209.
[9] J. Turnbull, *Monitoring with Prometheus*. Turnbull Press, 2018.
[10] Prometheus Project, "Visualization with grafana," https://prometheus.io/docs/visualization/grafana/, 2023, accessed: 2025-05-02.
[11] ——, "Prometheus http api documentation," https://prometheus.io/docs/prometheus/latest/querying/api/, 2023, accessed: 2025-05-02.
[12] D. Berman, "Prometheus vs. graphite: Which should you choose for time series or monitoring?" https://logz.io/blog/prometheus-vs-graphite/, 2023, accessed: 2025-05-02.
[13] MetricFire, "Monitoring with graphite: Architecture and concepts," https://medium.com/@MetricFire/monitoring-with-graphite-architecture-and-concepts-de744853badc, 2023, accessed: 2025-05-02.
[14] ——, "Graphite graphing and monitoring tool," https://www.metricfire.com/blog/graphite-graphing-and-monitoring-tool/, 2023, accessed: 2025-05-02.
[15] P. Sonpatki, "Graphite vs prometheus," https://last9.io/blog/graphite-vs-prometheus/, 2023, accessed: 2025-05-02.

Web Platform for Personal Finance Management

[1*]Nazarii Herasymchuk, [2*]Ondrej Kainz, [3*]Miroslav Murin,
[4*]Ivana Novakova, [5*]Miroslav Michalko

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

[1]nazarii.herasymchuk@student.tuke.sk, [2]ondrej.kainz@tuke.sk,
[3]miroslav.murin@tuke.sk, [4]ivana.novakova@tuke.sk, [5]miroslav.michalko@tuke.sk

Abstract **– This paper presents the design and implementation of a web-based platform for personal finance management, integrating modern approaches in data security and financial trend prediction. It focuses on the analysis of current solutions in the field of data protection, with an emphasis on safeguarding sensitive user information. Furthermore, the study explores the application of machine learning techniques in forecasting personal financial developments. A comparative analysis of existing personal finance planning tools is also conducted. The proposed platform provides users with real-time insights into the performance of funds and commodities, supporting more effective financial analysis and decision-making.**

Keywords **– Financial planning, personal budget, personal finance, web application**

I. Introduction

In today's dynamic world, the topic of investing is gaining increasing attention, as individuals seek to supplement their primary income through financial markets and equity ownership in specific companies. This investment approach not only enables personal capital growth but also supports corporate expansion and value appreciation. Investors benefit from the price differential between the initial and current value of their shares. While investment processes are more complex than they may initially appear, they offer considerable long-term potential for capital appreciation.

Contemporary investing no longer relies on physical securities or direct human interaction, as was common decades ago. Instead, digital investment platforms have transformed the landscape by offering online access to brokers and automated systems for buying and selling financial instruments. These platforms handle vast volumes of data, enabling fast and efficient transaction processing in near real-time. In line with these developments, current research explores how user behavior history can enhance security mechanisms in online financial applications, including investment platforms [1].

Despite this accessibility, many individuals still lack a clear understanding of how investing works, how to get started, where to allocate funds, for how long, and how much to invest. To address this gap, the system proposed in this work aims to assist users in managing their income and expenses, monitoring the prices of specific companies or funds, and visualizing potential market trends. The prediction features of this system, inspired by methodologies such as random forest regression, contribute to better foresight into price movement and investment outcomes [2]. This information will enable users to make informed decisions and develop personalized investment strategies.

Beyond its core financial management functionality, the system offers additional features such as transaction history tracking, price prediction, and investment history tracking. Although initially intended for individual users, the platform is scalable and can serve broader user groups, such as families, teams, or small organizations pursuing collective financial goals.

Furthermore, the platform holds significant educational potential. By incorporating essential economic concepts and investment terminology, it can act as a training tool for those new to investing. This dual purpose—practical application and education—positions the system as a valuable contribution to both personal finance management and financial literacy development.

II. System architecture analysis

The analytical component of this work focuses on key areas essential for system development: data security, finance forecasting, and a comparative review of existing financial planning tools and graph visualization libraries. The study begins by examining current methods for password storage and user authentication, highlighting prevailing trends and best practices in cybersecurity. It then explores the potential integration of AI, assessing its advantages and limitations within the context of personal finance management. Additionally, the functionality, flexibility, and integration capabilities of available financial data APIs are evaluated. A comparative analysis of existing investment planning systems is also conducted to inform the selection of relevant features. The goal is to establish a comprehensive overview of current technologies and trends that can enhance system design, data security, and user experience. The following sections describe in more detail the methods used in the work.

A. Data security

Security of user data is described in a few methods and the main of them are password security storing [3], security authorization principles, and authentication methods, references [4] and [5]. Every security method was described and used in the work. Storing passwords securely in the work was done by hashing with algorithms bcrypt. It is preferred over encryption or SHA following the next article [6]. Bcrypt offers strong protection and good performance for web use, reference [7], while Argon2 is more secure but resource-intensive following reference [8]. Plain-text storage is unsafe and must be avoided. Authorization is made for protecting users' sensitive financial data and ensuring privacy. Strong authentication methods—such as passwords, two-factor authentication, and biometrics—have been analysed to use to prevent unauthorized access. Finally, a standard authentication method with password and email was used. Plus, email verification was added to the process. Additionally, in the work, an analysis of JWT (JSON Web Token) as a user authentication method was conducted, allowing secure and stateless verification of user identity through signed tokens, reference [1]. Other methods were also considered. Traditional session-based authentication and integration with OAuth 2.0 and OpenID Connect for third-party login, as referenced in [9]. The simpler API key authentication was also reviewed but deemed less secure and not suitable for user-level access control.

B. Finance forecasting

As prediction of future stock market price, an analysis of simple models such as linear regression, random forest, and decision tree was conducted, reference [2]. Linear regression was used in the work as a simple predictive model to analyse price trends, references [10] and [11], specifically to identify upward or downward movement. Despite its simplicity, it provides a useful overview of the relationships between time periods. Due to its ease of implementation, it was applied as an initial prediction tool before moving on to more complex models such as decision trees.

C. Planning method and graph visualization

As a planning method for the system, the popular envelope budgeting method was chosen, reference [12], which is also used in similar systems like YNAB, reference [13], for example. The envelope budgeting system is a method that uses different spending categories to allocate cash into separate envelopes. Each envelope represents a budget category (e.g., groceries, rent, entertainment), and the method's purpose is to ensure that money is spent only from the designated envelope. Once it's empty, you can't spend more in that category unless you move money from another envelope. This system promotes disciplined, intentional spending and helps avoid overspending.

The envelope method remains very useful for this kind of system and is used in applications like YNAB and GoodBudget, reference [14], which were compared in this work. The comparison of these applications was based on several factors, covering a wide range of aspects from design interface to the available functionality for users. In the end, YNAB was chosen as the better system, offering more possibilities for tracking personal finances.

A graph visualization analysis was conducted to explore different approaches for representing graphs in a web system, comparing two main libraries: Chart.js and ApexCharts. The analysis showed that ApexCharts offers better interactivity, real-time data support, and extensive customization options, reference [15]. Chart.js, on the other hand, stands out for its simplicity and ease of use, particularly for beginners or smaller projects. It provides clean and visually appealing charts with minimal setup, while still offering enough flexibility for advanced configurations. As an open-source tool with an active community, it is a reliable and well-maintained option that can be used in both basic and more complex web applications.

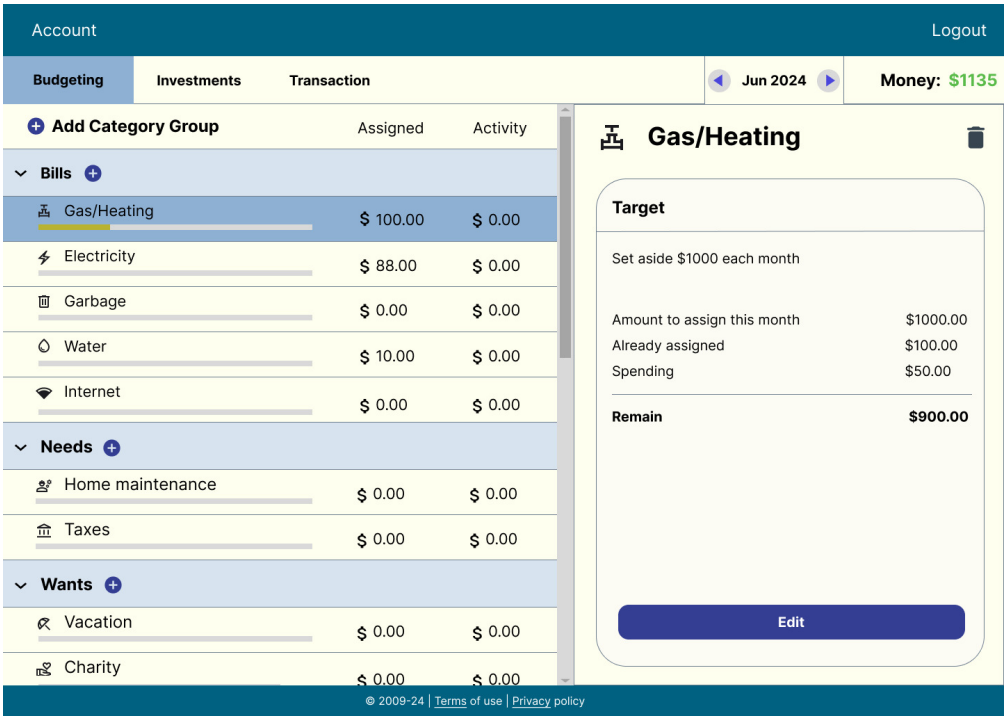


Fig. 1: Budgeting page prototype

III. Solution Design

In the design for this work, the goal was to create a visual design mockup of how the system will look to the end user.

The design includes login, registration, a budgeting interface based on the envelope method, investment planning, transaction history, prediction tools, and a user-friendly tutorial section.

A prototype was created using Figma, and a flowchart illustrates the navigation between key pages. The prototype is designed to support multiple languages (English, Slovak, Ukrainian), with English as the default. Language translations are handled via JSON files for scalability.

The application aims to simplify financial management compared to existing solutions like YNAB [13] by offering a cleaner, more intuitive interface. It reduces complexity by using only two budget columns (allocated and used funds), eliminating the often confusing third column (next month's balance). This simplified design is based on user reviews, which showed a preference for clarity and ease of use.

The prototype consists of several pages, one of them is shown in Figure 1, including the sign-up page, login page, and intro page. The prototype also includes a planning page where the user can manage their finances using the envelope budgeting method. Additionally, it features an investment page for tracking current market data, which is displayed as a graph.

Ultimately, the system prototype tries to provide essential budgeting and investment tools without overwhelming users, focusing on usability and accessibility for beginners.

Project design solution also described possible technologies that will be used further in work, for client and server side. For client side, technologies like Angular as the main framework for implementation were evaluated. The back end was evaluated using Node.js with Express.js for building REST APIs. Node.js allowed asynchronous request handling, which ensured fast and efficient server performance, and it provided smooth integration with frontend frameworks. Spring Boot was also evaluated as an alternative—especially for large-scale enterprise applications—Node.js was selected for its flexibility and suitability for rapid development.

IV. Project Implementation

This section presents all the technologies that were used in the project, including both the server-side and client-side components. The aim is to provide a clear overview of the tools, frameworks, and libraries that were selected to support the development of the system. Each technology is briefly described, along with the reasons for its inclusion and how it contributes to the overall functionality and performance of the application.

Also in the section described the graphical interface and the database that were used in the project.

A. Client side

The client side of the personal finance management system was developed separately from the server side logic to ensure better scalability and easier maintenance. It was built using Angular, Bootstrap, and HTML/CSS. HTML/SCSS was used to provide the structure and styling of the interface, making it possible to build a graphical user interface. Angular was used for creating dynamic, modular applications with reusable components, which also simplify and speed up the development process. Bootstrap simplifies UI development with ready-to-use components and helps maintain a consistent, responsive design across devices.

For data visualization in graphs, ApexCharts and Chart.js were used. While Chart.js is simpler, ApexCharts also offers more interactivity and customization, making it more suitable for real-time financial data. Both libraries were used in the system on different pages.

The system was also used "Twelve Data API" to access real-time and historical financial market data, allowing users to track prices of stocks, currencies, commodities, and indices.

B. Server side

The server side of the project used Node.js as the main framework for building server-side logic and handling communication with the database. In combination with Node.js, Express.js was also used, as it enables fast development of RESTful APIs with efficient, asynchronous request handling ideal for lightweight applications. It integrates well with frontend frameworks like Angular and provides easy database connectivity (e.g., PostgreSQL), making it highly effective.

Since the server side and client side are implemented as two separate parts of the system and run on different ports, CORS (Cross-Origin Resource Sharing) was used on the server side. CORS is an HTTP header-based mechanism that allows a server to specify which origins (domain, scheme, or port) are permitted to access its resources, enabling secure communication between the frontend and backend.

C. Database

The system used a PostgreSQL database to store user data. The main table, user, holds essential information such as name, password, and email. Additional tables store user activities related to saving and investing.

Key tables include:

- Preferences – stores user settings like theme and language.
- Saving Plan – a table to record details of the savings category. Includes basic data such as category name, description, amount of money allocated, target, dates and user activity (income and expenditure).
- Sections – this table is designed to facilitate the grouping of data in the savings plan table and contains only basic information such as the name of the section and a reference to the user to whom it belongs.
- Investment Plan – similar to saving plans but includes broker info, price, and quantity, and links directly to the user.
- Transactions – logs income and expenses, linked to users and saving plans. Each entry includes either an income or an expense field, but never both.

All tables are connected via foreign keys referencing the user ID. The structure is shown in the relational database model.

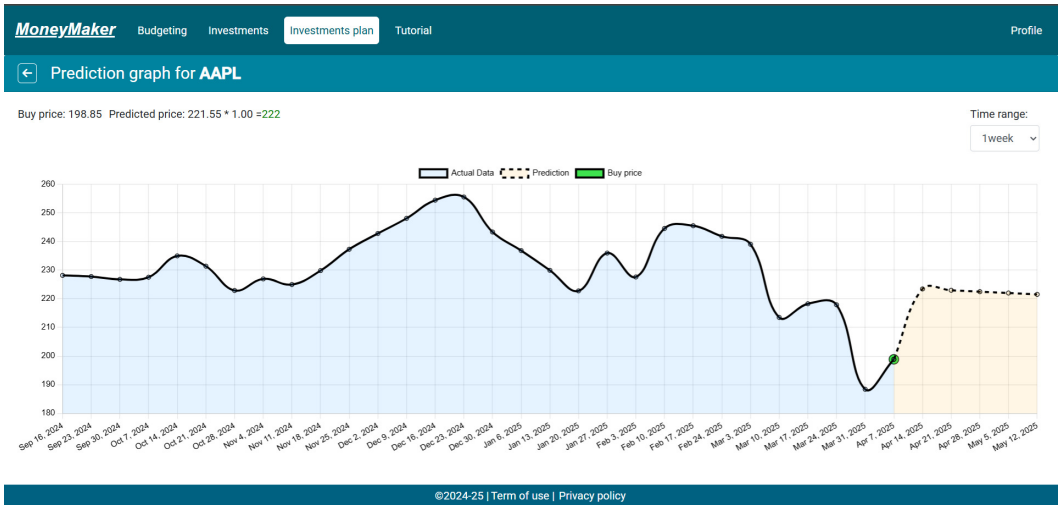
D. Graphical interface

The application's user interface was based on a Figma design created during the draft phase. The system includes eleven distinct pages, each serving a specific function. It also features several pop-up windows to display smaller information or gather user input without navigating away.

Upon entering the app, users land on the homepage, where they can register or log in. The system remembers the user's last visited page and login status, eliminating the need to log in repeatedly. The design and functionality are inspired by popular saving and investment platforms, with the budgeting page resembling YNAB and the investment and prediction pages featuring stock-style charts.

Each page (except login and registration) includes a consistent header for navigation and a footer for additional information. The color scheme uses a light cream or white background, teal for the header and footer, dark blue, and green for buttons to visually distinguish them.

Charts are present on the budgeting, investment, and prediction pages 2, implemented using Chart.js and ApexCharts. While both offer similar visual outputs, Chart.js proved to be simpler to work with.



(a) Prediction page

Fig. 2: The final design prediction page of the application.

V. Evaluation

The result of the project implementation is a functional prototype of the future application, including all planned features. However, thorough testing is needed to address identified issues. To improve the system, testing is conducted with students and regular users without prior programming experience.

The testing process is divided into three parts:

Test Design – Describes the chosen testing methods and scenarios.

Test Execution – Details the conditions, tools, and participants involved in testing.

Evaluation – Summarizes results using charts and user feedback, highlighting suggestions for improvement.

The goal is to identify areas for enhancement and propose solutions to improve usability, functionality, and overall user experience.

A. Test Design

In the testing scenario, participants were instructed to access the system via a web browser and complete the following tasks:

- Register within the application and verify their email address.
- Log in using the newly created credentials.
- Enter an initial financial amount (a value exceeding 1,000 was recommended).
- Review the provided tutorial to become familiar with the system.
- Navigate to the ”Budgeting” section, create a budget category and item, and define a target amount and deadline.
- From the ”Budgeting” interface, proceed to the ”Transactions” section and assign a monetary amount to the previously created budget item.
- Access the ”Investments” module and create an investment plan for a selected ETF or stock.
- Navigate to the ”Investment Plan” section to confirm the creation of the investment plan.
- Independently explore the system, define personal financial goals within the application, and attempt to meet them by allocating funds to appropriate categories.
- Complete the CSUQ (Computer System Usability Questionnaire), which assesses user satisfaction with system usability, quality of information, and interface design. Responses were rated on a 7-point scale ranging from ”Strongly agree” (7) to ”Strongly disagree” (1). An additional section was provided for reporting issues or submitting comments.

This scenario was designed to evaluate the core functionalities, overall usability, and user-friendliness of the system.

B. Test Execution

The testing process was conducted through both remote and in-person sessions. Participants involved in remote testing received a detailed set of instructions along with access to the system and a link to the evaluation questionnaire. In-person testing was carried out under supervision to

ensure a controlled environment, while avoiding direct intervention in order to replicate typical user behavior.

A total of 18 individuals participated in the testing, including 12 students from the Technical University of Košice (TUKE). The remaining participants represented a heterogeneous group with no prior experience in information technology or similar digital systems, providing a broader perspective on the application's usability.

Remote testing utilized platforms that supported screen sharing and video conferencing, enabling observation of user interactions. In-person testing took place in the TUKE library and other appropriate environments, with each session conducted individually to maintain focus and minimize external influences. All sessions were carried out using laptop or desktop computers to ensure consistency in the testing conditions.

The application is designed for a general audience; therefore, no specific target demographic was defined. Basic demographic information was collected to provide contextual insights but was not considered a primary evaluation criterion. The overarching objective of the testing was to evaluate the usability and core functionality of the system, gather user feedback, and identify opportunities for improving the overall user experience across diverse user groups.

C. Evaluation

The questionnaire included 18 questions and was completed by all participants. The tested system version contained all planned features.

Users had varying experience levels with similar apps. Interestingly, those without prior experience, rated the system more positively (average 6–7 points) than more experienced users (average 5–6 points). Overall feedback was positive, though some usability issues were noted.

The main issue identified was the unclear process of adding transactions, particularly confusion about the interaction between the budgeting and transaction pages. This was due to the system's design, which separated budget data into a different page. To improve this, a feature was added allowing users to assign funds directly through the transaction page via double-clicking. Users also requested a clearer overview of expenses by category, which led to the implementation of a summary chart with spending data and statistics.

Testing revealed several usability issues and bugs, which were fixed after testing. These findings helped to improve the project in the interface and functionality, making it more user friendly.

VI. Conclusion

This paper presented the design and implementation of a personal finance management system aimed at supporting users in effective saving, budgeting, and financial planning. The system consists of two primary modules: saving and investing. It also incorporates advanced functionalities such as investment tracking and predictive financial analytics.

The development focused on usability, featuring a clean user interface and an integrated help section to facilitate user onboarding. The system design was inspired by existing solutions such as YNAB for budgeting and various investment platforms for financial monitoring. Key technical features include secure authentication using JSON Web Tokens, password hashing, and route protection implemented in Angular. Predictive modeling was carried out using linear regression, while financial data visualization is achieved through ApexCharts and Chart.js. The chosen technology stack comprises Angular for the frontend, Node.js for the backend, and PostgreSQL as the database, with design prototypes created in Figma.

User experience was evaluated using the CSUQ questionnaire, administered both in person and online. The feedback indicated a generally positive user reception. The application supports multi-language functionality and was initially developed in English, with a scalable architecture for future language additions. Users are able to record categorized transactions, monitor investments with real-time market data, and receive predictions on the future performance of stocks, ETFs, and commodities. Additionally, the system serves as a simulation tool for personal savings and financial planning.

A potential direction for future improvement is the integration of banking APIs to automate expense tracking and further enhance user experience.

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. 004 ŽU-4/2024.

References

[1] A. Bucko, K. Vishi, B. Krasniqi, and B. Rexha, "Enhancing jwt authentication and authorization in web applications based on user behavior history," *Computers*, vol. 12, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2073-431X/12/4/78

[2] P. F. Smith, S. Ganesh, and P. Liu, "A comparison of random forest regression and multiple linear regression for prediction in neuroscience," *Journal of Neuroscience Methods*, vol. 220, no. 1, pp. 85–91, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027013003026

[3] G. Hatzivasilis, "Password-hashing status," *Cryptography*, vol. 1, no. 2, 2017. [Online]. Available: https://www.mdpi.com/2410-387X/1/2/10

[4] S. Z. Syed Idrus, E. Cherrier, C. Rosenberger, and J.-J. Schwartzmann, "A Review on Authentication Methods," *Australian Journal of Basic and Applied Sciences*, vol. 7, no. 5, pp. 95–107, Mar. 2013. [Online]. Available: https://hal.science/hal-00912435

[5] A. Ometov, S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen, and Y. Koucheryavy, "Multi-factor authentication: A survey," *Cryptography*, vol. 2, no. 1, 2018. [Online]. Available: https://www.mdpi.com/2410-387X/2/1/1

[6] A. Sadeghi-Nasab and V. Rafe, "A comprehensive review of the security flaws of hashing algorithms," *Journal of Computer Virology and Hacking Techniques*, vol. 19, no. 1, pp. 287–302, 2023.

[7] P. Sriramya and R. Karthika, "Providing password security by salted password hashing using bcrypt algorithm," *ARPN journal of engineering and applied sciences*, vol. 10, no. 13, pp. 5551–5556, 2015.

[8] A. Biryukov, D. Dinu, and D. Khovratovich, "Argon2: New generation of memory-hard functions for password hashing and other applications," in *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2016, pp. 292–302.

[9] D. Fett, R. Küsters, and G. Schmitz, "The web sso standard openid connect: In-depth formal security analysis and security guidelines," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 2017, pp. 189–202.

[10] X. Su, X. Yan, and C.-L. Tsai, "Linear regression," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 3, pp. 275–294, 2012.

[11] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012.

[12] J. McCaffery, "Canada's envelope budgeting system," *American Review of Canadian Studies*, vol. 14, no. 1, pp. 45–62, 1984. [Online]. Available: https://doi.org/10.1080/02722018409481437

[13] "You Need A Budget," Website, 2024. [Online]. Available: https://www.ynab.com/

[14] Goodbudget, "Goodbudget: Budget tracker & planner," 2024, accessed: 2024-03-23. [Online]. Available: https://www.goodbudget.com

[15] "Apexcharts: Interactive javascript charting library," JavaScript library, accessed on March 8, 2024. [Online]. Available: https://apexcharts.com/

Design and Implementation of an AI-Driven Mobile Fitness and Nutrition Assistant

[1*]*Filip Rzuhovský,* [2*]*Ondrej Kainz,* [3*]*Miroslav Murin,*
[4*]*Ivana Novakova*

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

[1]filip.rzuhovsky@student.tuke.sk, [2]ondrej.kainz@tuke.sk, [3]miroslav.murin@tuke.sk, [4]ivana.novakova@tuke.sk

Abstract **– This paper presents the design and implementation of an AI-driven mobile application for personalized fitness and nutrition guidance. Developed using the React Native framework, the application leverages large language models (LLMs) to enable interactive communication with users and deliver tailored recommendations in the domains of exercise and dietary planning. Core functionalities include the generation of individualized training plans, calorie intake calculations, progress tracking, and adaptive suggestions based on user data over time.**

In addition to its practical use as a personal health assistant, the application is designed with the potential to support scientific research in fitness and nutrition. While this research capability is not yet fully implemented, future development will enable data collection and analysis for academic purposes. This will allow researchers to access real-world user data, while users benefit from evidence-based recommendations delivered in a clear and accessible format.

Keywords **– Artificial Intelligence, Fitness, Mobile Application, Personalized Training Plans, Nutrition, Progress Tracking, Scientific Research.**

I. Introduction

Nowadays, AI is becoming increasingly important in the field of health, fitness and lifestyle. With the growing interest of athletes in personalized training plans and nutritional recommendations for fitness application important for a wide range of users - from beginners to professional users [1].

The traditional approach to fitness training has often relied on generic workout plans and standardized nutritional advice, which fails to account for individual differences in physiology, fitness levels, and personal goals [2]. As pointed out by Ferraro et al. [3], personalization is key to achieving optimal results in fitness and nutrition regimens. Mobile applications integrated with AI technologies are uniquely positioned to deliver this personalized experience at scale, providing users with adaptive recommendations that evolve based on their progress and feedback.

Recent advancements in machine learning algorithms and natural language processing have enabled more sophisticated interactions between users and fitness applications [4]. These technologies allow for real-time adjustments to training programs, intelligent analysis of performance data, and conversational interfaces that simulate the experience of having a personal trainer. Such features are particularly valuable in contexts where access to professional fitness guidance is limited by geographical constraints or financial considerations.

In the following chapters, we present the solution to the current state of the application, the technical system and the key functions of the application. We will also focus on evaluating its effectiveness and the possibility of further expansion.

II. CURRENT STATE OF FITNESS APPS AND ARTIFICIAL INTELLIGENCE INTEGRATION

Fitness apps have become a popular tool [5] for promoting a healthy lifestyle in recent years. Based on the analysis available in the paper, most existing solutions offer basic functionalities [6],

such as generating training plans, tracking physical activity, or counting calorie intake. Studies have shown that calorie tracking is among the most utilized features in fitness applications [7], with significant impact on user adherence to diet plans when implemented effectively [8].

A. Fitness application landscape and market trends

The fitness application landscape is dominated by several major players that have established themselves as industry standards. Among the most widely used applications are MyFitnessPal, recognized for its extensive food database, and Nike Training Club, which offers professionally designed workout programs. Other popular platforms include Strava, which focuses on the social aspects of fitness through activity sharing [9], and Freeletics, which pioneered AI-based training plans [8]. More specialized coaching applications like True Coach have carved out their niche by focusing on trainer-client relationships [10].

Mobile platforms have emerged as the dominant medium for fitness application delivery, with compelling statistics supporting this trend. In 2024, mobile users account for more than 50.48% of all web traffic [11], with 90% of mobile device time spent in specialized applications. The younger demographic (18-34 years) particularly favors mobile devices, with over 70% [12] using smartphones as their primary internet tool and 40% as their exclusive means of accessing online content. Furthermore, mobile devices serve as the primary platform for online transactions, facilitating 58% [12] of all purchases across various devices.

Beyond usage statistics, mobile platforms offer unique advantages for fitness solutions – they provide sufficient hardware capabilities, are ubiquitous and compact, and integrate exceptionally well with complementary technologies such as smartwatches and fitness trackers. These factors make mobile devices not merely an optimal choice, but practically essential for implementing intelligent fitness training solutions.

B. Essential attributes of effective fitness applications

Although current applications provide a wide range of functionalities [4], they often lack deeper personalization of training and nutrition plans based on the user's individual needs. Most of them operate on the principle of static recommendations that change only minimally based on manual data entry. This gap in personalization represents a significant opportunity for AI-powered solutions to differentiate themselves in the market [13].

Research into the attributes of high-quality fitness coaching reveals several essential capabilities that digital solutions should incorporate to effectively replace human trainers:

- **Personalized exercise programs:** Just as personal trainers design customized exercise programs based on an individual's fitness level, goals, physical limitations, and preferences, effective fitness applications must offer training plans that adapt to user data such as fitness level, progress, and health status.
- **Goal setting and progress tracking:** Personal trainers collaborate with clients to establish achievable fitness goals and regularly monitor their progress. Digital solutions should replicate this by enabling users to define fitness goals and automatically track their progress, offering real-time insights and suggestions based on performance data such as weight increases or completed repetitions.
- **Nutritional guidance:** While not always mandatory, personal trainers provide basic dietary advice to complement fitness goals, helping clients balance macronutrients and manage caloric intake. AI-powered applications can potentially exceed human capabilities by analyzing users' food intake and automatically comparing it with tracked attributes like weight to provide highly personalized nutritional recommendations.
- **Motivation and support:** A crucial role of personal trainers is keeping clients motivated through encouragement and counseling. Applications can emulate this by sending motivational prompts, progress updates, and reminders, creating a supportive virtual environment that encourages users.
- **Safe and effective exercise:** Personal trainers ensure clients perform exercises with proper form to avoid injuries and maximize effectiveness. Similarly, applications must provide detailed exercise instructions, video demonstrations, and real-time feedback to ensure users perform movements safely, with the ability to adjust exercises based on progress or limitations.
- **Adaptability:** Personal trainers must adapt workouts to various fitness levels, injuries, or lifestyle changes. Fitness applications need equal adaptability, modifying training plans based on user feedback, progress, or physical challenges.

- **Accountability:** Personal trainers foster accountability by providing structured training plans and regular monitoring. Likewise, applications should offer structured plans, send reminders, and track missed workouts.
- **Health and fitness knowledge:** Personal trainers maintain current knowledge of fitness trends, techniques, and research to provide scientifically-based advice. Applications should continuously update their exercise protocols and nutritional recommendations based on current research to offer users the most relevant and effective guidance.

C. Mobile application development approaches for fitness solutions

The development of fitness applications can follow several methodological approaches, each with distinct advantages and limitations. According to industry analyses [14], three primary approaches dominate the mobile application development landscape:

Native applications are developed and optimized for specific platforms using tools provided by platform manufacturers. For Android, this typically involves Java within Android Studio, while iOS development commonly uses Swift within XCode. This approach allows developers to effectively leverage the features and functionalities of the operating system and device, achieving better optimization and user satisfaction. Native apps offer high performance through direct access to hardware functions, consistent user experience aligned with platform design standards, direct access to native APIs, simple distribution via app stores, and robust debugging tools. However, their primary drawbacks include platform dependency, requiring separate applications for each platform, and more complex development processes requiring specialized teams for each platform.

Web applications, unlike native applications, run through web browsers rather than directly on the device operating system. Services like Google Docs or MS Office 365 exemplify this approach, allowing users to access applications directly through browsers without installation from app stores. Web applications offer advantages such as easy installation (requiring no device storage), minimal permission requirements, and cross-platform portability. However, they depend heavily on internet connectivity and browser functionality, have limited access to device hardware and APIs, and offer restricted user interface customization options.

Hybrid and progressive web applications combine elements of both approaches. Progressive web applications present themselves as standard websites but possess native mobile application characteristics, including offline functionality and access to native device APIs. These applications typically install directly from websites rather than app stores and update centrally on web servers. Hybrid applications, another subcategory of web applications that mimic native applications, are particularly suitable for developing simple, cross-platform compatible solutions. They offer multi-platform compatibility, lower development costs, access to native APIs, and offline mode capabilities, though they generally provide lower performance than native applications and may not support the complete suite of native API communications.

For applications prioritizing performance, such as mobile games, native development is recommended, while simpler applications may benefit from web architecture. Hybrid applications, however, offer a reasonable compromise that often outweighs other benefits, making this approach particularly suitable for fitness solutions targeting multiple platforms.

D. Integration of artificial intelligence in fitness applications

In recent years, there has been a growing trend of integrating artificial intelligence into fitness apps. Research by Li et al. [15] demonstrates that AI can analyze large amounts of data and generate dynamic recommendations, allowing the creation of highly personalized training plans. This finding is supported by industry analyses showing increased user engagement when AI elements are incorporated into fitness applications [16].

Modern AI integration in fitness applications follows several approaches:

- **On-device AI:** Processing occurs locally on the user's device, enhancing response speed and privacy protection. This approach is commonly used for motion analysis or voice recognition but faces limitations in computational power and device memory.
- **Cloud-based AI:** Data is sent to remote servers for computationally intensive processing. This model enables the use of advanced neural networks and performance scaling but requires stable internet connectivity and robust data security.
- **Hybrid approaches:** Combining advantages of both previous methods – basic tasks are processed locally while more complex analyses occur in the cloud. This model optimizes performance while minimizing the need for sensitive data transfer.
- **Personalized AI models:** Enabling algorithm adaptation to individual user needs through continuous learning based on their interactions and preferences. This approach is frequently used in recommendation systems or generating customized training plans.

Technologies such as large language models (LLMs) [1] enable interactive communication with users, providing feedback and continuously adapting plans based on progress and preferences, creating an experience that more closely resembles having a personal trainer [17]. LLMs have demonstrated particular effectiveness in analyzing users' textual data and generating personalized recommendations for achieving fitness goals, as well as providing context-aware suggestions that consider factors like environmental conditions or unusual caloric intake.

The Smart Fitness Coach application differentiates itself from traditional solutions by leveraging artificial intelligence on multiple levels. In addition to automatically generating training plans and calculating calorie intake, it also offers adaptive recommendations based on real-world data, interactive coaching, and progress analysis. AI also enables prediction of future results and provides personalized motivation, increasing engagement and long-term sustainability of users' fitness goals.

With the growing trend of using AI in the fitness sector, it can be expected that smart training apps will play an even more important role [18] in the field of health and sports in the future. Smart Fitness Coach aims to be part of this evolution by combining modern technological approaches with scientifically based knowledge, thus bringing an innovative solution to a wide group of users.

III. Application Design and Development

The development of the Smart Fitness Trainer application was carried out with an emphasis on personalization, intuitiveness and effective use of artificial intelligence. The main goal was to create a platform that would help users achieve their fitness goals through automated training planning, nutritional recommendations and interactive feedback.

A. Platform and technology stack selection

After analyzing the possibilities of developing mobile applications, it was decided to implement the application using React Native. This technology allows development for both Android and iOS, thereby reducing the time required for development. At the same time, it offers sufficient performance and a wide range of available libraries for integrating AI and backend services.

The backend solution was built on Firebase 1 in combination with Node.js, which allows secure data management, user authentication and fast data synchronization between devices. Artificial intelligence was integrated via API calls to LLM models (e.g. OpenAI GPT) and other specialized data processing services.

B. Application Architecture

The application was designed as a modular system with several main components:

- **Authentication and user account management** Data security and personalized recommendations.
- **Training plan generator** Automatic creation of plans based on user goals and preferences.
- **Nutrition calculation module** Calculating calorie intake and macronutrient distribution.
- **AI trainer and chatbot** Interactive advice and feedback through artificial intelligence.
- **Progress tracking** Historical data analysis and progress visualization.

Each of these parts was designed to allow for scalability and future expansion.

C. Integration of Artificial Intelligence

A key element of the application is the use of AI for adaptive recommendations. The AI model processes the user-entered data (e.g. weight, exercise intensity, eating habits) and generates dynamic plans. In addition, the AI analyzes trends in progress and provides motivation and feedback in real time. We used the OpenAI GPT-4 model for its cost-effectiveness and high accuracy.

D. User interface and UX design

The application was designed with an emphasis on simplicity and user comfort. The main parts of the application are intuitive and divided into logical sections, while attention was also paid to accessibility for a wide group of users.

IV. Key functionalities

The Smart Fitness Trainer application offers a wide range of functionalities designed to help users effectively achieve their fitness goals. Key features include personalized training plans, nutritional recommendations, progress tracking, and an interactive AI assistant.

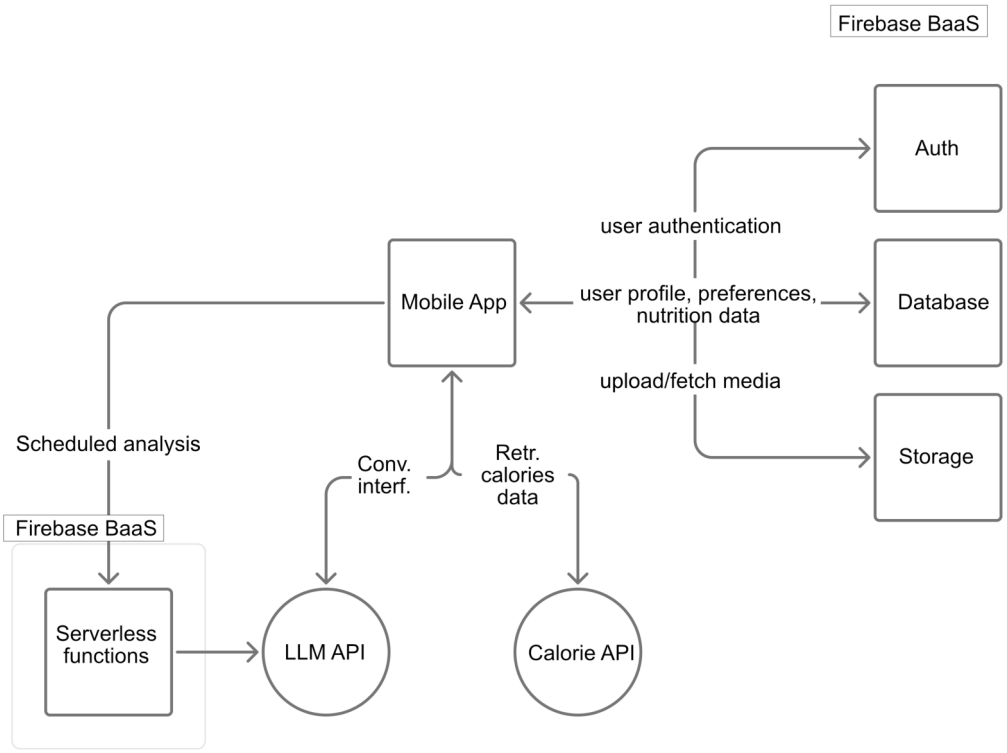


Fig. 1 App's architecture showing the relationships between key components and their interactions.

A. *Personalized Training Plans*

The application generates training plans based on the user's individual goals, fitness level, and available equipment. The main features of this functionality include:

- Automatic creation of training programs for various fitness goals (strength, endurance, weight loss, muscle gain). 2
- Adjustment of training intensity and volume based on the user's progress and preferences.
- Video instructions and detailed description of exercises for each workout.
- Possibility to edit and customize individual workouts.

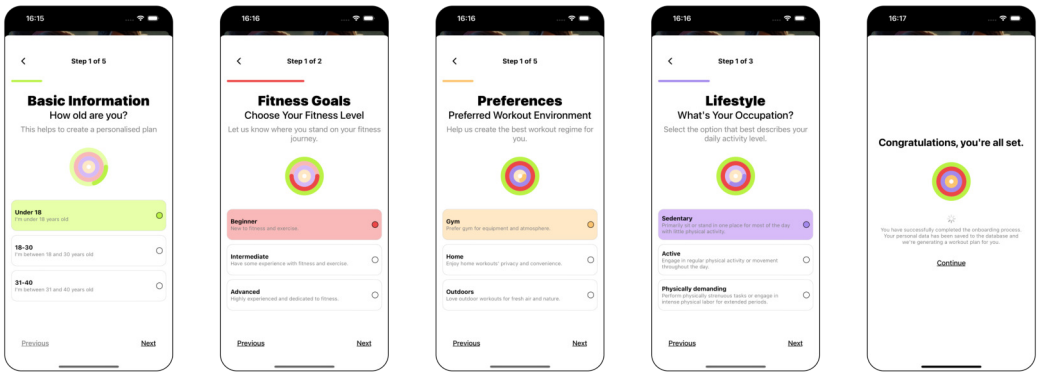


Fig. 2 App's onboarding process for user's preferences.

B. Nutritional recommendations and diet tracking

Diet plays a key role in achieving fitness goals, which is why the app provides AI-generated recommendations based on user preferences and calorie expenditure:

- Calculation of recommended calorie intake and macronutrient distribution 3.
- Generation of personalized meals based on dietary preferences (e.g. vegan, keto, paleo diet).
- Possibility of synchronization with fitness bracelets and other calorie tracking applications.

C. AI trainer and chatbot

Artificial intelligence serves as a virtual personal trainer, helping users with training, motivation and providing feedback:

- Instant answers to questions about exercise technique and nutrition 3.
- Customization of recommendations based on recorded progress.
- Motivation and psychological support to maintain discipline.

D. Progress tracking and analytics

The application allows detailed monitoring of the user's performance and progress using advanced analytical tools:

- Graphical display of training progress (strength increase, endurance, weight loss) 2.
- Comparison of historical data and suggestions for optimizing the training process.
- Ability to record body measurements (BMI, body fat percentage, body circumference).
- Integration with wearable devices (e.g. Apple Watch, Fitbit, Garmin).

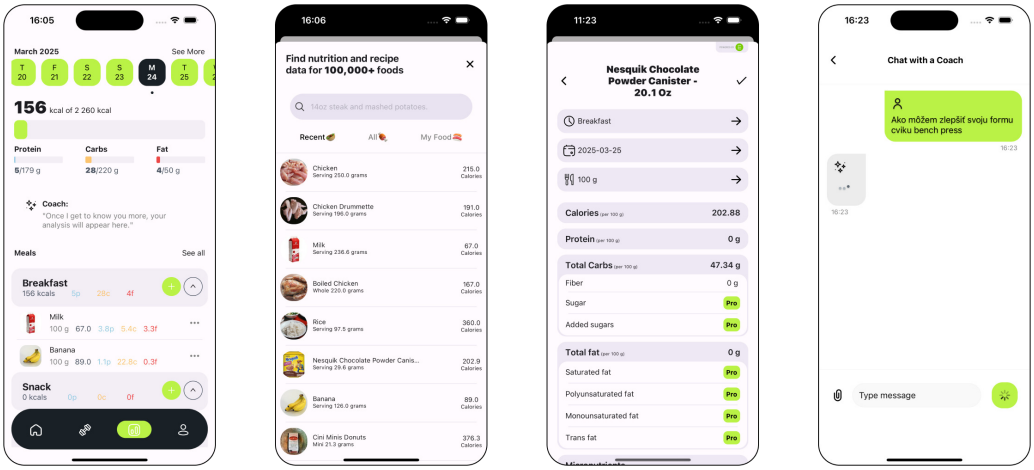


Fig. 3 In app calories tracker.

E. Integration with external services

To make the application as compatible as possible with existing solutions, it is planned to implement support for synchronization with multiple platforms:

- Connection with applications such as Apple Health, Google Fit, MyFitnessPal.
- Ability to export data to Excel or PDF for more detailed analysis.
- API for connecting to other fitness services and smart devices.

V. EVALUATION AND FUTURE DEVELOPMENT

The application evaluation was conducted using a combination of qualitative and quantitative methods to comprehensively assess various aspects of the Smart Fitness Trainer, while also identifying opportunities for future enhancements and expansion.

A. Methodology and Performance Analysis

We developed a multi-phase testing approach focusing on application functionality, AI recommendation quality, user experience, and real-world usage. Each area was analyzed using data collected from either real or simulated user interactions, with emphasis on obtaining objective results.

Performance testing evaluated both front-end responsiveness and backend efficiency under various conditions. Average application loading time was 2.8 seconds on high-end devices and 3.2 seconds on low-end devices, with API response times of 750 ms for LLM model queries and 450 ms for database operations. The application required an average of 150 MB RAM during normal use, with 90% of tested actions occurring without visible stuttering.

Scalability testing showed Firebase Realtime Database's excellent capabilities, supporting up to 200,000 simultaneous connections per instance. Our OpenAI API usage remained within the established limits of 30 requests per minute (RPM) and 100 tokens per minute (TPM). Primary challenges identified were optimization for older devices and potential API response slowdowns with higher numbers of simultaneous users.

B. AI Recommendations and User Experience

Certified fitness trainers Samuel Duleba and Peter Demeter evaluated the AI-generated exercise plans and nutritional guidance using criteria that assessed appropriateness, safety, and personalization. In case study testing, plans generally received high scores, particularly for appropriate intensity scaling based on user feedback, with AI recommendations achieving approximately 87% alignment with expert recommendations in controlled scenarios.

User experience testing with 12 respondents of varying fitness experience levels revealed that approximately 80% of users found the application navigation intuitive and easy to use, though some expressed confusion when looking for specific sections. 75% rated their experience with the AI trainer positively, particularly appreciating personalized recommendations, while 90% gave the highest score to the personalization quality of exercise and nutrition recommendations.

The application demonstrated robust performance in real-world conditions, with 93% of scheduled workouts successfully completed through the app and users reporting high satisfaction with the AI trainer's ability to adapt plans based on real-world constraints. Many specifically cited the "human-like understanding" of the AI as a key differentiator from other fitness applications they had used.

C. Limitations and Future Research Opportunities

While the application successfully meets most of its design requirements, several limitations were identified. The primary technical constraint is dependency on external API services, specifically OpenAI's API, which provides high-quality responses but could become financially demanding with a large user base. Implementation challenges included personalization algorithm optimization, balancing response accuracy with performance, and ensuring consistency in nutritional recommendations.

These limitations present opportunities for future research and development. The application has potential to become a valuable platform for scientific research in fitness and nutrition through:

- Introduction of structured research studies with voluntary user participation, allowing researchers to define experimental protocols and monitor results in real time
- Development of a continuously learning AI system based on the latest scientific knowledge and real user experiences
- Integration with physiotherapy methods and specialized programs for different age groups or specific health conditions
- Implementation of exercise form correction through live video analysis, the only major feature not yet included in the current version
- Expansion of the platform to allow generation of anonymized datasets that could be shared in academic and scientific circles
- Collaboration with experts in nutrition and sports medicine to create certified recommendations

By addressing these opportunities, the application could evolve beyond a personal fitness tool to become an important contributor to the broader health and fitness research community. The combination of artificial intelligence with fitness technologies has the potential to not only help individuals achieve their goals but also advance scientific understanding of personalized health interventions.

VI. Conclusion

The Intelligent Fitness Trainer project represents an innovative solution that combines modern artificial intelligence technologies with a personalized approach to fitness and healthy nutrition. The integration of large language models allows users to receive interactive and adaptive recommendations, thereby increasing the effectiveness of training and overall motivation for a healthy lifestyle.

The main contribution of this work is the creation of a mobile application that provides comprehensive support in the field of training, nutrition and advanced data analysis. Testing of the application has shown that the use of AI can significantly improve the personalization of recommendations and help users achieve their fitness goals more effectively.

Many possibilities open up in the future to expand the functionality of the application. One of the main areas is its use as a research platform, where AI can constantly improve its capabilities based on real data and scientific studies. In addition, the application can be enriched with new methods of progress tracking, smarter recommendations and even closer integration with wearable devices and medical databases.

The development of intelligent fitness trainers is a dynamic field with great potential. The combination of artificial intelligence and fitness technologies can in the future help not only individuals achieve their goals, but also the entire scientific community discover new approaches to health and fitness.

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. 004 ŽU-4/2024.

References

[1] L. Ajay, V. G. Biradar, M. Chandu, and B. Jb, "Ai tool as a fitness trainer using human pose estimation," 2023.
[2] W. R. Thompson, "Worldwide survey of fitness trends for 2021," *ACSM's Health & Fitness Journal*, vol. 25, no. 1, pp. 10–19, 2021.
[3] K. Ferraro, K. J. Grimm, and T. Kautz, "Estimating the technology of cognitive and noncognitive skill formation," *Quantitative Economics*, vol. 11, no. 2, pp. 579–614, 2020.
[4] J. Parashar, A. Jain, and S. Ali, "Artificial intelligence impact on human fitness: Exploring emerging trends," 2023.
[5] D. Gupta, "Fitness app development – step-by-step process, features, costs & more," March 2024. [Online]. Available: https://appinventiv.com/blog/fitnes-app-development/
[6] M. Adamakis, "Nike+ training club, an ultimate personal trainer: mobile app user guide," *British Journal of Sports Medicine*, vol. Online first, no. 13, 2017. [Online]. Available: https://www.researchgate.net/publication/319987781_Nike_Training_Club_an_ultimate_personal_trainer_mobile_app_user_guide
[7] E. Team. (2024) Ai in fitness coaching: A deep dive into real-world use cases. Accessed: 2025-04-16. [Online]. Available: https://blog.everfit.io/ai-in-fitness-coaching-a-deep-dive-into-real-world-use-cases
[8] M. S. Tredrea, V. J. Dalbo, and A. T. Scanlan, "Lifesum: easy and effective dietary and activity monitoring," *British Journal of Sports Medicine*, vol. 51, no. 13, pp. 1042–1043, 2017. [Online]. Available: https://bjsm.bmj.com/content/51/13/1042
[9] V. Giardina, "We reviewed headspace – here's why it's the best meditation app," 2025. [Online]. Available: https://nypost.com/article/headspace-meditation-app-review/
[10] F. Drum, "Truecoach review," 2025, accessed: Feb. 11, 2025. [Online]. Available: https://fitnessdrum.com/truecoach-review/
[11] P. I. Bouchrika, "Mobile vs desktop usage statistics for 2024," April 2024. [Online]. Available: https://www.statista.com/statistics/1111111/mobile-vs-desktop-usage-statistics/
[12] I. Taiwo, "24 + key mobile vs desktop usage statistics in 2024," March 2024. [Online]. Available: https://www.statista.com/statistics/1111111/mobile-vs-desktop-usage-statistics/
[13] R. Kraus. (2023) Artificial intelligence in personalized fitness gets smarter, for real. Accessed: 2025-04-16. [Online]. Available: https://www.wellandgood.com/ai-personalized-fitness/
[14] M. Binas, "Úvod do vývoja aplikácií pre chytré zariadenia," 2022. [Online]. Available: https://kurzy.kpi.fei.tuke.sk/iot1/2023/lecture.04.html
[15] M. Li, Y. Sun, and R. Zaid, "A health-focused system that blends machine learning and flutter to assist users in accomplishing their diet/fitness goals along with providing helpful advice," *Journal of Strength and Conditioning Research*, 2023.
[16] M. Betts, "Embracing the future: How ai is revolutionising fitness," December 2023. [Online]. Available: https://train.fitnes/personal-trainer-blogs/how-ai-is-revolutionising-fitnes
[17] TechRadar Staff, "I use chatgpt to hit my fitness and exercise goals – here are 7 prompts to help you get in shape using ai," 2025, accessed: 2025-04-16. [Online]. Available: https://www.techradar.com/health-fitness/i-use-chatgpt-to-hit-my-fitness-and-exercise-goals-here-are-7-prompts-to-help-you-get-in-shape-using-ai
[18] R. M. Nazarenko and Daniel, "Fitness application development guide: Best practices and case studies," 2024. [Online]. Available: https://mobidev.biz/blog/fitnes-application-development-guide-best-practices-and-case-studies

Web Platform for Personal Finance Management

[1*]*Nazarii Herasymchuk,* [2*]*Ondrej Kainz,* [3*]*Miroslav Murin,*
[4*]*Ivana Novakova,*

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

[1]nazarii.herasymchuk@student.tuke.sk, [2]ondrej.kainz@tuke.sk,
[3]miroslav.murin@tuke.sk, [4]ivana.novakova@tuke.sk

Abstract **– This paper presents the design and implementation of a web-based platform for personal finance management, integrating modern approaches in data security and financial trend prediction. It focuses on the analysis of current solutions in the field of data protection, with an emphasis on safeguarding sensitive user information. Furthermore, the study explores the application of machine learning techniques in forecasting personal financial developments. A comparative analysis of existing personal finance planning tools is also conducted. The proposed platform provides users with real-time insights into the performance of funds and commodities, supporting more effective financial analysis and decision-making.**

Keywords **– Financial planning, personal budget, personal finance, web application**

I. Introduction

In today's dynamic world, the topic of investing is gaining increasing attention, as individuals seek to supplement their primary income through financial markets and equity ownership in specific companies. This investment approach not only enables personal capital growth but also supports corporate expansion and value appreciation. Investors benefit from the price differential between the initial and current value of their shares. While investment processes are more complex than they may initially appear, they offer considerable long-term potential for capital appreciation.

Contemporary investing no longer relies on physical securities or direct human interaction, as was common decades ago. Instead, digital investment platforms have transformed the landscape by offering online access to brokers and automated systems for buying and selling financial instruments. These platforms handle vast volumes of data, enabling fast and efficient transaction processing in near real-time. In line with these developments, current research explores how user behavior history can enhance security mechanisms in online financial applications, including investment platforms [1].

Despite this accessibility, many individuals still lack a clear understanding of how investing works, how to get started, where to allocate funds, for how long, and how much to invest. To address this gap, the system proposed in this work aims to assist users in managing their income and expenses, monitoring the prices of specific companies or funds, and visualizing potential market trends. The prediction features of this system, inspired by methodologies such as random forest regression, contribute to better foresight into price movement and investment outcomes [2]. This information will enable users to make informed decisions and develop personalized investment strategies.

Beyond its core financial management functionality, the system offers additional features such as transaction history tracking, price prediction, and investment history tracking. Although initially intended for individual users, the platform is scalable and can serve broader user groups, such as families, teams, or small organizations pursuing collective financial goals.

Furthermore, the platform holds significant educational potential. By incorporating essential economic concepts and investment terminology, it can act as a training tool for those new to investing. This dual purpose—practical application and education—positions the system as a valuable contribution to both personal finance management and financial literacy development.

II. System architecture analysis

The analytical component of this work focuses on key areas essential for system development: data security, finance forecasting, and a comparative review of existing financial planning tools and graph visualization libraries. The study begins by examining current methods for password storage and user authentication, highlighting prevailing trends and best practices in cybersecurity. It then explores the potential integration of AI, assessing its advantages and limitations within the context of personal finance management. Additionally, the functionality, flexibility, and integration capabilities of available financial data APIs are evaluated. A comparative analysis of existing investment planning systems is also conducted to inform the selection of relevant features. The goal is to establish a comprehensive overview of current technologies and trends that can enhance system design, data security, and user experience. The following sections describe in more detail the methods used in the work.

A. Data security

Security of user data is described in a few methods and the main of them are password security storing [3], security authorization principles, and authentication methods, references [4] and [5]. Every security method was described and used in the work. Storing passwords securely in the work was done by hashing with algorithms bcrypt. It is preferred over encryption or SHA following the next article [6]. Bcrypt offers strong protection and good performance for web use, reference [7], while Argon2 is more secure but resource-intensive following reference [8]. Plain-text storage is unsafe and must be avoided. Authorization is made for protecting users' sensitive financial data and ensuring privacy. Strong authentication methods—such as passwords, two-factor authentication, and biometrics—have been analysed to use to prevent unauthorized access. Finally, a standard authentication method with password and email was used. Plus, email verification was added to the process. Additionally, in the work, an analysis of JWT (JSON Web Token) as a user authentication method was conducted, allowing secure and stateless verification of user identity through signed tokens, reference [1]. Other methods were also considered. Traditional session-based authentication and integration with OAuth 2.0 and OpenID Connect for third-party login, as referenced in [9]. The simpler API key authentication was also reviewed but deemed less secure and not suitable for user-level access control.

B. Finance forecasting

As prediction of future stock market price, an analysis of simple models such as linear regression, random forest, and decision tree was conducted, reference [2]. Linear regression was used in the work as a simple predictive model to analyse price trends, references [10] and [11], specifically to identify upward or downward movement. Despite its simplicity, it provides a useful overview of the relationships between time periods. Due to its ease of implementation, it was applied as an initial prediction tool before moving on to more complex models such as decision trees.

C. Planning method and graph visualization

As a planning method for the system, the popular envelope budgeting method was chosen, reference [12], which is also used in similar systems like YNAB, reference [13], for example. The envelope budgeting system is a method that uses different spending categories to allocate cash into separate envelopes. Each envelope represents a budget category (e.g., groceries, rent, entertainment), and the method's purpose is to ensure that money is spent only from the designated envelope. Once it's empty, you can't spend more in that category unless you move money from another envelope. This system promotes disciplined, intentional spending and helps avoid overspending.

The envelope method remains very useful for this kind of system and is used in applications like YNAB and GoodBudget, reference [14], which were compared in this work. The comparison of these applications was based on several factors, covering a wide range of aspects from design interface to the available functionality for users. In the end, YNAB was chosen as the better system, offering more possibilities for tracking personal finances.

A graph visualization analysis was conducted to explore different approaches for representing graphs in a web system, comparing two main libraries: Chart.js and ApexCharts. The analysis showed that ApexCharts offers better interactivity, real-time data support, and extensive customization options, reference [15]. Chart.js, on the other hand, stands out for its simplicity and ease of use, particularly for beginners or smaller projects. It provides clean and visually appealing charts with minimal setup, while still offering enough flexibility for advanced configurations. As an open-source tool with an active community, it is a reliable and well-maintained option that can be used in both basic and more complex web applications.

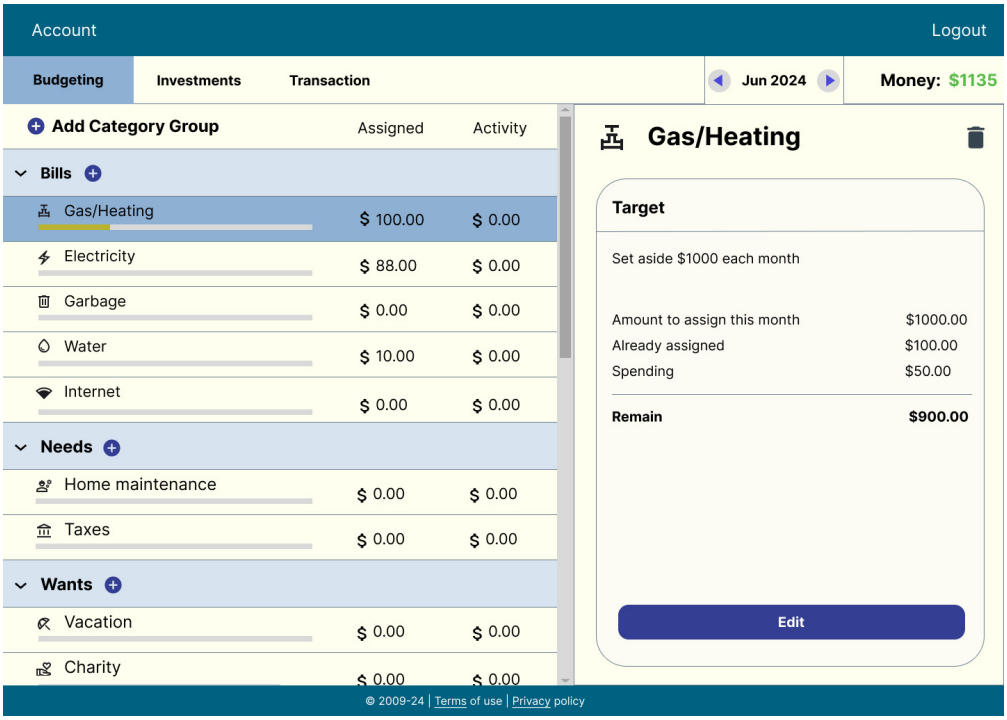


Fig. 1: Budgeting page prototype

III. Solution Design

In the design for this work, the goal was to create a visual design mockup of how the system will look to the end user.

The design includes login, registration, a budgeting interface based on the envelope method, investment planning, transaction history, prediction tools, and a user-friendly tutorial section.

A prototype was created using Figma, and a flowchart illustrates the navigation between key pages. The prototype is designed to support multiple languages (English, Slovak, Ukrainian), with English as the default. Language translations are handled via JSON files for scalability.

The application aims to simplify financial management compared to existing solutions like YNAB [13] by offering a cleaner, more intuitive interface. It reduces complexity by using only two budget columns (allocated and used funds), eliminating the often confusing third column (next month's balance). This simplified design is based on user reviews, which showed a preference for clarity and ease of use.

The prototype consists of several pages, one of them is shown in Figure 1, including the sign-up page, login page, and intro page. The prototype also includes a planning page where the user can manage their finances using the envelope budgeting method. Additionally, it features an investment page for tracking current market data, which is displayed as a graph.

Ultimately, the system prototype tries to provide essential budgeting and investment tools without overwhelming users, focusing on usability and accessibility for beginners.

Project design solution also described possible technologies that will be used further in work, for client and server side. For client side, technologies like Angular as the main framework for implementation were evaluated. The back end was evaluated using Node.js with Express.js for building REST APIs. Node.js allowed asynchronous request handling, which ensured fast and efficient server performance, and it provided smooth integration with frontend frameworks. Spring Boot was also evaluated as an alternative—especially for large-scale enterprise applications—Node.js was selected for its flexibility and suitability for rapid development.

IV. Project Implementation

This section presents all the technologies that were used in the project, including both the server-side and client-side components. The aim is to provide a clear overview of the tools, frameworks, and libraries that were selected to support the development of the system. Each technology is briefly described, along with the reasons for its inclusion and how it contributes to the overall functionality and performance of the application.

Also in the section described the graphical interface and the database that were used in the project.

A. Client side

The client side of the personal finance management system was developed separately from the server side logic to ensure better scalability and easier maintenance. It was built using Angular, Bootstrap, and HTML/CSS. HTML/SCSS was used to provide the structure and styling of the interface, making it possible to build a graphical user interface. Angular was used for creating dynamic, modular applications with reusable components, which also simplify and speed up the development process. Bootstrap simplifies UI development with ready-to-use components and helps maintain a consistent, responsive design across devices.

For data visualization in graphs, ApexCharts and Chart.js were used. While Chart.js is simpler, ApexCharts also offers more interactivity and customization, making it more suitable for real-time financial data. Both libraries were used in the system on different pages.

The system was also used "Twelve Data API" to access real-time and historical financial market data, allowing users to track prices of stocks, currencies, commodities, and indices.

B. Server side

The server side of the project used Node.js as the main framework for building server-side logic and handling communication with the database. In combination with Node.js, Express.js was also used, as it enables fast development of RESTful APIs with efficient, asynchronous request handling ideal for lightweight applications. It integrates well with frontend frameworks like Angular and provides easy database connectivity (e.g., PostgreSQL), making it highly effective.

Since the server side and client side are implemented as two separate parts of the system and run on different ports, CORS (Cross-Origin Resource Sharing) was used on the server side. CORS is an HTTP header-based mechanism that allows a server to specify which origins (domain, scheme, or port) are permitted to access its resources, enabling secure communication between the frontend and backend.

C. Database

The system used a PostgreSQL database to store user data. The main table, user, holds essential information such as name, password, and email. Additional tables store user activities related to saving and investing.

Key tables include:

- Preferences – stores user settings like theme and language.
- Saving Plan – a table to record details of the savings category. Includes basic data such as category name, description, amount of money allocated, target, dates and user activity (income and expenditure).
- Sections – this table is designed to facilitate the grouping of data in the savings plan table and contains only basic information such as the name of the section and a reference to the user to whom it belongs.
- Investment Plan – similar to saving plans but includes broker info, price, and quantity, and links directly to the user.
- Transactions – logs income and expenses, linked to users and saving plans. Each entry includes either an income or an expense field, but never both.

All tables are connected via foreign keys referencing the user ID. The structure is shown in the relational database model.

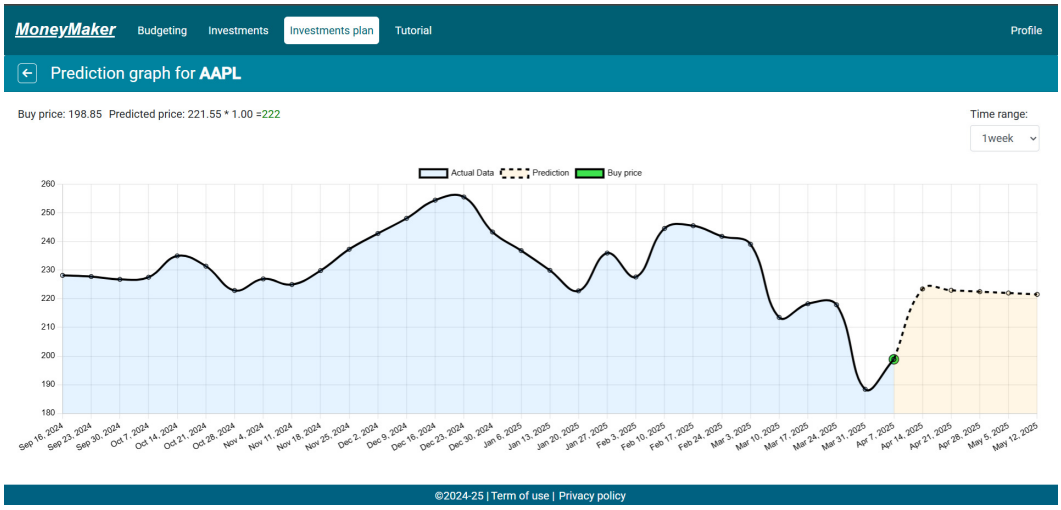
D. Graphical interface

The application's user interface was based on a Figma design created during the draft phase. The system includes eleven distinct pages, each serving a specific function. It also features several pop-up windows to display smaller information or gather user input without navigating away.

Upon entering the app, users land on the homepage, where they can register or log in. The system remembers the user's last visited page and login status, eliminating the need to log in repeatedly. The design and functionality are inspired by popular saving and investment platforms, with the budgeting page resembling YNAB and the investment and prediction pages featuring stock-style charts.

Each page (except login and registration) includes a consistent header for navigation and a footer for additional information. The color scheme uses a light cream or white background, teal for the header and footer, dark blue, and green for buttons to visually distinguish them.

Charts are present on the budgeting, investment, and prediction pages 2, implemented using Chart.js and ApexCharts. While both offer similar visual outputs, Chart.js proved to be simpler to work with.



(a) Prediction page

Fig. 2: The final design prediction page of the application.

V. Evaluation

The result of the project implementation is a functional prototype of the future application, including all planned features. However, thorough testing is needed to address identified issues. To improve the system, testing is conducted with students and regular users without prior programming experience.

The testing process is divided into three parts:

Test Design – Describes the chosen testing methods and scenarios.

Test Execution – Details the conditions, tools, and participants involved in testing.

Evaluation – Summarizes results using charts and user feedback, highlighting suggestions for improvement.

The goal is to identify areas for enhancement and propose solutions to improve usability, functionality, and overall user experience.

A. Test Design

In the testing scenario, participants were instructed to access the system via a web browser and complete the following tasks:

- Register within the application and verify their email address.
- Log in using the newly created credentials.
- Enter an initial financial amount (a value exceeding 1,000 was recommended).
- Review the provided tutorial to become familiar with the system.
- Navigate to the ”Budgeting” section, create a budget category and item, and define a target amount and deadline.
- From the ”Budgeting” interface, proceed to the ”Transactions” section and assign a monetary amount to the previously created budget item.
- Access the ”Investments” module and create an investment plan for a selected ETF or stock.
- Navigate to the ”Investment Plan” section to confirm the creation of the investment plan.
- Independently explore the system, define personal financial goals within the application, and attempt to meet them by allocating funds to appropriate categories.
- Complete the CSUQ (Computer System Usability Questionnaire), which assesses user satisfaction with system usability, quality of information, and interface design. Responses were rated on a 7-point scale ranging from ”Strongly agree” (7) to ”Strongly disagree” (1). An additional section was provided for reporting issues or submitting comments.

This scenario was designed to evaluate the core functionalities, overall usability, and user-friendliness of the system.

B. Test Execution

The testing process was conducted through both remote and in-person sessions. Participants involved in remote testing received a detailed set of instructions along with access to the system and a link to the evaluation questionnaire. In-person testing was carried out under supervision to

ensure a controlled environment, while avoiding direct intervention in order to replicate typical user behavior.

A total of 18 individuals participated in the testing, including 12 students from the Technical University of Košice (TUKE). The remaining participants represented a heterogeneous group with no prior experience in information technology or similar digital systems, providing a broader perspective on the application's usability.

Remote testing utilized platforms that supported screen sharing and video conferencing, enabling observation of user interactions. In-person testing took place in the TUKE library and other appropriate environments, with each session conducted individually to maintain focus and minimize external influences. All sessions were carried out using laptop or desktop computers to ensure consistency in the testing conditions.

The application is designed for a general audience; therefore, no specific target demographic was defined. Basic demographic information was collected to provide contextual insights but was not considered a primary evaluation criterion. The overarching objective of the testing was to evaluate the usability and core functionality of the system, gather user feedback, and identify opportunities for improving the overall user experience across diverse user groups.

C. Evaluation

The questionnaire included 18 questions and was completed by all participants. The tested system version contained all planned features.

Users had varying experience levels with similar apps. Interestingly, those without prior experience, rated the system more positively (average 6–7 points) than more experienced users (average 5–6 points). Overall feedback was positive, though some usability issues were noted.

The main issue identified was the unclear process of adding transactions, particularly confusion about the interaction between the budgeting and transaction pages. This was due to the system's design, which separated budget data into a different page. To improve this, a feature was added allowing users to assign funds directly through the transaction page via double-clicking. Users also requested a clearer overview of expenses by category, which led to the implementation of a summary chart with spending data and statistics.

Testing revealed several usability issues and bugs, which were fixed after testing. These findings helped to improve the project in the interface and functionality, making it more user friendly.

VI. Conclusion

This paper presented the design and implementation of a personal finance management system aimed at supporting users in effective saving, budgeting, and financial planning. The system consists of two primary modules: saving and investing. It also incorporates advanced functionalities such as investment tracking and predictive financial analytics.

The development focused on usability, featuring a clean user interface and an integrated help section to facilitate user onboarding. The system design was inspired by existing solutions such as YNAB for budgeting and various investment platforms for financial monitoring. Key technical features include secure authentication using JSON Web Tokens, password hashing, and route protection implemented in Angular. Predictive modeling was carried out using linear regression, while financial data visualization is achieved through ApexCharts and Chart.js. The chosen technology stack comprises Angular for the frontend, Node.js for the backend, and PostgreSQL as the database, with design prototypes created in Figma.

User experience was evaluated using the CSUQ questionnaire, administered both in person and online. The feedback indicated a generally positive user reception. The application supports multi-language functionality and was initially developed in English, with a scalable architecture for future language additions. Users are able to record categorized transactions, monitor investments with real-time market data, and receive predictions on the future performance of stocks, ETFs, and commodities. Additionally, the system serves as a simulation tool for personal savings and financial planning.

A potential direction for future improvement is the integration of banking APIs to automate expense tracking and further enhance user experience.

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. 004 ŽU-4/2024.

References

[1] A. Bucko, K. Vishi, B. Krasniqi, and B. Rexha, "Enhancing jwt authentication and authorization in web applications based on user behavior history," *Computers*, vol. 12, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2073-431X/12/4/78

[2] P. F. Smith, S. Ganesh, and P. Liu, "A comparison of random forest regression and multiple linear regression for prediction in neuroscience," *Journal of Neuroscience Methods*, vol. 220, no. 1, pp. 85–91, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027013003026

[3] G. Hatzivasilis, "Password-hashing status," *Cryptography*, vol. 1, no. 2, 2017. [Online]. Available: https://www.mdpi.com/2410-387X/1/2/10

[4] S. Z. Syed Idrus, E. Cherrier, C. Rosenberger, and J.-J. Schwartzmann, "A Review on Authentication Methods," *Australian Journal of Basic and Applied Sciences*, vol. 7, no. 5, pp. 95–107, Mar. 2013. [Online]. Available: https://hal.science/hal-00912435

[5] A. Ometov, S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen, and Y. Koucheryavy, "Multi-factor authentication: A survey," *Cryptography*, vol. 2, no. 1, 2018. [Online]. Available: https://www.mdpi.com/2410-387X/2/1/1

[6] A. Sadeghi-Nasab and V. Rafe, "A comprehensive review of the security flaws of hashing algorithms," *Journal of Computer Virology and Hacking Techniques*, vol. 19, no. 1, pp. 287–302, 2023.

[7] P. Sriramya and R. Karthika, "Providing password security by salted password hashing using bcrypt algorithm," *ARPN journal of engineering and applied sciences*, vol. 10, no. 13, pp. 5551–5556, 2015.

[8] A. Biryukov, D. Dinu, and D. Khovratovich, "Argon2: New generation of memory-hard functions for password hashing and other applications," in *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2016, pp. 292–302.

[9] D. Fett, R. Küsters, and G. Schmitz, "The web sso standard openid connect: In-depth formal security analysis and security guidelines," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 2017, pp. 189–202.

[10] X. Su, X. Yan, and C.-L. Tsai, "Linear regression," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 3, pp. 275–294, 2012.

[11] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012.

[12] J. McCaffery, "Canada's envelope budgeting system," *American Review of Canadian Studies*, vol. 14, no. 1, pp. 45–62, 1984. [Online]. Available: https://doi.org/10.1080/02722018409481437

[13] "You Need A Budget," Website, 2024. [Online]. Available: https://www.ynab.com/

[14] Goodbudget, "Goodbudget: Budget tracker & planner," 2024, accessed: 2024-03-23. [Online]. Available: https://www.goodbudget.com

[15] "Apexcharts: Interactive javascript charting library," JavaScript library, accessed on March 8, 2024. [Online]. Available: https://apexcharts.com/

Kompresné snímanie EKG pomocou učenia slovníkov

[1]*Andrii KHAINAS,* [1]*Ondrej KOVAČ,* [2]*Antónia KOVAČOVÁ*

[1] Katedra Technológii v Elektronike, [2]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenska Republika

[1]andrii.khainas@student.tuke.sk, [1]ondrej.kovac@tuke.sk, [2]antonia.kovacova@tuke.sk

***Abstrakt* — V tomto článku sa skúma efektívna kompresia EKG signálov s využitím metódy kompresného snímania (KS) v kombinácii s učením slovníkov. KS predstavuje moderný prístup, ktorý umožňuje znížiť počet vzoriek potrebných na rekonštrukciu signálu bez výraznej straty kvality. V práci sa používajú deterministické binárne blokovo diagonálne (DBBD) snímacie matice a slovníky získané pomocou algoritmu MMV-OMP (Multiple Measurement Vector - Orthogonal Matching Pursuit), ktorý je založený na greedy princípe. Pre každý zvod EKG bol vytvorený vlastný slovník optimalizovaný na základe reálnych signálov. Výsledky ukazujú, že navrhnutý systém dosahuje akceptovateľnú chybu rekonštrukcie aj pri vyšších kompresných pomeroch. Článok je výsledkom riešenia semestrálneho projektu predmetu Číslicové spracovanie obrazov, ktorý sa vyučuje v študijnom programe Automobilová elektronika.**

***Kľúčové slova* — kompresne snímanie, biomedicínske signály, greedy algoritmy**

I. Úvod

Elektrokardiogram (EKG) patrí medzi najvýznamnejšie diagnostické nástroje na sledovanie zdravotného stavu myokardu. Monitorovanie srdcovej aktivity prostredníctvom 12-zvodového EKG poskytuje komplexné informácie z rôznych uhlov, čo umožňuje presnú diagnostiku širokého spektra srdcových ochorení, ako sú kardiomyopatia, arytmia, infarkt myokardu a ďalšie kardiovaskulárne poruchy [1]. V klasickom 12-zvodovom EKG sa využíva desať elektród, ktoré sú umiestnené na tele pacienta v presne definovaných pozíciách. Tieto elektródy sa rozdeľujú na tri unipolárne končatinové zvody (VL, aVR, aVF), tri bipolárne končatinové zvody (I, II, III) a šesť hrudníkových zvodov (V1–V6) [2], [3]. Analýza signálov zo všetkých zvodov umožňuje kardiológovi diagnostikovať rôzne ochorenia srdcového svalu. Signál získaný štandardným vzorkovaním má vysoký objem dát, čo je obzvlášť problematické pri dlhodobých a kontinuálnych meraniach, napríklad pomocou EKG holtera. Takýto objem dát kladie nároky na ukladanie, spracovanie aj prenos. V závislosti od typu ochorenia môže byť potrebná rôzna dĺžka monitorovania elektrickej aktivity srdca. Dlhodobé sledovanie je možné aj prostredníctvom nositeľných bezdrôtových systémov [4]. Jedným z možných riešení uvedených výziev je kompresné snímanie (KS), ktoré predstavuje moderný prístup k spracovaniu signálov.

II. Podstata kompresného snímania

Princíp KS je založený na predpoklade, že v určitej doméne môžu byť signály reprezentované výrazne nižším počtom nenulových koeficientov. Dané signály môžu byt reprezentované s vysokým stupňom presnosti a zároveň s malým počtom koeficientov, čo vedie k úspore pamäte a k zjednodušenému výpočtu. EKG signál je dobre aproximovateľný v DWT a DCT doménach [5] [6]. Komprimovateľný signál $f \in \mathbb{R}^N$ (signál, ktorý vykazuje riedkosť v určitej doméne) po spracovaní môže byt reprezentovaný ako vektor $x \in \mathbb{R}^N$, ktorý obsahuje iba K nenulových koeficientov. Takýto signál sa nazýva K - riedky signál [5]. Toto spracovanie je možné vďaka ortogonálnym bázam alebo špeciálnym slovníkom. Tieto bázy sú zapísané v maticovom tvare $\psi \in \mathbb{R}^{N \times N}$, potom komprimovatelny signal je možné zapísať ako (1).

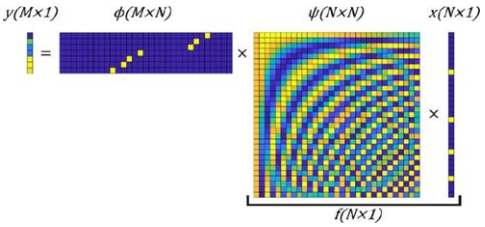
$$f = \psi x, \tag{1}$$

Platí, že $0 < K \ll N$ v inom prípade vektor x nepovažujeme za riedky a nie je vhodný na použitie. Tradičný prístup kompresných metód spočíva v tom, že sa najprv signál ovzorkuje podľa Shannonovo-Kotelnikovej teorémy a potom sa väčšina koeficientov v snímanom signálu vyradí a v komprimovanej verzii zostanú iba tie koeficienty, ktoré majú najvyššiu hodnotu. Namiesto toho je oveľa efektívnejšie získať komprimovanú verziu signálu priamo z podstatne menšieho počtu nameraných údajov. Inými slovami, signál sa nebude vzorkovať podľa Shannonovo-

Kotelnikovej teorémy a následne sa spracovať, ale bude sa snímať len také množstvo vzoriek, ktoré nesie esenciálnu časť informácie potrebnú pre obnovu celého signálu. KS je efektívnou technikou získavania údajov, v prípadoch kedy sú obmedzenia na príkon, dátove úložisko alebo šírku prenosového kanála. Na získavanie neadaptívnych meraní, ktoré sú reprezentovane náhodne nameranými údajmi v čase, sa používa meracia matica $\phi \in \mathbb{R}^{M \times N}$, ktorá určuje ktoré hodnoty budú zapísané a ktoré vynechané. Výsledkom daného merania je komprimovaný signál $y \in \mathbb{R}^{M}$. Na základe uvedeného je teda možné pre riedke signály [7] v kontexte KS uviesť vzťah (2):

$$y = \phi\psi x \tag{2}$$

Je zrejmé, že počet stĺpcov ϕ je oveľa väčší ako počet riadkov, teda existuje viac neznámych premenných ako počet rovníc čo znamená nekonečne množstvo možných riešení vzťahu (2). Vzťah (2) je možné zobraziť graficky tak ako je vidieť z Obr. 1.



Obr. 1 Vizualizácia princípu kompresného snímania

Snímaciu a transformačnú maticu je možné zapísať ako jednu kombinovanú maticu A, potom je možné (2) zapísať v tvare (3). Všetky x, ktoré spĺňajú (3) sa nazývajú prípustnými riešeniami, teda daná matica zaručuje možnosť rekonštrukcie pôvodného signálu [8].

$$y = Ax, \tag{3}$$

A. DBBD snímacia matica

Na získavanie komprimovanej verzie sa používajú snímacie matice, ktoré majú rozhodujúcu rolu pri redukcii rozmeru pôvodného signálu. Snímacia matica $\phi \in \mathbb{R}^{M \times N}$, kde $M < N$ vytvara subor komprimovanych merani, ktoré je možne zapísať v tvare (4). Aby snímacia matica bola použiteľná musí byt inkoherentna s bázovou maticou a spĺňať podmienku obmedzenej izometrie (Restricted Izometry Property). Inkoherencia dvoch matíc (v našom prípade snímacia ϕ a transformačná ψ) reprezentuje mieru nezávislosti [9][10] a je definovaná ako (5):

$$y = \phi f, \tag{4}$$

$$\mu(\phi, \psi) = \max_{i,j} \frac{|\phi_i, \psi_j|}{\|\phi_i\|_2 \|\psi_j\|_2}, \tag{5}$$

Existuje viacero druhov matíc, ktoré spĺňajú dane podmienky. Medzi nich patria náhodne Gaussovske, Bernouliho matice alebo deterministické také ako Hadamardove, Deterministicke Binarne Blokovo Diagonalne (DBBD) matice. V rámci tohto článku budeme uvažovať DBBD maticu s rôznymi kompresnými pomermi (CR). DBBD matica je tvorená binárnymi blokmi pozdĺž hlavnej diagonály (6). Obsahuje hodnoty 0 a 1, čo umožňuje jednoduchú implementáciu. Je definovaná deterministicky na rozdiel od náhodných matíc.

$$\Phi_{DBBD} = \begin{bmatrix} \overbrace{1 \cdots 1}^{CR} & 0 & 0 & 0 \cdots 0 \\ 0 \cdots 0 & \overbrace{1 \cdots 1}^{CR} & 0 & 0 \cdots 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 \cdots 0 & 0 & 0 & \overbrace{1 \cdots 1}^{CR} \end{bmatrix} \tag{6}$$

Každý jeden blok počtu jednotiek v riadku má pevnú dĺžku $m = {}^{N}/_{M} = CR,\ m \in \mathbb{N}$. Inými slovami, snímacia matica Φ_{DBBD} je navrhnutá takým spôsobom, aby každý riadok mal m-ty pocet jednotiek a každý jeden ďalší riadok bude mať blok s jednotkami posunutý presne o m vzoriek [11]. Kompresný pomer určuje, koľkokrát menej vzoriek je nasnímaných v porovnaní s originálnym signálom [12].

III. Učenie slovníkov

KS zvyčajne predpokladá, že signál f je komprimovatelny v určitej doméne pomocou správne zvolenej ortogonálnej bazy ψ. Dane bazy môžu reprezentovať signál s veľkou presnosťou, ale

nemusia byt optimálne pre každý typ signálu. Z hľadiska úspory dát a zvýšenia efektivity výpočtov výhodnejšie používať špeciálne zostavené bázy, tieto sa nazývajú slovníky [13]. Vzťah (1) je potom možne prepísať nasledovne:

$$f = Dx, \tag{7}$$

kde D je matica $\mathbb{R}^{N\times K}$ s ľubovoľným počtom stĺpcov K (atómov) oproti symetrickým ortogonálnym bázam. Taký typ matíc sa vola nadkompletná slovníková matica [14]. Rovnako ako aj pre (1) platí, že riešenie nie je jedinečne. Najredšie zastúpenie sa rovná najmenšiemu poctu nenulových koeficientov (8), ktorý sa určujú l_0 normou.

$$\min_x \|x_0\|, \qquad pouzitim\ f = Dx\,, \tag{8}$$

A. MOMP algoritmus na učenie slovníka

Existuje viacero metód na redukciu rozmeru transformačnej matice prostredníctvom učenia. Medzi najznámejšie patria: K-SVD (K-Singular Value Decomposition), ktorá iteratívne aktualizuje atómy slovníka [15]; MOD (Method of Optimal Directions), pri ktorej sa striedavo optimalizuje slovník a riedke koeficienty pomocou metódy najmenších štvorcov [16]; a MOMP (Multiply Orthogonal Matching Pursuit), ktorý patrí do skupiny greedy algoritmov. Tento prístup hľadá riešenie pomocou l_1 minimalizácie s cieľom nájsť optimálnu množinu atómov, ktoré najpresnejšie reprezentujú vstupný signál [17].

Algoritmus MMV-OMP
Vstup: $\psi \in \mathbb{R}^{N\times N}$, X$\in \mathbb{R}^{N\times k}$, $T^{[0]} = \emptyset$,
Inicializácia: $r^{[0]} = y, R = konstanta, i = 0$;
Hlavný cyklus: pokiaľ nie je splnená ukončovacia podmienka
výpočet korelácii medzi rezíduom a atómami slovníka;
vyber najlepšieho atómu;
riešenie riedkej reprezentácie;
aktualizácia rezídua
Výstup: $D \in \mathbb{R}^{N\times K}$

Kde ψ je transformacna matica; X predstavuje do vektory signálov zapísaných do matice, kde každý stĺpec predstavuje jeden signál; $r^{[0]}$ je inicializácia rezídua, ktoré sa na zatáčku algoritmu rovná vstupnému signálu; R je referencna hodnota rezídua, po prekročení ktorej sa zastaví hlavný cyklus; D je vystupny slovník. Algoritmus sa snaží nájsť atóm pôvodnej matice ψ, ktory má najväčšiu koreláciu s rezíduom. Pred ďalšou iteráciou sa vypočíta nová hodnota rezídua. Dany algoritmus sa opakuje, kým hodnota rezídua neklesne pod stanovenú prahovú hodnotu.

IV. Obnova signálu pomocou OMP

Výsledkom rekonštrukcie je riedky vektor x. Klasická spätná rekonštrukcia sa robí pomocou konvexných optimalizačných metód [8], ktoré hľadajú najredšie riešenie:

$$\min_x \|x\|_0, vzhladom\ k\ Ax = y\,, \tag{9}$$

Tento prístup k riešeniu problému spätnej rekonštrukcie má viacero nevýhod, ktoré ho v mnohých aplikáciách robia výpočtovo náročným a prakticky obmedzeným. Medzi hlavné nevýhody patrí NP-úplnosť, kombinatorická povaha, nelinearita a vysoká citlivosť na šum. NP-úplnosť vyplýva z použitia optimalizačného problému s l_0 normou, čo znamená, že neexistuje polynomiálny algoritmus na jeho presné riešenie [18]. Na aproximáciu l_0 normy sa preto v prípade splnenia podmienok kompresného snímania využívajú l_1a $l_2$normy. Okrem konvexných optimalizačných metód existujú aj tzv. greedy algoritmy, ktoré majú nižšiu výpočtovú náročnosť. Medzi tieto algoritmy patria Matching Pursuit [19], Orthogonal Matching Pursuit [20] a Iterative Hard Thresholding [21].V tejto práci sa zameriavame na algoritmus OMP (Orthogonal Matching Pursuit), ktorý iteratívne hľadá riedku reprezentáciu signálu. V každej iterácii vypočítava koreláciu medzi aktuálnym rezíduom a všetkými stĺpcami matice A, pričom vyberá stĺpec s najvyššou zhodou. Následne vykoná ortogonálnu projekciu vybraných atómov, čím vytvorí aproximačný vektor, ktorý slúži na aktualizáciu rezídua. Rovnaké stĺpce sa nevyberajú opakovane a nové rezíduum je vždy ortogonálne na vybrané atómy. Algoritmus sa ukončí po dosiahnutí maximálneho počtu iterácií alebo po prekročení nastaveného prahového kritéria pre rezíduum.

OMP
Vstup: $A \in \mathbb{R}^{N\times M}$, y a podmienky ukončenia
Inicializácia: Inicializácia: $r^{[0]} = y,\ \hat{x}^{[0]} = 0, T^{[0]} = \emptyset, i = 0$;
Hlavný cyklus: pokiaľ nie je splnená ukončovacia podmienka
$g^{[i]} = A^T r^{[i-1]}$

$j^{[i]} = argmax \frac{|g_j^{[i]}|}{\|A_j\|_2}$
$T^{[i]} = T^{[i]} \cup j^{[i]}$
$\hat{x}_{T^{[i]}}^{[i]} = A_{T^{[i]}}^{\dagger} y$
$r^{[i]} = y - A\hat{x}^{[i]}$
Výstup: $r^{[i]}, \hat{x}^{[i]}$;

V. EXPERIMENTÁLNE VÝSLEDKY

Na vyhodnotenie analyzovanej metódy boli použité EKG záznamy z databázy PTB Diagnostic ECG Database, ktoré boli získané pri vzorkovacej frekvencii 1 kHz a s rozlíšením 16 bitov [22]. Okrem samotných EKG signálov databáza obsahuje aj doplňujúce informácie o veku pacienta a detailné popisy jeho zdravotného stavu. Na testovanie metódy bola použitá DBBD snímacia matica. Redukovaná bázová matica bola zostavená z DCT funkcií a následne optimalizovaná pomocou algoritmu MMV-OMP, ktorý bol trénovaný na EKG záznamoch pochádzajúcich od zdravých aj chorých pacientov. Pre každý jednotlivý zvod bol vytvorený samostatný slovník prispôsobený špecifickým charakteristikám príslušného signálu.

Tab. 1. Výsledne rozmery slovníkov po učení pomocou MMV-OMP

1	2	3	4	5	6
1800x631	1800x515	1800x645	1800x717	1800x606	1800x460
7	8	9	10	11	12
1800x302	1800x468	1800x377	1800x339	1800x319	1800x355

Kvalita rekonštrukcie bola vyhodnotenia pomocou PRD:

$$PRD(\%) = \frac{\|x - \hat{x}\|_2}{\|x\|_2} \times 100, \tag{10}$$

kde x je originálny signál a $\hat{x}$ je rekonštruovaný signál. Vzorky EKG signálov boli testované pri rôznych kompresných pomeroch (CR = 2, 4, 6, 8, 10). Každý EKG signál bol rozdelený na rámce rovnakej dĺžky, pričom dĺžka každého rámca bola 1800 vzoriek. Pre každý jednotlivý prípad bolo vykonaných päť nezávislých výpočtov, z ktorých boli do tabuliek zaznamenané priemerné hodnoty a smerodajné odchýlky chyby rekonštrukcie.

Tab. 2 Výsledky kompresného snímania pre rôzne CR pre pacienta č.1 – infarkt myokardu

CR	2		4		6		8		10	
Zvod	PRD	σ^2	PRD	σ^2	PRD	σ^2	PRD	σ^2	PRD	σ^2
1	1,2658	0,0129	1,9498	0,0050	2,5636	0,0173	3,1650	0,0415	3,7215	0,0663
2	1,1586	0,0008	1,8957	0,0014	2,5711	0,0021	3,0953	0,0051	3,7666	0,0116
3	2,4990	0,0226	2,7727	0,0316	3,4349	0,0469	4,2993	0,0503	5,7462	0,1591
4	1,5847	0,5344	1,9154	0,0031	2,5280	0,0130	3,1043	0,0198	3,6340	0,0246
5	2,1513	0,3755	2,4918	0,2004	3,0506	0,2636	3,8742	0,4208	4,7630	0,7886
6	1,4603	0,0004	2,0188	0,0009	2,7142	0,0025	3,2037	0,0017	4,1319	0,0086
7	1,5328	0,0017	2,1106	0,0132	2,7474	0,0176	3,2635	0,0075	4,4410	0,0259
8	0,7579	0,0020	1,0236	0,0028	1,3671	0,0024	2,0586	0,0104	3,2216	0,0019
9	0,8568	0,0025	1,1380	0,0258	1,4241	0,0368	2,5038	0,0993	3,8145	0,0967
10	0,9461	0,0003	1,4183	0,0041	2,0115	0,0058	3,0227	0,0411	4,3244	0,1453
11	1,1764	0,0007	1,6234	0,0038	2,1704	0,0065	2,5345	0,0086	3,5649	0,0369
12	1,3886	0,0040	1,9073	0,0023	2,4730	0,0042	3,0298	0,0116	3,5714	0,0133

Tab. 3 Výsledky kompresného snímania pre rôzne CR pre pacienta č.2 – zdravý pacient

CR	2		4		6		8		10	
Zvod	PRD	σ^2	PRD	σ^2	PRD	σ^2	PRD	σ^2	PRD	σ^2
1	3,3583	0,0781	5,7961	0,2402	7,3044	0,4688	8,8940	0,7330	10,1068	1,3278
2	1,9132	0,0239	3,1860	0,0660	3,9938	0,1851	5,2673	0,3315	5,6096	0,2549
3	2,1659	0,1315	3,8134	0,3222	4,7946	1,0099	6,0967	1,4075	6,6920	1,6396
4	2,2963	0,0193	3,8083	0,0389	4,7083	0,1103	6,0688	0,0756	6,5271	0,1993
5	3,0509	0,3522	5,0057	0,7097	6,4981	1,7556	7,9464	2,9979	8,9244	3,7932
6	1,9602	0,0567	3,3128	0,2294	4,0418	0,3229	5,4190	0,6277	5,7338	0,7306
7	1,1233	0,0001	1,8905	0,0015	2,5151	0,0023	3,3862	0,0038	3,8383	0,0028
8	0,9259	0,0001	1,5835	0,0004	2,1804	0,0028	2,8696	0,0017	3,2629	0,0004
9	0,8575	0,0001	1,4028	0,0010	1,9976	0,0019	2,9159	0,0032	3,3254	0,0029
10	1,0859	0,0012	1,8582	0,0028	2,4833	0,0015	2,9356	0,0035	3,4817	0,0074
11	1,2537	0,0147	2,1277	0,0382	2,9483	0,0434	3,8597	0,1032	4,0866	0,0806
12	2,2047	0,0963	3,2389	0,3113	4,4157	0,4539	5,8859	0,9809	6,0232	0,8169

Na základe experimentálnych údajov možno pozorovať, že kvalita rekonštrukcie EKG signálu výrazne závisí od hodnoty kompresného pomeru (CR) a zároveň aj od typu patologického stavu pacienta. Pre oboch pacientov boli dosiahnuté priemerné hodnoty PRD pri CR = 2 a 4 výrazne pod hranicou PRD = 9%, čo zaručuje vysokú diagnostickú presnosť. Pri CR = 6 zostali hodnoty PRD u väčšiny zvodov stále v akceptovateľnom rozmedzí, avšak s narastajúcou variabilitou v závislosti od jednotlivých pacientov a zvodov. Zvýšenie kompresného pomeru na CR = 8 a 10 viedlo k nárastu PRD nad hranicu 9 v prípade viacerých zvodov, najmä u zdravého pacienta.

Najnižšie PRD hodnoty boli dosahované vo zvodoch 7 až 12. Naopak, zvody 1 až 6 vykazovali vyššiu chybovosť. Výsledky potvrdzujú, že navrhovaná metóda kompresného snímania v kombinácii s učením slovníkov poskytuje dostatočnú presnosť rekonštrukcie pre CR ≤ 6 a je vhodná pre klinické aplikácie, pričom odporúčaná horná hranica CR pre zachovanie diagnostickej spoľahlivosti by nemala prekročiť hodnotu 6.

VI. Záver

Táto práca sa zaoberala optimalizáciou kompresného snímania EKG signálov s využitím učenia slovníkov, konkrétne pomocou algoritmu MMV-OMP. Pre každý zvod bol vytvorený samostatný slovník, ktorý bol trénovaný na základe reálnych signálov. Využitie DBBD snímacích matíc v kombinácii s učenými slovníkmi umožnilo znížiť množstvo potrebných vzoriek pri zachovaní dostatočnej kvality rekonštrukcie. Kvalita bola hodnotená metrikou PRD, pričom za hranicu akceptovateľnej diagnostickej presnosti sa považuje hodnota PRD < 9. Experimentálne výsledky potvrdili, že pri kompresných pomeroch CR = 2 až 6 zostáva väčšina rekonštruovaných signálov pod touto hranicou, čo znamená, že signály sú dostatočne presné pre potreby diagnostiky kardiológom. Pri vyšších CR (napr. CR = 8 alebo 10) sa však niektoré zvody u určitých pacientov dostali nad túto hranicu, čo môže viesť k zníženej spoľahlivosti klinickej interpretácie.

Použitá literatúra

[1] R. Jain, R. Singh, S. Yamini, and M. Das, 'Fragmented ECG as a Risk Marker in Cardiovascular Diseases', Curr. Cardiol. Rev., vol. 10, no. 3, pp. 277–286, May 2014, doi: 10.2174/1573403X10666140514103451.
[2] T. B. Garcia and Neil E. Holtz, '12 Lead ECG: The Art of Interpretation'. Accessed: Jan. 25, 2025. [Online].
[3] I. Tomasic and R. Trobec, 'Electrocardiographic Systems With Reduced Numbers of Leads—Synthesis of the 12-Lead ECG', IEEE Rev. Biomed. Eng., vol. 7, pp. 126–142, 2014, doi: 10.1109/RBME.2013.2264282.
[4] S. Majumder, L. Chen, O. Marinov, C.-H. Chen, T. Mondal, and M. J. Deen, 'Noncontact Wearable Wireless ECG Systems for Long-Term Monitoring', IEEE Rev. Biomed. Eng., vol. 11, pp. 306–321, 2018.
[5] O. Kováč, J. Kromka, J. Šaliga, and A. Jusková, 'Multiwavelet-based ECG compressed sensing', Measurement, vol. 220, p. 113393, Oct. 2023, doi: 10.1016/j.measurement.2023.113393.
[6] A. Pandey, 'ECG data compression using the formation of QRS-complex segment bank and integer DCT-based plateau region processing', Biomed. Signal Process. Control, vol. 85, p. 104823, Aug. 2023, doi: 10.1016/j.bspc.2023.104823.
[7] F. Picariello, G. Iadarola, E. Balestrieri, I. Tudosa, and L. De Vito, 'A novel compressive sampling method for ECG wearable measurement systems', Measurement, vol. 167, p. 108259, Jan. 2021.
[8] Y. C. Eldar and G. Kutyniok, Eds., Compressed Sensing: Theory and Applications, 1st ed. Cambridge University Press, 2012. doi: 10.1017/CBO9780511794308.
[9] S. Yao, T. Wang, W. Shen, P. Shaoming, and Y. Chong, 'Research of incoherence rotated chaotic measurement matrix in compressed sensing', Multimed. Tools Appl., vol. 76, no. 17, pp. 17699–17717, Sep. 2017, doi: 10.1007/s11042-015-2953-2.
[10] G. Wang, M.-Y. Niu, and F.-W. Fu, 'Deterministic construction of compressed sensing matrices from constant dimension codes', Discrete Appl. Math., vol. 285, pp. 9–17, Oct. 2020, doi: 10.1016/j.dam.2020.04.032.
[11] P. Daponte, L. De Vito, G. Iadarola, F. Picariello, and S. Rapuano, 'Deterministic Compressed Sensing of heart sound signals', in 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Lausanne, Switzerland: IEEE, Jun. 2021, pp. 1–6. doi: 10.1109/MeMeA52024.2021.9478766.
[12] M. Z. Shakir and N. Ramzan, Eds., 'Combining deterministic compressed sensing and machine learning for data reduction in connected health', in AI for Emerging Verticals: Human-robot computing, sensing and networking, Institution of Engineering and Technology, 2020, pp. 301–314. doi: 10.1049/PBPC034E_ch15.
[13] K. Luo, L. Ou, M. Zhang, S. Liao, and C. Zhang, 'A dictionary learning based unsupervised neural network for single image compressed sensing', Image Vis. Comput., vol. 151, p. 105281, Nov. 2024, doi: 10.1016/j.imavis.2024.105281.
[14] K. Okoudjou, Ed., Finite Frame Theory, vol. 73. in Proceedings of Symposia in Applied Mathematics, vol. 73. Providence, Rhode Island: American Mathematical Society, 2016. doi: 10.1090/psapm/073.
[15] M. Aharon, M. Elad, and A. Bruckstein, 'K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation', IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006, doi: 10.1109/TSP.2006.881199.
[16] K. Engan, S. O. Aase, and J. Hakon Husoy, 'Method of optimal directions for frame design', in 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258), Phoenix, AZ, USA: IEEE, 1999, pp. 2443–2446 vol.5. doi: 10.1109/ICASSP.1999.760624.
[17] M. Elad, Sparse and redundant representations: from theory to applications in signal and image processing. New York: Springer, 2010.
[18] I. Orović, V. Papić, C. Ioana, X. Li, and S. Stanković, 'Compressive Sensing in Signal Processing: Algorithms and Transform Domain Formulations', Math. Probl. Eng., vol. 2016, pp. 1–16, 2016, doi: 10.1155/2016/7616393.
[19] S. G. Mallat and Zhifeng Zhang, 'Matching pursuits with time-frequency dictionaries', IEEE Trans. Signal Process., vol. 41, no. 12, pp. 3397–3415, Dec. 1993, doi: 10.1109/78.258082.
[20] K. Schnass, 'Average Performance of Orthogonal Matching Pursuit (OMP) for Sparse Approximation', IEEE Signal Process. Lett., vol. 25, no. 12, pp. 1865–1869, Dec. 2018, doi: 10.1109/LSP.2018.2878061.
[21] R. E. Carrillo, L. F. Polania, and K. E. Barner, 'Iterative hard thresholding for compressed sensing with partially known support', in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic: IEEE, May 2011, pp. 4028–4031. doi: 10.1109/ICASSP.2011.5947236.
[22] R.-D. Bousseljot, D. Kreiseler, and A. Schnabel, 'The PTB Diagnostic ECG Database'. physionet.org, 2004.

Poďakovanie

Táto práca bola podporená Vedeckou grantovou agentúrou - projekt VEGA č. 1/0413/22.

Základné vlastnosti signálov globálneho lokalizačného systému

[1]*Tomáš KANÁLIK,* [1]*Ondrej KOVAČ,* [2]*Antónia KOVAČOVÁ*

[1] Katedra Technológii v Elektronike, [2]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, Slovenska Republika

[1]tomas.kanalik@student.tuke.sk, [1]ondrej.kovac@tuke.sk, [2]antonia.kovacova@tuke.sk

Abstrakt - Globálny lokalizačný systém (GPS) je satelitný systém, ktorý poskytuje presnú lokalizáciu a údaje o čase pre široké spektrum aplikácií, od každodenného používania v smartfónoch až po pokročilé vedecké a vojenské využitie. Táto práca poskytuje prehľad princípov fungovania GPS, s dôrazom na signálové pásma L1 a L2, ich štruktúru, techniky modulácie a procesy trilaterácie. Opisované sú základné charakteristiky a technológie využívané v GPS, vrátane softvérového spracovania signálov, ktoré umožňuje presné určenie polohy prijímača. Analyzované sú niektoré technické parametre signálov. Tento príspevok vznikol ako súčasť riešenia semestrálnej práce na predmete Číslicové spracovanie signálov, ktorý sa vyučuje v študijnom programe Automobilová elektronika.

***Kľúčové slova* - GPS, družica, satelit, trilaterácia, lokalizácia**

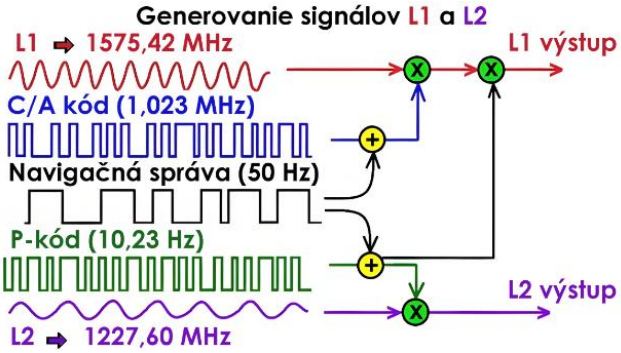
I. Úvod

Globálny lokalizačný systém (v ang. jazyku Global Positioning System, GPS) je satelitný systém umožňujúci určovanie geografickej polohy na Zemi alebo obežnej dráhe Zeme. Systém pozostáva z 24 satelitov nachádzajúcich sa na strednej obežnej dráhe, ktoré sú pre verejnosť nepretržite dostupné. Vďaka dosiahnuteľnej presnosti prevažne do niekoľkých jednotiek až desiatok metrov a univerzálnemu využitiu sa GPS stal neoddeliteľnou súčasťou moderných navigačných systémov či inteligentných telefónov [4]. Prevádzka GPS je zabezpečovaná ministerstvom obrany Spojených štátov amerických, pričom jeho primárnym účelom pôvodne bolo presné určovanie polohy vojenských jednotiek. S postupným vývojom a rozširovaním jeho funkcií bol systém sprístupnený aj civilnému sektoru, hoci s určitými obmedzeniami v presnosti, ktoré do značnej miery závisia od technických a environmentálnych podmienok. Tento technologický pokrok umožnil GPS preniknúť do širokej škály aplikácií, od komerčných navigačných prístrojov až po špecializované vedecké a priemyselné riešenia. História GPS siaha do 60. rokov 20. storočia, keď bol systém vyvíjaný ako súčasť spoločného vojensko-civilného projektu. Výraznejší pokrok však nastal v roku 1973, kedy sa GPS stal základom modernej navigácie. GPS taktiež predstavuje zdroj presných hodín, keďže súčasťou každého satelitu sú atómové hodiny [8]. Hoci GPS predstavuje globálne rozšírený systém, na trhu existujú aj jeho alternatívy, ako napríklad ruský systém GLONASS alebo európsky Galileo [5][6]. Táto práca sa primárne zameriava na analýzu GPS, pričom pozornosť bude venovaná parametrom jeho signálov a taktiež aj metódam modulácie, ktoré umožňujú praktické využitie tohto satelitného systému v rôznych odvetviach. V druhej kapitole článku sa zaoberáme všeobecnou signálovou štruktúrou GPS so zreteľom na jednotlivé pásma, trilateráciu a princíp generácie signálov. Objasňuje podstata používaných kódov a bližšie špecifikácie GPS signálov. V tretej kapitole sa nachádza analýza modulácií signálov – BPSK a BOS, ktoré sú využívané práve v navigačných (ale aj iných) systémoch.

II. Signálová štruktúra GPS

Verejnosti sú aktuálne prístupné dve GPS pásma, L1 a L2. Pásmo L1 využíva pre určenie polohy C/A (Coarse / Acquisition) kód, ktorý je určený na používanie verejnosťou, a P-kód (Precision Code), ktorého účely sú výhradne armádne vďaka vysokej presnosti lokalizácie. Každý satelit má pridelený svoj vlastný, unikátny C/A kód, ktorý je tvorený 1023 prvkami - čipmi. V kontexte modulácií používaných v GPS a iných navigačných systémoch, termín „čip“ (chip) označuje základnú jednotku pseudonáhodného kódu. Tie sú posielané rýchlosťou 1,023 Mb/s. Sekvencia kódu je teda opakovaná každú milisekundu. Ide o PRN (Pseudo-Random Noise) kódy s veľmi nízkou vzájomnou koreláciou, teda sú takmer ortogonálne. Táto vlastnosť povoľuje použitie CDMA (Code Division Multiple Access). P-kód taktiež predstavuje sériu jednotiek a núl (čipov) generovaných o frekvencii 10,23 MHz. Počas sledovacieho procesu je stredná frekvencia

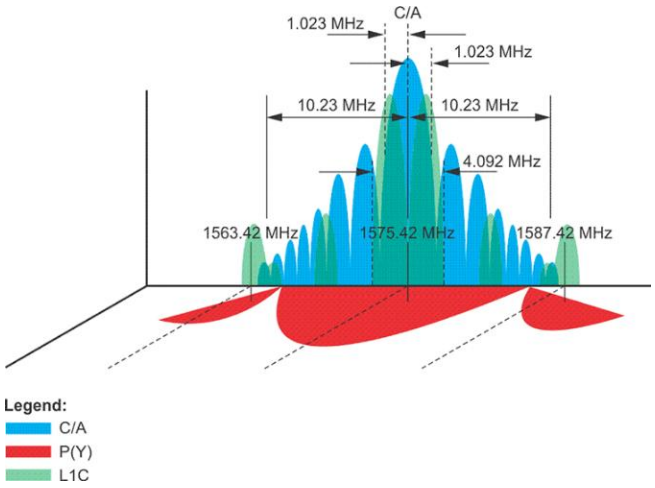
úzkopásmového filtra pevná, zatiaľ čo lokálne generovaný signál sleduje frekvenciu vstupného signálu. Na Obr. 1 sa nachádza stručné znázornenie princípu generovania signálov L1 s C/A kódom a L2 s P-kódom, aj s obsahom navigačnej správy. Znázornené sú konkrétne frekvencie daných signálov.



Obr. 1 Princíp generovania GPS signálu L1 pásma [11].

Fázy oboch signálov sú porovnávané prostredníctvom fázového komparátora, pričom výsledný signál prechádza cez úzkopásmový filter. Na sledovanie GPS signálu sú potrebné dve sledovacie slučky: jedna na sledovanie nosnej frekvencie a druhá na sledovanie C/A kódu [7][8].

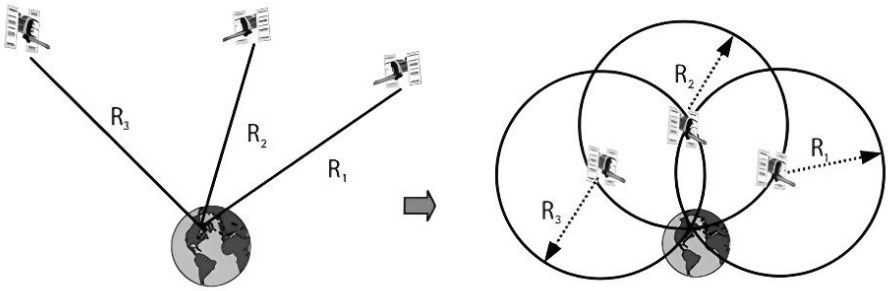
Informácie pochádzajúce z GPS sú kódované prostredníctvom zmeny fáz nosných vĺn. Táto metóda modulácie, známa ako fázová modulácia, umožňuje signálu obsadiť širšie pásmo, ako by bolo inak možné. GPS signály preto disponujú tzv. rozprestretým spektrom (Obr. 2), čo znamená, že celková šírka pásma signálu je úmyselne väčšia než šírka pásma samotnej prenášanej informácie. Nosná frekvencia pásma L1 je 1575,42 MHz, L2 potom 1227,60 MHz a pásmo L5 je vysielané na frekvencii 1176,45 MHz. Šírka signálov však presahuje pásmo okolo týchto frekvencií. Napríklad, signál C/A kódu sa rozprestiera na 2,046 MHz a signál P(Y) kódu až na 20,46 MHz v rámci L1. Pásmo L1C je rozprestreté na frekvencii 4,092 MHz [9].



Obr. 2 Ukážka rozprestretého spektra GPS signálu (pásmo L1) [9].

Najbežnejšie používanou technikou modulácie rozprestretého spektra je BPSK – Binary-Phase Shift Keying, prípadne pokročilejšia alternatívna BOC – Binary Offset Carrier. Ich princíp bude objasnený v 3. kapitole. Pokiaľ sú známe vzdialenosti z určitého bodu na Zemi k minimálne trom GPS satelitom a zároveň sú známe polohy satelitov, potom je možné určiť polohu tohto bodu (GPS prijímača) prostredníctvom trilaterácie. Najprv je však potrebné poznať vhodný spôsob na získanie vzdialenosti od GPS satelitov a ich polohu na obežnej dráhe Zeme. Každý GPS satelit nepretržite vysiela mikrovlnný rádiový signál, ktorý je tvorený z dvoch nosných frekvencií, dvoch kódov a navigačnej správy. GPS prijímač zachytáva signály prostredníctvom jeho vstavanej antény [7][8]. Po zachytení signálu nasleduje jeho softvérové spracovanie. Čiastočný výsledok prebiehajúcich výpočtov obsahuje informáciu o vzdialenosti k GPS satelitom, známe ako pseudovzdialenosti. Súradnice satelitov sú obsiahnuté v navigačnej správe. Teoreticky je postačujúca len informácia o vzdialenosti GPS prijímača od troch satelitov. V tomto prípade by sa prijímač nachádzal v priesečníku troch sfér, z ktorých každá má polomer rovnajúci sa vzdialenosti prijímača od satelitu, ako je znázornené na Obr. 3. Z praktického hľadiska je však potrebný štvrtý satelit a informácia z jeho vstavaných, extrémne presných atómových hodín, na

kompenzáciu chyby v hodinách prijímača. Jednotlivé vzdialenosti sú vypočítavané na základe časového rozdielu medzi vyslaním signálu zo satelitu a prijatím signálu na strane prijímača [7][8].



Obr. 3 Znázornenie princípu trilaterácie [7].

Na presné určenie požadovaných súradníc je potrebné vytvoriť pomyselnú sféru s polomerom r_i okolo každého zo štyroch satelitov. Sféry je možné vyjadriť rovnicami: $(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = r_i^2$, kde n predstavuje daný satelit ($i \in \langle 1,4 \rangle$). x_i, y_i, z_i sú známe súradnice satelitov v priestore a x, y, z sú neznáme súradnice GPS prijímača (pozorovateľa) na Zemi. Následne je potrebný odpočet prvej rovnice od každej z nasledujúcich troch pre štvoricu družíc. Tým sú eliminované kvadratické členy v x, y, z a vznikne sústava lineárnych rovníc:

$$2(x_2 - x_1)x + 2(y_2 - y_1)y + 2(z_2 - z_1)z = r_1^2 - x_1^2 - r_2^2 + x_2^2 \quad (1)$$

$$2(x_3 - x_1)x + 2(y_3 - y_1)y + 2(z_3 - z_1)z = r_1^2 - x_1^2 - r_3^2 + x_3^2 \quad (2)$$

$$2(x_4 - x_1)x + 2(y_4 - y_1)y + 2(z_4 - z_1)z = r_1^2 - x_1^2 - r_4^2 + x_4^2 \quad (3)$$

Vyjadrením súradníc x, y a z zo sústavy (1-3) sú získané súradnice pozorovateľa vzhľadom na pevne stanovený, známy bod na Zemi. Tento príklad ukazuje aplikáciu jednoduchých vzorcov odvodených z analytickej geometrie sfér a elementárne použite Gaussovej eliminácie. Tri uvedené rovnice by boli dostatočné na výpočet x, y a z súradníc. (t.j. použitie len troch satelitov, hoci ich môže byť pri výpočte zapojených viac). V praxi, keď je pri určovaní polohy zapojených viac satelitov, sa využívajú pokročilé počítačové algoritmy [7]. V takom prípade sa často využíva výpočet polohy pre všetky možné kombinácie štyroch satelitov, pričom sa následne vyberá kombinácia s najnižším faktorom DOP (Dilution of Precision), čo zlepšuje presnosť určenia polohy. Okrem toho sa v praxi využívajú aj metódy ako diferenciálne GPS (DGPS), ktoré korigujú chyby merania pomocou referenčných staníc, a metódy založené na fázových meraniach nosných vĺn, ktoré umožňujú dosiahnuť vyššiu presnosť lokalizácie. Dodatočne, štruktúra AGPS dokáže určiť približnú polohu pozorovateľa na Zemi len za pomoci dvoch satelitov, pričom aproximácii napomáha pozemná sieť GPRS (General Packet Radio Service).

Pásmo L2 má stredovú frekvenciu 1227,60 MHz a šírku pásma 11 MHz (rozsah frekvencií 1221,1 až 1233,1 MHz). Keďže L2 pracuje v porovnaní s L1 (1575,42 MHz) na nižšej frekvencii, umožnený je lepší prenos signálu cez fyzické prekážky (oblaky, častice v atmosfére, resp. ionosfére, veľké objekty na zemskom povrchu, ...). Pásmo L2 bolo pôvodne navrhnuté výhradne na armádne účely, no neskôr bol navrhnutý signál L2C (C – „Civilian“), ktorý znamenal počiatok dvojpásmových GPS prijímačov určených pre verejnosť. Dvojpásmové GNSS (Global Navigation Satellite System) prijímače sa vyznačujú vyššou presnosťou. Kým jednokanálové prijímače dosahujú presnosť v rozsahu 3 až 15 metrov (za určitých podmienok, ako napr. za jasného počasia a bez fyzických prekážok v okolí, 1 až 3 metre), profesionálne dvojkanálové prijímače je možné lokalizovať do pár desiatok centimetrov vďaka možnosti eliminácie chýb zapríčinených napríklad časticami v ionosfére a objektami v blízkom okolí [14]. Signály z pásma L2 sú tiež preto vhodné pri monitorovaní ionosféry, počasia a pre účely vedeckých aktivít. Využívané sú štyri druhy kódov: L2CM, L2CL, P(Y) kód a M kód [1].

Predtým, než začneme analyzovať jednotlivé L2 kódy, je potrebné pochopiť základné charakteristiky a kontext blokov GPS satelitov. Pásma L2 sú využívané v rámci blokov IIR-M, IIF a III. Satelity bloku IIR-M predstavujú vylepšenú verziu predchádzajúcich satelitov IIR. Zavádzajú vojenský signál M-Code dostupný na frekvenciách L1 a L2, ktorý zvyšuje odolnosť voči rušeniu a vyznačuje sa lepším zabezpečením. Pre civilné aplikácie bol na frekvencii L2 zavedený robustnejší signál L2C, ktorý zlepšuje presnosť a spoľahlivosť služieb určovania geografickej polohy. Blok IIF predstavuje ďalšiu generáciu GPS satelitov a obsahujú vylepšenia v podobe dlhšej životnosti (12 rokov), rýchlejších procesorov a väčších pamätí. Zavádzajú taktiež tretiu frekvenciu, L5. Satelity GPS IIIA zavádzajú ďalší civilný signál (L1C), ktorý je navrhnutý tak, aby bol vysoko interoperabilný so signálmi európskeho systému Galileo a taktiež je

zabezpečená kompatibilita so signálmi z japonského systému Quasi-Zenith Satellite System (QZSS). Pre vojenské aplikácie, GPS IIIA prinášajú výrazné zvýšenie odolnosti proti rušeniu prostredníctvom vylepšeného M-kód signálu [10].

L2CM kód (CM – Civil Moderate) je vysielaný satelitmi blokov IIR-M, IIF a III, pričom slúži na prenos informácií o vzdialenosti medzi satelitom a prijímačom. Tento kód využíva techniku časového multiplexovania (TDM - Time Division Multiplexing), čo znamená, že je vysielaný v unikátnych časových slotoch (intervaloch). TDM minimalizuje interferencie medzi rôznymi signálmi a zabezpečuje, že celý frekvenčný rozsah je efektívne využívaný v príslušných časových slotoch. Každý L2CM kód pozostáva z lineárne generovanej pseudonáhodnej sekvencie binárnych hodnôt. Táto sekvencia je navrhnutá tak, aby poskytovala nízku pravdepodobnosť kolízie a zlepšovala presnosť pri určovaní polohy [1].

L2CL kód (Civil Long) je podobne ako L2CM generovaný satelitmi blokov IIR-M, IIF a III, no jeho štruktúra je charakteristická relatívne dlhšou sekvenciou. Tento kód je rovnako časovo multiplexovaný, čo zabezpečuje efektívnu synchronizáciu a prenos údajov. Okrem svojej hlavnej funkcie, ktorou je prenos navigačných údajov, umožňuje L2CL kód pridávať ďalšie informácie určované pozemnými stanicami. V kombinácii s L2CM kódom poskytuje vyššiu presnosť a robustnosť pri určovaní polohy [1].

P(Y) kód je navrhnutý špecificky pre vojenské alebo vládne aplikácie. Tento kód je rozšírenou verziou P-kódu, pričom komponent Y sa používa na zabezpečenie ochrany proti „spoofingu“ a rušeniu. Spoofing v kontexte GPS signálov označuje útok, pri ktorom útočník vysiela falošné GPS signály s cieľom oklamať prijímač, aby vypočítal nesprávnu polohu alebo čas. Takéto signály napodobňujú legitímne GPS signály, ale obsahujú upravené údaje, čo môže viesť k dezorientácii navigačných systémov, narušeniu logistiky alebo ohrozeniu bezpečnosti. Tento kód využíva pokročilé modulačné techniky, akou je BOC, o ktorej sa ešte budeme zmieňovať. V Tab. 1 sú podrobne uvedené vlastnosti a parametre vyššie uvedených kódov pre pásmo L2 [1].

Tab. 1 Porovnanie vlastností kódov v rámci pásma L2

Technológia GNSS	GPS	GPS	GPS	GPS
Typ kódu	L2 CM	L2 CL	P (Y)	M
Stredová frekvencia	1227.60 MHz	1227.60 MHz	1227.60 MHz	1227.60 MHz
Kódová frekvencia	511.5 kHz	511.5 kHz	10.23 MHz	5.115 MHz
Modulačná schéma	BPSK	BPSK	BPSK	BOC
Typ použitého signálu	Dátový	Pilotný (referenčný)	Dátový	N/A
Frekvenčné pásmo	L2	L2	L2	L2
Dátový tok	25 alebo 50 bps	N/A	50 bps	N/A
Hlavné využitie	Civilné navigačné účely	Rozšírené civilné aplikácie	Vojenské a bezpečnostné účely	Vysoko presné vojenské účely
Presnosť	Stredná	Vysoká	Vysoká	Najvyššia
Odolnosť voči rušeniu	Stredná	Stredná	Vysoká	Najvyššia

III. MODULÁCIE GPS SIGNÁLU – BPSK A BOC

Modulácia signálu predstavuje základný aspekt úspešného prenosu a interpretácie dát. BPSK patrí k jednoduchým a osvedčeným typom modulácie. Hlavnou výhodou BPSK je jej efektívnosť pri spracovaní signálu a nenáročné požiadavky na hardvér. Využívaný je princíp zmeny fázy nosného signálu na reprezentáciu binárnych hodnôt. Ide o efektívny variant pre väčšinu bežných aplikácií, no nevýhodou je nižšia odolnosť voči šumu a obmedzený dátový tok. Pre účely zlepšenia presnosti a odolnosti voči šumu bol navrhnutý algoritmus na odhad pomeru SNR (Signal to Noise Ratio). Ten umožňuje korekciu chýb a zvýšenie efektivity. Odhad SNR je opísaný vzťahom:

$$SNR = P\left(\frac{1}{L}\sum_{k=0}^{L-1}|r_k|^2 - P\right)^{-1} \tag{4}$$

Jednotlivé parametre majú nasledujúci význam. Priemerný výkon signálu P predstavuje užitočný výkon prijatého signálu, ktorý sa vypočíta na základe vzoriek signálu. Parameter L označuje počet vzoriek prijatého signálu počas daného časového intervalu, pričom každá vzorka je označená ako r_k. Suma reprezentuje súčet okamžitých výkonov všetkých vzoriek v analyzovanom intervale. Výkon šumu v signáli sa vypočíta ako rozdiel medzi priemerným celkovým výkonom prijatého signálu a priemerným výkonom užitočného P. Tento vzťah je možné zapísať ako:

$$P_{noise} = \frac{1}{L}\sum_{k=0}^{L-1}|r_k|^2 - P \tag{5}$$

Odhad SNR je následne vyjadrený ako pomer medzi užitočným výkonom signálu a výkonom šumu. Tento vzťah umožňuje spoľahlivé hodnotenie kvality prijímaného signálu aj v prostredí s vysokým šumom alebo pri nepriaznivých podmienkach prenosu. Prijímaný signál je typicky ovplyvnený šumom a môže byť popísaný ako:

$$r_k = A_k a_k e^{j\theta} + n_k \tag{6}$$

kde A_k je amplitúda signálu, a_k je hodnota binárneho čísla – ak je binárne číslo rovné 0, potom $a_k = 1$. Naopak, ak je jeho hodnota rovná 1, potom $a_k = -1$. θ predstavuje fázový posun a n_k je šumová zložka, pričom sa predpokladá, že má vlastnosti Gaussovho bieleho šumu. Okamžitý výkon každej vzorky r_k sa vypočíta:

$$|r_k|^2 = (Re(r_k))^2 + (Im(r_k))^2 \tag{7}$$

pričom $Re(r_k)$ a $Im(r_k)$ predstavujú reálnu a imaginárnu časť vzorky r_k. Priemerný výkon signálu sa vypočíta spriemerovaním okamžitých výkonov všetkých vzoriek:

$$P = \frac{1}{L}\sum_{k=0}^{L-1} |r_k|^2 \tag{8}$$

BPSK mení fázu nosnej vlny medzi hodnotami 0° a 180° v závislosti od charakteristiky vstupného binárneho signálu. Ide o najčastejšie používanú moduláciu v oblasti rozprestretých spektier. BPSK je typ modulácie využívaný na tvorbu navigačnej správy, P(Y) kódu a C/A kódu. Rýchlosť všetkých súčastí GPS signálov je násobkom štandardnej rýchlosti oscilátorov. Táto štandardná rýchlosť je 10,23 MHz a je známa ako základná hodinová frekvencia. Nosné vlny GPS sú 154-násobkom (1575,42 MHz - L1), 120-násobkom (1227,60 MHz - L2), a 115-násobkom (1176,45 MHz - L5) tejto frekvencie. Za čas, kedy je vygenerovaný jeden čip C/A kódu, sa vygeneruje desať čipov P(Y) kódu. Toto je jeden z dôvodov, prečo pseudovzdialenosť odvádzaná od P(Y) kódu je presnejšia, ako pseudovzdialenosť odvádzaná od C/A kódu [9]. Počas procesu modulácie sa fáza sínusovej nosnej vlny mení v závislosti od základného binárneho pásmového signálu, čím vzniká BPSK signál. Jeho časovú doménu, vyjadrenú funkciou $S_{BPSK}(t)$, možno zapísať nasledovne:

$$S_{BPSK}(t) = \sum_k a_k g(t - nT_s).\cos(\omega_0 t) \tag{9}$$

kde a_k predstavuje štatistickú premennú charakterizujúcu binárne číslo (jej podstatu sme objasnili pri analýze odhadu SNR BPSK modulovaného signálu). T_s predstavuje šírku impulzu [5].

Na druhej strane, BOC je pokročilejšou moduláciou s vyššou presnosťou, ktorá rozdeľuje spektrum signálu do dvoch symetrických lalokov okolo centrálnej frekvencie. Nevýhodou je požiadavka na použitie drahších, pokročilejších prijímačov. Tento typ modulácie zavádza pseudonáhodný kód a pomocnú nosnú vlnu. V porovnaní s tradičnou BPSK moduláciou, BOC lepšie odoláva rušeniu a vyznačuje sa viaccestným šírením signálu. Táto zlepšená výkonnosť vyplýva z vyššej koncentrácie energetického spektra na okrajoch prideleného frekvenčného pásma. BOC modulácia znižuje interferenciu s existujúcimi navigačnými signálmi v spoločných pásmach, akým je napríklad pásmo L1, vďaka odlišnému obsadeniu v spektrálnej oblasti. Výstupný BOC modulovaný signál $S_{BOC}(t)$ možno definovať rovnicou:

$$S_{BOC}(t) = \sum_{k=-\infty}^{\infty} c_k \mu_{T_c}(t - kT_c).sc(t) \tag{10}$$

kde c_k je pseudonáhodná kódová sekvencia modulovaných dátových hodnôt, T_c je šírka čipu, μ_{T_c} je obdĺžnikový impulzný signál so šírkou T_c a $sc(t)$ je offsetová binárna nosná vlna. Rozdelenie spektra umožňuje znížiť interferenciu s inými signálmi, no prijímač musí byť schopný tento zložitejší signál spracovať. Na prekonanie týchto výziev bol vyvinutý algoritmus paralelného vyhľadávania kódových fáz [5].

IV. Záver

Tento článok poskytuje prehľad základných princípov fungovania globálneho lokalizačného systému (GPS), s dôrazom na parametre signálov a metódy ich spracovania v pásmach L1 a L2. Rozbor signalizačných pásiem poukázal na rozdiely medzi civilnými a vojenskými aplikáciami a

analyzoval charakteristiky kódov, ako sú C/A, P(Y), L2C, či pokročilý M-kód. Bolo zdôraznené, že GPS dosahuje presné určenie polohy a času vďaka sofistikovaným technikám, ako je trilaterácia a modulácie. Dôraz bol upriamený na BPSK a BOC. BPSK modulácia, využívaná pre základné navigačné služby, ponúka jednoduchú implementáciu a spoľahlivé výsledky v bežných podmienkach. Na druhej strane, BOC modulácia, vyvinutá pre vojenské a moderné civilné aplikácie, prináša vyššiu odolnosť voči šumu a rušeniu, čo z nej robí vhodnú voľbu do náročnejších prostredí. Nevýhodou BOC modulácie sú však zvýšené požiadavky na výpočtový výkon prijímača. V budúcnosti sa očakáva širšie využitie multikonštelačných systémov, ako Galileo, GLONASS či QZSS, ktoré spolu s GPS vytvoria komplexný a robustný globálny navigačný systém.

POUŽITÁ LITERATÚRA

[1] everythingRF.com; What is the GPS L2 Band? [Online] [cit 30. 11. 2024] https://www.everythingrf.com/community/what-is-the-gps-l2-band.
[2] gps.gov; What is GPS? [Online] [cit. 13. 12. 2024] https://www.gps.gov/systems/gps/.
[3] glonass-iac.ru; About GLONASS [Online] [cit. 14. 12. 2024] https://www.glonass-iac.ru/en/about_glonass/.
[4] euspa.europa.eu; Galileo is the European global satellite-based navigation system [Online] [cit. 14. 12. 2024] https://www.euspa.europa.eu/european-space/galileo/What-Galileo.
[5] Haochen Liu; BPSK/BOC Modulation Signal System for GPS Satellite Navigation Signals.
[6] faa.gov; Satellite Navigation – Global Positioning System (GPS) [Online] [cit. 15. 12. 2024] https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/gps.
[7] Mounchili Mama; Mathematical Modelling of The Global Positioning System Tracking Signals.“ [Online] [Dátum: Máj 2008]. https://www.diva-portal.org/smash/get/diva2:831645/FULLTEXT01.pdf.
[8] Geoconnect.com.au; GPS signals (L1, L2, L5) [Online] [cit 30. 11. 2024] http://geoconnect.com.au/gps-signals-l1-l2-l5/.
[9] e-education.psu.edu; Spread Spectrum and Code Modulation of L1 GPS Carrier [Online] [cit 30. 11. 2024] https://www.e-education.psu.edu/geog862/node/1753.
[10] Los Angeles Air Force Base; Global Positioning System [Online] https://web.archive.org/web/20071116230801/http://www.losangeles.af.mil/library/factsheets/factsheet.asp?id=5325.
[11] YouTube – Michel van Biezen; Special Topics – GPS (5 of 100) Satellite Transmission Signals on L1 and L2 [Online] https://www.youtube.com/watch?v=IcxrGIdOGYo.
[12] Ji-zhong Li; Two satellites positioning algorithm based on AGPS system with two clock bias [Online] [Dátum: 17. jún 2010] https://ieeexplore.ieee.org/document/5485519.
[13] Tomáš Kučera; Algoritmy výpočtu polohy, rychlosti a času z GNSS signálů [Online] [Dátum: 24. máj 2013] https://dspace.vut.cz/server/api/core/bitstreams/941c774f-3fea-4bfc-8374-bf4acb5cfaff/content.
[14] Slavomír Labant, Lucia Kalatovičová, Pavel Kukučka, Erik Weiss; Porovnanie presnosti statickou metódou v reálnom čase pre prístroje globálnych navigačných satelitných systémov [Online] [Rok publikácie: 2009] https://actamont.tuke.sk/pdf/2009/n1/8labant.pdf.

POĎAKOVANIE

Táto práca bola podporená Vedeckou grantovou agentúrou - projekt VEGA č. 1/0413/22.

Nízkoenergetický statický zdroj pre elektrosystémy

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

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

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

Abstract — **Pre výskum a testovanie rôznych elektrotechnických zariadení pomocou vopred definovaných prúdových a napäťových priebehov jednotlivých veličín je častou úlohou ich výrobcov kvôli testovaniu efektívnosti výroby a tiež z dôvodov dodržiavania technických noriem platných pre tieto výrobky. Pretože sortiment takýchto výrobkov je dnes veľmi rozsiahly a často sa mení podľa požiadaviek zákazníkov, bolo by neefektívne vyvíjať výkonové zdroje zvlášť pre každý jednotlivý výrobok. Článok sa zaoberá vývojom univerzálneho inteligentného výkonového zdroja pre testovanie tepelnej spúšte ističov s popisom jeho požiadaviek, štruktúry a funkcií. Uvedená štruktúra výkonovej časti a riadiaceho systému výkonového zdroja je použitá pri konkrétnej aplikácií vo firme pre výrobu ističov SEZ Krompachy a.s.**

***Keywords* — istič, testovacie pracovisko, tepelná spúšť, výkonový zdroj**

I. Úvod

Celosvetový prechod na zelené energie a európsky trend elektrifikácie dopravy v Európe si nutne bude vyžadovať rapídne rozširovanie počtu elektrických rozvodných zariadení (napr. nabíjacích staníc pre elektromobily, rozvádzačov vo fabrikách, domácnostiach a pod.). Najpoužívanejším bezpečnostným prvkom nadprúdovej ochrany je v každom takomto zariadení istič elektrickej siete navrhnutý a inštalovaný na príslušný výkon [1-2]. Pre testovanie týchto ističov sa používajú testovacie zariadenia na testovanie tepelnej a skratovej (magnetickej) spúšte ističov, ktoré sú jedinečné a nikde sa v súčasnosti sériovo nevyrábajú, iba sa vyvíjajú na zákazku podľa potrieb zákazníka [3].

Základnou časťou oboch testovacích zariadení je špecializovaný prúdový zdroj. Pretože ističe predstavujú z elektrického hľadiska spotrebiče s veľmi nízkym odporom, musí prúdový zdroj zabezpečiť veľmi vysoké prúdy s predpísaným (sínusovým) tvarom počas krátkodobého trvania celého testu, čo nie je možné zabezpečiť napäťovými meničmi so štandardnými elektrickými parametrami (napätiami a prúdmi) [4-5]. Okrem toho ceny výkonových súčiastok (výkonových tranzistorov) pre zabezpečenie požadovaných veľmi vysokých testovacích prúdov trvajúcich krátky čas sa v meničoch šplhajú do obrovských rozmerov, čím sa cena testovacieho pracoviska navyšuje [6].

V článku je popísaný návrh nízkoenergetického statického zdroja pre testovacie pracovisko tepelnej spúšte ističov priamo využiteľného pri ich výrobe a zabezpečovaní kvality ich parametrov minimálne na úrovni požadovanej európskymi normami kvality.

II. Popis činnosti a požiadaviek na tester tepelnej spúšte ističov

Hlavnou úlohou tepelnej spúšte ističa je zabezpečiť istené elektrotechnické zariadenie pred krátkodobým nadmerným tepelným preťažením elektrickým prúdom, pričom nesmie vypnúť pri dlhodobom nominálnom prúdovom zaťažení zariadenia. Hodnoty časov a prúdov pre konkrétnu triedu ističov sú definované v príslušnej norme. Ďalej sú uvedené hlavné technické požiadavky, kladené na testovacie zariadenie ističov:

- Maximálny počet testovaných ističov pri teste: 12
- Prevedenie ističov (počet fáz) 1 až 4
- Rozsah menovitých prúdov testovaných ističov (I_n) 0,2 až 125 A
- Hodnota nastavenia vypínacieho prúdu 1,3 až 1,45 I_n
- Hodnoty nastavenia pomerných testovacích prúdov 105%, 113%, 130%, 145%, 255%

- Požadovaný rozsah testovacích prúdov 0,2 až 320 A
- Požadovaná tolerancia presnosti nastavenia prúdu ±0,5%
- Trvanie testu nevypínacím (približne I_n)prúdom max. 60 až 120 minút
- Trvanie testu vypínacím prúdom max. 60 až 120 minút
- Zálohovanie riadiaceho systému pri výpadku napájania
- zo siete a možnosťou pokračovania testu min. 1 minúta

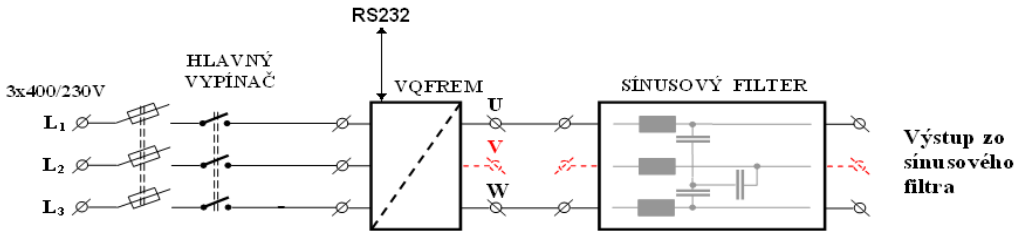
Prídavnou administratívnou požiadavkou je spracovanie protokolov o priebehu jednotlivých testov.

Vysoká varibilita uvedených požiadaviek vedie na použitie inteligentného výkonového zdroja, zloženého z výkonovej časti a riadiaceho systému.

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

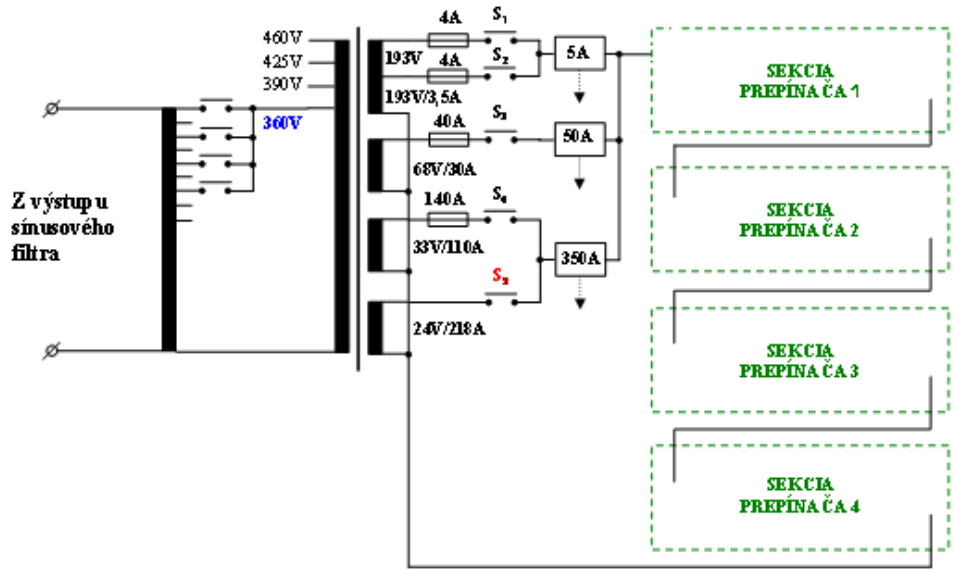
III. VÝKONOVÁ ČASŤ TESTERA TEPELNEJ SPÚŠTE ISTIČOV

Pre výkonovú časť testera hodinových skúšok bola prijatá alternatíva riešenia s jedným meničom pre napájanie všetkých testovacích sekcií. Menič napája sekcie cez odbočkový autotransformátor a hlavný tranformátor. Bloková schéma napájacej výkonovej časti testera je na Obr.1. Výstupné napätie je generované meničom, riadeným signálovým procesorom a schopným komunikovať s riadiacim systémom pomocou RS232 linky. Toto napätie je filtrované sínusovým filtrom. Sínusový filter zabezpečí napájanie transformátorovej sekcie testera variabilným napätím (a prúdom) s takmer ideálnym neskresleným priebehom a presne udržiavanou frekvenciou 50 Hz.



Obr.1 Výkonová časť testera tepelnej spúšte ističov

Transformátorová sekcia, uvedená na Obr.2 je navrhnutá tak, aby umožnila čo najlepšie nastavenie východiskových napäťových pomerov pred zahájením testu. Vzhľadom na extrémne veľký rozsah impedancií ističov s menovitými prúdmi 0,2 až 125A, ktorý je navyše znásobený požiadavkou na vykonávanie testov 1 až 4 pólových ističov, je nutné meniť napätie na testovanej skupine ističov vo veľmi širokom rozsahu.



Obr.2 Transformátorová časť výkonového zdroja

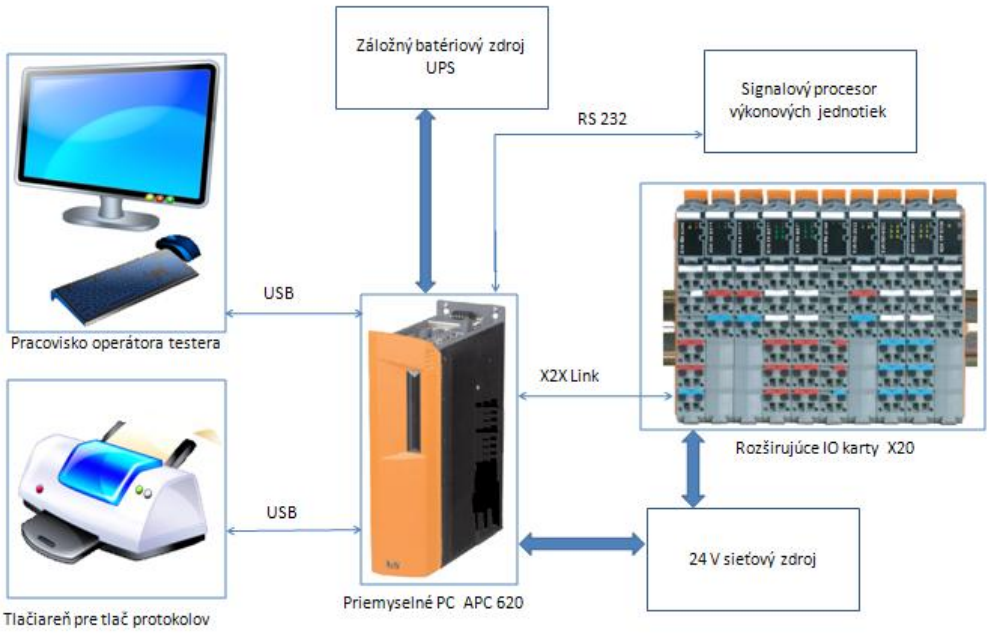
Využitím stupňovitej predregulácie napätia regulačným transformátorom sa dosiahne lepšie výkonové využite hlavného transformátora aj meniča. Prispôsobenie napäťových a prúdových rozsahov celému vyrábanému radu ističov sa uskutočňuje prepínaním sekundárnych sekcií hlavného transformátora. Uvedenými spôsobmi prednastavenia napäťových pomerov sa znižujú nároky na regulačný rozsah meniča, čo prispieva k presnosti regulácie nastavených testovacích prúdov.

IV. RIADIACI SYSTÉM PRE TESTER TEPELNEJ SPÚŠTE ISTIČOV

Funkciou riadiaceho systému testera ističov je automatizovať vykonávanie testu skupiny ističov a zaevidovať výsledky testov vo zvolenej forme pre ďalšie spracovanie. Riadiaci systém zadáva príkazy a prijíma informácie od signálového procesora, ktorý riadi výkonové jednotky testera, ďalej prijíma a spracováva povely od obsluhy testera a nakoniec spracováva a archivuje, prípadne aj tlačí údaje o priebehu jednotlivých testov vo forme protokolov.

A. Technické prostriedky RS (hardvér)

Bloková schéma riadiaceho systému testovacieho pracoviska tepelnej spúšte ističov je uvedená na Obr.3.



Obr.3 Riadiaci systém tepelnej spúšte ističov

Centrom riadiaceho systému je priemyselné PC firmy B&R, ktoré riadi testovanie ističov, a na ktorom sú ukladané všetky údaje z testov. Toto PC je zálohované batériovým záložným zdrojom. PC je vybavené štandardnými komunikačnými linkami USB, RS232 a Ethernet. Ďalej obsahuje špeciálnu komunikačnú kartu pre priemyselnú zbernicu X2X Link pre komunikáciu s rozširujúcimi IO kartami. Samotné PC a IO moduly sú napájané nezávislým 24 V zdrojom, ktorý súčasne slúži na priebežné dobíjanie batérie UPS. Operátor komunikuje so systémom iba pomocou štandardnej myši a LCD displeja, systém však umožňuje v prípade servisu aj pripojenie štandardnej PC klávesnice cez USB port. V prípade potreby vytlačenia protokolov je možné na PC pripojiť štandardnú tlačiareň s USB rozhraním. K dispozícii pre budúce rozširovanie systému je aj Ethernet rozhranie, ktoré je možné využiť napr. pre diaľkový prenos protokolov po lokálnej podnikovej sieti alebo cez internet.

B. Programové prostriedky RS (softvér)

Priemyselné PC pracuje pod operačným systémom Windows 11 a obsahuje:

- Systémové softvérové prostriedky
- Aplikačné programové vybavenie APV

Základným systémovým programom testera je runtime program AR010, ktorý vykonáva činnosť programovateľného automatu, riadiaceho priebeh testovania a ukladania údajov na HDD. Pre vizualizáciu procesu testovania a nastavovanie parametrov testovania obsluhou slúži program VNC Viewer4, komunikujúci s AR010 cez virtuálnu ethernetovú sieť. Pretože PC pracuje pod operačným systémom Windows, pre čítanie, úpravu a spracovanie protokolov je možné použiť

štandardné programy balíka OffIce, ako sú Word, Excel a Access.

Medzi dôležité vlastnosti testovacieho pracoviska bude patriť:

- možnosť súčasného testu 6 ističov (tepelnej alebo magnetickej spúšte),
- meranie prúdu v rozsahu 0,5 A-500 A (tepelná spúšť) a 1 A-2 500 A (magnetická spúšť),
- meranie počtu cyklov s možnosťou nastavenia časov,
- meranie času otvárania a zatvárania ističov,
- nastavovanie parametrov podľa katalógu ističov,
- možnosť archivovania výsledkov testov,
- možnosť generovania protokolov testov podľa navolenej štruktúry,
- schopnosť identifikácie poruchových stavov testovaných ističov,
- pokročilé HMI s možnosťou pripojenia na internet,
- možnosť jeho pripojenia do plne automatizovanej výrobnej prevádzky.

Realizáciou navrhnutého testovacieho pracoviska tepelnej spúšte ističov je možné očakávať nasledujúce prínosy pre prax:

- Výrazné ekonomické úspory pre domáceho veľkovýrobcu ističov.
- Úspory elektrickej energie vzhľadom na použitie kvalitných výkonových súčiastok a zariadení.
- Zvýšenie alebo minimálne udržanie zamestnanosti v rámci menej rozvinutého regiónu na východe Slovenska.
- Zvýšenie konkurencieschopnosti slovenského výrobcu ističov.
- Dodatočné ekonomické prínosy pri možnosti výroby a dodávok testovacích pracovísk pre zahraničné subjekty.
- Predpokladané zvýšenie výroby vzhľadom na aktuálny rozvoj elektromobility a s ňou súvisiacich nabíjacích staníc.

V. Záver

V tomto článku je stručne popísaná koncepcia inteligentného zdroja pre testovanie tepelnej spúšte ističov. Na základe analýzy požiadaviek na takýto zdroj bude postavený univerzálny voľne programovateľný nízkoenergetický zdroj, určený na prúdové a napäťové testy najrôznejších elektrotechnických zariadení, ktorý bude použitý priamo v prevádzke podniku SEZ Krompachy a.s.

Presné testovanie a nastavovanie parametrov ističov podľa ich vypínacej charakteristiky zvyšuje ich konkurencieschopnosť nielen na Slovensku, ale ja na európskom trhu, kam sa plánujú vyvážať, čím sa podporí trvalo udržateľný rozvoj spoločnosti SEZ Krompachy a.s. a celého regiónu.

Poďakovanie

Článok bol publikovaný s podporou projektu APVV-19-0210.

Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00018.

Literatúra

[1] C. H. Flurscheim: Development of circuit breakers. In: Power Circuit Breaker Theory and Design. London: IET, p. 1-19, 1982. https://doi.org/10.1049/PBPO001E_ch1.

[2] E. Aherzadeh, H. Radmanesh, S. Javadi, et al. „Circuit breakers in HVDC systems: state-of-the-art review and future trends“, Prot Control Mod Power Syst vol.8, no.38, 2023. https://doi.org/10.1186/s41601-023-00304-y.

[3] S. M. Sanzad Lumen, R. Kannan and N. Z. Yahaya, "DC Circuit Breaker: A Comprehensive Review of Solid State Topologies," *2020 IEEE International Conference on Power and Energy (PECon)*, Penang, Malaysia, 2020, pp. 1-6, doi: 10.1109/PECon48942.2020.9314300.

[4] S. Rahimpour, O. Husev, D. Vinnikov, „Design and Analysis of a DC Solid-State Circuit Breaker for Residential Energy Router Application“, *Energies, vol.15*, 9434, 2022. https://doi.org/10.3390/en15249434.

[5] M. Moradian, T.T. Lie, K. Gunawardane, „DC Circuit Breaker Evolution, Design, and Analysis, *Energies, vol.16,* 6130, 2023. https://doi.org/10.3390/en16176130.

[6] B. Perea-Mena, J.A. Valencia-Velasquez, J.M. López-Lezama, J.B. Cano-Quintero, N. Muñoz-Galeano, „ Circuit Breakers in Low- and Medium-Voltage DC Microgrids for Protection against Short-Circuit Electrical Faults: Evolution and Future Challenges, *Appl. Sci., vol.12, 15,* 2022. https://doi.org/10.3390/app12010015.

Digital Twin of a Humanoid Robot in a Virtual Reality Environment

[1]*Branislav SOBOTA,* [2]*Denis REDAJ,* [3]*Štefan KOREČKO*

[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* — This article presents a prototype of a digital twin for the pre-existing physical robot NICO (Neuro-Inspired Companion) within a virtual environment. In this work we seek to develop a tool that allows the user to interact with a digital twin of the robot NICO for conducting experiments without the need to directly manipulate with the physical model of the robot. Efforts are made to implement features that would empower the user to experiment with virtual objects in the virtual environment with inverse kinematics implementation as well as to provide the necessary functionalities to save and load movements of the robot digital twin.**

***Keywords* — humanoid robot, virtual reality, digital twin, inverse kinematics**

I. INTRODUCTION

Robots and robotics have become an integral part of modern life. Although the primary deployment of robots is in industry, there is a specialized field of robotics that focuses on humanoid robots. These robots are intended to resemble humans in both appearance and behavior and they are expected to become part of everyday human life in the future. Therefore, such robots must be capable of successfully communicating with humans and assisting them in various situations. Consequently, it is necessary to address human-robot interaction (HRI) [1]. Over the past decade, significant progress in HRI has been achieved, largely due to advancements in artificial intelligence (AI). Standard human interaction paradigms have thus been extended to include interactions between humans and robots. This interaction encompasses three "co-actions": cooperation, communication, and collaboration. Early humanoid robots were designed also as credible, controllable tools for studying social interactions [2], applicable to various tasks and contexts [3]. Additionally, humanoid robots must be able to recognize human emotions and make decisions based on them (e.g. by using mentioned AI). They will need to know when a human is happy, when he is afraid or in pain. This part is very difficult from a psychological point of view, because not all people express emotions in the same way, and sometimes we don't even know if a given human is happy, in pain or stressed until we know the cause of his expression. Transparent HRI should therefore support the ability to infer the intentions of the interaction partner (a common feature of human communication), leading to the development of theory of mind, which involves a mentalizing network [4]. Equally important is ensuring the safety of robot operations for human partners [5].

Such defined activities must also be included in the process of robot "learning". However, there remains a probability of damage to the robot's components or the execution of an action that could lead to injury or even endanger a human interacting with the robot. In this context, virtual reality technologies [6] and digital twins (DT) [7] of robots appear to be highly beneficial. This has led to the development of tools, primarily in the form of simulators operating within virtual environments. A robot's digital twin can be imported into these simulators, allowing e.g. its movements to be tested without posing any safety risks. The most commonly used applications for robot simulation include e.g. *CoppeliaSim* [8], *Gazebo* (`https://gazebosim.org/home`), or the *Unity engine* (`https://unity.com/download`). Proficiency in and use of these tools can contribute significantly to the development of more advanced and efficient robotic systems, including humanoid robots.

The field of DT technologies is already being utilized in practice by a vast amount of companies seeking to reduce initial costs for prototyping as well as maintaining an organized and efficient real-time monitoring environment for machines and facilities during their lifecycle [9][10]. Through this research they seek to apply the technology paradigm of digital twins onto the real model of a pre-existing physical robot to observe the benefits this approach provides, while also

creating a product that would enable the user to interact with the robot in a non-destructive environment.

The home laboratory of the authors (LIRKIS, Laboratory of Intelligent inteRfaces of Communication and Information Systems, Department of Computers and Informatics, FEEI TU Košice) participates also in the project of trustworthy robot-human interaction (cooperating organizations: the Institute of Measurement Science of Slovak Academy of Sciences (SAS) and the Centre for Cognitive Science Comenius University of Bratislava (CCS)). The project uses virtual reality (VR) technologies for testing/simulation of the humanoid robot NICO (Neuro-Inspired Companion [11][12][13], available in CCS (Fig. 1)).

Fig. 1 Real NICO Robot (leg-less version available in CCS)

II. DIGITAL TWINS AND VIRTUAL REALITY TECHNOLOGIES

Digital Twin (DT) is a virtual representation (virtual model) of a physical object or process that mirrors its real-world counterpart in a digital (virtual) environment [9]. This virtual model is continuously updated with real-time data from sensors, Internet of Things (IoT) devices, and other sources to reflect the current state and behavior of the physical system. Digital twin:

- enable monitoring, analysis, and simulation of physical systems, allowing for predictive maintenance, performance optimization, and scenario testing,
- serve as a bridge between the physical and digital worlds,
- provide insights for decision-making,
- enhance operational efficiency,
- reduce downtime, and
- support innovation across all sectors.

The concept and idea of DT were developed to monitor, analyse, and simulate physical objects in a virtual environment [7]. The motivation for this development arose from damage to a robotic component during testing, which disrupted its lifecycle. By linking physical systems with data collection, computational platforms, and communication networks, digital twins enable the execution of identical testing cycles without any risk to the physical model. After testing motion operations on the virtual object, operators can estimate the behavior of the real counterpart, minimizing the risk of damage.

As stated in [14], a digital twin consists of three main components:

a) *Physical Reality* - Its key components are the Physical System (PS) and the Physical Environment (PE). The PS represents a specific part of physical reality selected for modeling and it includes all relevant entities forming a unified whole. It serves as the central focus of the DT, capturing the essence of the replicated physical entity or system. The PE comprises external factors that influence chosen the PS. Understanding and incorporating these external influences is key to accurate modeling and simulation within the DT environment. Physical processes refer to the functions performed by the PS, which lead to changes in its state. These processes govern the behavior and performance of the PS. Their accurate modeling is essential for the DT to reflect real-world dynamics.

b) *Virtual Representation* - Its key components are the Virtual System (VS) and the Virtual Environment (VE) (see Fig. 2). The VS contains data models that retain all variables/parameters describing reality at a chosen level of abstraction. These data models provide a structured representation of the attributes, properties, and states of the PS. In addition to data models, the VS includes computational behavioral models that describe how variables/parameters interact and change over time. These models are often based on physical laws or data-driven approaches that express relationships among various variables/parameters. The VE digitally replicates the PE, allowing simulations and analyses to be performed in a controlled virtual space. This digital environment enables virtual representation to interact with external factors and accurately simulate real-world scenarios.

Virtual processes describe how the VS is expressed at a selected level of abstraction. These processes typically take the form of computational models that simulate the underlying physical processes. By modeling how the PS undergoes state transitions, insights can be generated to support decision-making and control processes, potentially enhanced by AI.

c) *Information Interconnection* - This component ensures the integration and synchronization between the physical and virtual components, enabling real-time data exchange and interaction between both previous components.

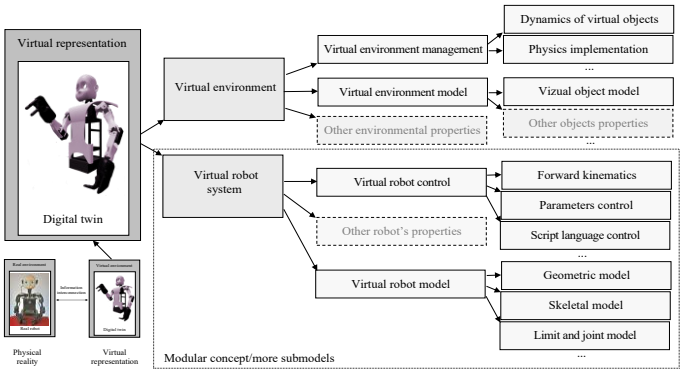


Fig. 2 Example of simplified virtual representation structure of humanoid robot digital twin [6]

One of the significant challenges when working with humanoid robots is the risk of damaging their physical components during experimentation and testing. In addition, repairing or replacing damaged parts can be time-consuming and expensive, potentially delaying research and development efforts. The testing process of these robots often involves pushing the boundaries of their capabilities, exploring new behaviors, and interacting with dynamic environments. Furthermore, the iterative nature of robotics development requires frequent testing and refinement of control algorithms and behaviors. To mitigate these risks, it is beneficial, as mentioned earlier, to utilize testing/simulation systems that allow for thorough verification of scripts and robot behaviors before deployment on physical robots. The technologies of digital twins and virtual reality, as previously mentioned, offer a potential solution to this challenge. The part of these technologies implementation within mentioned project is described in next subchapters.

III. System Design and Implementation

A. Model and Environment

The environment was created in Unity Engine and uses simple primitives configured with physics colliders that would allow the imported model of NICO to perform experiments involving interaction with objects in the scene. The intent was to create a simple controlled environment to demonstrate the capabilities of the model during various stages of development. To create a robot DT, in addition to the parameters shown in Fig. 2, it is necessary to take into account the following minimum criteria and functions that are considered critical for our implementation:

- The DT must be able to lift up a virtual 3D object from a virtual table.
- The DT must be controllable by the user by using the intuitive user interface.
- The program must generate code in DSL that, when entered into a text field, would be able to put the virtual robot into a specified pose.
- The DT must be able to execute pre-created commands and they sequences.

The robot DT is positioned upon a plane primitive, which contains a table in the form of a cylinder primitive and a sphere placed on top of the table (Fig. 3). The table and sphere both possess physics colliders, allowing them to be used for experiments involving demonstrations of NICO's dexterity and precision.

The model of NICO was created in authors' home laboratory. The rigging of the robot was implemented using the 3D modeling and animation tool Blender. The armature (skeleton) was created in a way so that the joints mirror the limited degrees of freedom (DOF) of the real robot. The DOF on each joint of the robot were manually adjusted in Unity using an individual script to lock each bone along a certain axis.

Within Unity, individual colliders are set using several primitive shapes per part of various sizes to account for the non-primitive topology of the model. The collision capable parts of the robot are mostly the hands of the robot, on account that the hands are the primary point of contact with other physical objects like the sphere.

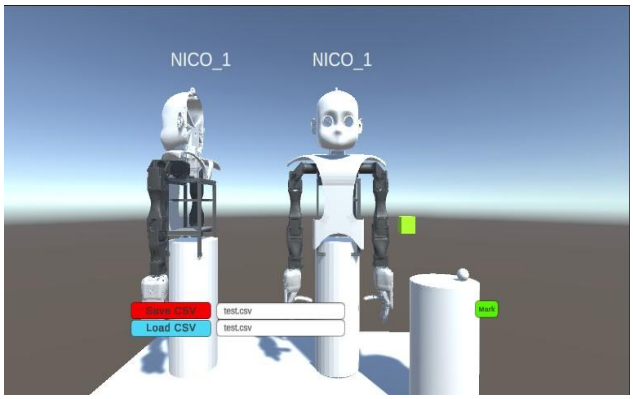


Fig. 3 Model of humanoid robot NICO in a virtual 3D environment.

B. User Interface

To ensure a correct and effective interaction between the user and model, a user interface (see bottom of Fig. 3) was developed to allow the user ample interaction possibilities with the robot DT and its VE. The user's primary form of interaction with the VE is using of mouse and keyboard. The user's camera moves them along the 2D horizontal axes using the standard *WASD* keys, while also permitting the user to move along the vertical axis using the keys O and P. Right-clicking on an empty piece of the canvas and moving the mouse moves the view of the user using inverted controls.

A script exists that monitors using a ray-caster method whether the mouse is hovering on a certain part of the robot model or whether the mouse clicked on part of the robot, at which point the robot's part is either colored green if hovered over or red if clicked on.

C. CSV Saving and Loading

The user interface provides the users with the option to create and read *.csv* files with rotation angles for named joints. The *CSV* format was selected due to its universal application in most libraries dedicated to the robot NICO [11] as well as its easy readability and minimal overhead. The user is presented with the buttons labelled "*Save CSV*" and "*Load CSV*" (see bottom of Fig. 3), which will either load or save files from the desktop with a name identical to the name within the input fields. By default the names within the inputs are "*test.csv*" for both options. These can be changed by the user at any time.

The current configuration permits files to only be loaded from and saved to the user's desktop as the base, but through adding forward slashes to the file path in the input boxes, the user can load and save files to their desired directories. The independent input fields provide flexibility during run-time allowing swift and efficient changes to the file names. For example, if the user likes a certain configuration the robot model is in, they can save the same pose configuration into a separate file under a descriptive name. This goes for loading as well, where the user can load multiple csv files after one another to construct a custom sequence of movements.

The "*Mark*" button on the right of the user interface panel saves the current joint rotations of the robot model into a sequence. The user can mark multiple joint configurations into a sequence, after which if they press the "*Save CSV*" button, they can save the entire sequence into a *.csv* file of their choice.

The angles are acquired by calling a function that extracts the directional vectors between a bone and its parent and calculates the angle they form.

D. Visual Feedback

For the purposes of experimentation, such as gauging distance or seeing if the robots limbs are in view, the user may require to see the visual feedback the robot is receiving.

By toggling the key "*t*" on the keyboard, the user can look through the eyes of the robot DT and perceive what it sees. This is done through implementing two cameras into the model of the robot and situating them within where the eye sockets would be, one for each eye. The view from each eye is then projected onto two separate planes in front of the user (Fig. 4). These two planes represent the visual input for possible stereoscopic display in VR headset.

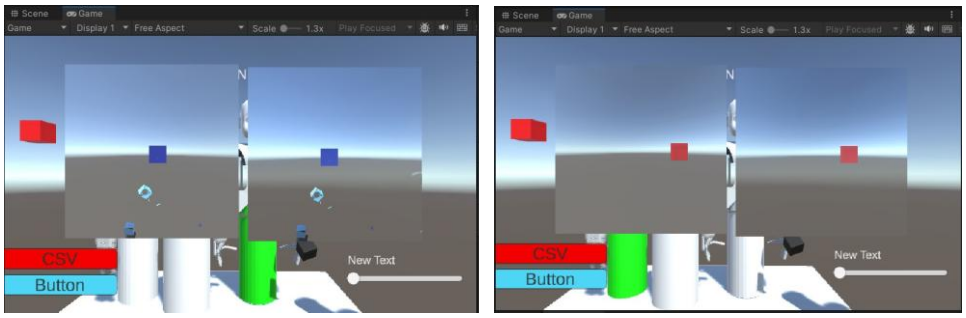


Fig. 4 Visual (stereoscopic) feedback from models one and two. Model one is perceiving the blue cube while model two is perceiving the red cube.

E. Dehierarchization

Many systems utilizing kinematic chains require a complex model (DT) to have all its parts on the same level of hierarchy. Nested objects are difficult to adjust properly due to the increasing need to convert their local positions and rotations to their world equivalents. The complexity increases with the depth of nesting, therefore a necessary step in the initialization process is to take a full model and place all parts of the model onto the same level within the object hierarchy.

Information about the hierarchy still needs to be retained, therefore a special structure called *userdata* is created to retain important information about the object, like its own name, its parent within the hierarchy and the specific robot the part belongs to, in the case that there are multiple robot models in the scene.

F. Inverse Kinematics

Inverse kinematics (IK) is a calculation method in robotics and animation used to calculate the configuration of joints needed to place a manipulated part, also known as an end effector, into a desired position. The work implements an IK chain system where before initialization, the user selects a set of bones in a hierarchy from which they want to create an IK chain. The chain must contain only bones which have a direct relation of child-parent and must be organized from the end effector to the base. The bones also require a restrict value to be set, which will constrain the bones from accepting rotations along specific axes.

During runtime, once the end effector is moved for the first time, a set of helper bones is spawned that mimics the given IK chain and using a process consisting of multiple iterations of backward and forward passes attempts to converge upon an achievable IK chain that also adheres to the given constraints (Fig. 5).

The helper bones each consist of 3 primitives:

- *Primitive A* - The end of the bone. Used to gauge length of bone.
- *Primitive B* - Accounts for minor offsets and helps during the alignment process
- *Primitive C* - The base of the bone and the part that is pivoted.

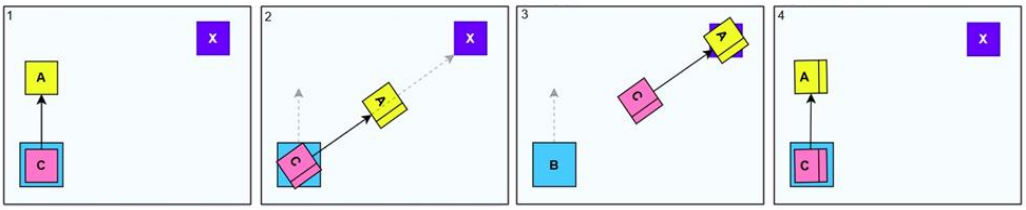


Fig. 5 Visual Simplified illustration of the functionality of a helper bone. *A* bone orients itself towards a target *X* with its *C* primitive and then proceeds to align itself using the *B* primitive as reference. The final result is a rotated helper bone that adheres to its own alignment constraints

In the scene, the end effector is represented by a green cube primitive (Fig. 6). The cube can be selected and moved around the scene, which will create and manipulate the chain of helper bones from the previously described IK chain. The solve function activates every time the end effector changes position, causing it to run a predefined amount of iterations every time the end effector changes position. By default, 13 iterations have through empirical testing yielded an optimal result while preserving computational efficiency. Past 20 iterations, the user may observe significant performance degradation.

The IK solve function starts from the end effector moving down to the base, orienting the end bone towards the end effector, after which every subsequent bone is oriented towards its previous bone. This is the backward pass and it creates an unconstrained IK chain. The IK chain then moves to the forward pass, starting from the base, where every bone is then corrected and aligned

along its restrict axis, adding the restriction (Fig. 7). After a few iterations of the solve function, the function converges on an achievable constrained IK chain from the base to the effector. In situations where the effector is out of reach of the IK chain, the chain gives the nearest possible solvable IK that points towards the end effector.

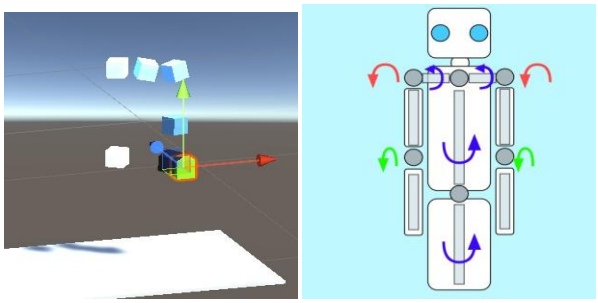


Fig. 6 Kinematic chain of the robot without the model. It represents an IK chain leading from the base to the left hand of the robot model.

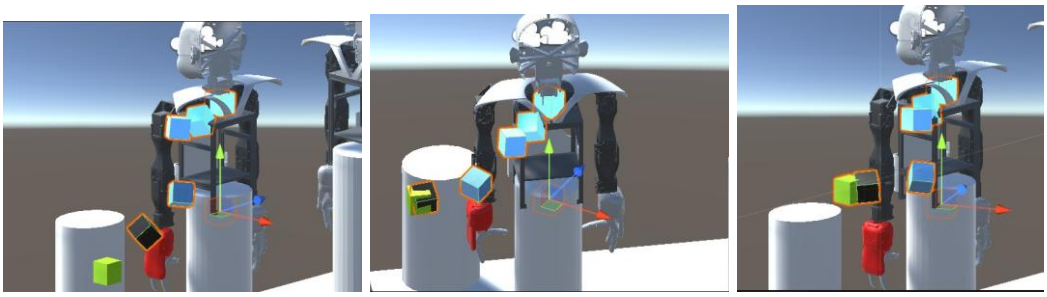


Fig. 7 A graphical representation of the generated helper bones using inverse kinematics. The green cube represents the end effector, and the highlighted cubes represent the chain of helper bones.

A non-IK option for manipulating the joints of the robot exists that allows the user to control the joints of a robot even once all the parts of the robot are dehierarchized, permitting the robot to smoothly move a joint along its permitted axis along with all the children.

IV. Conclusion

Using an iterative method has proven itself to be effective at properly estimating a kinematic chain configuration achievable even with limited degrees of freedom in a timely manner with computational efficiency. Given that the solve function is called every time the end effector changes position, the chain, under correctly set conditions, yields a highly accurate result.

The work serves as a suitable baseline for further innovation and can be improved on by implementing more generalization of the scripts to make it more flexible and reusable. The need for manual configuration during and before experimentation creates unnecessary complications that can be amended by implementing automated during the initialization process. The user interface provides adequate functionality for rudimentary experimentation. To facilitate smoother interaction, more UI/UX testing should be conducted to enhance the user's experience.

Overall, such a system is an excellent tool for training and testing humanoid robots in the form of their DT in a virtual reality environment. Its potential could be used in various areas, not only in the intentions of the project being solved. Therefore, it is important to pay attention to its development and improvement, so that it is even more effective and increases its impact on modern technological solutions in the field of robotics and virtual reality.

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] Goodrich M.A., Schultz A.C. (2008). Human-Robot Interaction: A Survey. Foundations and Trends in Human-Computer Interaction, 1(3), pp. 203-275.

[2] Sciutti A., Sandini G. (2017). Interacting with robots to investigate the bases of social interaction. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(12), 2295-2304.

[3] Bahishti A.A. (2017). Humanoid Robots and Human Society. Advanced Journal of Social Science, 1(1), pp. 60-63

[4] Vinanzi S., Cangelosi A., Goerick C. (2021). The collaborative mind: intention reading and trust in human-robot interaction. Iscience, 24, https://doi.org/10.1016/j.isci.2021.102130

[5] Lasota P.A., Song T., Shah J.A. (2017). A Survey of Methods for Safe Human-Robot Interaction. Now Foundations and Trends, https://doi.org/10.1561/2300000052

[6] Sobota B. et al. (2024), 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 s. 365-369, ISBN 979-8-3503-1720-6/, url: http://dx.doi.org/10.1109/sami60510.2024.10432918.

[7] Jones D., Snider C., Nassehi A., Yon J., Hicks B. (2020). Characterising the digital twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, pp. 36–52.

[8] Rooban S., Suraj S. D., Vali S. B., and Dhanush N., "Coppeliasim: Adaptable modular robot and its different locomotions simulation framework," Materials Today: Proceedings, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221478532100078X

[9] Jia W., Wang W., Zhang Z. (2022). From simple digital twin to complex digital twin Part I: A novel modeling method for multi-scale and multi-scenario digital twin. Advanced Engineering Informatics, 53. https://doi.org/10.1016/j.aei.2022.101706

[10] Katona, J. A Review of Human–Computer Interaction and Virtual Reality Research Fields in Cognitive InfoCommunications. Appl. Sci. 2021, 11, 2646.

[11] Kerzel M., Strahl E., Magg S., Navarro-Guerrero N., Heinrich S., and Wermter S. (2017): "Nico - neuro-inspired companion: A developmental humanoid robot platform for multimodal interaction," in Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Aug 2017, pp. 113–120.

[12] Kerzel M., Pekarek-Rosin T., Strahl E., Heinrich S., and Wermter S., "Teaching NICO how to grasp: An empirical study on crossmodal social interaction as a key factor for robots learning from humans," Frontiers in Neurorobotics, vol. 14, 2020. url: https://www.frontiersin.org/article/10.3389/fnbot.2020.00028

[13] Kerzel M., Allgeuer P., Strahl E., Frick N., Habekost J.-G., Eppe M., and Wermter S. (2023): "Nicol: A neuro-inspired collaborative semi-humanoid robot that bridges social interaction and reliable manipulation," IEEE access, vol. 11, pp. 123531–123542, 2023.

[14] Mahadevan E. V. S. (2021) Digital twin: Generalization, characterization and implementation. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923621000348

Generátor internetových portálov s využitím Model Driven Architecture

[1]Dominik IVANOC, [2]Branislav MADOŠ

Katedra počítačov a informatiky, Technická univerzita v Košiciach, Košice, Slovensko
[1]dominik.ivanoc@student.tuke.sk, [2]branislav.mados@tuke.sk

***Abstract* – Článok sa zaoberá problematikou modelovo riadenej architektúry, ktorá bola využitá pri implementácií webovej služby zameranej na tvorbu personalizovaných portálov. Na základe analýzy architektúr softvéru bola navrhnutá služba, ktorá umožňuje používateľom vytvárať univerzálne portály podľa vlastných požiadaviek. Služba je využiteľná v plnom rozsahu až po úspešnom autentifikovaní používateľa. Po vstupe do služby je možné v admin rozhraní spravovať a vytvárať viacero portálov s dátami. Cieľom článku bolo navrhnúť a implementovať službu vytvárajúcu personalizované portály bez prítomnosti programátora či používateľských znalostí v danej doméne. Služba prostredníctvom formulára na to určeného prijme vytvorený model. Model obsahuje dáta a konfiguračnú štruktúru nového portálu. MDA generátor na základe modelu vygeneruje personalizovaný portál, ktorý v menu manažmente zobrazuje tabuľkové entity s dátami a ponúka nástroje pre ich spracovanie a analýzu. Implementované riešenie bolo v článku otestované a vyhodnotené, čím sa overila jeho funkčnosť.**

***Kľúčové slová* – MDA architektúra, MDA generátor, personalizovaný web-portál, XML model**

I. Úvod

Dáta zohrávajú kľúčovú úlohu. Postupným rozvojom sa však zmenila ich forma a spôsob uchovávania. Kedysi to bola hlavne ústna forma, pomocou ktorej sa dáta odovzdávali z generácie na generáciu. S príchodom písma sa dáta začali zaznamenávať na rôzne materiály, ako napríklad kameň, hlinené steny alebo papier. Technologickým rozvojom sa tento spôsob uchovávania dát zmenil. V dnešnej dobe sú dáta zaznamenávané elektronicky a uchovávajú sa v počítačoch či veľkých databázových serveroch.

Cennosť dát bola kľúčovým faktorom v mnohých etapách ľudskej histórie. Čím citlivejšiu informáciu tieto dáta obsahovali, tým boli dôležitejšie a mali vyššiu hodnotu. Dôležitosť dát sa ešte prehĺbila s rozvojom informačných technológií a internetu. Niektoré dáta sú záznamy, ktoré samy o sebe nemajú žiadnu hodnotu. Hodnotu nadobúdajú až vtedy, keď sú spracované a analyzované. Vďaka spracovaniu a analýze je možné získať informácie, ktoré by inak zostali nepoznané. Dáta sú preto v dnešnej dobe jedným z najdôležitejších zdrojov s veľkým finančným potenciálom.

Motiváciou tohto článku je vytvoriť nástroj, ktorý umožní používateľom takéto dáta spravovať, analyzovať a uchovať v bezpečí pred stratou alebo zneužitím. Nakoľko každý používateľ má iné požiadavky a preferencie pre správu týchto dát, je potrebné navrhnúť a vytvoriť nástroj, ktorý bude schopný tieto požiadavky zohľadniť. Prístup MDA (angl. Model Driven Architecture) je riešením, ktoré umožní vytvárať personalizované portály pre používateľov na mieru. Funguje na princípe modelom riadených architektúr, ktoré pomocou modelu vytvárajú a generujú usporiadaný zdrojový kód do funkčných aplikácií. Tento prístup bude využitý pri tvorbe MDA generátora pre webovú službu na generovanie personalizovaných portálov.

II. Súvisiace diela architektúr softvéru

K vývoju softvéru sa pristupuje rôznymi spôsobmi. Každý z týchto spôsobov má pri tvorbe špecifické fázy a charakteristické postupy. Vývoj definuje najmä softvérová architektúra, ktorá ďalej delí vývoj na jednotlivé fázy. Architektúra softvéru tvorí základný kameň pre efektívne fungovanie softvéru. Správnym návrhom je možné eliminovať časté zmeny, znížiť finančnú či časovú náročnosť. Ako príklad sa dá uviesť obmena softvérovej platformy. Táto obmena softvérovej platformy, ale aj programovacieho jazyka môže nastať z dôvodu technologického pokroku, pri modernizácií alebo na základe požiadavky zákazníka.

Americký softvérový inžinier Robert C. Martin sa v citovanom blogu vyjadril, že: *"Dôležité rozhodnutia, ktoré robí softvérový architekt, sú tie, ktoré vám umožňujú nerobiť rozhodnutia o databáze, webovom serveri a rámcoch"* [1]. Citácia poukazuje na to, že pri správnom počiatočnom rozhodnutí o návrhu architektúry je možné predísť zbytočným úvahám o výbere konkrétnych softvérových platforiem do architektúry. Ak je softvérová architektúra navrhnutá správne dosadenie konkrétnej softvérovej platformy nebude problém, nakoľko samotná komunikácia a spojitosť medzi jednotlivými modulmi v architektúre bude vždy rovnaká.

Problematiku architektúry softvéru popisujú aj iní významní vedci svojimi prínosmi. Definície sú odlišné, no hlavnú podstatu veci zachytávajú rovnako. Rovnako tak aj Fred Brooks pripomína, že cesta k úspechu v tejto oblasti vedie cez starostlivo navrhnutú a dobre spravovanú architektúru [2]. Ako spomína autor citácie, veľmi podstatnou vecou je navrhnutý softvér ďalej spravovať a starať sa oň, či už pravidelným pridávaním nových funkcií a vlastností alebo opravami vo forme aktualizácií.

"Architektúra softvéru predstavuje abstrakciu na zvládnutie komplexity systému a slúži tiež na vytvorenie komunikácie a koordinácie medzi menšími časťami systému" [3]. Citácia popisuje architektúru softvéru ako isté abstraktné zaobalenie systému. Pod zaobalením systému sa dá rozumieť spojenie viacerých častí do jedného celku. Tento abstraktný celok poskytuje základné návrhy usporiadania a komunikácie medzi jednotlivými komponentami systému. Pomocou navrhnutej abstraktnej architektúry softvéru sa ďalej rozvíjajú jednotlivé komponenty softvéru a komunikácia medzi nimi.

A. Štýly architektúry softvéru

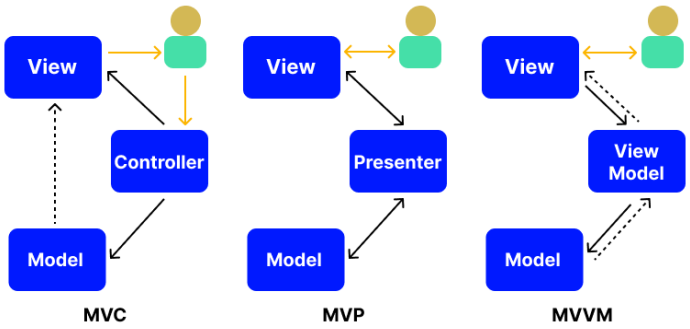
Existuje viacero typov architektúr softvéru. Líšia sa od seba najmä spôsobom, akým je architektúra navrhnutá a organizáciou jednotlivých komponentov. Nie je možné jednoznačne určiť, ktorá architektúra je lepšia, pretože každá má svoje výhody, nevýhody a unikátne využitie na konkrétne prípady. Základné typy štýlov architektúr pre vývoj softvéru sú [4]: **monolitická architektúra**, **architektúra mikro služieb**, **serverovo-orientovaná architektúra**, **udalosťami riadená architektúra**, **N-vrstvová architektúra** a **distribuovaná architektúra**.

Jednotlivé štýly softvérovej architektúry je možné aj kombinovať a používať ako podštýly pre komplexnejšiu implementáciu. Každý štýl má svoje výhody a nevýhody, ktoré je potrebné zvážiť pri výbere na konkrétnom projekte. Výber správneho štýlu architektúry môže závisieť napríklad od požiadaviek na softvér, jeho škálovateľnosti, bezpečnosti a ďalších faktorov.

B. Vzory architektúry softvéru

Zatiaľ čo predchádzajúce delenie *Štýlov architektúry softvéru* II-A popisuje najvyššiu úroveň abstrakcie, základnú štruktúru softvéru a poskytuje vysokoúrovňové rozdelenie do modulov a prepojení medzi nim, *Vzory architektúry softvéru* II-B popisujú organizáciu kódu, komponentov a celkový spôsob implementácie týchto architektonických štýlov vo vnútri softvéru.

"Každý architektonický vzor popisuje nejaký opakujúci sa problém a predstavuje jeho možné riešenie" [5]. Vzory zoskupujú opakujúce sa vlastnosti architektúry do abstraktnej predlohy a zároveň riešia časté chyby s ňou spojené. Taktiež sa zameriavajú na oddelenie používateľského rozhrania od hlavnej logiky a dát softvéru.



Obr. 1: Architektonické vzory [6]

Tri základné architektonické vzory sú znázornené na obrázku 1. Sú nimi **Model-View-Controller (MVC)**, **Model-View-Presenter (MVP)** a **Model-View-ViewModel (MVVM)**. Tieto vzory sa od

seba líšia spôsobom, akým sú navrhnuté, komunikáciou medzi jednotlivými modulmi a organizáciou komponentov aplikácie. Každý z týchto vzorov obsahuje tri moduly, ktoré sú zodpovedné za rôzne úlohy.

Vzory architektúry softvéru sa taktiež zameriavajú na oddelenie používateľského rozhrania od hlavnej logiky a dát softvéru. Toto rozdelenie zabezpečuje lepšiu kontrolu nad softvérom a obmedzovanie prístupu k údajom. Vzory taktiež dotvárajú štruktúru softvéru a správanie modulov či ich vzťahov v architektúre.

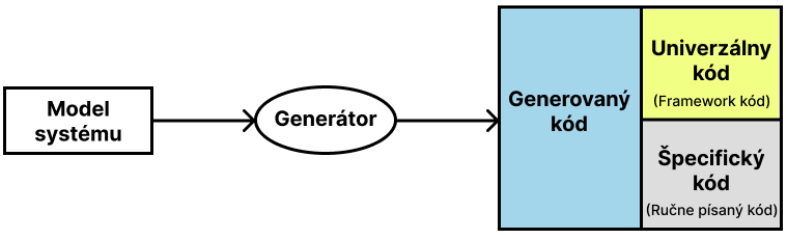
III. Model Driven Architecture (MDA)

MDA je prístup k vývoju softvéru, ktorý definuje štandardy pre modelovanie softvérových systémov a ich transformáciu na konkrétne implementácie. Uvádza sa, že MDA prístup je založený na klasickom Objektovo Orientovanom prístupe (angl. Object-Oriented Programming (OOP)) [1]. Opiera sa o jeho základy a rozširuje ho ďalšími technikami.

Ako sa uvádza v článku: *"Hlavnou myšlienkou modelovo riadenej architektúry je oddeliť biznis a aplikačnú logiku od technológií"* [7]. Prakticky to znamená, že softvér navrhnutý prostredníctvom MDA prístupu sa dokáže realizovať na rôznych platformách a v rôznych programovacích jazykoch. Tento prístup dokáže správne rozdeliť a zároveň izolovať jednotlivé časti aplikácie, ako je napríklad biznis a aplikačná logika. Ak sú časti aplikácie definované správne, je ich implementácia možná prostredníctvom širokej škály technológií. Technológie je možné medzi jednotlivými časťami aplikácie aj obmieňať a tým dosiahnuť lepšiu modularitu či prispôsobiteľnosť systému na nové požiadavky.

MDA v procese vývoja softvéru využíva modely, ako hlavné prvky pri návrhu aplikácie. *"MDA predstavuje základnú paradigmu vývoja systémov, založenú na tvorbe modelov a následných transformáciách medzi týmito modelmi"* [8]. Dá sa tvrdiť, že za pomoci abstraktných modelov sa vytvára konkrétna implementácia softvéru. Modely tvoria takzvanú kostru MDA architektúry. Dôležitou podstatou je to, aby aplikácia dokázala prostredníctvom modelov generovať funkčný a spustiteľný kód. Tento kód by mal vytvárať a prepájať logické časti aplikácie. MDA štandardy zabezpečujú konzistentnosť a interoperabilitu medzi jej komponentami.

Väčšina konfiguračných nastavení sa vykonáva v samotnom modeli pred vygenerovaním aplikácie. Táto konfigurácia môže byť zadaná buď analytikom či produktovým špecialistom na základe požiadaviek zákazníka. Po vyplnení konfigurácií sa za pomerne krátky čas dokáže generovať spustiteľný program, na ktorý by inak bolo potrebné dlhodobé plánovanie a vývoj.



Obr. 2: Architektúra modelovo riadeného generátora [9]

Generovaný kód sa vytvára v istých fázach. Tieto fázy sú spolu so samotným generátorom znázornené na obrázku 2, popisujúcom jeho jednotlivé časti. Tieto časti sa skladajú z *"modelu systému"*. Tento model obsahuje všetky potrebné konfiguračné nastavenia a štruktúru aplikácie. Ďalšou častou je samotný *"generátor"*, ktorý na základe preddefinovaných podmienok, pravidiel a konfigurácií vygeneruje z modelu spustiteľný kód aplikácie. Poslednou časťou je samotná aplikácia alebo aj *"kód"*, ktorý sa skladá z troch častí. Prvou z nich je univerzálna časť kódu, spoločného pre všetky aplikácie generované daným generátorom. Druhou časťou je generovaný kód, ktorý je vytvorený na základe modelu a konfigurácií samotným generátorom. Poslednou časťou je špecifický kód, nazývaný aj ručne písaný kód. Tento kód je možne doplniť pre každú aplikáciu samostatne na základe požiadaviek, ktoré sa nedali definovať v generátore.

IV. Navrhované riešenie

Hlavným cieľom tohto článku je navrhnúť a vytvoriť funkčné riešenie generátora internetových portálov s využitím MDA. Tento generátor bude slúžiť na vytváranie personalizovaných portálov, ktoré budú zobrazovať dáta zadané používateľmi.

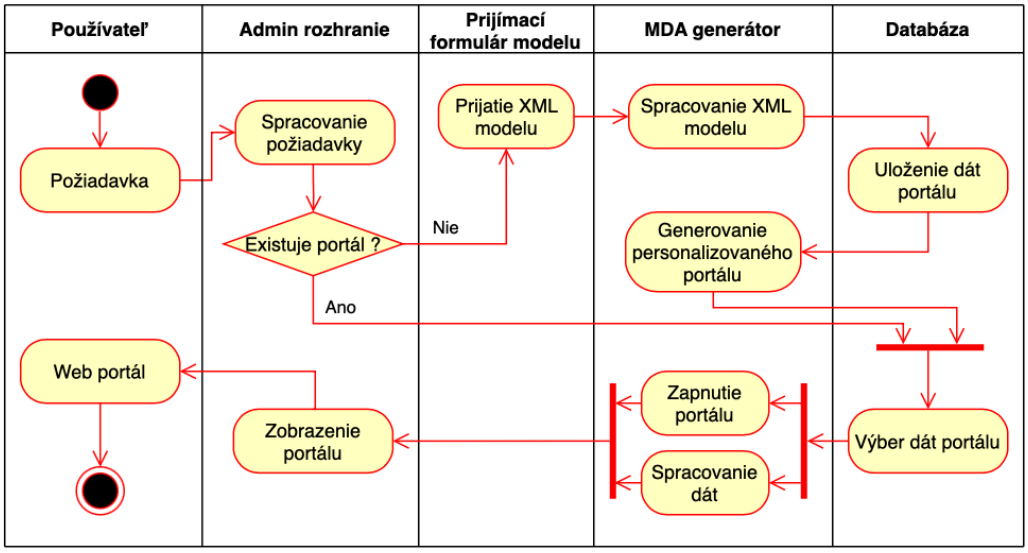
[1] https://www.geeksforgeeks.org/introduction-of-object-oriented-programming/

Základnou časťou webovej služby MDA generátora je spracovanie XML modelov a následné funkčné vygenerovanie internetových portálov na základe špecifikácií zadaných v XML modeli. Vygenerované portály budú samostatné webové aplikácie, ktoré budú schopné komunikovať s databázou a zobrazovať dáta používateľovi v prehľadnej forme. Používateľ bude mať možnosť vytvárať, upravovať a vymazať dáta v databáze, pričom všetky tieto operácie budú vykonávané prostredníctvom webu.

Pre zabezpečenie bezpečného a efektívneho využívania služby je nevyhnutné autentifikovať používateľov prostredníctvom procesu registrácie alebo prihlásenia. Tento proces je vykonávaný prostredníctvom formulára, ktorý obsahuje polia pre zadávanie prihlasovacích údajov. Autentifikácia zabezpečí, že prístup k jednotlivým dátam v databáze bude medzi používateľmi oddelený. Práva na prácu s dátami majú iba autorizované osoby, čím sa predchádza neoprávnenému prístupu a úniku citlivých dát.

A. MDA generátor personalizovaných portálov

Vytváranie personalizovaných portálov je realizované prostredníctvom používateľských XML modelov, ktoré sú prijaté formulárom na to určeným. Formulár obsahuje dve možnosti, ktoré môžu používatelia využiť na vloženie modelu. Pred samotným odoslaním dát z formulára sa model validuje regulárnymi výrazmi, ktoré kontrolujú správnosť XML formátu a modelu samotného. Po prijatí XML modelu z formulára na to určeného je model poslaný do MDA generátora na backende. Generátor je implementovaný v jazyku Java, pričom pozostáva z viacerých častí. Hlavnou časťou generátora je spracovanie modelu a jeho transformácie. Ďalšou časťou je vygenerovanie personalizovaného portálu na základe spracovaných dát z modelu. Tretia časť generátora pozostáva zo správy portálov a ich zobrazenia používateľom.



Obr. 3: UML diagram aktivít opisujúci postup spracovania portálov

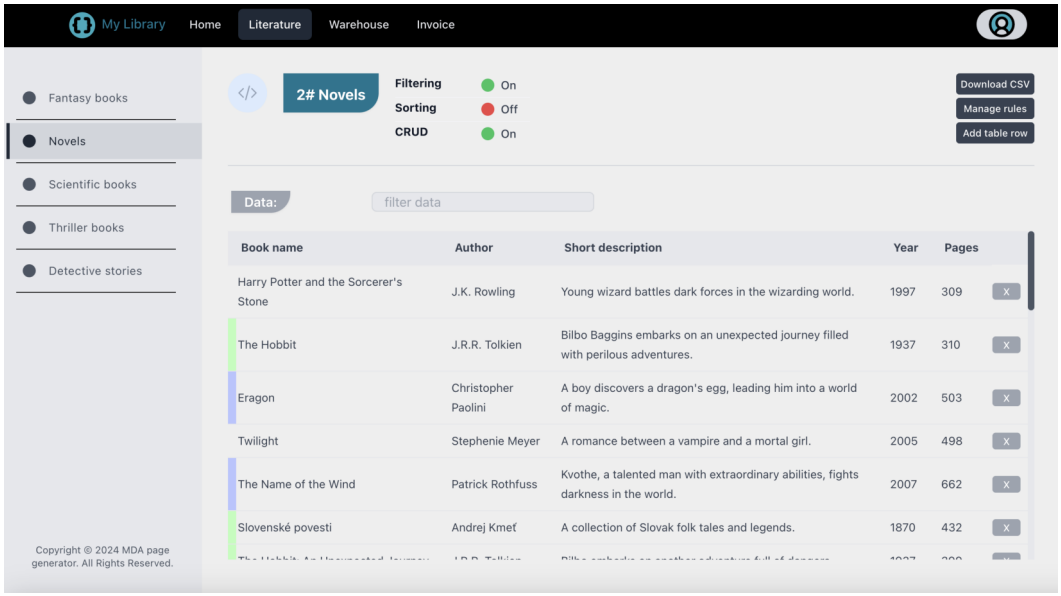
Na obrázku 3 je zobrazený UML diagram aktivít, ktorý znázorňuje proces spracovania používateľských portálov. Diagram aktivít prezentuje kroky vykonávané službou, rozdelené medzi jej funkčné časti. Požiadavka používateľa je spracovaná v admin rozhraní, a to buď v podobe generovania nového portálu alebo práce s existujúcim portálom.

Základný priebeh pri spracovaní používateľskej požiadavky je nasledovný. Používateľ, ktorý chce pracovať s už existujúcim portálom odošle požiadavku na zapnutie portálu cez admin rozhranie. MDA generátor na backende spracuje potrebné dáta a portál zapne. Ak chce používateľ vygenerovať nový portál, musí vytvoriť XML model a vložiť ho do formulára. Po spracovaní a analýze modelu MDA generátorom sa dáta uložia do databázy, z ktorej sa v priebehu generovania načítajú. Vygenerovaný portál sa zapne a je pripravený na použitie.

B. Personalizovaný web portál

Po úspešnom spracovaní modelu a vygenerovaní personalizovaného portálu je možné s týmto portálom a dátami v ňom pracovať. Webový portál je vygenerovaný MDA generátorom na základe požiadaviek používateľa zadaných v modeli. Celý proces generovania je vykonaný bez zásahu a prítomnosti programátora, nakoľko je proces plne automatizovaný.

Obrázok 4 znázorňuje ukážku vygenerovaného web portálu. Webové rozhranie portálu pozostáva z viacerých komponentov, a to hlavného menu v hornej časti, vedľajšieho menu na ľavej strane a tabuľkovej entity s dátami, ktorá je kategorizovaná v menu manažmente. Tabuľka obsahuje dáta čítane z databázy portálu. Zobrazenie je plne prispôsobiteľné a poskytuje užitočné funkcionality pre prácu s dátami. Pre tabuľkové entity sa definuje presný počet stĺpcov, ich dátové typy a názvy. Portál je plne univerzálny a prispôsobiteľný používateľským požiadavkám.



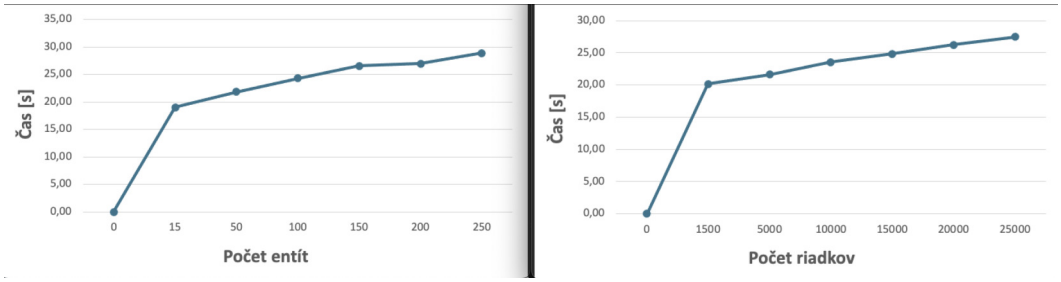
Obr. 4: Zobrazenie personalizovaného web portálu

Personalizované web portály taktiež poskytujú rôzne nástroje pre správu a analýzu dát. Základom sú prepínače zobrazené nad tabuľkou pre filtrovanie, triedenie a CRUD operácie. Vygenerované portály ponúkajú možnosť exportovania dát do CSV súboru či farbenie riadkov tabuľky na základe pravidiel a predvypočítavanie hodnôt stĺpcov tabuľky. Riešenie rovnako disponuje moderným dizajnom používateľského rozhrania či responzívnosťou, ktorá je zabezpečená použitím moderných technológií. Webová služba je unikátny nástroj s možnosťou personalizácie a prispôsobenia vlastných portálov bez nutnosti programovania.

V. Testovanie riešenia

Testovanie je dôležitou súčasťou vývoja softvéru, pretože umožňuje odhaliť chyby a zlepšiť kvalitu výsledného produktu. Testovanie sa uskutočnilo na zariadení MacBook Pro vyrobenom v roku 2020 s 2,3 GHz 4-jadrovým procesorom Intel i7 desiatej generácie a pamäťou o veľkosti 16 GB RAM.

Nakoľko je MDA generátor hlavným komponentom služby, je dôležité otestovať jeho výkonnosť pod rôznym typom zaťaženia v závislosti od veľkosti a zložitosti XML modelu. Testovaním sa overí či je generátor schopný spracovať požiadavku v prijateľnom čase a s pozitívnym výsledkom.



Obr. 5: Výsledky záťažových meraní v grafoch

Testovanie výkonnosti prebehlo v dvoch fázach. Cieľom testovania bolo zaznamenať čas generovania portálu a zároveň zvyšovať záťaž postupným vkladaním zložitejších XML modelov.

Tieto modely boli navyšované o počty tabuľkových entít alebo o riadky v tabuľkách. Záťaž bola zvyšované postupne v závislosti od druhu vykonávaného testu. Výsledky realizácie záťažových testov dokumentuje obrázok 5.

VI. Záver

Tento článok sa venoval problematike *generovania internetových portálov s využitím MDA* technológie. Jeho cieľom je nielen demonštrovať praktickú implementáciu generátora, ale aj poskytnúť webovú službu, ktorá používateľom umožní vytvárať personalizované web portály na spracovanie, uchovanie a analýzu dát v tabuľkových entitách dostupných cez menu manažment.

Na základe analýzy súčasných trendov, dôležitosti dát a architektúry softvéru bola identifikovaná potreba pre vytvorenie služby, ktorá by umožnila používateľom vytvárať univerzálne portály na mieru podľa vlastných požiadaviek. Vďaka personalizácií web portálu a automatizovanému prístupu k vývoju sa táto služba stáva efektívnou a užitočnou pre používateľov. Modelom riadená architektúra umožňuje prepojiť abstraktné modely vytvorené predstavou používateľa do konkrétnej implementácie architektúry, čím sa vývoj portálu zjednoduší. Tento prístup umožňuje používateľovi vytvárať a spravovať svoje vlastné portály bez nutnosti zásahu programátora.

Nakoľko ide o modelom riadený prístup, bolo nutné vytvoriť MDA generátor, ktorý bude obsahovať potrebnú konfiguráciu na spracovanie modelu. Vytváranie portálov prebieha prostredníctvom implementovaného MDA generátora, ktorý príjme od používateľa nakonfigurovaný model vo forme XML štruktúry.

Navrhované riešenie svojím zameraním pomôže zjednodušiť správu a analýzu dát v tabuľkových entitách. Tabuľkové entity budú intuitívne na použitie prostredníctvom implementovaných komponentov pre filtrovanie, triedenie a CRUD operácií dát. Táto hlavná funkcionalita je doplnená o pomocné funkcie, ktoré dopĺňajú celkovú funkčnosť a vytvárajú z používateľského portálu jeden plnohodnotný nástroj pre správu dát. Implementované riešenie napĺňa očakávania a požiadavky. Vytvorený MDA generátor je schopný spracovať model a vygenerovať personalizovaný portál podľa špecifikácií modelu. Úspešným otestovaním služby sa potvrdila spoľahlivosť a správnosť implementácie, ktorá prináša nové a unikátne riešenie.

Pridanou hodnotou je to, že samotní používatelia nemusia disponovať programátorskými znalosťami. Jedinou podmienkou na používanie personalizovaného portálu, je správne vytvorenie a nakonfigurovanie XML modelu tak, aby obsahoval všetky potrebné informácie. Po validácii a spracovaní modelu MDA generátorom sa zostaví portál na mieru.

Výsledky riešenia poskytujú pevný základ pre ďalší rozvoj výskumu v modelovo riadenom vývoji internetových portálov a predstavujú funkčné MDA riešenie v dynamickom prostredí moderných systémov. Moderné technológie Java a React zabezpečujú načítavanie a prácu s dátami bez obnovenia stránky, čo zvyšuje používateľský komfort.

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] R. C. Martin, “The clean code blog,” 4.1.2016. [Online], citované: 31.5.2025, dostupné: http://blog.cleancoder.com/uncle-bob/2016/01/04/ALittleArchitecture.html
[2] F. P. Brooks, “The mythical man-month, essays on software engineering,” *Addison-Wesley Publishing Company*, p. 212, 1975, ISBN 0-201-00650-2.
[3] B. D. Holec, “Návrh architektúry pre univerzálnu portálovú pltformu,” 2021. [Online], citované: 31.5.2025, dostupné: https://is.muni.cz/th/lcug5/DP_Holec.pdf
[4] D. M. Milanovič, “Most common software architecture styles,” 22.12.2023. [Online], citované: 31.5.2025, dostupné: https://medium.com/@techworldwithmilan/most-common-software-architecture-styles-86881d779683
[5] B. M. Kolomazník, “Architektonické vzory v moderním vývoji pro ios,” 28.7.2020. [Online], citované: 31.5.2025, dostupné: https://vskp.vse.cz/85017_architektonicke-vzory-v-modernim-vyvoji-pro-ios??page=1
[6] A. Sinhal, “Obr. 2.7: Mvc, mvp and mvvm design pattern,” 3.1.2017. [Online], citované: 31.5.2025, dostupné: https://medium.com/@ankit.sinhal/mvc-mvp-and-mvvm-design-pattern-6e169567bbad
[7] O. M. Group, “Mda® - the architecture of choice for a changing world,” 2024. [Online], citované: 31.5.2025, dostupné: https://www.omg.org/mda/
[8] B. Bučko, K. Zábovská, and M. Zábovský, “Modelom riadený vývoj informačných systémov,” *Ikaros, elektronický časopis o informační společnosti*, 2019, ročník 23 číslo 3, https://ikaros.cz/modelom-riadeny-vyvoj-informacnych-systemov.
[9] S. Chodarev, “Obr. 2.8: Metaprogramovanie - generovanie kódu,” 2024. [Online], citované: 31.5.2025, dostupné: https://kurzy.kpi.fei.tuke.sk/meta/slides/7-generovanie.html?full#20

Software for testing and managing HTTP requests

[1*]*Tomáš VARGOŠKO,* [2]*Miroslav MURIN,* [3]*Ondrej KAINZ,*
[4]*Miroslav MICHALKO*

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

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

Abstract **– Testing is now an essential part of any software development. The aim of this work is to design and develop a compact software to manage and test the requests of hypertext transfer protocol. Initially, the thesis focuses on the market analysis of existing solutions in the field of hypertext transfer protocol request testing. From the analysis, the shortcomings of the existing solutions and their possible improvements were then deduced. Low fidelity screens were used to design the desktop application and its user interface. The resulting software was then developed from the initial design. The resulting desktop application contains various functions such as login, registration, project dialing, writing and managing tests in the project, managing users on the project, real-time communication between users within the project, and user account settings. Testing itself offers various options such as adding a request body, predefined headers, and the ability to create custom headers. The software has an inbuilt feature to automatically evaluate the correctness of tests based on the expected return response.**

Keywords **– API, HTTP, Java, JavaFX, Software, Spring Boot, Testing**

I. Introduction

Software development is one of the most desirable professions nowadays. People cannot imagine everyday life without the various applications that make it possible. Many factors affect the quality of the final product during the development process. One of these is testing HTTP requests, often referred to as application programming interface (API) testing.

Nowadays, there are many tools that facilitate software development. Without them, the creation of applications would be extremely challenging for companies and developers alike. When it comes to testing HTTP requests, there are many popular solutions that are often used, but even these suffer from shortcomings such as poor organisation of work on certain projects. This leads to a situation where teams are forced to use another external application or solution to organise their work on a project. Existing solutions offer the option of creating organisations and collections, but this is very limited in the free versions, and better solutions can cost companies more money, which can be complicated.

The result of the problem is that developers have to use lots of different applications for a single project. These software products can require training for individual teams. This is another thing that companies need to think about to make sure that teams work well together when they are developing software. In a case that users want to chat with each other in real-time, for example, to agree on work, discuss various errors and critical situations that need to be tested, they need to use a different tool. Some tools for creating tests for HTTP requests offer a wide number of features. This can take time to learn, and in the worst case, training may be needed.

II. API testing software

Nowadays, there are many ways to test HTTP requests. Examples include web applications or programs that users install on their devices. Nowadays, this problem is well covered and every tool offers something new that other tools do not. But some tools just copy others and generally have nothing new to offer. Most of these tools have user interfaces that are not very easy to use,

which can make them hard to use. Most user interfaces are very similar, and this has good and bad points. If users know how to use one tool, they will find the other tool easy to use too. Most of these tools do not really help with teamwork and interaction when testing HTTP requests. They do not have features like real-time communication for teams or projects. This problem is usually solved by using third-party applications. This chapter looks at the different ways of testing HTTP requests and the advantages and disadvantages of each.

A. Postman

Postman is now one of the most popular tools for managing and testing HTTP requests. It is intended for developers and people who test backend services. There is a free version and three paid versions. It supports testing request methods like POST, PUT, GET, DELETE, PATCH and OPTIONS. Once a user has logged in, the application's user interface changes and various new functions are offered, making the user interface more complicated. It supports adding of testing headers, authorization, body and various parameters directly by using user interface. Postman lets users add notes to written tests, but it does not allow team members to chat with each other. Even though it does not have better team communication, it is a great tool for testing and managing HTTP requests.

B. Insomnia

Insomnia is an application developed by company King Inc. It is a multi-platform tool and one of the biggest advantage of Insomnia is that it is Open-Source. It works in a very similar way to Postman, especially when it comes to creating collections and grouping tests inside those collections. It also allows for scheduling the sending of requests as part of adding tests. After logging into Insomnia, the user interface changes and becomes less simple, like it does on Postman. The free version of Insomnia does not have the same team-wide collection sharing as Postman.

C. HoppSotch

HoppSotch is a multi-platform Open-Source application developed for testing API. It has a web application that can be used straight away, as well as a desktop version. It lets developers test real-time communication using web sockets, which is easier than previous methods because this function was hidden in the user interface and users had to search for it using the search bar. Since this software has a web application, it has its disadvantages. Potential problem is that the server on which the application is running may be located in a different domain to that of the application which the user is about to test. The disadvantage of this approach is that it does not permit the testing of an application that is running on a local device. This is due to the issue of cross-origin resource sharing. A more effective approach would be to download and install the desktop version, or alternatively, to disable cross-origin resource sharing on a tested backend. According to [1] comparison, HoppSotch does not contain as many features as Postman does.

D. SoapUI

Another existing solution for testing HTTP requests is Open-Source tool called SoapUI developed by SmartBear Software. According to [2] this software is developed in Java and Swing. This tool is primary focused on Simple Object Access Protocol (SOAP) and Representational State Transfer (REST) testing. The software under consideration features a built-in function for project creation, which is comparable to the collection creation functionality in previous iterations. One of the biggest disadvantage is that it does not conatain any option of inviting users to created project. For this reason there is not possible interaction between users. The user interface design is reminiscent of older desktop application designs. It has also weaker responsivness. During project creation users are requested to enter some kind of address and new users do not have to know what address should they enter. It is not a bad tool but there are better alternatives to use.

III. Application creation tools

During the development of software, it is important to select the correct tools for the job. When making this choice, it is also important to consider application management and future updates. This section focuses on the existing programming languages and frameworks used for software development.

A. Java programming language

Java is an object-oriented language which is according to [3] derived from the C language. It is primarily used for developing desktop and mobile applications. In the context of desktop applications Java has several frameworks that support the creation of graphical components. One of these frameworks is Java Swing, which, according to [4] makes it easier to create difficult graphical applications. Authors in [4] further state that it contains predefined classes that facilitate the creation of graphical components and their properties. A newer and more modern framework for desktop application creation is JavaFX. It was released according to [5] in 2008 as a script language and in 2011 it was written into Java.

B. Python programming language

According to book [6] Python is an Open-Source programming language that is used in areas of user interface development, numeric programming, and various systems. This programming language is currently one of the most widely used programming languages, with a simple syntax and fast software development. It is possible to develop desktop applications in Python using different frameworks. One of the frameworks is Tkinter, which, according to [7] is part of the standard library.

C. C# programming language

According to the author [8], C# is an object-oriented language, which is developed by Microsoft and runs on the .NET platform. It is used to create graphical desktop applications as well as games. According to the author in [8] it is also usable for web services. According to [9], it provides a Xamarin framework for creating mobile and desktop applications, but according to [10] it is not supported anymore. The ASP.NET framework is available for web application development, which was created according to [11] in 2002. C# also includes the ML.NET framework for machine learning.

D. Spring Boot framework

Spring Boot is Java framework which is primary used for developing backend services and API programming. It can be used for developing backend for web applications and also for desktop applications. According to [12] Netflix is running Spring Boot on its backend.

E. Data Storage

Security of storing data is very important because data can contain sensitive user information like passwords. Passwords in general can not be stored in plain text so it is necessary to use hash algorithms. According to [13] one of the most used hash algorithms are PBKDF2, BCrypt and Scrypt. All of these mentioned algorithms use salt during hash which significantly improves security.

IV. Design of system for testing HTTP requests

The aim of the work is to create application for managing and testing HTTP requests with focus on teamwork. Before the implementation itself, it is important to design the entire system and how the system will work. The entire implementation is implemented in the Java programming language using its 2 frameworks:

- **JavaFX** - frontend
- **Spring Boot** - backend

A. Database design

Our application will have many different entities, which will be stored in the system database. There will be a login and registration form. First, the user has to register in the system. Then, they have to enter their email address and password to log in. One of the functionalities of the system is the grouping of tests inside projects. It is evident that one of the entities contained within the database will be **Project**, which will has its name. It is possible for users to create multiple projects. The most significant entity is the **Test**, which is representative of the application's core. Every test has different properties, and each test must be linked to a certain project. The **Invite** entity represents an invitation that the user can either accept or decline. If users want to work together, they should be able to send each other invitations through the app. It is very important that users can communicate with each other. The **ChatMessage** entity represents message which

is sent for users inside project. Users can send several messages for a single project, and each message belongs to a different project. This column is a number in the form of a primary key. The **AppUser** entity represents users who can send multiple messages, own multiple projects and work on multiple projects, as well as send multiple invitations to specific projects. Projects can have multiple tests, one owner and multiple messages.

B. REST API design

Communication will be achieved via REST API endpoints. These are what REST controllers look like in our system:

- **/api** - REST controller for login and register
- **/api/chat** - REST controller for sending and receiving chat messages.
- **/api/invite** - REST controller for accepting or rejecting invitations to projects.
- **/api/project** - REST controller that provides operations with the Project entity, including the ability to create, open, delete, abandon, and invite users to projects.
- **/api/settings** - REST controller that secures the change of user settings
- **/api/test** - REST controller for adding tests to projects, editing and displaying them
- **/api/user** - REST controller that provides searching for users via a search field and displaying invitations that have been delivered to the user

C. User interface design

The user interface is one of the most important factors for desktop applications. We started by creating a low-fidelity design using the Miro tool. The application includes screens for logging in and registering, the home screen, the settings tab, the invites tab, the project dashboard, user management in the project, the interface for adding tests and the project chat screen.

V. Implementation of a system for testing HTTP requests

After analysing the project, we decided that the whole system will be developed using the Java programming language. The application will consist of two parts: client and server. We chose the JavaFX framework for the client side and the Spring Boot application for the server side.

A. System architecture

Our solution consists of a client server implementation of a desktop application. The server is constantly in communication with the database and thus the system architecture consists of 3 layers:

- **Client** - Contains configuration files for the application design and provides interaction with graphical elements
- **Server** - Consists of a 3 layer architecture, serves as a middle point between the client and the database
- **Database** - Contains data and interacts with the backend

The system architecture is shown in fig. 5. A connection is established between the client and the server over which requests are sent via HTTP protocol in JSON format.

B. Backend

The system must have a backend that can store data permanently in a database and can communicate with clients. The first thing we did at this stage was to make a database based on the database design. We chose the PostgreSQL database system for this purpose. We used Spring Boot to set up the database. This involved creating tables that will be included in the database. The backend is made up of more than just the database and the data storage. It also includes the system that describes how the client will access the data. To make sure that data could be accessed, we used REST controllers. These allowed the client to communicate with the server using endpoints. We have created a service interface for each REST controller. Our backend architecture has 3 layers:

- **REST controller** - contains endpoints
- **Service interface** - performs the required services
- **Repository** - direct interaction with database like saving, getting, deleting and editing data.

Fig. 6 shows the backend architecture, including the processing of requests from the client. Endpoints receive data from clients, after which the controller calls a service method to execute the business logic and perform the necessary database interactions.

C. Security

Security is very important, especially when it comes to using or storing data. After registering, it is important to protect sensitive information, like passwords. Before storing a user's password, we use a special way of hashing called BCrypt. This is very secure because it uses a 'salt' during the hashing process. When it comes to security, it is also important to protect each individual REST endpoint. This means only authorised users can access specific resources. In order to ensure authorisation, the method with JavaScript Object Notation (JSON) web tokens (JWT) was utilised. After the user has logged in, the application will send the user an authorisation token. The user is then able to use the application with this token. The system is also compatible with emails, and it was necessary to ensure the validity of the emails. The system uses mail exchange (MX) to domain name system (DNS) of specific mail domains to verify the validity of said domains.

The aim of the application is to ensure teamwork during testing HTTP requests. It is necessary to implement real-time communication between users without using third-party applications. The system does not require the use of voice or video communication. From this point it was decided that the Simple Text Oriented Messaging Protocol (STOMP) would be implemented, as it is designed for the transmission of text messages. STOMP does not include real-time communication functionality. Therefore, it was used in combination with web sockets.

D. Realtime communication

SimpMessageBroker is implemented on the server which ensures the creation of the webSocket endpoint and the subsequent connection of individual clients. The client connects to the websocket with a unique postfix that represents the user and the project to which the message belongs. In the request body, data such as the sender's email and the target id for the project will be sent along with the message body. This data is sent to the REST endpoint and then the server stores the message in the database and sends this message via web sockets to the target project. This method allows tests to be shared in real time between users.

E. Implementation of user interface

Once we had finished setting up the database and the rest of the backend for the server to work, we started to use JavaFX framework to add the client part. We started with a low-fidelity design and added some extra features. The creation of the screen was completed using SceneBuilder, a program that was selected due to its simplicity and its ability to be used with JavaFX. This tool does not require the writing and knowledge of FXML Markup Language (FXML) code. The tool was utilised to create the layout of elements on individual screens. For the purpose of adding styles, CSS files were used, which are a common method for styling applications and web pages. Fig. 1 shows the implementation of login form. The screen after login is shown in Fig. 2 . The project chat is shown in Fig. 4 . The most important function, which is used to add tests, is shown in Fig. 3.

VI. Conclusion

The aim of the project was to create a desktop application for managing and testing HTTP requests. The application should be focused on teamwork and organisation while writing individual tests. This testing software has lots of features that help teams work together on testing HTTP requests. The most important thing during development was to design a user interface that is clear, modern and consistent. This interface should be easy to use and not lose any functionality. The system has a simple, consistent layout of graphics on the screen.

The most significant functionality of the system is the capacity to add tests to an opened project. All written tests are displayed in the project dashboard. The process of adding new tests within the system has been designed to be both straightforward and intuitive. The tool designed for the purpose of testing HTTP requests is capable of supporting the testing of methods such as POST, PUT, GET and DELETE. In the context of writing tests, full support is provided for adding request bodies in JSON format using a built-in editor for JSON. Another added functionality is the capacity to add request headers inside HTTP requests. The system contains predefined headers that can be edited, as well as the option to add custom headers. There are two options for submitting the test in the system. One available option is the save test, in which the user enters the test name and the expected return code, and the test is automatically evaluated according to the code. The second option is to send the test without saving it to the project. The system has an additional function that enables all tests in the project to be executed. It is therefore possible to perform test-driven development using the system. The software also supports testing HTTPS requests, because the java client supports it natively.

Given that the application's core functionality is team-based, it has been necessary to implement a user management system within the project. The project owners have the capacity to invite other users to participate in the project, who may then accept or decline the invitation. After accepting the invitation, they will be automatically included in the project and will be able to interact in the project chat and add tests in real time. The project owner can also remove users from the project. It is important to note that users within the system possess the capability to modify their user password and also to remove their user account from the system. The system was tested using Postman software throughout the development process.

The advantages of our solution over existing solutions include sharing tests in real time, writing automated tests without the need to write JavaScript code, the presence of team chat within the project, the system of managing users on the project. The software works on a similar principle as, for example, Jetbrains development environments, where the user first creates a project and then writes the source files. In our case the user writes tests and not source files. The backend of the application is deployed on docker and the client part is built as an .exe application.

The system is currently in a continuous state of development, with ongoing consideration being given to potential future updates. Updates may include changes to user profile pictures, real-time communication testing, connection session testing, load test support, and scheduling of send requests.

Abbreviations

API	Application Programming Interface
HTTP	Hypertext Transfer Protocol
SOAP	Simple Object Access Protocol
JSON	JavaScript Object Notation
JWT	JSON Web Token
REST	Representational State Transfer
MX	Mail Exchange
DNS	Domain Name System
STOMP	Simple Text Oriented Messaging Protocol
FXML	FXML Markup Language

Figures and tables

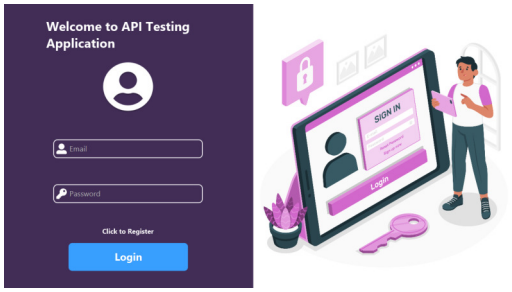


Fig. 1: Login implementation

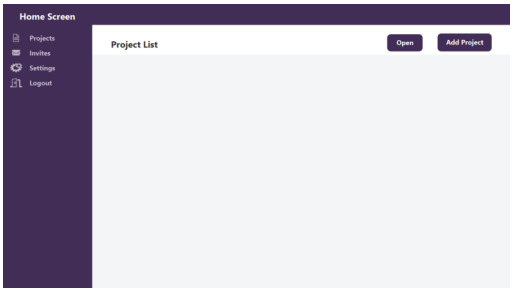


Fig. 2: Home screen implementation

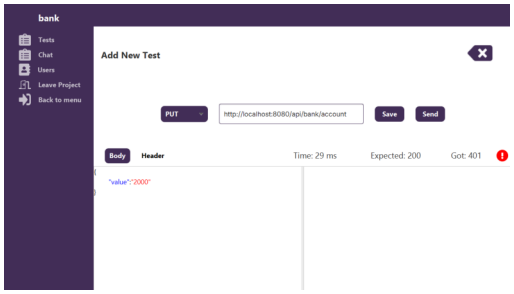


Fig. 3: User interface for adding tests

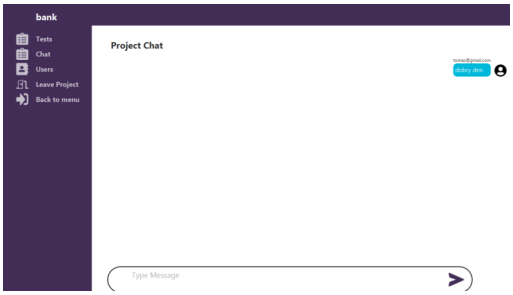


Fig. 4: User interface for project chat

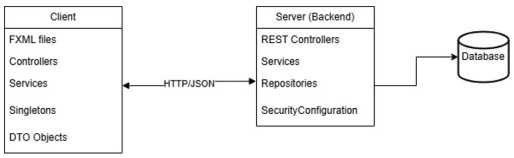


Fig. 5: System architecture

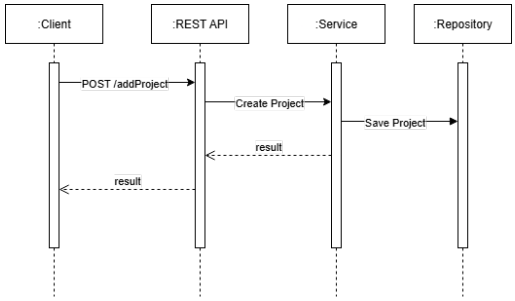


Fig. 6: Backend architecture

REFERENCES

[1] I. Chadere, "Postman vs hoppscotch: What you must know," 2024. [Online]. Available: https://apidog.com/blog/postman-vs-hoppscotch/
[2] A. S. Sekhon, "Soapui vs jmeter," 2015. [Online]. Available: https://www.researchgate.net/profile/Armaan-Sekhon/publication/285982412_Load_Testing_tools_evaluation_-_SOAP_UI_vs_Jmeter/links/566527f608ae418a786d7769/Load-Testing-tools-evaluation-SOAP-UI-vs-Jmeter.pdf
[3] H. Schildt, *Java: the complete reference*. Mc Graw Hill, 2007, ISBN: 978-0-07-163177-8.
[4] B. Cole, R. Eckstein, J. Elliott, M. Loy, and D. Wood, *Java Swing*. O'Reilly, 2002, ISBN: 0-596-00408-7.
[5] C. Dea, *JavaFX 2.0: introduction by example*. Apress, 2012, ISBN: 978-1430242574.
[6] M. Lutz, *Programming python*. " O'Reilly Media, Inc.", 2011, ISBN: 978-0596158101.
[7] A. D. Moore, *Python GUI Programming with Tkinter: Design and build functional and user-friendly GUI applications*. Packt Publishing Ltd, 2021, ISBN: 978-1801815925.
[8] J. Liberty, *Programming C#: building. NET applications with C.* " O'Reilly Media, Inc.", 2005, ISBN: 978-0596006990.
[9] J. Skeet, *C# in Depth*. Simon and Schuster, 2019, ISBN: 978-1617294532.
[10] Microsoft Corporation. (2024) Xamarin Support Policy. [Online]. Available: https://dotnet.microsoft.com/en-us/platform/support/policy/xamarin
[11] J. Galloway, P. Haack, B. Wilson, and K. S. Allen, *Professional ASP. NET MVC 4*. John Wiley & Sons, 2012, ISBN: 978-1118348468.
[12] O. Oliinyk, V. Vyshniak, and P. Berest, "Building a microservices architecture with high-level programming languages and frameworks on the example of using spring boot at netflix," *Scientific Collection InterConf*, no. 200, pp. 226–228, 2024.
[13] G. Hatzivasilis, "Password-hashing status," *Cryptography*, vol. 1, no. 2, p. 10, 2017. [Online]. Available: https://www.mdpi.com/2410-387X/1/2/10

Finance Management Application

[1]*Matúš PENIČKA,* [2]*Miroslav MURIN,* [3]*Lenka BUBEŇKOVÁ,*
[4]*Miroslav MICHALKO*

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

[1]matus.penicka@student.tuke.sk, [2]miroslav.murin@tuke.sk, [3]lenka.bubenkova@tuke.sk, [4]miroslav.michalko@tuke.sk

Abstract **– The objective of this final project was the design and implementation a web application to aimed at enhancing personal financial management, identify cost-efficiency opportunities, and provide an overview of financial behavior trends. The application development process includes a thorough analysis of existing solutions, system architecture design, and implementation for usability and scalability. Software consists primarily of four main parts that encompass core system functionalities – expense tracking, loan management, budget planning, and investment monitoring. Individual sections of the system offer various data visualization elements, tabular representations, and information that are presented in an user-friendly manner that prioritizes usability. One of the primary goals was to create a data entry interface that would offer a streamlined yet at the same time comprehensive input mechanism with the possibility of future expension based on user feedback. The software incorporates several security layers to enforced protection of sensitive financial data. The software itself was subsequently deployed and tested in a cloud-based environment. The result of testing was highly positive, from which we can conclude that deployed solution is a reliable and well-designed, which helps to obtain information about one's financial habits and increase the user's financial literacy.**

Keywords **– finance management, web application, ASP.NET Core, React, PostgreSQL**

I. Introduction

Money is an inseparable aspect of everyday life and plays a significant role in shaping our decisions. Many individuals face challenges related to a lack of financial resources or their suboptimal allocation. In response, people often seek higher-paying jobs, look for additional income sources, or explore investment strategies. According to Robert Kiyosaki [1], the core problem does not lie in the lack of money itself, but rather in the inability to manage, allocate, and preserve wealth over the long term.

A conventional way to manage finances inefficiently is the manual recording of income and expenses on paper. This approach is not only time-consuming, labor-intensive, and vulnerable to human error, but also often leads to disorganized documentation and lost records, with overall reduced effectiveness of financial oversight.

The solution to this problem lies in digital transformation and process automation. The system should streamline the process of adding new expenses while offering flexibility and providing sufficient detail for all users, from novices to advanced users. It should be designed with a focus on intuitive user experience and functional adaptability, offering both simplified and more detailed options to accommodate different user needs.

Each expense should be classified into categories, and the system should allow users to dynamically define new categories directly within the expense entry form. The software should also support filtering mechanisms and selecting transactions for visual analytics. An important feature of the application should be the implementation of loan management, including the possibility of refinancing. In addition to expense tracking, the system should also focus on income optimization. Therefore, integrated investment monitoring is considered beneficial, allowing users to track profit and loss dynamics for individual stocks. A cornerstone of the system should be the implementation of an efficient retrieval function, providing users with rapid access to relevant financial data. The application should be designed for visual clarity, with data clearly and visually distinguished. From the user's perspective, it should be possible to modify transaction attributes or choose a specific time period for visualization. The primary focus is to ensure that the developed system

delivers a user-friendly and intuitive solution for all users, giving them a comprehensive financial overview and an opportunity to improve their budgeting strategies.

II. Analysis of Financial Management Applications

This chapter examines existing systems that have been designed for the purpose of efficiently processing financial data, managing financial resources, and optimizing their allocation. This analysis aims to identify their advantages and limiting factors, and thereby gain a deeper understanding of the issue.

A. Commercial Systems

Commercial systems often provide a wide range of tools for financial management. However, they frequently focus on a single financial domain overlooking the overall financial landscape. Additionally, many fail to provide a robust customization framework that can meet the diverse requirements of users.

YNAB (You Need A Budget) is a well-known financial management application focused primarily target around its web interface, with more simplified versions available for Android and iOS. Its main strength lies in its extensive features, supporting both income and expense tracking as well as basic functionality for loan management, investment tracking, and budgeting. The application is based on a subscription-based licensing model with no free version, although it offers a free trial, which helps determine whether it is alignment with user expectations. A notable advantage is Auto-Assign, which enables a degree of automation that can reduce the time needed for manual input. The design may feel overwhelmingly complex, with tools for hierarchical categorization, zero-based budgeting, bank account synchronization, compatibility with CSV, QFX, QIF, and OFX file imports, and loan management with a simulator.

On the other hand, the main weakness is the insufficient granularity of individual expense records, as the system specializes in managing finances at the category level. Another shortcoming is suboptimal support for investments, which fails to provide an effective way to monitor and gain insights about your portfolio and investments.

Monefy is a widely recognized personal finance management mobile application, with over 5 million downloads and a 4.1/5 rating, indicating its usability and popularity among users. It is principally designed to offer a user-friendly interface focused on basic expense and category tracking, making it suitable for less demanding users. Compared to previously analyzed management applications, it provides fewer advanced features and prioritizes simplicity and visual clarity through its pie chart representation of spending categories.

The application is available in both free and premium versions. While the free version facilitates core tracking capabilities, it includes ads and functional restrictions. The subscription-based version offers extended options such as custom categories, currency selection, and synchronization via Dropbox and Google Drive.

However, even the premium version lacks bank integration and offers limited data analysis, which indicates a prioritization of simplicity rather than advanced financial tracking. This may not align with the users seeking a more exhaustive financial overview of their expenses.

Despite these constraints, Monefy stands out for its ease of use, integrated calculator, and visually appealing design, offering an accessible and effective solution for straightforward personal finance tracking.

B. Open-source Systems

In contrast, open-source solutions provide a broad spectrum of functionalities, although they often require certain level of technical proficiency in software engineering, which consequently leads to a more complicated installation process. Moreover, open-source software absence of reliable support and may suffer from stability issues.

Firefly III is an open-source personal finance management application, which is available on the official GitHub repository. The application is based primarily on PHP and incorporates technologies such as Twig, Vue.js, JavaScript, Blade, Alpine.js, and Tailwind CSS for the user interface.

As mentioned, one of its principal drawbacks is its complex installation process, which pose a barrier especially for non-technical users. The system requires self-hosting and server configuration. However, the developers supply thorough installation guides and documentation, along with a live demonstration instance to help users decide whether the software fits their needs.

Overall, the user interface is generally intuitive, although when managing large amount of data, the interface may become difficult to read. A major upside of the system is extensive is

customization options, which includes multi-language support, customizable transaction fields, and more.

The application supports data classification through categories, groups, and tags, which enhances the user ability to systematically organize their financial data. It also supports multi-currency and allows users to create and define their own custom currencies.

Firefly III offers detailed visual output and simplified procedure for recording transactions. It offers limited bank support, which integration relies on third-party services such as Salt Edge and GoCardless. However, the system lacks of tool in area like loan and investment management. The application is oriented toward budgeting and transaction tracking, making it an optimal and suitable solution for individuals seeking a clear overview of their income and expenses.

III. Development of Web Applications

We will concentrate on the critical aspects involved in the selection of appropriate technologies and conduct a comparative analysis of various programming languages. This evaluation will assess the respective strengths and limitations of each language in the context of the application requirements. Choosing the right software environment constitutes a decisive factor in effective and efficient development. The system should be designed to deliver a high-performance and robust solution simultaneously ensuring low computational overhead on the underlying hardware infrastructure.

A. Server-Side Of Web Applications

The server-side component of a program forms the fundamental pillar for the development of a web application. Separation from the client ensures better protection of sensitive data and ensures data confidentiality and integrity, which is indispensable for financial management applications. According to the authors [2], from a security perspective of websites, the client-side environment cannot be fully trusted. Nevertheless, due to modern and feature-rich applications, certain privileges may exist on the client side as well.

Popular languages for server-side development include JavaScript (Node.js), Python, Java, PHP, and C#. As the author notes [3], Node.js excels at handling concurrent requests and real-time outputs. It also offers resource-efficient data processing for communication with the client-server communication. In contrast, Python is well-known for its performance and code readability. It is primarily utilized in data analytics and artificial intelligence, where processing efficiency is crucial. The Python programming language is generally considered a suboptimal choice for mobile applications.

An alternative approach is the use of the ASP.NET framework, developed by Microsoft as part of the .NET ecosystem. It primarily supports the C# programming language, which is optimally compatible. C#, a statically typed, object-oriented language is quite similar to the Java programming language, which facilitates the transition for developers with Java experience. It offers developers various official extensions that speed up project initialization and development. According to the authors, ASP.NET MVC, compared to Web Forms and Web Pages, serves as the most fitting option for web application development due to its greater efficiency and maintainable architecture. This architectural pattern helps developers separate concerns, leading to easier management and scalability.

B. Client-Side Of Web Applications

The client side of a website serves as a crucial part of the application and directly impacts the overall user experience. Consequently, the application must be developed with an emphasis on simplicity, clarity, and visual sophistication. According to the authors [4], the user interface is one of the decisive factors not only for distinguishing software from competitors but also for significantly influencing the user's decision to choose the right software. The user interface should not concentrate solely on visual design but also on user interactions, which are critical to the overall user experience. Furthermore, a thorough detailed analysis of the user needs and behaviors is important in order to understand their actual interests rather than relying on assumptions. The user should not spend time thinking about how the program works, but rather on solving their problems.

Bootstrap is among the most well-known frameworks for web development. It offers a wide range of functions and components, as well-developed HTML and CSS templates, which streamline and accelerate website development, even for beginners. Another popular CSS framework is Tailwind. It is a "utility-first" responsive CSS framework that gives developers full control over component customization and does not provide predefined components. It speeds up interface

styling according to specific software needs and ensures efficiency and maintainability, as stated by the authors [5].

Angular is a framework built on TypeScript that brings numerous advantages, including a component-based architecture. The architecture enables the reuse of individual components, thereby simplifying user testing and maintenance. Angular performs quickly due to its ability to load only parts of the page. However, its complexity and component management make it difficult. Additionally, its overall complexity and scale can be challenging for newcomers [6].

React is a JavaScript library focused on the creation of stateful and reusable User Interface (UI) components. Leveraging a virtual Document Object Model (DOM), enhances rendering performance by updating only the modified nodes. Strengths include server-side rendering and lowered demands on user resources during loading, thanks to support for code merging and removal of unused code. Due to the lack of MVC support, it is necessary to import libraries for states and models [7].

C. Database Servers

For financial management, a secure and efficient database is essential, providing users with rapid query processing and secure, long-term storage of sensitive data. The database architecture must be designed to handle large volumes of data while ensuring their integrity, consistency, and availability. Currently, there are a variety of database technologies and associated data structures. The most commonly used data architectures: are relational databases, object-oriented databases, and NoSQL databases. Relational databases organize data into rows and columns(attributes), which use a method called Structured Query Language (SQL) to perform commands to access and retrieve data from the database. A core benefit of this architecture is ease of use, efficient access, and the ability to quickly create new categories, as stated by the authors [8]. Object-oriented databases are noted for processing high-dimensional data and intricate inter-object relationships. These databases are more suitable for interpreting real-world objects. According to the authors [8], the disadvantages of object-based databases may be: lower efficiency with simple data, slow data access, and lack of standardization compared to relational databases.

NoSQL databases are not intended to replace relational databases but to coexist as two architectures that offer alternative solutions for data processing. These newly implemented solutions come with their own shortcomings of relational databases and bring both new advantages and disadvantages. When comparing MongoDB as an example of NoSQL with Microsoft SQL (a relational database example), the authors [9] found that MongoDB is more efficient in inserting and updating with simple queries. On the other hand, Microsoft SQL showed better performance in insert and update operations that do not require keys and in aggregated queries.

PostgreSQL is a widely recognized management system renowned for its performance, flexibility, and efficient data processing. It is one of the oldest systems still in use, thanks to ongoing updates and broad support. It is an open-source system based on an object-relational Database Management System (DBMS). It is designed with a focus on efficiency and integration with various software. One of PostgreSQL's most important features is Write-Ahead Logging (WAL). This feature allows online backups and recovery within the time interval covered by WAL data [10].

D. Security of Web Applications

Modern systems are constantly exposed to various security threats, with increasingly sophisticated methods to obtain confidential data that threaten software integrity. Security risks should be considered from the earliest stages of the development lifecycle. It is also critical to ensure secure communication between the server and the client.

One of the most prevalent and dangerous to system security is Cross-Site Scripting (XSS). The threat consists of the attacker's attempt to inject malicious scripts into a web application. This script can serve various purposes – obtaining personal data, session cookies, or impersonating the user. These risks continue to pose significant a threat to current systems.

The authors indicate that [11], another security threat is Cross-Site Request Forgery (CSRF). The attack consists of deceiving the user in order to perform an action on a trusted site that abuses the rights of the logged-in user and bypassing authorization mechanisms.

Man-in-the-Middle (MitM) represents a potential vulnerability to the system. The attack consists of malicious software designed to eavesdrop on communication between the program and the user. This is caused primarily due improper implementation of secure communication protocols, such as HTTPS.

The part of system security also involves user inputs and verifying their validation. SQL injection remains a common technique used by attackers to insert malicious SQL statements into the input fields, intended to violate the integrity of the software [11].

APS.NET, a web framework developed by Microsoft, that offers developers an extensive array of services for developers. The tool delivers various security features and extensions, which are constantly updated. According to the authors [12], APS.NET supplies tools for security against XSS attacks and CSRF using HTML encoding, AntiForgeryToken, and several libraries such as ASP.NET Identity and AntiXSS Library.

IV. Design of the Financial Management System

Application design plays a vital role in the web application development. The strategic goal of the web application is to provide the user with a complete assessment of his finances from several perspectives. During the designing and developing phases, it is extremely important to focus on the compatibility and scalability of the system. These features will ensure the sustainability and further growth of the application.

A. Design of the System Architecture

The architecture design of the system must ensure flexibility, modularity, and scalability, simplifying developer support and enabling future development. Modularity divides the system into manageable components for easier maintenance and updates. Scalability allows gradual expansion and integration of new components. Security is prioritized through well-designed architecture, component communication, and the use of Docker containers.

The financial management system structure is based on the client, server, and services model. Modules covering client and server parts are deployed in Docker containers, which results in fast deployment and easier implementation, improving communication and flexibility. Users interact only with the React-based client interface, which contains no sensitive logic or data to protect system integrity.

The client communicates with the ASP.NET server, which manages application components and communication, handling requests and returning data as JSON responses.

The PostgreSQL database stores user data and financial behavior patterns, protected with encryption, secure communication protocols, request validation, anti-forgery tokens, or Docker container isolation.

The system storage service manages image data efficiently and scalably through the integration of Microsoft Azure and ASP.NET.

B. User Interface Design

The user interface design must be carefully and thoughtfully designed, with essential factors being simplicity and accessibility for users. The environment must be clear and efficient to deliver a positive user experience.

The financial management tools will be immediately visible without requiring scrolling, providing the user with an immediate overview and streamlining access to various services. The interface will include a sidebar offering options for switching between tools and clear navigation. The panel should include profiles, settings, expenses, loans, and other necessary functions.

The environment for visualizing financial tools will include extensive graph display options. For Expenses, Investments, Loans, and Budgets, there will be tables with relevant data and graphs to support comparing or interpreting user finances. The implementation will leverage AG Grid and Shadzn for flexibility and rapid implementation.

Expenses, including cash and electronic transactions, comprise a significant importance that requires simplicity, readability, and visual representation. Data visualization will include pie charts of monthly expenses by category alongside historical comparisons with previous months.

Loan data will be visualized through graphs to represent historical values, current status, and repayment progress. Users should be able to compare multiple loans to gain additional overview and insights about their financial situation.

The budget tool should focus on simplicity and practicality, an assigned table, and a summarized block overview of short-term budgets. The summary block will contain a graph showing budget shares, a fulfillment indicator, and a graphical representation of expenses.

The investment tool will report on individual investments and overall portfolio profits. With goal is to deliver a detailed analysis of the portfolio, with the aim of improving investments through visually attractive representations such as graphs, tables, and another element.

Fig. 1 Expenses Page forms

V. Implementation of a Financial Management Application

In this chapter, we concentrate on the implementation of the software solution. The financial management system uses advanced technologies to maintain high performance, scalability, and flexibility. The backend part is developed using ASP.NET with C#, offering security and native Docker support. The frontend utilizes React implemented by Typescript and JavaScript for data manipulation and visualization. Persistent data storage is managed by PostgreSQL, a reliable relational database. The software aims to provide easy and powerful personal finance management, enabling detailed tracking of finances, and giving users an overview to improve financial decisions.

A. Form for Adding Expenses

One of the system requirements is to integrate input forms for expense creation, covering various types such as transactions and cash expenses. It is designed to enable users to access to overview, thereby improving control and an easy way to identify and correct errors. Data validation is ensured by the Yup library to avoid errors or irrelevant data. The form initially provides basic expense information for fast processing, while also supporting an extended version for detailed and complex data entry. In case of errors, users are informed by the toast library notifications.

In figures 3 the form contains two labels positioned in the upper left corner with different functionalities. One label allows adding multiple cash or transaction expenses under one overarching expense. The other label contains a table displaying entered data for efficient overview and correction, ensuring flexibility and control.

B. Loan Data Visualization

The Loan functionality primarily consists of the graph, which offers three interactive options for in-depth analysis while optimizing space layout. Additionally, the program presents a table and a block of information related to the loan.

The main visualization element is a graph, with the first option focused on comparing current loans, showing loan amount, remaining balance, monthly income, interest rate, and repayment period, helping users identify advantageous loan attributes or refinancing needs.

The second set of graphs includes pie charts showing the amount left to pay off loans, monthly payments and interest, and relative loan lengths for planning.

The last primary chart is a line graph that shows loan progression over time, reflecting new borrowings, repayments, or refinancing activities.

The block of information related to the loan displays the active borrowed amount, total loan balance, and highest interest rate.

The table allows advanced searching and contains fields loan name, providers, loan amount, interest rate, repayment period with a graphical ratio of remaining vs elapsed time, the monthly payment (highlighted in red), loan status, and description for notes. The “display” button lets users show selected data on the main graph.

Fig. 2 Investment Page Overview

C. *Visualization of Budget Data*

The application offers a core table designed for intuitiveness and clarity, containing critical data for optimal budget monitoring. The table structure includes a name attribute with two parts: name (highlighted in white) and description (in gray) to improve user experience. It shows progress as percentages, total amounts, and spent funds, giving a clear overview of the budget. Other attributes include remaining time and budget duration, updated based on repeatability settings. Clicking the pie chart icon visualizes expenses for the selected period.

The secondary block features a pie chart displaying short-term budgets with limits for clarity, along with a legend for budget categories. It also shows the total overdrawn amount, and current overrun (capped at 200%), and highlights the largest budget and its highest expenditure this month, using progress indicators from the Shadzn library.

D. *Investment*

This subsection will focus primarily on the implementation of investment data visualization and analysis that is intended to support to improve user portfolios and investment plans.

Data collection will be mediated using the Alpaca tool, which provides a flexible environment suitable for our software, but also for direct investment in the financial market. Alpaca offers documentation of APIs and WebSockets that convey information about the user's portfolio and the current state of the market.

As illustrated in figure 3 the implementation of the "Investments" page consists of three block units. Investment data is analyzed from both long and short-term perspectives. The system is designed to provide a simple overview of necessary and fundamental information about investing.

Visualization of balance, including a percentage representation of profit/loss. It also shows the total amount invested in a given year. This data allows the user to monitor and adjust the volume of investment expenses.

Identification of the most profitable investment is another functionality of the system and from two perspectives – net and percentage profit. The user can select the desired item by clicking on it. The selected item will be enlarged. The user is then provided with a card with data on the most profitable stock. The default stock is set based on net profit, but the user can change it to percentage profit. Net and percentage profits are implemented with respect to the current year. The user is provided with buttons to switch between the options.

Another function of this system deliver functions related to the overview of individual stocks. It consists of stocks owned by the user, the cards display data such as the invested amount, profit, and current value, which corresponds to market prices. A possible disadvantage of the Alpaca website (free version) is the provision of information only during open market hours. Tabs can be scrolled using the horizontal mouse wheel or navigation buttons. These buttons disappear after reaching the border.

Visual analysis of the portfolio as a preview of the best actions, diversification, and investment trends. The dominant feature is a line graph that shows 5 actions and provides a visual analysis of results.

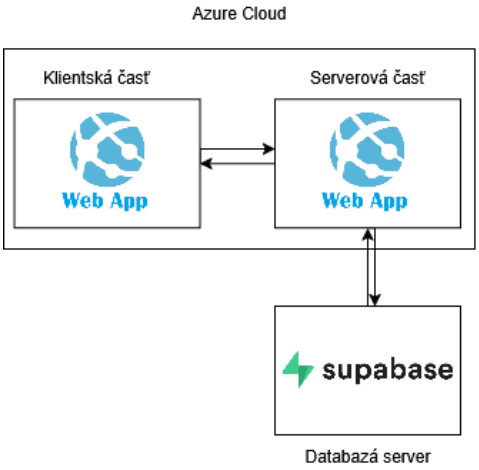


Fig. 3 Cloud architecture

E. Azure cloud

An essential phase of development is the deployment of the application on the web environment. This phase enables testing with real users to detect system errors or issues on weaker hardware. It helps better understand the application behavior before full-scale deployment.

Cloud services offer advantages but require careful consideration of cost-effectiveness and operational efficiency. They simplify deployment and allow scaling based on real-time demands.

We chose the Azure cloud system for its compatibility with ASP.NET and scalable environment for managing application parts.

The system, as shown in figures 3, is divided into two web servers for better control and separation of client-server parts, thereby improving security. The drawback may be higher costs and complex system maintenance.

To reduce costs, we separated the database from Azure and used Supabase, which allows free non-commercial use and easily integrates with Azure and .NET.

F. Conclusion

The primary objective of this work was to design and implement a personal finance management system that would integrate extensive functionalities reflecting the user's needs. Nowadays, users are increasingly demanding powerful, flexible, and at the same time simple solutions that offer strategic analysis and financial management. As part of the survey, we analyzed various available solutions, including commercial and open-source software. As a result of the analysis, we identified the positive and negative aspects of the software and their most relevant tools for financial management. A common limitation among these tools is a narrow focus on one segment of finances or a problem with suboptimal user experience and low flexibility. The outcome of the analysis was the creation of our system, which tries to achieve a balance between extensive functionalities and user-friendly design.

The software offered several visually attractive elements that were intended to improve the overview and monitoring of finances. The system provides flexible and scalable forms designed for users who prefer simplicity and efficient data entry, as well as the possibility of expanding the form for those requiring more complex data, and many other practical functions. The system is secured by several protective layers such as JWT token, HyperText Transfer Protocol Secure (HTTPS) communication, Cookies storage, and password hashing. The overall system is focused on intuitiveness, simplicity, and consistent integration of individual components. The software is structured into 4 major financial sections – expenses, budget, loans, and investments. Each section performs a specific task to meet the user's requirements.

Expenses are focused on monitoring and analyzing personal financial expenses. The system offers clear functions such as a graphical representation of target months in relation to categories, monthly expenses for years, or a comparison of common categories from previous months. In addition to these functionalities, it provides a table display of all transactions and the possibility of deleting, editing, or displaying expense details in the form of a graph.

Loans provide an overview of the management of both short-term and long-term liabilities. It enables a comparison of individual loans based on their key attributes – loan amount, amount repaid, monthly installment, repayment duration, or interest amount. Additional functionalities

include the visualization of data through pie charts and line graphs, providing a history of loan expenses and block of information related to the loan.

The budget section is used to set spending limits in relation to categories, which will then notify the user when they are exceeded. The budget also contains a table that supports functionalities such as displaying the percentage of budget exhaustion and its planned duration. Furthermore, offers functions such as a graphical display of the share of individual budgets.

Investments bring the possibility of tracking current stock market data in real-time, as well as monitoring your own portfolio through external integration with the Alpaca application. The software then processes the data into functionalities, such as calculating monthly and annual returns, analyzing portfolio diversification, and comparing stocks and their returns in order to identify investment problems.

The system was deployed in a simulated environment, where 18 users tested the functionality of the application and the goal was to provide feedback from direct communication or the SUS questionnaire. Based on the results, we achieved a SUS score of 87.94 and received the highest possible grade of A.

Further development of the application may include advanced features such as full integration with banks, email notifications for various critical events, and a chatbot. A general-purpose chatbot would be designed to respond to questions relating to financial advice and queries. A secondary chatbot would analyze and process the user data in order to offer advice and information about their finances. Another enhancement could involve the integration of ”drag and drop”, which would allow the user to personalize the layout, add new features, or, conversely, remove them from the interface.

References

[1] R. Kiyosaki, *Rich Dad Poor Dad*. Scottsdale, AZ: Plata Publishing, 2000.

[2] J. Rantanen *et al.*, “Isolation mechanisms for web frontend application architectures,” Master’s thesis, 2015, 2024-11-03.

[3] S. Bangare, S. Gupta, M. Dalal, and A. Inamdar, “Using node. js to build high speed and scalable backend database server,” in *Proc. NCPCI. Conf*, vol. 2016, 2016, pp. 61–62, 2024-11-03.

[4] D. S. Platt, *The joy of UX: user experience and interactive design for developers*. Addison-Wesley Professional, 2016, 2024-12-07.

[5] V. Levashenko, G. Ainura, and G. Kamalova, “Features of using the bootstrap 4 framework in the design of an adaptive website,” 2023, 2024-12-07.

[6] O. Kovalchuk, “Top front-end frameworks in 2024,” online, 2024, 2024-12-07. [Online]. Available: https://existek.com/blog/top-front-end-frameworks-2021/#most_popular_frontend_frameworks

[7] R. Vyas, “Comparative analysis on front-end frameworks for web applications,” *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 7, pp. 298–307, 2022, 2024-12-07.

[8] Z. S. Zubi, “On distributed database security aspects,” in *2009 International Conference on Multimedia Computing and Systems*, 2009, pp. 231–235, 2024-12-07.

[9] K. Sahatqija, J. Ajdari, X. Zenuni, B. Raufi, and F. Ismaili, “Comparison between relational and nosql databases,” in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018, pp. 0216–0221, 2024-12-07.

[10] Ünal, H. T., Mete, S., Vurgun, Ö. U., Mendi, A. F., Özkan, Ö., and Nacar, M. A., “Postgresql Database Management System: ODAK,” in *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2023, doi: 10.1109/ASYU58738.2023.10296600.

[11] H. O. Ekpobimi, R. C. Kandekere, and A. A. Fasanmade, “Front-end development and cybersecurity: A conceptual approach to building secure web applications,” vol. 5, no. 9, 2024, pp. 2154–2168, 2024-12-09.

[12] R. Thorson, “Asp.net security best practices: Protecting your web applications,” online, 2023, 2024-12-09. [Online]. Available: https://medium.com/@robert0321/asp-net-security-best-practices-protecting-your-web-applications-a9ea012d7434

Web Application with Booking System for Service Providers

[1]Matúš VAŠKO, [2]Miroslav MURIN, [3]Lenka BUBEŇKOVÁ, [4]Miroslav MICHALKO

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

[1]matus.vasko@student.tuke.sk, [2]miroslav.murin@tuke.sk, [3]lenka.bubenkova@tuke.sk, [4]miroslav.michalko@tuke.sk

Abstract **– This paper focuses on the conceptualization, implementation, and evaluation of a web application featuring an integrated booking system that combines intuitive reservation functionalities with effective digital representation of businesses in the online environment. The foundation for application design was established through analysis of existing technological solutions and research examining user preferences regarding currently operating reservation platforms. The implemented solution provides a business-oriented administrative interface for service management and enterprise presentation, while simultaneously offering a customer-facing interface facilitating seamless service discovery and reservation processes. The developed web application successfully addresses identified deficiencies in current market solutions, reflects established user preference patterns, and introduces new functionalities that enhance the digital presence of service providers within competitive online marketplaces.**

Keywords **– Django, providing services, React, reservation system, web application**

I. Introduction

In today's dynamic environment, efficient organization of time and resources represents a crucial success factor for businesses across all sectors. Systematic planning, process digitalization, and simple access to services have become standards that customers expect. Reservation management and brand building constitute key aspects of successful operations in this domain.

Manual reservation recording exposes businesses to various challenges, including erroneous records, complex customer database management, and reduced staff productivity, which can lead to customer dissatisfaction and decreased competitiveness.

The optimal solution may represent an implementation of a web application with a reservation system. A comprehensive booking system should enable not only basic appointment management but also provide advanced functionalities such as detailed booking and revenue analysis, as well as comprehensive business presentation. While multiple market solutions address these needs, many exhibit significant shortcomings. A common issue is unintuitive user interfaces that complicate navigation. Another limiting factor is insufficient flexibility in modifying existing reservations or the absence of specific time slot sharing capabilities.

Currently, we observe a significant trend in adapting social media elements into various online services. This trend presents an interesting opportunity for the booking system domain, where familiar user interfaces inspired by social networks can substantially improve user experience while offering businesses effective tools for online presentation.

The objective of this paper is to design and implement a web application solution with a booking system that meets the needs of modern service providers while providing a user-friendly environment for customers. The work encompasses analysis of existing booking systems, user needs surveys, and architecture design utilizing modern technologies. The implemented solution includes an intuitive customer interface for searching businesses and services, efficient administrative tools for service and customer database management, public business profiles inspired by social networks with comprehensive business presentation capabilities, automated email notification systems, and innovative service sharing capabilities through unique links. The final result is a solution that not only fulfills practical reservation system requirements but creates new value for businesses through effective development of their digital identity.

II. Booking Systems for Service Providers

A. Commercial Booking Systems

One of the most popular solutions in Slovakia is the Bookio platform, which provides extensive functionality including automated reminders, paid reservations, gift vouchers, customer feedback, and Google calendar integration. The system enables multiple establishments under one account with intuitive operating hours configuration, though it lacks temporal gaps like lunch breaks. The reservation form is intuitive with morning/afternoon time blocks enhancing user experience, but the page lacks visual branding elements. The administrative panel offers customer exports and bulk SMS but has visual shortcomings including unattractive color schemes, pixelated icons, outdated appearance, and confusing navigation with identical icons for different subpages. While navigation can be complex and less intuitive, this is partially offset by advanced features for experienced users.

Another option is the Reservio platform, which, beyond standard reservation functions offers a cash register system for smaller businesses. Reservio features modern, user-friendly design and Google account login capability unlike Bookio. It provides easy company profile setup with photo, slogan, description, and location, though it lacks gallery or review features and limiting URL links to Facebook or business website only. Customer reservation is straightforward but requires clicking through to services for details, making some information less immediately accessible.

B. Open-source Booking Systems

One of the freely available and highly customizable reservation system solutions is Easy! Appointments, whose PHP-based source code is available on GitHub and can be deployed on Apache or Nginx servers, also available as a WordPress extension. The system requires considerable customization as the standard settings lack basic service information during selection, fonts are too small, and time selection buttons are unnecessarily wide. The administrative panel is simple with good navigation but has problematic color combinations. While main advantages include regular GitHub maintenance, WordPress updates, community support, and well-structured code, basic version improvements would save implementation time.

The Cal.com application is another freely available solution focusing on meeting scheduling but applicable to service booking, offering extensive integration with calendars, CRM systems, email platforms, and project management tools. Built with modern technologies like Next.js, React.js, and PostgreSQL, the GitHub-available code can be deployed via Docker or PaaS solutions like Heroku. User advantages include clear, simple appearance with interface, event hiding capabilities, and extensive settings including the ability to customize layouts. Customer navigation is intuitive with well-placed information. While offering broad functionality, businesses providing physical services may need extensive modifications as it primarily targets online meetings and lacks pricing and branding features essential for physical service providers.

C. Analysis of User Experience with Booking Systems

An important step before designing an application is understanding what potential users actually need and prefer. To gather relevant information, we created an online survey with both closed and open-ended questions about people's experiences with reservation systems. We managed to get responses from 21 people between the ages of 18 and 34, which represents our target group of future users.

In the first part of the survey, we looked at what problems people run into when using reservation systems, and found two main issues. The majority of users, as many as 55%, indicated the inability to cancel reservations as the main problem. The second most common issue was confusing or unfriendly user interfaces, mentioned by 40% of respondents.

The second part focused on what features and characteristics users find most important in reservation systems. Respondents rated these on a scale from 1 (least important) to 5 (most important). The results are shown in figure 1, displaying the percentage breakdown of responses for each rating.

Our findings clearly show that users value a clear and simple interface above everything else, giving it an average rating of 4.67. An impressive 76.2% of people gave this the highest possible score. Easy access to service provider contact information was also highly valued, with an average rating of 4.48 and the top score from 57.1% of respondents. The ability to cancel reservations scored equally high at 4.48, which aligns with our earlier finding that this was a major pain point for users. Quick reservation creation earned an average of 4.29, with nearly half of respondents (47.6%) giving it the maximum rating, showing people expect booking processes to be both functional and fast. Features like sharing specific services or available slots with friends and

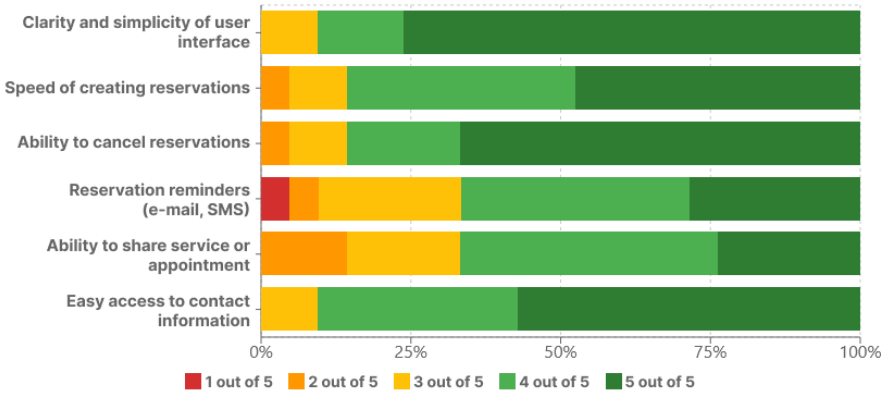


Fig. 1 Percentage distribution of ratings for individual functionalities of reservation systems

reservation reminders got lower but still solid ratings of 3.76 and 3.8. For sharing features, most people (42.9%) gave a rating of 4, indicating that although these aren't critical functions, they add meaningful value for a significant portion of users.

In the final section, we asked respondents if there were any features they'd like to see in reservation systems that they haven't encountered before. The responses revealed they'd appreciate more flexible reservation management - not just cancellation, but the ability to modify bookings as needed. They also wanted to see how many people had already booked or how many slots were still available on specific days, helping them choose less crowded times. These insights are valuable for both designing and implementing our system.

III. Development of Web Application

Web application development currently ranks among the most widespread areas of software engineering, as authors state in [1], with its popularity continuously growing due to flexibility, accessibility, and scalability possibilities. In application development, the design and implementation of appropriate solutions for both server-side and client-side components are crucial. It is important to focus on ease of use, reliability, and application security, which includes data protection during transmission and storage, proper access management, and prevention against potential security threats. As Li and Xue state in [2], web applications are very often targets of cyber attacks. This statement is confirmed by the authors in [3], who indicate that on average, up to 33 vulnerabilities are identified in each web application, of which six were classified as high severity.

A. Server-side Development Technologies

Django represents a freely available Python-based web framework. It follows Python's "batteries included" philosophy, meaning it comes with numerous built-in tools and functionalities essential for web development, such as authentication, user administration, database management, and form creation. Django emphasizes security with built-in mechanisms protecting against common threats like SQL injection and cross-site scripting, and offers an automatically generated administrative panel for managing users and access rights. According to the authors in [4], its advantages include simplicity and dynamic application development capabilities, making it popular for AI-connected web applications due to Python's familiarity and code readability.

Another popular option is the Spring framework for Java application development. The authors in [5] describe it as less demanding and focused on facilitating Java development through its Inversion of Control container, which manages object instantiation, configuration, and lifecycle via dependency injection. This approach increases modularity and testability by isolating components and making them easily interchangeable, following Dijkstra's Separation of Concerns philosophy from [6]. The extensive Spring framework includes key components like Spring Boot for simplified configuration, Spring MVC for Model-View-Controller architecture, Spring Data for database operations using ORM, and Spring Security for authentication and authorization.

A third analyzed option is Microsoft's modern, open-source ASP.NET Core framework, part of the .NET ecosystem. Like Spring, it utilizes IoC and dependency injection principles and is designed primarily for web applications and APIs. Compared to its predecessor, ASP.NET, ASP.NET Core is more modern and flexible with cross-platform support for Windows, Linux, and macOS. According to the author in [7], it's suitable for web development due to its performance

focus, supporting multiple .NET languages like C#, F#, and VB.NET. The framework emphasizes modularity and scalability through NuGet packages and integrates well with other .NET tools like Entity Framework Core and Blazor, with easy deployment on Microsoft's Azure cloud platform.

B. Database Systems

SQL database systems, such as MySQL, PostgreSQL, or Microsoft SQL Server, are based on the relational model organizing data into connected tables through primary and foreign keys. Meier and Kaufmann describe this in [8] as "one of the simplest and most intuitive ways to collect and present data". These systems are used where data accuracy, reliability, and secure transaction processing are important, making them common in e-commerce, banking, reservation systems, and enterprise applications.

NoSQL database systems, such as MongoDB and Redis, allow storing data in various formats like documents, key-value, columns, and graphs. According to authors in [9], NoSQL systems provide more flexible approaches suitable for unstructured data, with MongoDB allowing JSON format storage where records can have different structures, making them advantageous for varying data types and often used in social networks, analytical tools, big data processing, and caching systems requiring greater storage flexibility.

C. Client-side Development Technologies

One of the most widely used client-side development technologies today is React, a library developed by Facebook (now Meta) based on creating reusable components that logically separate interface parts and improve code navigation. React utilizes virtual DOM and state management concepts, which according to authors in [10] optimize user interface loading by creating new virtual DOM representations, comparing them with previous versions, and performing only minimal necessary modifications in the real DOM.

Angular, developed by Google, is another popular option that differs from React by being a framework utilizing MVC architecture rather than just a library. The authors in [11] highlight this approach for simplifying development, maintenance, and testing by clearly separating application logic, interface, and data. A key feature of Angular is two-way data binding, which enables automatic synchronization between user interface templates and component properties.

Vue is another popular JavaScript framework that Saks describes in [12] as lightweight and easy to use. This framework is based on combining Angular and React concepts like component-based approach and data binding. However, Vue provides a more intuitive API than complex frameworks, gaining popularity among both beginner and experienced developers.

D. Security

Broken Access Control is the most common security problem according to OWASP Top 10 2021 [13], occurring when attackers bypass access controls through URL manipulation to access protected resources. Solutions include strict authentication and authorization controls, with the authors in [14] recommending avoiding sensitive parameters in URLs or using valid session IDs, proper server configuration restricting access to sensitive resources, short-term session tokens and logging unsuccessful attempts.

Injection attacks involve inserting malicious code through input fields or URLs, with SQL injection and XSS being most common. SQL injection, described by the authors in [15], executes malicious SQL code on database servers, leading to unauthorized data access. This can be prevented through whitelisting, prepared statements, and stored procedures. XSS attacks insert malicious scripts into web pages executed in browsers, with the authors in [16] noting that 40% of web attacks result from XSS targeting sensitive data or content manipulation. These attacks can be prevented through input validation and sanitization of HTML tags.

IV. Design

Based on key properties and principles, Django was selected for the server-side, React for the client-side, PostgreSQL as the database system, and MinIO as the object storage. The simplified system architecture showing relationships between these technologies is illustrated in Figure 2.

Django is ideal for our server-side due to its modularity and comprehensive ecosystem. Its "batteries-included" philosophy provides integrated tools for authentication, authorization, and database operations. The Python-based Django contributes to code readability and maintenance, while Django ORM enables efficient object-relational database manipulation and migration management. The application will use Celery for asynchronous task processing (like email sending) and Redis as a message broker in a Publish-Subscribe architecture.

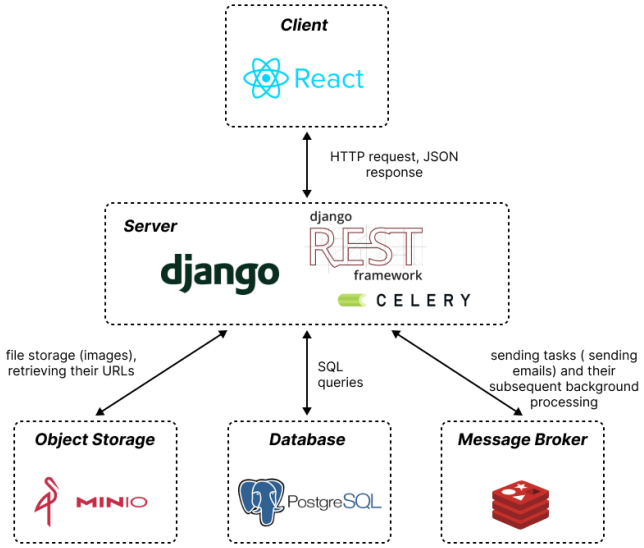


Fig. 2 Simplified diagram of system architecture

React was chosen for the client-side as it facilitates interactive user interface creation with a component-based approach enabling efficient code reuse. React's large community support provides numerous libraries and tools, and its widespread industry adoption is confirmed by the State of JS 2024 survey.

For data storage within our solution, the popular open-source database system PostgreSQL was selected. For photograph storage, we chose the MinIO object storage solution, which is also an open-source implementation. MinIO maintains compatibility with the S3 protocol, which represents the standard for object storage developed by Amazon.

The user interface, a key platform design element, emphasizes simplicity and efficiency through three main sections: a homepage with business search functionality, an administrative control panel, and public business profiles. The homepage serves as the customer entry point with a prominent search field. The control panel offers business management tools across various functional areas. Business profiles display essential information in a social media-inspired layout that creates a personal connection while improving usability, as shown in Figure 3, with sections for business details, quick actions, services, photos, and reviews.

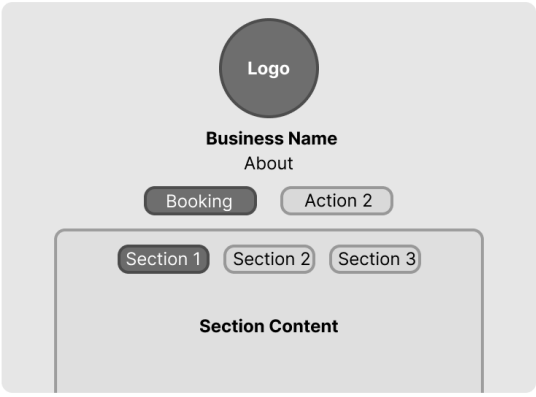


Fig. 3 Wireframe of business profile interface

V. Implementation

A. *Customer Application Interface*

The customer part of the application was designed for simplicity while providing a pleasant user experience. A key functionality is a straightforward search for services and businesses directly within the application. Users can enter service names, business names, or relevant keywords into the search field. The search utilizes Django's search filter to examine service and business names

and descriptions. Results are sorted by relevance, displaying exact matches first, followed by partial matches.

For customer reservation functionality, we divided the process from service selection to confirmation into logical steps beginning at the business profile. The reservation interface proceeds through several steps. First is service selection, highlighting administrator-chosen services to attract customer attention, with each service assigned to a specific category.

After service selection comes date selection, featuring a clear calendar with color-distinguished availability and the number of available time slots highlighted above each day. Upon selecting a specific day, dynamically generated time options appear based on service availability. The formula for calculating the set of available times A can be expressed as:

$$A = \{a_k | a_k = t_{od} + k \cdot t, a_k + t \leq t_{do}, \forall (b_s, b_e) \in B, (a_k + t \leq b_s \vee a_k \geq b_e)\} \quad \text{where } t = d + g \text{ and } k \in \mathcal{N}_0 \tag{1}$$

This formula identifies all available time points a_k beginning from initial time t_{od} and advancing in increments t (service duration plus time gap) up to final time t_{do}. Each a_k is available only when not overlapping with existing reservations, avoiding overlap with intervals (b_s, b_e) representing reserved times. Pricing for each time point is displayed, as rates may vary within a service by day and time. Selected appointments are stored server-side in UTC format to ensure time data consistency.

At any point during reservation creation, users can easily share the entire reservation including selected service, date, and time via a generated link.

B. Business Administrator Application Interface

The administrative section of the web application was implemented to provide business administrators with essential functions for service provision, emphasizing clarity, ease of use, and customization capabilities across the core modules: authentication, calendar visualization, service management, customer database administration, and configuration settings.

The authentication system features registration and login functionality with real-time validation, including username availability checking and password strength visualization.

The calendar module represents a central element of the administrative interface, developed using the FullCalendar library, as shown in Figure 4. This implementation offers dual display modes—weekly and daily views—with adjustable time intervals ranging from 10 to 60 minutes. Each reservation appears as a color-coded block, with colors indicating service categories to enhance visual differentiation when managing multiple service types. The calendar provides interactive functionality where selecting a reservation opens a modal window with booking details and modification options.

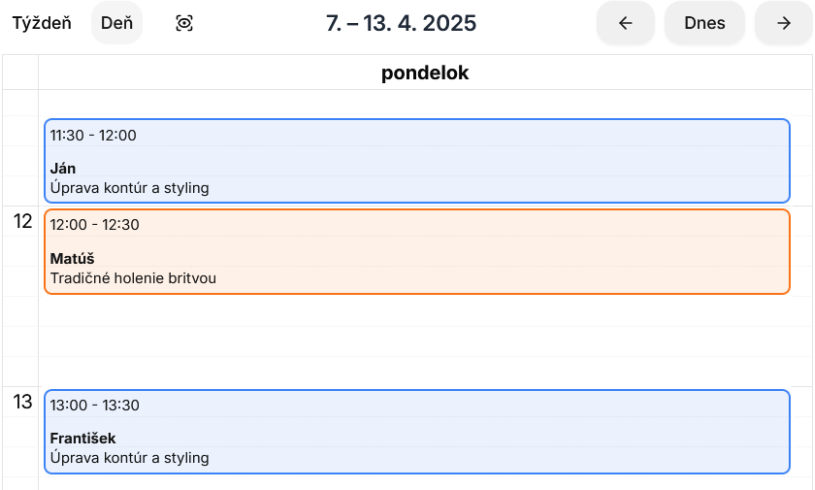


Fig. 4 Calendar interface module

Service management functionality was implemented as a tabular interface with filtering and sorting capabilities. The system supports service creation through a multi-step form process that collects essential information, pricing parameters, and temporal availability. A notable feature is the ability to designate services as recommended using a star icon toggle, which enhances their visibility in the customer interface. Additionally, the implementation includes flexible pricing

mechanisms allowing administrators to define standard and special rates applicable to specific time periods, thereby facilitating demand-based pricing strategies.

The customer database management module provides an overview of client information with advanced filtering and sortability across all columns. The implementation includes CSV export functionality that preserves applied filters, allowing administrators to extract precisely targeted customer subsets for marketing purposes or external analysis. This feature significantly enhances the application's utility for targeted customer relationship management initiatives.

Configuration settings implementation enables extensive system personalization according to business requirements. The settings interface is logically segmented into account security, business profile, media management, and general operational parameters. Particularly significant is the implementation of time-related settings including reservation creation advance period, modification deadlines, email reminder timing, and time zone selection which ensures consistent temporal representation despite data being stored in UTC format in the database.

C. Notification Automation

Our system features an automated email notification functionality built using Celery, Redis, and Django's HTML templates. The system delivers three notification types: reservation confirmations sent immediately after booking with all essential appointment details; appointment reminders sent before scheduled appointments with configurable timing and modification notifications that automatically inform both customers and administrators about any changes or cancellations.

Administrators can customize notification settings according to specific business needs, including reminder timing and notification recipients. This system helps minimize missed appointments while maintaining clear communication between service providers and their customers.

VI. Conclusion

Based on an analysis of existing solutions and a survey of user preferences, we have developed a functional system with reservation capabilities for service providers that fulfills established objectives, addresses deficiencies in existing solutions, and introduces new, innovative functionalities.

The developed application is logically divided into customer and administrative sections. The customer section provides intuitive search functionality for services and businesses, transparent reservation processes in logical steps, and visualization of available appointment slots on specific days. The administrative section offers tools for service management, a clear calendar display of reservations with modification capabilities, and a customer management module.

Our solution significantly differs from existing systems through the implementation of public business profiles inspired by social media platforms. This approach creates a familiar environment for users while enabling businesses to present their services, contact information, opening hours, and photo gallery in one location. These profiles thus serve not only for service reservations but also as an effective tool for brand building and establishing online identity, potentially providing a competitive advantage in the digital landscape.

Additional advantages of our system include high flexibility in setting available time intervals for services, allowing the definition of multiple time segments within a single day, including breaks. The system also offers the possibility of modifying existing reservations on the customer side and setting special pricing for specific time periods, which helps optimize utilization during less popular hours. We have also implemented a unique system for sharing services or specific appointments through URL links.

The solution features an automated email notification system that includes sending confirmations after reservation creation, reminders before upcoming appointments, and notifications about changes or cancellations. Administrators can customize the time interval for sending reminders and target recipients of notifications according to their specific business requirements.

Based on the implemented functionalities, we can conclude that the created application fulfills the requirements of the target group identified in the survey. The system successfully integrates all basic functionalities expected from a modern reservation system while adding several innovative elements that can enhance its value for service providers and their customers.

REFERENCES

[1] J. Molina-Ríos and N. Pedreira-Souto, "Comparison of development methodologies in web applications," *Information and Software Technology*, vol. 119, p. 106238, 2020.
[2] X. Li and Y. Xue, "A survey on web application security," *Nashville, TN USA*, vol. 25, no. 5, pp. 1–14, 2011.
[3] J. Shahid, M. K. Hameed, I. T. Javed, K. N. Qureshi, M. Ali, and N. Crespi, "A comparative study of web application security parameters: Current trends and future directions," *Applied Sciences*, vol. 12, no. 8, p. 4077, 2022.
[4] P. Jaiswal and S. Heliwal, "Competitive analysis of web development frameworks," *Sustainable Communication Networks and Application: Proceedings of ICSCN 2021*, pp. 709–717, 2022.
[5] D. Mane, N. Ojha, and K. Chitnis, "The spring framework: An open source java platform for developing robust java applications," *International Journal of Innovative Technology and Exploring Engineering*, vol. 3, no. 2, 2013.
[6] E. W. Dijkstra and E. W. Dijkstra, "On the role of scientific thought," *Selected writings on computing: a personal perspective*, pp. 60–66, 1982.
[7] A. Lock, *ASP. NET core in Action*. Simon and Schuster, 2023.
[8] A. Meier and M. Kaufmann, *SQL & NoSQL databases*. Springer, 2019.
[9] W. Khan, T. Kumar, C. Zhang, K. Raj, A. M. Roy, and B. Luo, "Sql and nosql database software architecture performance analysis and assessments—a systematic literature review," *Big Data and Cognitive Computing*, vol. 7, no. 2, p. 97, 2023.
[10] S. Chen, U. R. Thaduri, and V. K. R. Ballamudi, "Front-end development in react: an overview," *Engineering International*, vol. 7, no. 2, pp. 117–126, 2019.
[11] A. Kumar and R. K. Singh, "Comparative analysis of angularjs and reactjs," *International Journal of Latest Trends in Engineering and Technology*, vol. 7, no. 4, pp. 225–227, 2016.
[12] E. Saks, "Javascript frameworks: Angular vs react vs vue." 2019.
[13] OWASP Foundation, "Owasp top 10 2021," https://owasp.org/Top10/, 2021, accessed: May 2025.
[14] P. Sharma and B. Nagpal, "A study on url manipulation attack methods and their countermeasures," 2015.
[15] 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.
[16] A. Mishra and S. Juneja, "Prevention of website from cross site scripting," in *2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, 2023, pp. 471–473.

Simulácia decentralizovaných zdrojov v rámci distribučnej siete

[1]*Jozef Király,* [1]*Peter Daniel Kelemen*

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

Jozef.kiraly@tuke.sk, peter.daniel.kelemen@student.tuke.sk

***Abstract* — Tento článok sa zaoberá simuláciou prevádzky decentralizovaných zdrojov energie v rámci distribučnej siete. Cieľom je analyzovať vplyv fotovoltických systémov (FV) a technológie elektrických vozidiel v režimoch G2V (Grid-To-Vehicle) a V2G (Vehicle-To-Grid) na výkonovú bilanciu a stabilitu uzla v nízkonapäťovej distribučnej sústave. Pomocou simulačného modelu v prostredí MATLAB Simulink boli analyzované viaceré prevádzkové scenáre. Výsledky ukazujú, že zapojenie FV a V2G systémov dochádza k odľahčeniu siete a stabilizácii výkonových pomerov v uzle.**

***Keywords* — Decentralizované zdroje, Distribučná sieť, Elektromobil, Fotovoltika, MATLAB Simulink**

I. Úvod

Tradičné distribučné sústavy boli pôvodne navrhnuté ako pasívne siete s jednosmerným tokom energie od zdroja ku koncových odoberateľov, postupne sa však pretransformujú na aktívne systémy s obojsmerným tokom energie v reálnom čase [1][2]. V takýchto sieťach zohrávajú dôležitú úlohu decentralizované zdroje energie, ako sú FV systémy a elektrické vozidlá s technológiou G2V a V2G.

Kombinácia FV systémov a elektromobilov ponúka nový rozmer flexibility pre vyrovnávanie výkonovej bilancie, umožňuje lokálne pokrývanie spotreby a potenciálne zníženie zaťaženia siete. Tieto výhody sú však podmienené efektívnym riadením a pochopením dopadov na kvalitu elektrickej energie a stabilitu uzlov distribučnej siete [4][5].

Cieľom článku bolo analyzovať vplyv zapojenia FV systému a elektromobilov v režimoch G2V a V2G na vybratý uzol distribučnej siete. Simulácia bola realizovaná pomocou simulačného modelu v prostredí MATLAB Simulink, vyhodnotené parametre boli napäťové a prúdové pomery a tok výkonu v rôznych prevádzkových režimoch.

II. Popis modelu a metódy simulácie

Simulácia je realizovaná v softvérom prostredí MATLAB Simulink, ktoré umožňuje detailné modelovanie elektroenergetických systémov a analýzu ich správania v rôznych scenároch. Navrhnutý model predstavuje časť distribučnej siete s napäťovou úrovňou 22 kV / 400 V. Model zahŕňa transformátor, dynamickú trojfázovú záťaž, trojfázový zdroj, FV systém a elektrické vozidlá v režimoch V2G a G2V.

Časť modelu s elektromobilmi je založený na architektúre prezentovanej v článku [5], ktorý sa zaoberá V2G technológiou v mikrogrid systéme s rýchlo nabíjaním DC architektúrou. FV systém bol namodelovaný podľa článku [6], kde je detailne vysvetlený návrh FV systému a jeho riadiacich obvodov.

A. Distribučná sieť

Sieť pozostáva z trojfázového sínusového zdroja napätia 22kV (združené, RMS), ktorý napája ideálny transformátor s menovitým prevodom 22 kV/400V. Na nízkonapäťovej strane je napojená dynamicky sa meniaca záťaž s výkonom od 800kW do 1,4MW simulujúca denný diagram zaťaženia. Priebehy prúdu, napätia a výkonu sa vyhodnocujú pomocou meracích blokov na vysokonapäťovej a na nízkonapäťovej strane, na ktorých sú pripojené bloky typu Scope.

B. Fotovoltický systém (FV)

FV systém má menovitý výkon 80 kW a je tvorení 18 paralelnými vetvami s 25 sériovo zapojenými modulmi. Výstupné jednosmerné napätie je prevádzaný cez DC-AC menič s LCL filtrom a je pripojený na nízkonapäťovú stranu siete. Riadenie je zabezpečené pomocou MPPT algoritmu, ktorý sleduje maximálny výkonový bod FV generácie podľa vstupného osvetlenia a teploty.

C. Elektromobily a technológia V2G/G2V

Model obsahuje 4 EV batériové bloky (10,8 kWh), každý z nich je pripojený cez Buck-Boost konvertor umožňujúci obojsmerný tok energie. V simulácii EV vozidlá sú schopné pracovať v režime G2V (nabíjanie batérie z siete) a V2G (spätná dodávka energie z batérie do siete) [5].

Jednoduché prepínanie medzi režimami je možná pomocou prepínača (manual switch) pre každý EV blok.

D. Vstupné signály

Pre zabezpečenie realistického priebehu simulácie boli do modelu zavedené časovo závislé vstupné signály (dynamická záťaž, teplota a slnečné žiarenie), ktoré reprezentujú premenné reálneho prostredia a správania systému. Signály sú načítané z .mat súborov a definované pomocou bloku Signal Editor.

Tento spôsob umožňuje jednoduchú zmenu simulovaných podmienok, je možné testovať rôzne profily bez zásahu do štruktúry samostatného modelu.

III. SCENÁRE SIMULÁCIE

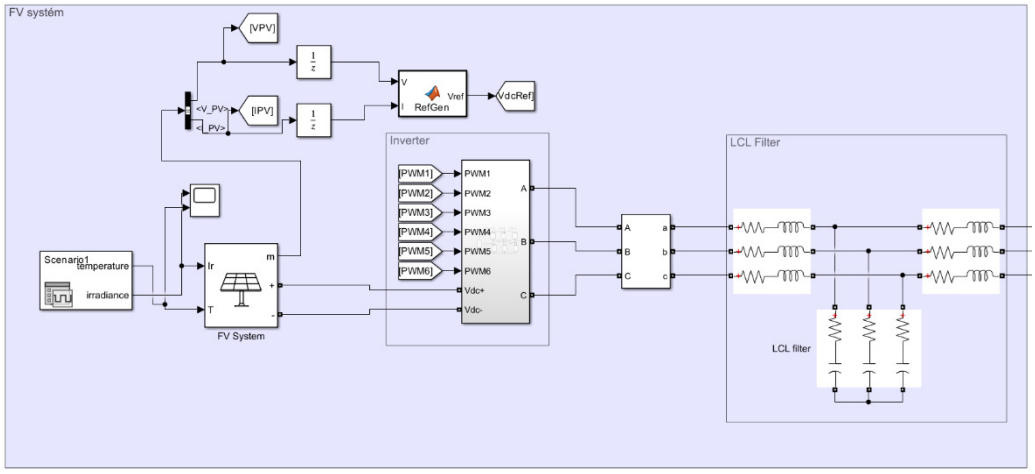
Na overenie vplyvu decentralizovaných zdrojov boli definované štyri rôzne prevádzkové scenáre. Každý scenár predstavuje odlišnú konfiguráciu aktívnych prvkov.

Scenár S1 predstavuje referenčný stav siete, záťaž je napájaná výhradne z distribučnej siete. Tento scenár slúži ako východiskový bod pre porovnávanie s ostatnými scenármi.

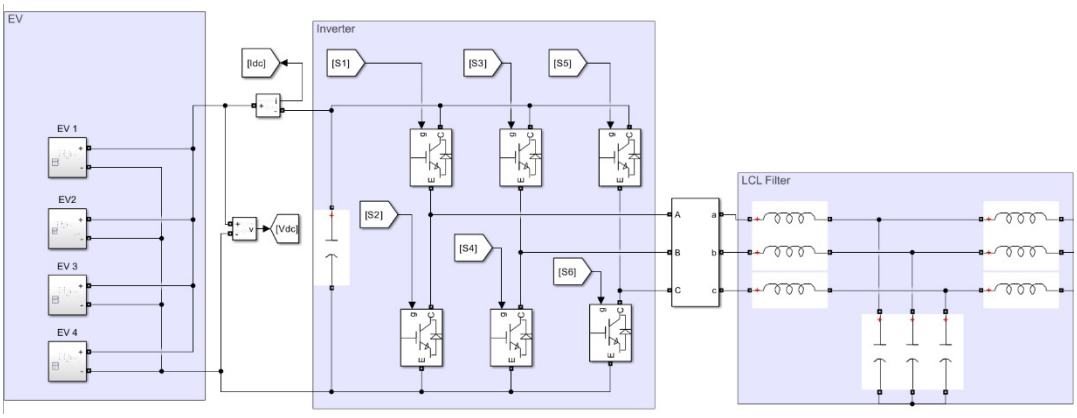
Scenár S2 simuluje stav siete, keď je aktívny len FV systém. Elektrická energia z FV panelov smeruje primárne do lokálnej záťaže, v prípade prebytku energia je dodávaná späť do distribučnej siete. Cieľom je overiť do akej miery dokáže FV zdroj pokryť spotrebu siete.

Scenár S3 kombinuje prevádzku FV systému s nabíjaním elektrických vozidiel (G2V). FV systém napája záťaž a zároveň dodáva výkon do batérie elektromobilov. Táto konfigurácia umožňuje analyzovať interakciu medzi lokálnou výrobou a dynamickou spotrebou.

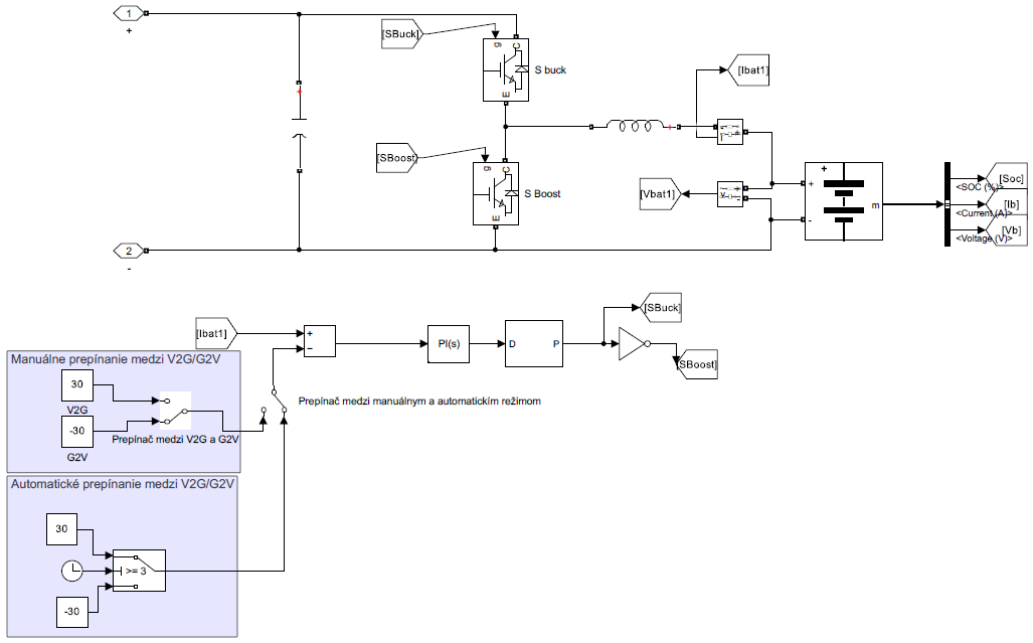
Scenár S4 je najkomplexnejší prevádzkový režim, v ktorom je FV systém aktívny a zároveň prebieha kombinovaná činnosť elektromobilov. Počas simulácie sa časť EV prepne z režimu V2G do G2V. Scenár umožňuje posúdiť schopnosť systému pružne reagovať na aktuálne potreby uzla a demonštruje výhody flexibilného riadenia decentralizovaných zdrojov.



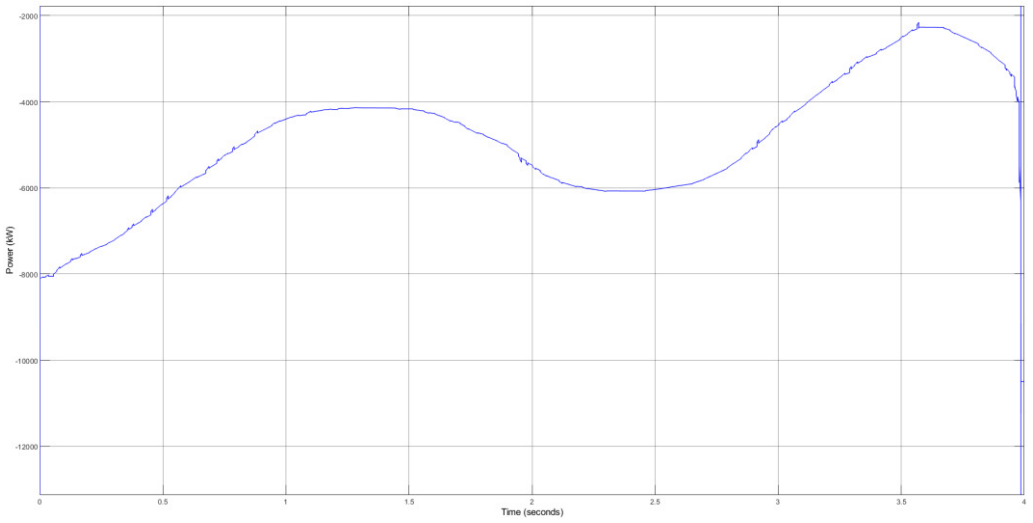
Obr. 1 Schéma zapojenia FV systému



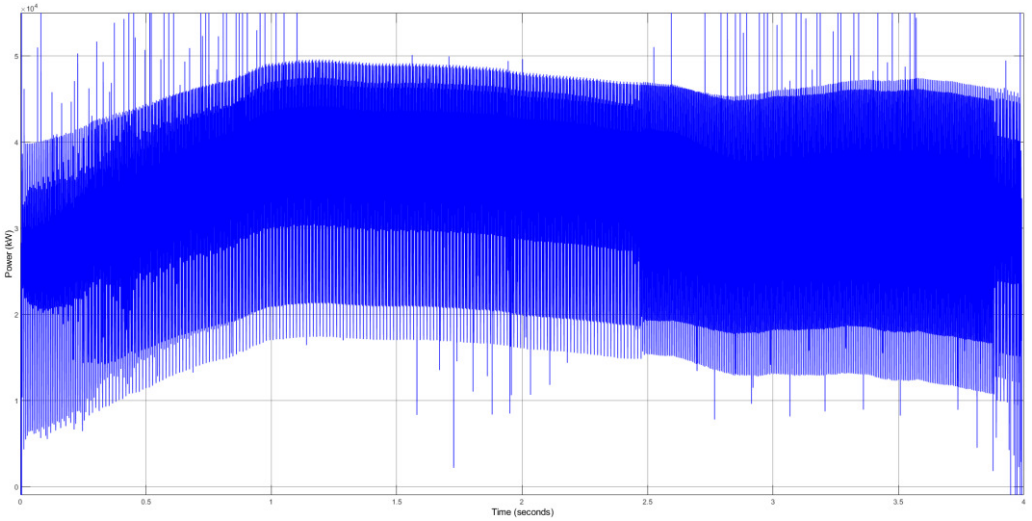
Obr. 2 Schéma zapojenia EV



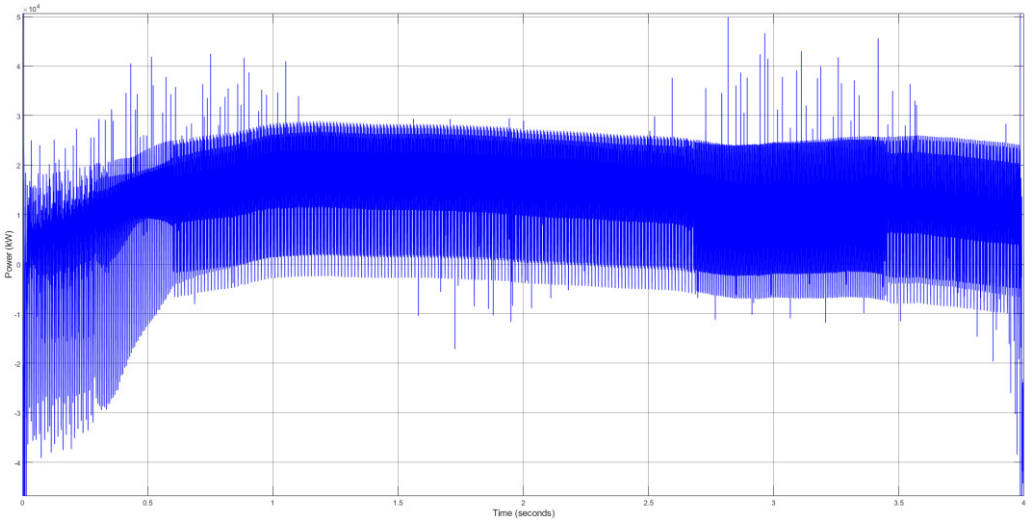
Obr. 3 Scháma Buck-Boost konvertora



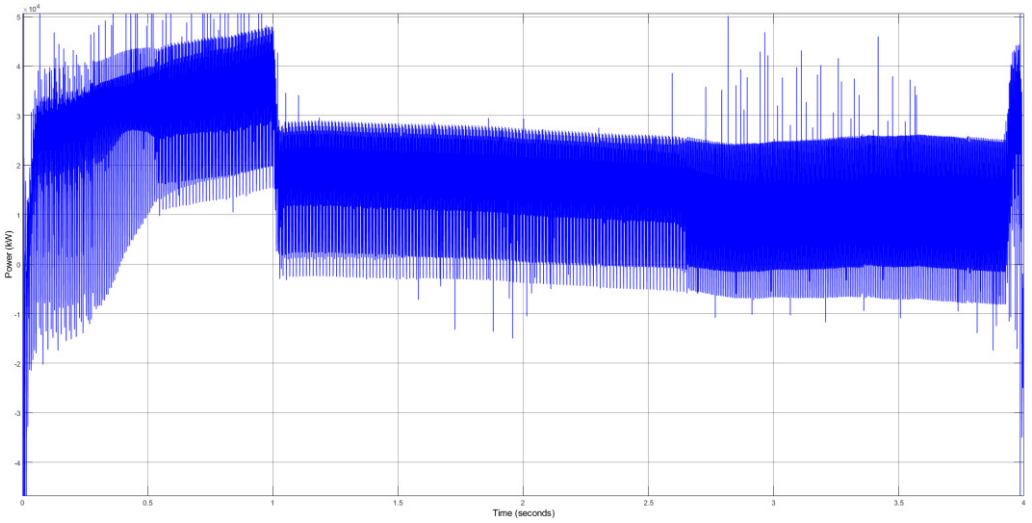
Obr. 4 Priebeh výkonu v scenári S1



Obr. 5 Priebeh výkonu v scenári S2



Obr. 6 Priebeh výkonu v scenári S3



Obr. 7 Priebeh výkonu v scenári S4

IV. VÝSLEDKY A DISKUSIA

Simulácia bola realizovaná pre štyri prevádzkové scenáre, ktoré reprezentujú rôzne konfigurácie zapojenia decentralizovaných zdrojov. Porovnané boli prúdové a napäťové pomery, ako aj smer a veľkosť činného výkonu.

Scenár S1 predstavoval stav siete bez zapojenia decentralizovaných zdrojov, dynamická záťaž bola napájaná len zo siete. Napätie v uzle ostávalo stabilné a blízke menovitej hodnote na nízkonapäťovej aj na vysokonapäťovej strane. Prúd na vysokonapäťovej strane stúpal až na 90 A. Výkon mal záporný charakter od -8 kW do -1,5 kW, ukazuje na jednosmerný tok energie zo siete do záťaže.

Scenár S2, kde bol aktívny len FV systém s výkonom do 80 kW, sa výrazne zmenila výkonová bilancia. Napätie na nízkonapäťovej strane ostávalo stabilné, priebehy boli jemne zdeformované, ale prúd dosahoval hodnoty 100-110 A, čo je dôsledkom lokálneho pokrytia spotreby. Krivky prúdov boli časovo posunuté a zdeformované. Na vysokonapäťovej strane hodnoty ostali približne rovnaké, ako v predchádzajúcom scenári. Výkon mal kladnú hodnotu, 35-45 kW, časť energie bola dodaná späť do siete. Sieť bola v tomto prípade výrazne odľahčená.

Scenár S3 modeloval stav, keď FV systém napájal záťaž a zároveň prebiehalo nabíjanie elektromobilov (G2V). Napätie na nízkonapäťovej strane ostalo rovnaké ako v scenári S2, avšak prúd klesol na hodnoty 75-85 A, krivky boli zdeformované a časovo posunuté. Na vysokonapäťovej strane sa priebehy nezmenili. Výkon sieti sa stabilizoval okolo 20 kW, časť energie potrebné na nabíjanie EV bola pokrytá z lokálneho FV systému.

V scenári S4 bolo súčasne aktívny FV systém a prebiehala kombinovaná prevádzka elektromobilov (dve vozidlá pracovali v režime G2V a dve v režime V2G). Napätie ostávalo stabilné, prúdové zaťaženie na nízkonapäťovej strane bolo spočiatku 100 A, neskôr postupne klesalo na 80 až 70 A v dôsledku prepínania EV z V2G na G2V. Výkon sa stabilizoval okolo hodnoty 35-40 kW, po prepnutí EV výkon klesol na 20 kW. Scenár ukázal schopnosť systému pružne reagovať na zmenu režimu EV a vyrovnať tok energie.

Z výsledkov vyplýva, že integrácia decentralizovaných zdrojov energie do distribučnej siete prináša značný potenciál na zlepšenie výkonovej bilancie, zníženie záťaže z vyššieho napäťového systému a podporu stability siete. Simulačné scenáre preukázali, že FV systémy a elektromobily so schopnosťou obojsmerného toku energie dokážu výrazne ovplyvniť prevádzkové podmienky v siete a môžu sa stať kľúčovými prvkami v prechode na inteligentné a flexibilné energetické systémy. Pre ich efektívne nasadenie je však nevyhnutné zabezpečiť pokročilé riadiace algoritmy, technickú koordináciu zdrojov a dôslednú analýzu ekonomických a regulačných aspektov.

ZOZNAM LITERATÚRY

[1] R. H. Lasseter, *"Smart Distribution: Coupled Microgrids,"* in *Proceedings of the IEEE*, vol. 99, no. 6, pp. 1074-1082, June 2011.

[2] Hatziargyriou, Nikos, ed. *"Microgrids: architectures and control,"* John Wiley & Sons, 2014.

[3] Kempton, Willett, and Jasna Tomić, *"Vehicle-to-grid power fundamentals: Calculating capacity and net revenue."* Journal of power sources 144.1 (2005): 268-279

[4] V. Gupta, A. K. Maurya and H. Ahuja, *"Vehicle to Grid Technologies Resolving Micro Grid Power Intermittencies Issues," 2023 9th IEEE India International Conference on Power Electronics (IICPE)*, SONIPAT, India, 2023, pp. 1-5.

[5] F. M. Shakeel and O. P. Malik, *"Vehicle-To-Grid Technology in a Micro-grid Using DC Fast Charging Architecture," 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, Edmonton, AB, Canada, 2019, pp. 1-4

[6] Das, Shamik Kumar, M. K. Sinha, and A. K. Prasad*, "Designing and simulation of double stage grid-linked PV system using MATLAB/simulink.",* Recent advances in Power Systems: Select Proceedings of EPREC 2022, Singapore: Springer Nature Singapore, 2023, 169-195.

Procedural Generation of Polygonal Surface Models of Plants for Computer Games

[1]*Tomáš ČIGÁŠ,* [2]*Branislav MADOŠ*

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

[1]tomas.cigas@student.tuke.sk, [2]branislav.mados @tuke.sk

***Abstract* — This paper focuses on procedural generation of content, with a focus on procedural generation of polygonal surface models of plants for computer games. Special attention is given to the use of L-systems and their stochastic variants as a tool for modeling the branching structure and simulation of plant growth. The practical application consists of creating a structural skeleton, polygonal mesh and various components a plant has, such as flowers and leaves. The work includes various types of plants, which this algorithm allows to generate, as well as framework, which can be used to animate them. Program implementation is created using Unity engine, however it is possible to use this algorithm in any game engine including Unreal.**

***Keywords* — procedural content generation, model of plant, L-systems, Unity**

I. Introduction

In the last decade the gaming industry has become a large field in the creation of digital entertainment. The creation of games has also introduced many advanced modeling techniques as well as complex algorithms, which not only have an impact on the creation of games, but also many other programming fields, which use graphic interface. The goal of this paper is to introduce a way how L-systems can create complex models of plantation, mainly for the purpose of filling the environment of games. L-systems are largely used to create vegetation in many industries. Many, however, do not use a fully procedural system, which can create multitudes of same plant, without the need for additional tweaks to the configuration. This paper hence serves as a form of guide for anyone who wishes to make their own generator of content in games.

II. Concepts required for implementation

This section of the paper details concepts that are utilized for implementation of algorithm for procedural content generation of polygonal surface models of plants.

A. Game Engines

Game engines are comprehensive software frameworks that streamline video game development by providing essential tools such as rendering, physics, input management, and scripting. Among the leading engines today are Unreal Engine, Unity, and Godot, each catering to different development needs.

Unity is a versatile, cross-platform engine widely used in both 2D and 3D game development. It supports C# scripting and offers an extensive asset store, making it particularly popular for mobile games, indie projects, and virtual or augmented reality applications [1].

Unreal Engine, developed by Epic Games, is known for its high-end graphics and is frequently used in AAA games and real-time simulations. It supports C++ programming language and a visual scripting system called Blueprints, offering powerful tools for creating detailed and immersive 3D environments [2].

Godot is an open-source engine favored for its lightweight structure and strong 2D capabilities. It features GDScript, a Python-like language, and supports C# as well. Its open development model and permissive license make it a popular choice among independent developers seeking flexibility and control [3].

B. Polygonal mesh

A polygonal mesh is a fundamental building block of 3D graphics used in the creation of objects in three-dimensional games. It consists of a collection of polygons that together form the surface of a 3D object. Each polygon is defined by vertices, which have coordinates in space. Specifically positions along the x, y, and z axes. These vertices are then connected by edges to form polygonal surfaces [4].

C. L-systems

The Lindenmayer system (L-system) is a formal grammar used to model multicellular organisms. However, its application extends beyond plant growth to the generation of fractals and other natural structures. It was developed in 1968 by Hungarian biologist Aristid Lindenmayer. L-systems are classified into different types based on their rules and properties used in the creation of objects [5].

III. IMPLEMENTATION OF PROCEDURAL PLANT GENERATION

Procedural content generator of polygonal surface models of plants, designed and implemented within our research, is described in this section of the paper.

A. L-system grammar

Grammar used in this approach stems from stochastic L-systems. In principle, the plant is described as a string of characters, where each character represents a particular part of a plant. The complete configuration of a plant consists of an axiom, or the initial string of characters. This axiom is then extended by applying user defined rules, which replace an existing character or string of the configuration. This process applies the rules on an axiom multiple times. This approach allows to create a complex structure of a plant, with only a couple of simple rules [5].

For example, with an axiom:

$$\mathbf{F} \tag{1}$$

And a rule:

$$\mathbf{F = F+F-F} \tag{2}$$

The complete configuration with multiple recursions, or the count of rule application will expand as follows:

1. **F+F-F**
2. **F+F-F+F+F-F-F+F-F** (3)
3. **F+F-F+F+F-F-F+F-F+F+F-F+F+F-F- F+F-F-F+F-F+F+F-F-F+F-F**

L-systems ensure an exponential growth of the plant structure even with a simple axiom and a single rule.

To enforce the randomness of plant growth into the algorithm, a stochastic L-system is introduced. In other words, each rule has a probability of appearing on the resulting configuration. With this approach the algorithm can create different plants without changing any of the rules. Each rule is therefore described as:

$$A(w) = B \tag{4}$$

Where:

- A – is the character on which the rule applies, or simply which will be replaced
- B – is the string which will replace the A
- w – is the weight of the rule. Meaning how often this rule will appear in relation to other rules with same A

In this implementation there are three predefined symbols which correspond to a certain part of flower. Character *F* is the stalk, *K* is the flower and *L* is a leaf. Even this simple trio can create various plants simply by changing a part of configuration.

Each rule also has supporting characters, which further define the relationship between each part of the plant. The first pair of such characters are " + " and " - " signs. These signs describe the deviation of the next part by a certain degree from the parent node in either one or the other direction. This ensures a more natural look of a plant. The second pair of supporting characters are " [" and "] ". These describe the branching of the plant. Since many plants do not consist of

only one stalk and one flower this functionality accomplishes further organic look and description of plant. The character " [" is the beginning of the new branch and "] " is its end.

B. Non-grammatical parameters

Several parameters contribute to the overall generation, determining what the final plant will look like, independently of the grammar itself. Since the grammar describes how the parts of plant will correlate to each other. These non-grammatical parameters determine how each part of the plant will appear. These parameters are split into four categories depending on what they control

The first category describes the position of the new parts in relation to the previous one. This is accomplished using azimuth and elevation angles. Their relation and how they create the new position is visible on the *Figure III-1*. Where α is the azimuth and β is the elevation angle.

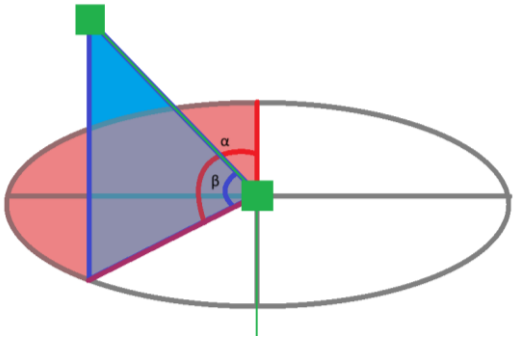


Figure III-1 Azimuth and Elevation angles.

The second category is dedicated to the distances between various nodes of a plant. Meaning the length of stalk or size of leaves and flowers can vary between each configuration of plant, depending on the preferences and goals of the user.

The third category consists of the rendering parameters of the plant. These attributes define which textures the flowers, leaves and stalk will have, as well as which type of flowers and leaves will be used. There are multiple detail parameters which define how the stalk will behave, when it approaches the top of the plant, as well as how the leaves will be attached to the stalk itself. The resolution of the plant can be modified in this category as well.

To ensure the stochastic functionality some sort of randomness needs to be implemented. Complete randomness can however lead to various problems when the user needs to create an identical plant for testing purposes or if they want to introduce a level of detail functionality into their system. Hence the fourth category handles whether the randomness will be used in generation, as well as the random seed of the plant. The seed is a simple integer which defines the particular created plant.

IV. PROCESS OF CREATION

When the user configures the complete description of a plant the creation of a model can begin. The whole process encapsulates the transformation of configuration into 3D space with all required nuances such as how the stalk, flowers and leaves will look like in the final result.

A. Skeleton structure

The process begins by creating a skeleton of a plant. Each part of a plant is described as a node with its position and type. This way a complete structure can be created before the polygonal mesh is created as well as leaves and flowers are placed on the plant. It also accomplishes the possibility of animating the plant simply by moving the nodes of the plant. For the purposes of visualizing this skeleton the *Figure IV-1* contains multiple plants with exactly same configuration. The cubes represent nodes of the plant. Green color represents the stalk, blue represents flowers and light green represents leaves.

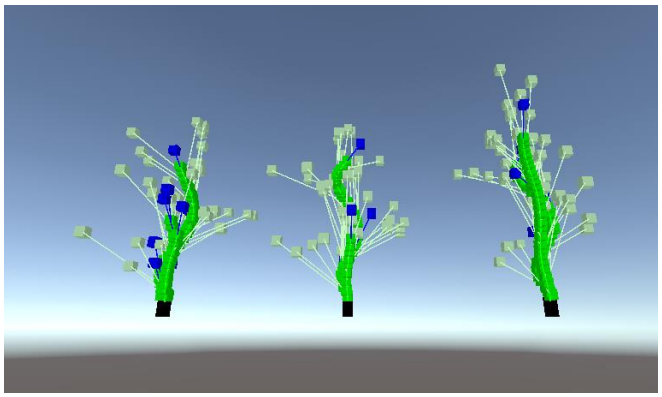


Figure IV-1 Skeleton structure of plants.

B. Ring of vertices

Upon completion of the skeleton, the algorithm creates a mesh around stalk nodes. Each node creates a ring of vertices which are then connected via triangles, hence completing the mesh. Since the stalk grows dynamically, the ring needs to adjust to the direction and branching of the plant. For this purpose, each ring is rotated towards the next node. The complete process of creating a ring with a single child is quite simple. A ring of vertices is created using goniometrical functions. The ring is therefore created parallel to the ground. Then each vertex is rotated using a quaternion. Quaternions describe rotation in 3D space, whilst ensuring no side effects, such as gimbal-lock, hence why they were used. A simple representation of how the quaternion describes the rotation of the whole ring is shown on *Figure IV-2*. Where **U** is the vector perpendicular to the ground, vector **F** describes the relative position of the child node and q is the quaternion that represents the rotation. The ring is represented as line k and k' is the original ring of vertices.

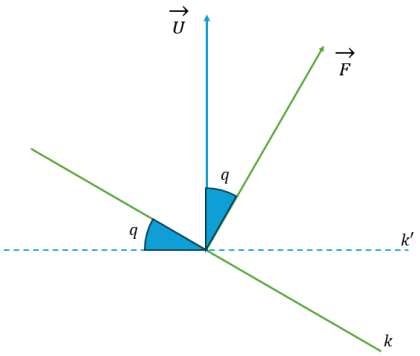


Figure IV-2 Rotation of ring using quaternion q.

While the rotation of ring with single child is simple, when introduced to the branching mechanism it creates an unnatural look in junctions of the nodes. The workaround consists of creating a separate rotation of each child node and rotating the vertex closest to the child by its corresponding quaternion. The child stalks can create unnatural looking junctions. Hence supporting vertices which define how the children will connect to each other in this junction need to be created. *Figure IV-3* shows the black and gray vertices belonging to each child. The supporting vertices are shown as red.

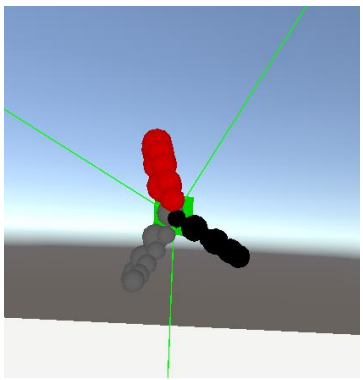


Figure IV-3 Vertices in branching of stalk nodes.

For better modeling of plant growth, the stalk can reduce towards the top of the plant dynamically. For better control of reduction, the user can control how much the stalk will reduce on branching, in its natural growth (meaning when a node has only one child) and on its end, as well as its minimum reduction, so the stalk does not disappear. The complete application of all steps mentioned above result in a natural looking stalk showcased on *Figure IV-4*.

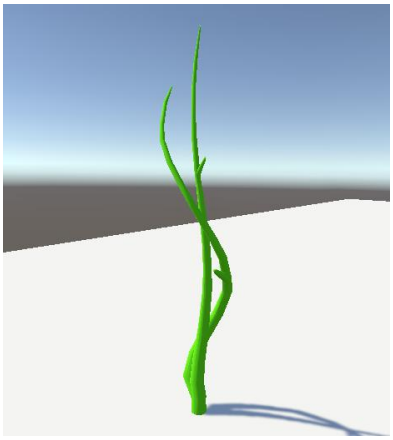


Figure IV-4 Complete stalk.

C. Auxiliary objects

Upon completion of the stalk the process of creating and placing leaves and flowers may begin. The process of placement is similar for both parts.

Differences are in the meshes of those objects. Leaves can be represented by a simple plane or a slanted plane, which has a V shape. Such leaves can be seen on *Figure IV-5*.

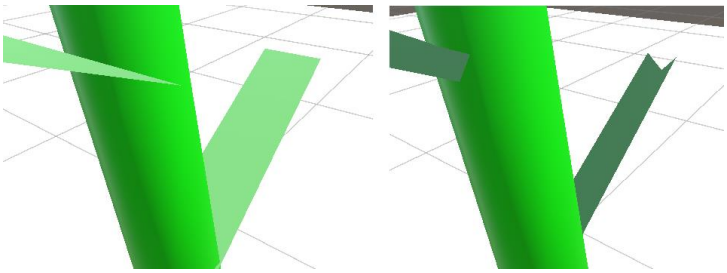


Figure IV-5 Representation of leaves.

The flowers can be represented via a billboarding technique used in many computer games, and other kinds of visualizations, to represent the whole plants. This approach allows to create realistically looking flowers without using complex shapes, which worsen the load of the whole system [6]. The creation of such billboard flower consists of creating intersecting planes on which a texture can be applied. The algorithm how to create these planes in a circle is similar to the creation of vertex ring described in the definition of the stalk. The flower will in result acquire shape shown on *Figure IV-6*.

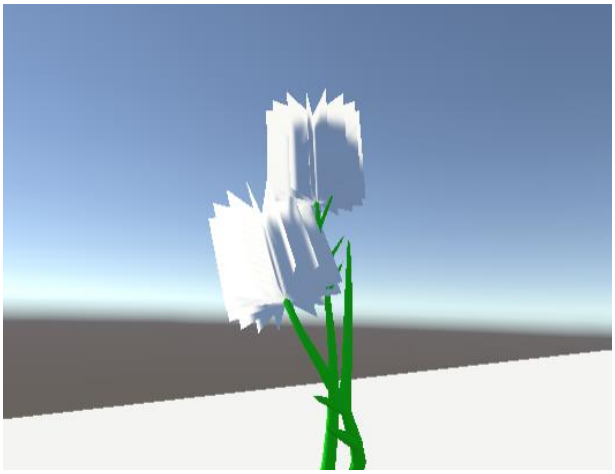


Figure IV-6 Billboard flower.

To further expand the visual capabilities of plants a spherical flower has been added to the repertoire of this algorithm. Since many flowers have multiple petals not only on the outside, but also on the inside, numerous spheres need to be inserted into the center of the main sphere, creating a flower with more depth. Any sphere algorithm can be used to create such flowers. The main functionality must be however, to create a triangle mesh on the outside as well as on the inside of the sphere to ensure the flower is visible from any angle. Such depiction of multiple spheres can be seen on *Figure IV-7*. A simple texture of rose flower has been applied to further showcase the double-sided spheres.

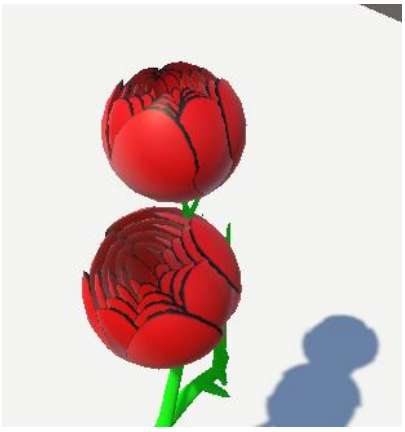


Figure IV-7 Sphere flower with rose texture.

D. Performance

The algorithm for procedurally generating plants is slower in performance compared to conventional methods of inserting pre-defined plants, which is expected since the program must simulate the entire growth process, the positioning of parts, and the overall shape of the plant. To assess the algorithm's efficiency, the time (in milliseconds) required to generate individual plant components was measured, along with the total number of generated vertices and triangles in the polygonal mesh. This testing was done on 50 copies of a single model with the average time required for their generation being included in the calculation.

Figure IV-8 showcases the results of testing stalk generation. The generation depends on the resolution of the stalk itself, meaning how many vertices are located on each ring. Time of generation as well as number of vertices and triangles grows linearly with increasing resolution.

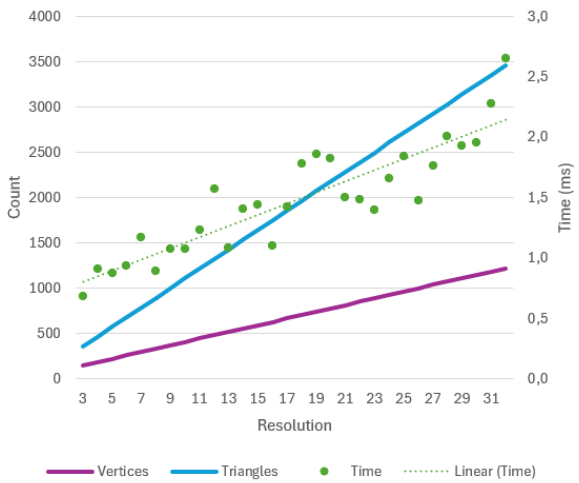


Figure IV-8 The dependence of geometric complexity and computation time on the resolution of the generated stalk.

The performance of creation of the two types of flower models the algorithm can create is shown on *Figure IV-9* (intersecting planes) and *Figure IV-10* (spheres). The plane flower results increase linearly with increasing number of planes, while results of spherical flower increase exponentially. The algorithm used for sphere generation describes resolution as number of vertices on the circumference of the sphere, hence the exponential growth.

In case of leaves there is no resolution, since each leaf consists of a single plane. There is a slight difference between the types of leaves shown on *Table IV-1*. Since the algorithm was constructed in Unity engine, the creation of the leaf objects hinders the performance by 0.02 milliseconds in the average.

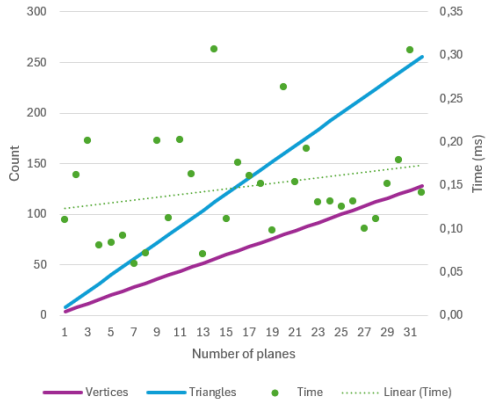


Figure IV-9 The dependence of geometric complexity and computation time on the resolution of the billboard flower.

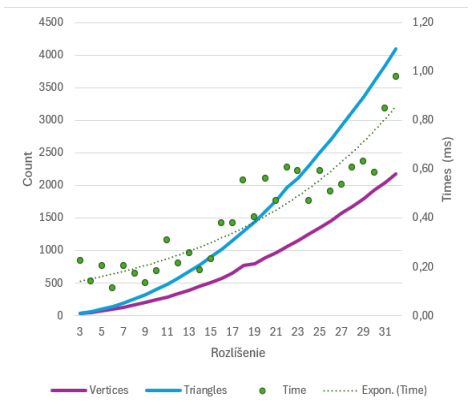


Figure IV-10 The dependence of geometric complexity and computation time on the resolution of the spherical flower.

Table IV-1 Type of leaf and its impact on performance

Type	Time (ms)	Vertices	Triangles
Plane	0.032013	4	4
V-shape	0.039601	6	8

The generation of the grammar and the skeleton itself — that is, the positioning of individual parts of the stem, flowers, and leaves — also needs to be considered in terms of performance and efficiency. For comparison, a model of complete plant was used, while the number of recursions the algorithm should perform during generation has varied. In other words, how long it takes the algorithm to convert the grammar into the overall configuration, and subsequently, how much time does the transformation of this grammar into positions in 3D space take. This relation can be seen on *Figure IV-9*.

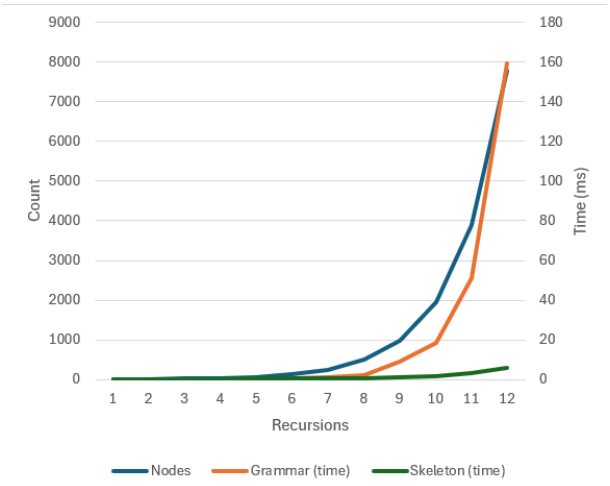


Figure IV-9 The dependence of computation time on the number of recursions in the generated plant.

V. Conclusion

The result of the work is a complex system that enables the creation of diverse plants, which can be integrated into a game environment and further developed in the context of interactive applications, games, or other types of visualizations. A couple of examples of full plants can be found on *Figure V-1*. The system can handle multiple types of plants depending on the rules and initial parameters the user wishes to implement.

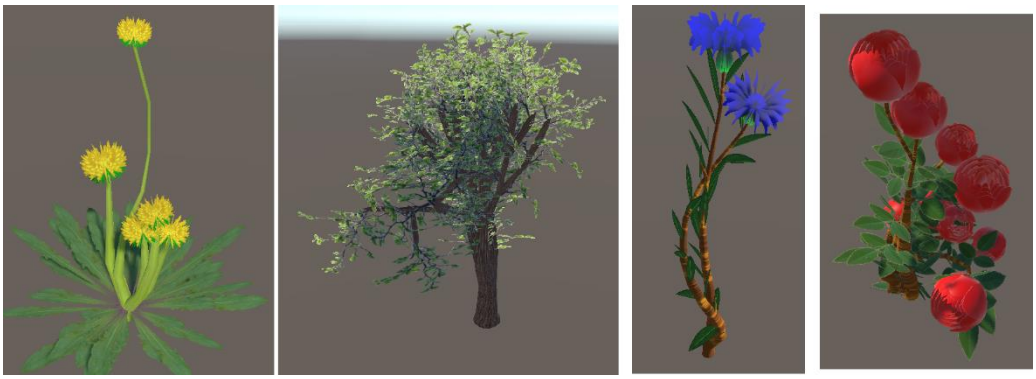


Figure V-1 Examples of flowers that the system can create.

In conclusion this system can create many variants of plants without further modification of code. There are many improvements which can be added on top of this system, such as better grammar conversion, more types of leaves or flowers, stalk smoothening algorithms etc.

References

[1] Unity Technologies, “Unity Documentation”, [Online], cited: 31.5.2025, available: https://docs.unity.com/.
[2] Epic Games, “Unreal engine 5.5 documentation”, [Online], cited 31.5.2025, available: https://dev.epicgames.com/documentation/en-us/unreal-engine/unreal-engine-5-5-documentation.
[3] J. Linietsky and A. Manzur, "Godot Docs", [Online], cited: 31.5.2025, available: https://docs.godotengine.org/en/stable/.
[4] M. Botsch, L. Kobbelt, M. Pauly, P. Alliez and B. Levy, “Polygon Mesh Processing”, New York: A K Peters/CRC Press, 2010, pp 250, eISBN: 9780429195709, https://doi.org/10.1201/b10688.
[5] P. Bourke, “L-System User Notes”, 1991. [Online], cited: 31.5.2025, available: https://paulbourke.net/fractals/lsys/.
[6] A. R. Fernandes, "Billboarding Tutorial", 2009, [Online], cited: 31.5.2025, available: http://www.lighthouse3d.com/opengl/billboarding/

Analýza stability elektrického odporu vnorených spájkovaných spojov

Tomáš LENGER, Peter LUKÁCS

Katedra technológií v elektronike, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Košice, Slovenská republika

tomas.lenger@tuke.sk

***Abstrakt* — Táto práca je zameraná na stabilitu elektrického odporu vnorených spájkovaných spojov. Spájkované spoje sú neoddeliteľnou súčasťou každej dosky plošných spojov a stabilita ich elektrických vlastností je kľúčovým prvkom spoľahlivosti dosiek plošných spojov. Táto práca obsahuje porovnávaciu štúdiu stability elektrického odporu spájkovaných spojov realizovaných na povrchu a vnorených spájkovaných spojov. Pre tento výskum boli použité tri rôzne spájkovacie pasty (Sn96.5Ag3Cu0.5 SAC305, Sn96.5Ag3.5 a Sn89.2Ag3.5Bi0.5In6Cu0.8). Stabilita elektrického odporu bola analyzovaná počas tepelných šokovania. Rozsah teploty pri tepelnom šokovaní bol od -45°C do 125°C a realizovaných bolo 1000 teplotných šokov. Meranie elektrického odporu bolo realizované pomocou 4-vodičovej metódy (Milliohmmeter Agilent 4338B). Meranie elektrického odporu bolo vykonané pred teplotným šokovaním, ako aj po každých 100 teplotných šokoch.**

***Kľúčová slová* — elektrický odpor, vnorené súčiastky, spájkované spoje**

I. Úvod

Nezastaviteľný vývoj elektroniky si vyžaduje čoraz zložitejšie a miniaturizované elektronické zariadenia. Zložitosť Dosiek Plošných Spojov (DPS) zvyšuje počet elektronických súčiastok, ktoré sú na nich zostavené. Zvyšujúci sa počet elektronických súčiastok zvyšuje aj počet spájkovaných spojov. Vnáranie súčiastok do substrátu umožňuje zvyšovanie zložitosti elektronických zariadení bez zväčšovania ich rozmerov [1], [2], [3].

V posledných rokoch hľadali dizajnéri inovatívne spôsoby osádzania súčiastok na DPS. Hoci počet spojov realizovaných vodivými lepidlami rastie, najmä vo flexibilnej elektronike, spájkované spoje sú stále najspoľahlivejšími vodivými spojmi pre elektronické zariadenia. Spájkované spoje musia zabezpečiť mechanické a elektrické spojenie medzi elektronickými súčiastkami a kvalita spájkovaných spojov je rozhodujúca pre spoľahlivosť každého elektronického zariadenia. Elektronické zariadenia s vysokou hustotou vyžadujú spoľahlivé spoje, aby sa zabezpečil vysokorýchlostný prenos dát a integrita signálu [4], [5].

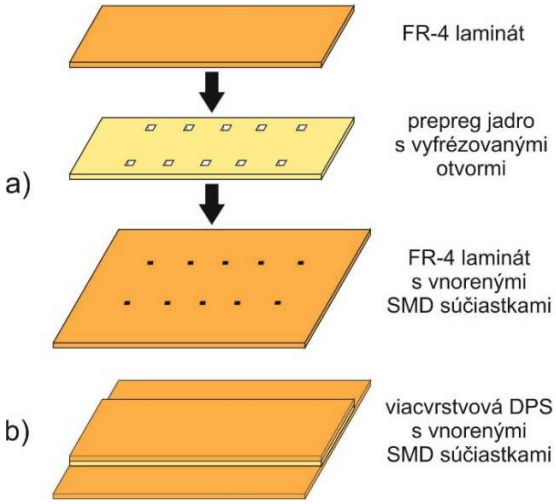
Miniaturizácia elektronických zariadení a vnáranie elektronických súčiastok do substrátov prináša nové technologické výzvy. Vnorené spájkované spoje, ako aj vnorené elektronické súčiastky zvyšujú počet materiálov zahrnutých v objeme substrátu. Na druhej strane, súčiastky pre povrchovú montáž (SMD) osadené na povrchu, ktoré sú následne laminované do viacvrstvových DPS nevyžadujú drahé inovácie pre výrobu [6], [7].

Vnorené súčiastky zvyšujú hustotu osadených súčiastok a umožňujú miniaturizáciu elektronických zariadení. Spájkované spoje a SMD súčiastky však obsahujú odlišné materiály ako lamináty na báze epoxidových substrátov. Rozdiely v tepelných vlastnostiach použitých materiálov, najmä v koeficiente teplotnej rozťažnosti (CTE), môžu viesť k zvýšenému mechanickému namáhaniu vo vnútri DPS s vnorenými súčiastkami. Vnútorné mechanické namáhanie sa môže znásobiť rôznymi tepelnými vlastnosťami použitých materiálov. Vnorené súčiastky nie je možné vymeniť počas životnosti elektronických zariadení, čo zvyšuje dôležitosť ich spoľahlivosti [8], [9], [10].

Spoľahlivosť elektronických zariadení a stabilita ich vlastností je testovaná pomocou testov zrýchleného starnutia [11]. Cieľom tejto práce je analyzovať stabilitu elektrického odporu troch rôznych spájkovacích pást počas teplotného šokovania. Stabilita elektrického odporu je analyzovaná porovnávacou štúdiou medzi spájkovanými spojmi realizovanými na povrchu a vnorenými do viacvrstvovej DPS.

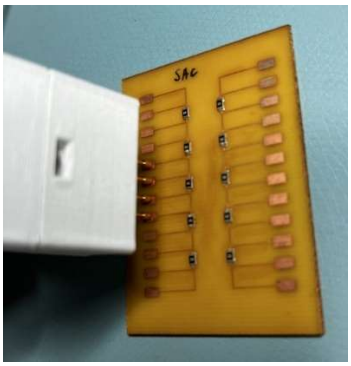
II. NÁVRH A PRÍPRAVA VZORIEK

Táto práca sa zameriava na stabilitu spájkovaných spojov vnorených do viacvrstvovej štruktúry. Viacvrstvová štruktúra pozostáva z dvoch FR-4 laminátov a jadra z prepregu medzi nimi (Obr. 1a). SMD súčiastky boli osadené na vnútornej strane spodného laminátu.



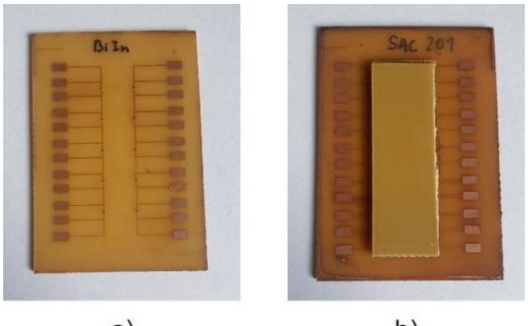
Obr. 1 Štruktúra viacvrstvovej DPS s vnorenými súčiastkami: a) pred laminiáciou, b) po laminácii

Viacvrstvová štruktúra bola laminovaná po osadení SMD súčiastok. Ako substrátový materiál pre tento experiment bol použitý materiál Isola DE104. Viacvrstvová štruktúra bola laminovaná konštantným tlakom (2 MPa) a tepelným profilom podľa odporúčania výrobcu. Rýchlosť nábehu teploty bola 2°C/min až do teploty vytvrdzovania 170°C. Teplota 170°C bola udržiavaná počas 50 minút, po ktorej nasledovalo ochladzovanie rýchlosťou približne 3°C/min. Laminácia spôsobí vytvrdzovanie prepreg materiálu a viedla k vytvoreniu viacvrstvovej štruktúry s vnorenými súčiastkami (Obr. 1b).



Obr. 2 Meranie elektrického odporu 4-vodičovou metódou

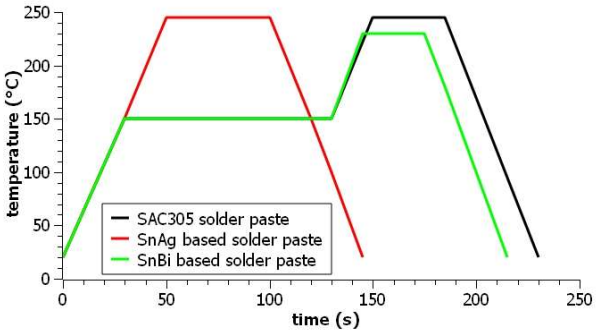
Dizajn vzoriek bol prispôsobený na meranie elektrického odporu 4-vodičovou metódou. Rozmery spodného laminátu boli väčšie ako rozmery jadra a vrchného laminátu. Navrhnutá viacvrstvová DPS umožňuje meranie elektrického odporu vnorených spájkovaných spojov bez zvýšenia zložitosti štruktúry viacvrstvovej DPS. Meranie elektrického odporu bolo realizované miliohmmetrom Agilent 4338B. Merací prípravok pre 4-vodičovú metódu merania bol navrhnutý a realizovaný pomocou 3D tlače (Obr. 2), aby sa zabezpečili konštantné podmienky merania a opakovateľnosť experimentu.



Obr. 3 Realizované vzorky: a) osadené na povrchu substrátu, b) vnorené do viacvrstvovej DPS

Stabilita elektrického odporu bola analyzovaná porovnaním SMD súčiastok osadených na povrchu (Obr. 3a) a vnorených do viacvrstvovej štruktúry (Obr. 3b). V tomto experimente boli použité rezistory s nulovým odporom v SMD puzdre 0201. Každá realizovaná DPS obsahovala desať osadených SMD súčiastok a pre každý typ vzorky boli realizované tri DPS.

Okrem toho boli použité tri rôzne zloženia spájkovacích pást. Sn96.5Ag3Cu0.5 (SAC305) predstavuje najpoužívanejšiu spájkovaciu pastu vo výrobnom procese DPS. Boli použité dve ďalšie spájkovacie pasty s porovnateľnými teplotami tavenia. Spájkovacia pasta Sn96.5Ag3.5 (spájkovacia pasta na báze SnAg) má rovnakú teplotu pretavenia ako SAC305 (245°C), táto pasta ale nevyžaduje fázu predohrevu, čo skracuje celkový čas pretavenia. Sn89.2Ag3.5Bi0.5In6Cu0.8 (spájkovacia pasta na báze SnBi) má porovnateľný profil spájkovania ako SAC305 avšak s nižšou maximálnou teplotou. Spájkovacie pasty boli pretavené podľa odporúčaní výrobcov spájkovacích pást (Obr. 4).

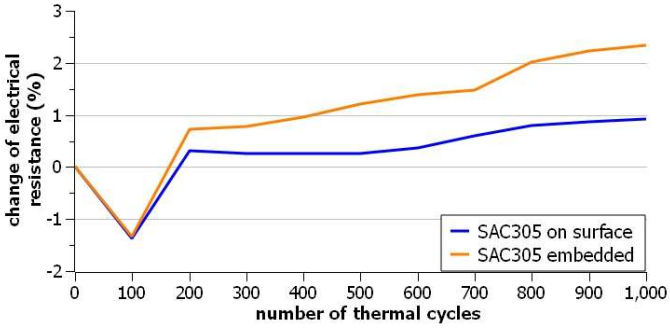


Obr. 4 Profily pretavenia použitých spájkovacích pást

Stabilita elektrického odporu bola analyzovaná počas testov zrýchleného starnutia. Zrýchlené starnutie navrhnutých DPS bolo realizované teplotným šokovaním. Teplotné šokovanie bolo realizované 20 minút v horúcej komore (125°C) a následne 20 minút v studenej komore (-45°C) v systéme Clima Temperatur System TSS – 70 podľa normy IPC-9701. Meranie elektrických vlastností bolo vykonané po každých 100 tepelných šokoch, až kým nebolo dosiahnutých 1000 tepelných šokov.

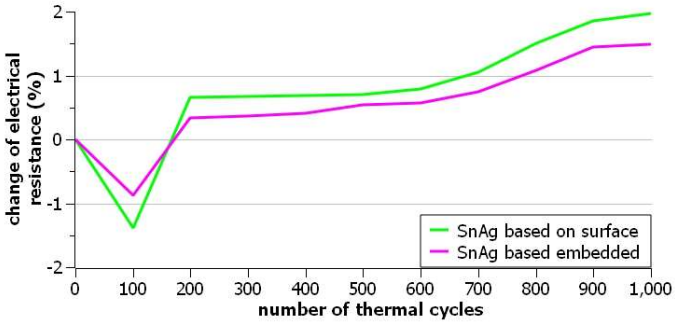
III. Výsledky a diskusia

Dosiahnuté výsledky predstavujú trendy zmien elektrického odporu počas teplotného šokovania vzoriek. Obr. 5 znázorňuje zmenu elektrického odporu spájkovaných spojov na báze spájkovacej pasty SAC305 počas teplotného šokovania. Počiatočná hodnota pre spájkované spoje SAC305 realizované na povrchu bola 31,12 mΩ, zatiaľ čo pre vnorené spájkované spoje to bolo 33,11 mΩ. Hodnoty elektrického odporu predstavujú priemer 30 meraných SMD súčiastok. Elektrický odpor vzoriek osadených na povrchu sa po 100 tepelných šokoch znížil o 1,37%. Pokles elektrického odporu bol takmer rovnaký pre vnorené spájkované spoje (1,34 %). Pokles elektrického odporu je pravdepodobne spôsobený nárastom intermetalických zlúčenín (IMC) medzi spájkovacou zliatinou a medenou spájkovacou podložkou. Väčší nárast elektrického odporu bol pozorovaný medzi 100 a 200 teplotnými šokmi. Elektrický odpor spájkovaných spojov sa po 200 tepelných šokoch zvýšil na vyššie ako počiatočné hodnoty. Po 200 tepelných šokoch elektrický odpor testovaných spájkovaných spojov kontinuálne stúpal až do 1000 teplotných šokov pre oba typy vzoriek. Rozdiel v elektrickom odpore (od 0 teplotných šokov do 1000 teplotných šokov) bol vyšší v prípade vnorených spájkovaných spojov (2,33 %), zatiaľ čo pre spájkované spoje realizované na povrchu bol rozdiel iba 0,92 %. Vnorenie spájkovaných spojov zvýšilo zmenu elektrického odporu viac ako 2-násobne v porovnaní so spájkovanými spojmi na báze SAC305.



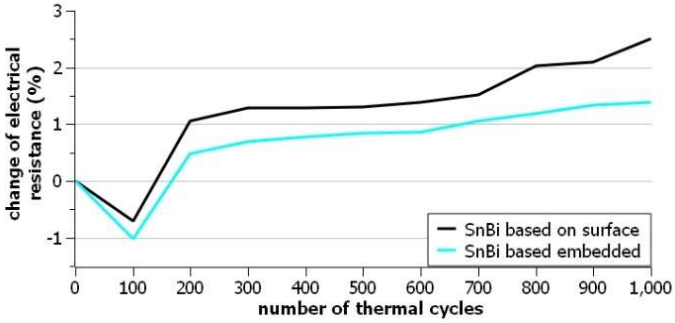
Obr. 5 Zmena elektrického odporu spájkovaných spojov na báze SAC305

Obr. 6 znázorňuje zmenu elektrického odporu spájkovacej pasty na báze SnAg počas teplotného šokovania. Počiatočná hodnota pre spájkované spoje na báze SnAg realizované na povrchu bola 33,05 mΩ, zatiaľ čo pre vnorené spájkované spoje bola 32,74 mΩ. Pokles elektrického odporu spájkovaných spojov na báze SnAg realizovaných na povrchu bol 1,39% po 100 teplotných šokoch, zatiaľ čo pri vnorených bola zmena 0,87%. Testované vzorky vykazovali najvyšší nárast elektrického odporu medzi 100 a 200 tepelnými šokmi. Zvýšenie elektrického odporu spájkovaných spojov na báze SnAg bolo 2,05% medzi 100 a 200 tepelnými šokmi, zatiaľ čo vnorené spájkované spoje mali zmenu elektrického odporu o 1,21%. Elektrický odpor vykazuje minimálnu zmenu pre obe testované vzorky medzi 200 a 600 teplotnými šokmi. Zvýšenie elektrického odporu bolo výraznejšie po 600 tepelných šokoch pre oba typy vzoriek, pričom najvyšší rozdiel od počiatočnej hodnoty bol pozorovaný po 1000 teplotných šokoch. Zmena elektrického odporu bola 1,97% pre spájkované spoje na báze SnAg realizované na povrchu a 1,49% pre vnorené spájkované spoje na báze SnAg.



Obr. 6 Zmena elektrického odporu spájkovaných spojov na báze SnAg

Obr. 7 znázorňuje zmenu elektrického odporu spájkovacej pasty na báze SnBi počas teplotného šokovania. Počiatočná hodnota pre spájkované spoje na báze SnBi realizované na povrchu bola 32,43 mΩ, zatiaľ čo pre vnorené spájkované spoje bola 35,67 mΩ. Spájkované spoje na báze SnBi vykazovali rovnaký trend poklesu elektrického odporu po 100 teplotných šokoch. Pokles elektrického odporu bol najmenší zo všetkých testovaných vzoriek, a to 0,71% pre spájkované spoje na báze SnBi realizované na povrchu a 1,02% pre vnorené spájkované spoje. Elektrický odpor oboch typov spájkovaných spojov na báze SnBi bol vyšší ako počiatočná hodnota po 200 teplotných šokoch. Po 200 teplotných šokoch rozdiel v elektrickom odpore vnorených spájkovaných spojov na báze SnBi kontinuálne stúpa z 0,48% na 1,39% po 1000 teplotných šokoch. Zmeny elektrického odporu spájkovaných spojov na báze SnBi realizovaných na povrchu však boli výraznejšie. Zmena elektrického odporu sa pri tomto type vzorky zvýšila z 1,05% (po 200 teplotných šokoch) na 2,51% (po 1000 teplotných šokoch).



Obr. 7 Zmena elektrického odporu spájkovaných spojov na báze SnBi

Dosiahnuté výsledky prezentujú stabilitu elektrického odporu počas tepelných šokov. Odpor spájkovaných spojov ovplyvňuje mnoho technologických aspektov, najmä v rozsahu mΩ. Pre tento experiment boli použité tri rôzne zloženia spájkovacích pást. IMC predstavujú kľúčové prvky mechanických a elektrických vlastností spájkovaných spojov. Je potrebné vziať do úvahy, že rôzne zloženia spájkovacej pasty ovplyvňujú zloženie IMC. Hrúbka IMC by tiež mohla ovplyvniť elektrický odpor. Hrúbka IMC sa mení počas životnosti elektronických zariadení a ovplyvňuje ich vlastnosti. Pokles elektrického odporu po 100 tepelných šokoch môže byť spôsobený zvýšením hrúbky IMC na vhodnejšiu hodnotu. Ďalšie zvýšenie hrúbky IMC však môže spôsobiť zvýšenie elektrického odporu. Vnorenie spájkovaných spojov do laminátu zvyšuje počet rôznych materiálov obsiahnutých v substráte. Použité materiály majú rôzne tepelné vlastnosti a počas zmien teploty podliehajú rôznym deformáciám. Okrem toho majú substrátové

materiály na báze epoxidu v každej osi rôzne tepelné vlastnosti. Rozdiely v tepelných vlastnostiach použitého materiálu môžu spôsobiť mechanické namáhanie vo vnútri substrátu a potenciálne spôsobiť mikrotrhliny vo vnorených spájkovaných spojoch. Zvýšený počet mikrotrhlín vo vnútri spájkovaných spojov ovplyvňuje mechanické a elektrické vlastnosti spájkovaných spojov.

IV. ZHRNUTIE

Táto práca analyzuje stabilitu elektrického odporu vnorených spájkovaných spojov počas teplotného šokovania. Stabilita spájkovaných spojov bola analyzovaná porovnaním medzi spájkovanými spojmi realizovanými na povrchu a zapustenými spájkovanými spojmi. Zároveň boli použité aj 3 rôzne zloženia spájkovacích pást.

Najmenšiu počiatočnú hodnotu elektrického odporu vykazovali spájkované spoje SAC305 realizované na povrchu (31,12 mΩ). Vnorenie spájkovaných spojov SAC305 do substrátu zvýšilo počiatočnú hodnotu elektrického odporu na 33,11 mΩ. Menej výrazný nárast elektrického odporu vykázali spájkované spoje SAC305 realizované na povrchu. Rozdiel medzi počiatočnou hodnotou elektrického odporu a elektrickým odporom po 1000 tepelných šokoch bol menší ako 1%. Vnorenie spájkovaných spojov SAC305 však zvýšilo zmenu elektrického odporu na 2,33%, čo bol najvyšší rozdiel (1,41%) spôsobený umiestnením spájkovaných spojov pre všetky použité spájkovacie pasty. Hoci vnorené spájkované spoje na báze SnBi vykazovali najvyššiu počiatočnú hodnotu elektrického odporu (35,67 mΩ), zmena elektrického odporu po 1000 tepelných šokoch bola najmenšia zo všetkých vnorených spájkovaných spojov (1,39%). Vnorenie spájkovaných spojov na báze SnAg spôsobilo mierny pokles počiatočných hodnôt elektrického odporu (z 33,05 mΩ na 32,74 mΩ). Spájkované spoje na báze SnAg boli jediným zložením spájkovacej pasty, pri ktorých vnorenie spájkovaných spojov viedlo k poklesu elektrického odporu. Rovnaké výsledky boli dosiahnuté po 1000 teplotných šokoch.

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.

ZOZNAM POUŽITEJ LITERATÚRY

[1] [1]W. Stęplewski *et al.*, “Thermal stability analysis of passive components embedded into printed circuit boards”, CW, roč. 44, č. 1, s. 29–36, feb. 2018, doi: 10.1108/CW-10-2017-0057.

[2] [2] S. Cho a Y. Ko, “Finite Element Analysis for Reliability of Solder Joints Materials in the Embedded Package”, Electron. Mater. Lett., roč. 15, č. 3, s. 287–296, máj. 2019, doi: 10.1007/s13391-019-00122-1.

[3] [3] K. Janeczek et al., “Investigations of temperature sensor embedded into PCB”, MI, roč. 37, č. 4, s. 199–204, sep. 2020, doi: 10.1108/MI-04-2020-0019.

[4] [4] Z. Zhao, J. Wei, L. Zhang, a Z. Ren, “Microstructural response characteristics of SAC305/Cu column joint under friction embedded welding conditions”, Materials Characterization, roč. 211, s. 113903, máj. 2024, doi: 10.1016/j.matchar.2024.113903.

[5] [5] P. Rous, A.-D. Pavlozas, a F. Steiner, “Influence of PCB Surface Finishes on Intermetallic Compound Growth”, v 2022 45th International Spring Seminar on Electronics Technology (ISSE), Vienna, Austria: IEEE, máj. 2022, s. 1–6. doi: 10.1109/ISSE54558.2022.9812788.

[6] [6] T. Kitamura, T. Furukawa, S. Yamada, K. Himeda, a A. Yamamoto, “Extremely Advanced Substrate as an integrated Package Solution (iPaS) for Next Generation High Performance Computing (HPC) applications”, v 2024 IEEE 74th Electronic Components and Technology Conference (ECTC), Denver, CO, USA: IEEE, máj. 2024, s. 2053–2057. doi: 10.1109/ECTC51529.2024.00350.

[7] [7] L. Giambuzzi, M. Cirillo, D. Schütze, M. Hempel, a T. Löher, “Integration of Microwave SMD Components into Organic Multilayer PCBs”, v 2024 IEEE 10th Electronics System-Integration Technology Conference (ESTC), Berlin, Germany: IEEE, sep. 2024, s. 1–4. doi: 10.1109/ESTC60143.2024.10712013.

[8] [8] M. Jannoun, Y. Aoues, E. Pagnacco, P. Pougnet, a A. El-Hami, “Probabilistic fatigue damage estimation of embedded electronic solder joints under random vibration”, Microelectronics Reliability, roč. 78, s. 249–257, nov. 2017, doi: 10.1016/j.microrel.2017.08.005.

[9] [9] Y. Qiao, D. Liu, H. Liang, a N. Zhao, “Trace of Ag atoms migration induced by temperature gradient in Cu/Sn-5.0Ag/Cu solder joints”, Materials Characterization, roč. 208, s. 113639, feb. 2024, doi: 10.1016/j.matchar.2024.113639.

[10] [10] P. Veselý, I. Králová, D. Pilnaj, M. Plaček, a K. Dušek, “SnAgCu Solder Joint Microstructure Evolution During Thermal Aging: Influence of Flux”, Adv Eng Mater, roč. 26, č. 23, s. 2401366, dec. 2024, doi: 10.1002/adem.202401366.

[11] [11] T. Lenger, A. Pietrikova, P. Lukacs, I. Vehec, a M. Jarkovsky, “Stability of OhmegaPly Resistors with Different Shapes”, v 2024 47th International Spring Seminar on Electronics Technology (ISSE), Prague, Czech Republic: IEEE, máj. 2024, s. 1–4. doi: 10.1109/ISSE61612.2024.10604181.

Optimal Haptic Feedback for Spatial Navigation and Orientation of Visually Impaired Individuals

[1]*Richard KRÁL,* [2]*Tibor VINCE*

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

[1]richard.kral@tuke.sk, [2]tibor.vince@tuke.sk

***Abstract* — This study presents an in-depth analysis of the specific needs of visually impaired individuals during navigation and spatial orientation, with the goal of developing a promising system for conveying essential information during independent mobility. Recent approaches were reviewed, and their major limitations—such as insufficient feedback resolution, cognitive overload, or interference with other sensory modalities—were critically addressed.**

Based on these insights, we propose a novel haptic feedback design that combines rapid response, information richness, and relevance filtering, ensuring efficient and context-sensitive guidance. The proposed solution is also cost-effective and aesthetically discreet, supporting practical implementation and user acceptance.

***Keywords* — haptic feedback, visually impaired, navigation assistance, wearable technology**

I. INTRODUCTION

Undoubtedly, individuals with visual impairments face significant disadvantages in their daily lives. Numerous efforts have been made to develop assistive devices aimed at alleviating the challenges they encounter. The functionalities of these devices vary widely, ranging from color, currency, and text recognition to voice-controlled systems, internet voice search, scene analysis, and spatial navigation. While some commercially available devices successfully integrate many of these features at a high level of performance, there is currently no solution that enables real-time navigation and spatial orientation without the aid of a white cane, guide dog, or human assistant [1].

Visually impaired individuals often experience frustration when required to navigate unfamiliar environments. To design an effective and independent navigation system, it is essential to first understand how visually impaired people currently navigate and orient themselves in space—or at least how they are trained to do so. In essence, we must understand how they perceive the world around them. In this regard, reviewing literature and guidelines for orientation and mobility (O&M) instructors can provide valuable insights [1].

One of the most critical considerations is that constant focus on feedback can be mentally exhausting. Simultaneously, disorientation itself is a major source of stress and fatigue. The more fatigued an individual becomes, the more likely they are to lose orientation. Therefore, suitable feedback must be intuitive and require minimal cognitive effort. Users should become accustomed to the system through practice and be able to quickly interpret the information provided. Moreover, feedback should be delivered selectively—only when necessary—rather than continuously [2].

Another key factor is that obstacles in the environment are often dynamic. The system must be capable of distinguishing not only the direction and speed of moving objects but also the user's own direction and velocity. This allows the system to predict potential collisions with moving obstacles and alert the user accordingly. Additionally, the system must consider how visually impaired individuals perceive distances. For example, studies have shown that visually impaired users often overestimate short distances (e.g., one meter) and underestimate longer ones (e.g., five meters). Similarly, rotational orientation tends to be inaccurate: a 90° turn is often executed as 100–105°, and a 180° turn as approximately 165°. A well-designed system can recognize these systematic errors and respond accordingly in real time [2].

Furthermore, the system should be able to inform users about changes in the terrain slope to prevent confusion. For instance, after walking uphill for an extended period, a flat surface may be perceived as a downhill slope, and even a slight incline as level. The reverse perception occurs after prolonged downhill walking. Users may also struggle to detect curvature in pathways. If a sidewalk forms part of a circle with a radius of 12 meters or less, the curvature is unlikely to be recognized. There also tends to be a perceptual bias: after following a sharp curve to the right, the subsequent straight path may be perceived as curving to the left; conversely, after a sharp curve to the left, the following straight section may appear to curve to the right [2].

When navigating, visually impaired individuals rely on information such as [2]:

- Distance (in meters)
- Horizontal orientation (directional bearings)
- Vertical changes (ascents and descents)
- Landmarks and orientation cues (e.g., localized sounds, texture of terrain, smells, wind direction, slope).

Understanding the types of cues that users are accustomed to—whether navigating independently or with assistance—enables the development of a user-friendly system. Such a system could serve as supplemental aid alongside the white cane or even act as a standalone replacement.

Beyond usability, several other aspects must be considered. From a technical perspective, the system must be capable of accurately and promptly conveying spatial information, including obstacle detection, user location, and direction. From an aesthetic and practical standpoint, visually impaired users often prefer discreet designs that do not draw attention to their disability. Finally, production costs are a significant concern, given that the target user group is relatively small. High manufacturing costs could render the device inaccessible.

With these considerations in mind, the next section provides a review of current state-of-the-art solutions. We describe selected approaches from recent studies and analyze their key limitations.

II. Existing Haptic-Based Navigation devices

Devices designed to assist visually impaired individuals with navigation and spatial orientation can be classified based either on the type of input data they acquire (i.e., how they perceive spatial information), or on the modality of feedback used to convey this information to the user. Most commonly, such devices acquire environmental data using depth cameras, LiDAR, or other forms of distance-measuring sensors.

The objective of this study is to propose an optimal design for conveying spatial information to the user. Therefore, our analysis primarily focuses on feedback mechanisms rather than on data acquisition methods. Typically, information is conveyed to users through auditory (verbal or non-verbal) or haptic feedback, or a combination of both.

Based on the considerations outlined above, we argue that verbal auditory feedback is suboptimal, primarily due to its inefficiency in high-speed or real-time navigation contexts. Although modern speech synthesis technologies are capable of delivering speech rapidly, providing detailed verbal descriptions of directions, obstacles, and their spatial relationships can be too time-consuming to ensure safe and autonomous mobility—particularly in complex or dynamic environments such as crowded urban areas.

Non-verbal auditory feedback, in contrast, is more efficient when distinct sounds are mapped to specific navigational cues. However, the appropriateness of auditory feedback as a whole is questionable, since visually impaired individuals heavily rely on their auditory sense for echolocation and environmental awareness. Introducing competing auditory signals may interfere with these natural orientation strategies.

For this reason, haptic feedback — the transmission of information through tactile stimuli such as vibrations or pressure, enabling users to perceive spatial or contextual cues via the sense of touch — is considered the most promising modality. It allows spatial information to be delivered in a silent and unobtrusive manner, without overloading the auditory channel. Consequently, the remainder of this study focuses on devices employing haptic feedback. These systems are further categorized according to the anatomical location at which the feedback is delivered. The following section presents and analyzes representative examples, with a particular emphasis on the nature and effectiveness of their haptic feedback mechanisms.

A. Head-Mounted Haptic Feedback

Gao et al. [3] developed a smart assistive system composed of wearable glasses integrated with a smartphone application. The system features three vibration motors positioned on the user's forehead, each corresponding to a designated angular sector in the field of view. Variations in vibration frequency are used to encode the distance to an obstacle detected within the respective

sector. The primary mode of feedback is haptic, with auditory alerts provided as supplementary cues. Fig. 1 Gao et al. — a wearable obstacle avoidance device [3].

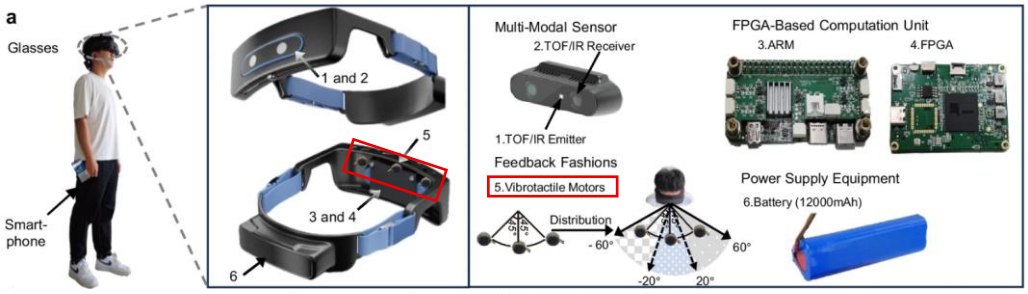


Fig. 1 Gao et al., a wearable obstacle avoidance device

From a feedback quality perspective, this approach presents several limitations. Each motor is responsible for a relatively large spatial area, which reduces the precision of directional information. Moreover, the system does not convey detailed characteristics of the detected obstacles, such as their exact location within the sector, size, or shape. The encoding of distance through vibration frequency is also problematic, as subtle frequency variations can be difficult for users to perceive and accurately associate with specific distance ranges. In addition to these functional concerns, the aesthetic design of the device is suboptimal, it is conspicuous and lacks discretion, which may discourage users from adopting it in public settings.

Another head-mounted vibrotactile system was developed by Fumihiko et al. [4], utilizing robotic arms attached to a head-mounted display to deliver targeted tactile stimulation to specific areas of the user's cheeks. Each stimulation point is associated with a corresponding spatial direction in the environment. Although this method offers improved spatial mapping compared to forehead-mounted solutions, it shares similar limitations. It lacks detailed information about obstacle characteristics, and its mechanical complexity and visible design raises questions regarding user comfort and social acceptability. Fig. *2* Fumihiko et al., vibrotactile device [4].

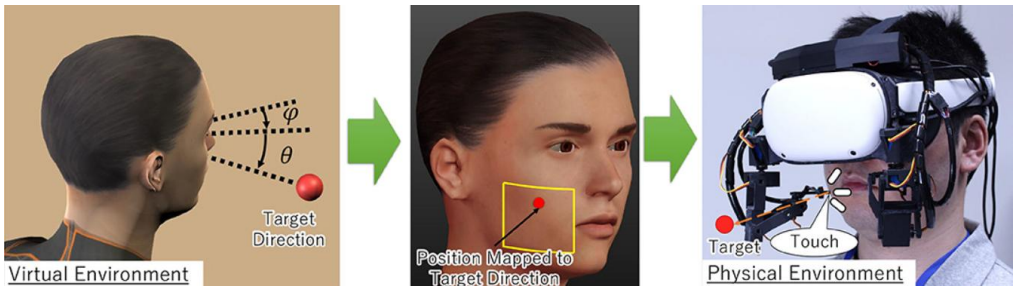


Fig. 2 Fumihiko et al., vibrotactile device

B. Torso-Based Haptic Feedback

Several studies [5]-[8] have explored the use of arrays of vibrotactile actuators positioned around the torso, either on the waist or the abdomen, to convey spatial information to visually impaired users. These systems aim to provide intuitive, body-centered feedback for navigation or obstacle avoidance.

De Paz et al. [5] proposed a system (Fig. 3,A) consisting of three horizontal rows of 16 vibration motors arranged on the abdomen. Each pair of vertical columns corresponds to a distinct spatial sector in front of the user. When an obstacle enters a given sector, the respective motors are activated, with vibration intensity modulated according to the proximity of the detected object—closer obstacles result in stronger feedback.

In another study, Brandebusemeyer et al. [6] implemented a belt (Fig. 3, B) containing 16 vibration motors evenly distributed around the waist, with each motor representing a 22.5° angular segment. Unlike obstacle detection systems, this device was used solely for orientation, with a single motor activated at any time to indicate the direction of geographic North. The purpose was to support egocentric alignment during navigation rather than avoid nearby obstacles.

Feng et al. [7] introduced a more dynamic system called Virtual Whiskers (Fig. 3, C) which utilized a 2×5 matrix of vibration motors. The system could operate in two distinct feedback modes. In the open path mode, the most navigable direction is identified, and the corresponding column of motors is activated, typically a single motor in the presence of obstacles, or two motors if the area is clear. In depth mode, all columns are active simultaneously, with each motor's vibration intensity reflecting the distance to the nearest obstacle in its corresponding spatial zone.

A similar approach was presented by Katzschmann et al. [8], who developed a haptic system (Fig. 3, D) consisting of six vibration motors, each mapped to an individual time-of-flight distance sensor oriented toward a critical spatial direction. Again, vibration frequency and intensity increased as obstacles approached, offering users a real-time representation of spatial constraints.

While torso-based systems offer improvements in terms of aesthetic acceptability compared to head-mounted designs, several limitations persist. Notably, these systems typically fail to convey detailed information about obstacle characteristics such as shape, height, or width. Furthermore, the continuous stream of vibrotactile information, especially when multiple actuators are active simultaneously—can lead to cognitive overload and user fatigue, potentially diminishing the overall usability and effectiveness of the device. Fig. 3 Torso-Based Haptic devices [5]-[8].

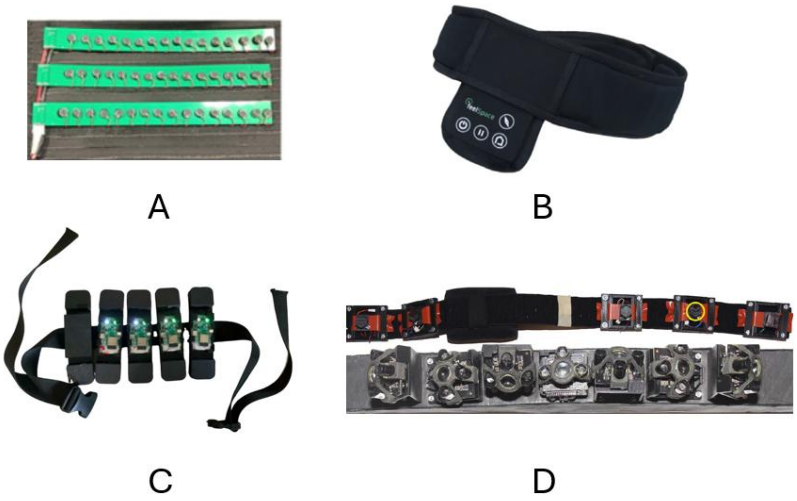


Fig. 3 A- De Paz et al.- vibrotactile device, B- Brandebusemeyer et al.-orientation belt , C- Feng et al.- Virtual Whiskers, D- Katzschmann et al.- haptic feedback device

C. Hand and Finger-Based Haptic Feedback

Khusro et al. [9] proposed a solution in which haptic feedback is delivered through smartphone-generated vibration patterns. Distinct and easily recognizable patterns—such as those resembling a heartbeat, door knocking, danger alert, or Morse code (using long and short vibrations)—are assigned to specific tasks, including navigational commands (e.g., "turn left") and hazard alerts (e.g., fall, collision, or burn). This method is considered highly advantageous due to its ability to convey spatial and contextual information rapidly and with a high degree of specificity. However, a potential limitation lies in the limited number of distinguishable vibration patterns; if the set of patterns is too small, the system may lack the expressive capacity to represent all necessary information. Furthermore, the system is not hands-free, which may constrain usability in dynamic real-world scenarios.

Another approach in the same category was introduced by Voutsakelis et al. [10], who developed a system that communicates information via the smartphone's touchscreen. The visually impaired user moves a finger across the screen, and when the finger reaches a region representing an obstacle, the phone vibrates. The vibration ceases when the user leaves that region. This method enables the user to obtain precise information regarding the location and dimensions of nearby obstacles. Notably, the system does not provide continuous feedback, which can be advantageous when users prefer to receive information only upon request, thereby minimizing cognitive load. However, similar to the previous system, this approach also lacks hands-free functionality.

III. Haptic Interface Design

As outlined in the introduction of this study, the design of an effective haptic feedback system for navigation and orientation must take into account three critical factors. First, the system must be designed to minimize cognitive overload. Feedback, once learned through training, should be intuitive and unobtrusive in daily use. This includes providing information selectively—only when necessary or upon user request—so as not to overwhelm the user with constant stimulation.

Second, the feedback must be delivered rapidly enough to support real-time navigation, even in crowded or dynamic environments, while still conveying detailed and relevant spatial information, as previously discussed. Finally, aesthetic acceptability and production cost are non-negligible considerations. The device should be affordable and discreet enough to encourage

adoption, especially among users who may be sensitive to the stigma associated with visible assistive technology.

In the studies reviewed above, haptic feedback is most commonly implemented by modulating the intensity of vibration to reflect the distance between the user and obstacles. However, this approach lacks precision. Sensitivity to vibration varies significantly between individuals, and vibration intensity alone communicates only approximate position—not the shape, size, or movement of the obstacle.

Regarding actuator placement, hand- and finger-based systems are not hands-free and can interfere with other tasks. Head-mounted systems tend to be aesthetically unappealing or socially intrusive. Among the reviewed approaches, torso-based haptic interfaces appear the most beneficial due to their balance between discreteness and spatial resolution. However, these too have limitations: they may be uncomfortable in varying weather conditions, can induce sweating, and are generally bulky.

Our proposed solution synthesizes these insights while addressing their limitations. Instead of encoding distance through vibration intensity, and instead of placing actuators on the head, torso, or hand palm, we propose a compact system composed of nine vibration motors arranged in a 3×3 matrix located on the back of the hand. The actuators are spaced sufficiently to avoid two-point discrimination errors and are intended to deliver structured vibration patterns (see Figure 4) that represent spatial and navigational information.

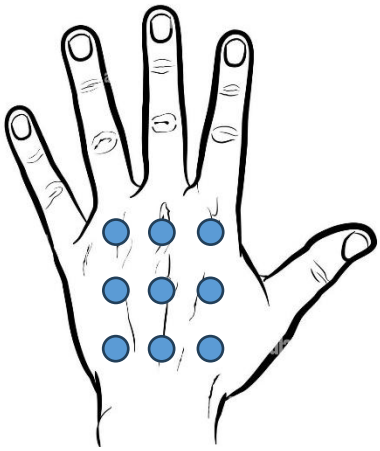


Fig. 4 Purpose solution of haptic feedback for visually impaired people

This haptic interface functions as a tactile language composed of two main types of signals:

1. Recognizable spatial patterns conveying direct environmental feedback. Examples include:
    - Three rapid pulses from motors in the upper-left corner: the user has deviated 90° from the intended direction (Fig. *5*-A).
    - A single pulse from the upper-left motor: an obstacle is located approximately 5 meters away and at 40–60° to the left of the user's current heading (Fig. *5*-B).
    - Three rapid pulses from the same motor: the obstacle in that direction is approaching at high speed (Fig. *5*-C).

2. Abstract patterns conveying contextual or navigational status. Examples include:
    - A vertical line traced across the top three motors (left to right): the previously indicated obstacle is identified as a human (Fig. *5*-D).
    - A cross-shaped activation pattern: the user has reached the desired destination (Fig. *5*-E).

This tactile language enables the system to deliver feedback that is both rapid and information-rich, while also being context-sensitive. Feedback is provided only when necessary or upon request, with a focus on conveying relevant and actionable spatial cues.

The choice of the back of the hand as the feedback site is deliberate. It offers a small, discreet, and cost-effective platform for actuator placement. Moreover, visually impaired individuals are already accustomed to using their hands to "feel the space" via a white cane. This system is intended to augment the white cane, not replace it, providing an additional layer of spatial awareness that is seamlessly integrated into the user's natural interaction with their environment.

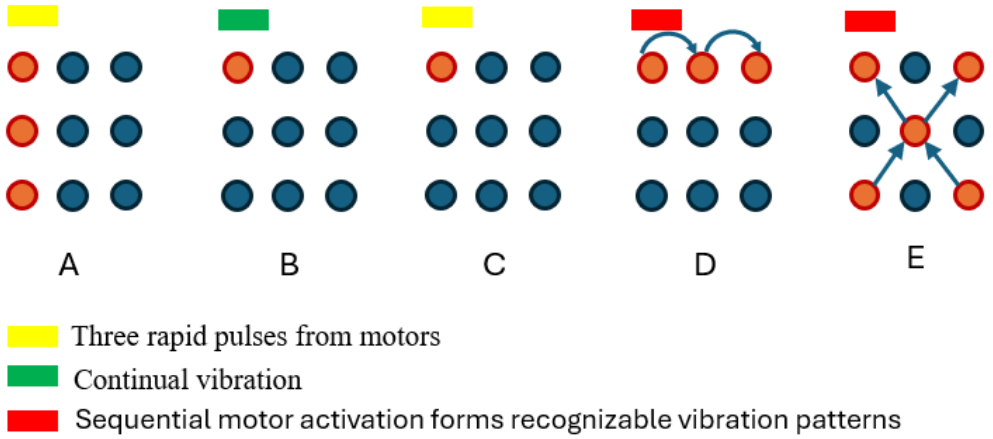


Fig. 5 Vibration patterns, A- Deviation of 90° from the desired direction, B- Obstacle detected in a specific area of space, C- Obstacle in a specific area is rapidly approaching the user, D- The obstacle mentioned in the previous command is a human, E- Desired destination has been reached.

IV. Conclusion

This study investigated the optimal form of feedback for visually impaired individuals during navigation and spatial orientation. A comprehensive analysis was conducted to gain a deeper understanding of the specific needs of visually impaired users. Based on this understanding, inappropriate feedback modalities were excluded, primarily verbal feedback, due to its slow delivery of information, and non-verbal auditory feedback, due to its potential to interfere with echolocation, a crucial strategy used by visually impaired individuals.

Among the various modalities, haptic feedback emerged as the most promising. Existing studies employing haptic feedback were reviewed and critically analyzed to identify their primary limitations. Building upon these findings, we proposed a novel solution that delivers information-rich and context-sensitive feedback with sufficient speed to support real-time navigation and orientation. The system is designed to provide feedback only when necessary or upon user request, thereby minimizing cognitive overload.

The proposed design is aesthetically discreet and cost-effective to manufacture, increasing its practicality and potential for real-world adoption.

References

[1] R. Král, P. Jacko and T. Vince, "Low-Cost Multifunctional Assistive Device for Visually Impaired Individuals," in *IEEE Access*, vol. 13, pp. 56326-56337, 2025, doi: 10.1109/ACCESS.2025.3554366.

[2] Pavel Wiener, “Prostorová orientace zrakově postižených”, Praha 2006, ISBN 80-239-6775-4

[3] Gao, Y., Wu, D., Song, J. et al. A wearable obstacle avoidance device for visually impaired individuals with cross-modal learning. Nat Commun 16, 2857 (2025). https://doi.org/10.1038/s41467-025-58085-x

[4] Nakamura Fumihiko , Verhulst Adrien , Sakurada Kuniharu , Fukuoka Masaaki , Sugimoto Maki, Evaluation of Spatial Directional Guidance Using Cheek Haptic Stimulation in a Virtual Environment, Frontiers in Computer Science, 2022, DOI 10.3389/fcomp.2022.733844

[5] de Paz, C., Huertas, J. A., Ibáñez-Gijón, J., Martín-Gonzalo, J. A., Varas, A. B., & Travieso, D. (2024). Enhancing navigation and obstacle avoidance with a vibrotactile device as secondary electronic travel aid. *Disability and Rehabilitation: Assistive Technology*, *20*(4), 1140–1150. https://doi.org/10.1080/17483107.2024.2417264

[6] Brandebusemeyer C, Luther AR, König SU, König P, Kärcher SM. Impact of a Vibrotactile Belt on Emotionally Challenging Everyday Situations of the Blind. Sensors (Basel). 2021 Nov 6;21(21):7384. doi: 10.3390/s21217384. PMID: 34770689; PMCID: PMC8587958.

[7] Junchi Feng, Giles Hamilton-Fletcher, Todd E Hudson, Mahya Beheshti, Maurizio Porfiri, John-Ross Rizzo, Haptics-based, higher-order Sensory Substitution designed for Object Negotiation in Blindness and Low Vision: Virtual Whiskers, https://doi.org/10.48550/arXiv.2408.14550

[8] Katzschmann RK, Araki B, Rus D. Safe Local Navigation for Visually Impaired Users With a Time-of-Flight and Haptic Feedback Device. IEEE Trans Neural Syst Rehabil Eng. 2018 Mar;26(3):583-593. doi: 10.1109/TNSRE.2018.2800665. PMID: 29522402.

[9] Khusro, S.; Shah, B.; Khan, I.; Rahman, S. Haptic Feedback to Assist Blind People in Indoor Environment Using Vibration Patterns. *Sensors* **2022**, *22*, 361. https://doi.org/10.3390/s22010361

[10] Voutsakelis, G.; Dimkaros, I.; Tzimos, N.; Kokkonis, G.; Kontogiannis, S. Development and Evaluation of a Tool for Blind Users Utilizing AI Object Detection and Haptic Feedback. *Machines* **2025**, *13*, 398. https://doi.org/10.3390/machines13050398

Agent Reasoning Capabilities of GPT-4.1 in Algorithmic Problem Solving on CodeWars Platform

[1]*Martin NGUYEN*

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

[1]martin.nguyen@student.tuke.sk

Abstract **– This study evaluates the problem-solving performance of Copilot and WindSurf GPT-4.1-based agentic models on algorithmic challenges. For this evaluation we use the Codewars platform, which categorizes algorithmical problems by increasing complexity across 8 Kyu to 1 Kyu levels. We tested twelve tasks using GitHub Copilot and Windsurf AI agent modes. Both models were powered by new GPT-4.1 large language model (LLM). We tested 8 Kyu, 6 Kyu, 3 Kyu and 1 Kyu level problems, three from each difficulty level in a C# 12.0 language. Tasks were considered solved if they passed Codewars built-in test suite within eight agentic retries. If the agent required more than eight attempts to achieve a correct solution, we considered the attempt as failed. The results demonstrate high accuracy and consistency at beginner level (8 and 6 Kyu). At intermediate level (3 Kyu), models required more retries to solve a task and had problems with recursive and combinatorial logic, but still performed well. At 1 Kyu (equal to senior developer level), performance degraded sharply where two of three tasks failed due to logical hallucinations or excessive recursion.**

Keywords **– Agentic AI, Algorithmical Reasoning, Code Generation, Large Language Models**

I. Introduction

It is no surprise that large language models (LLMs) has revolutionized code generation and problem-solving tasks across domains. Big software development companies integrate agentic variants of LLM across various stages of the development lifecycle. Software developers use these technologies to assist in solving programming tasks, to solve issues provided by business analysts, and even supporting automated code review during pull request reviews. Major software companies offering agentic variants and chat interfaces are competing for users between each other. Examples include GitHub Copilot, Windsurf AI, and Cursor, all of which provide similar coding assistance solutions. For that reason, there is a growing need to evaluate their problem-solving capabilities beyond single-prompt responses. This paper examines how a new GPT 4.1-based agentic system performs when solving algorithmic challenges from the Codewars platform. This platform offers a variety of algorithmical tasks with ranging difficulty through its Kyu system (ranging from 8 Kyu, equal to beginner/junior developer, to 1 Kyu, equal to expert/senior developer).

Traditional benchmarks like HumanEval [1] or MBPP [2] often evaluate code synthesis in isolated and short-function tasks. On the other hand, Codewars[1] offers algorithmical challenges generated by community. These problems have narrative and contextual complexity that tests both understanding and long-term reasoning of the human developer. These tasks are suitable for assessing not just correctness, but also the reasoning flexibility of agentic LLMs under retry loops and corrective feedback.

We explore how well can GPT-4.1-based agent tools iteratively reason and self-correct themselves across a set of 12 tasks with different difficulty. We used two LLM-powered tools in initial testing, GitHub Copilot and Windsurf AI, both using GPT-4.1 large language model. Based on very similar performance results, we chose results from GitHub Copilot for the evaluations in this paper. All problems were solved in a C# 12.0 environment within the JetBrains Rider IDE.

[1]https://www.codewars.com/

We validated the correctness through Codewars native test suites. Tasks were considered solved if they passed Codewars built-in test suite within eight agentic retries. If the agent required more than eight attempts to achieve a correct solution, we considered the attempt as failed.

The goal of this study is to identify how performance degrades with difficulty to see where current agentic models struggle most. These findings aim to inform researchers and developers of both the potential and the current boundaries of agentic LLMs in solving real-world algorithmic problems.

II. Related Work

A study made by Zhang et al. [3] examined the programming proficiency of GPT-3.5 and GPT-4 using Codewars problems ranging from 8kyu to 1kyu. Their findings indicated that while GPT-4 could reliably solve problems up to 4kyu with feedback, it struggled with challenges at the 3kyu level and beyond. That showed limitations in handling more complex tasks. Since the study was published, two new GPT models were released by OpenAI, GPT-4o [4] and GPT-4.1 [5]. As of June 2025, OpenAI's GPT-4.1 has the newest improvements in code generation, particularly in handling long-context scenarios and following complex instructions. The model achieved a 54.6% score on the SWE-bench benchmark which surpasses previous versions of GPT large language models in coding tasks [5].

The evaluation of AI-assisted code generation tools has gathered attention in research since it was first introduced. One notable study is by Yetistiren et al. [6], who assessed the code generation capabilities of GitHub Copilot, Amazon CodeWhisperer, and ChatGPT using the HumanEval dataset [1]. The study evaluated the generated code based on metrics such as code validity, correctness, security, reliability, and maintainability. At that time, ChatGPT achieved a higher correctness rate than GitHub Copilot and Amazon CodeWhisperer.

Liu et al. [7] introduced a framework designed to benchmark large language models called EvalPlus. It tests the functional correctness of code synthesized by LLMs. It augments the HumanEval dataset with a multiple automatically generated test cases. The study revealed that previous evaluations might have overestimated the correctness of LLM-generated code. The enhanced benchmark, HumanEval+, demonstrated a reduction in pass rates by up to 28.9% for some models, which highlighted the need for more comprehensive testing methodologies.

Zhou et al. [8] proposed an evaluation metric called CodeBERTScore for code generation tasks. It builds upon the BERTScore metric and incorporates both the generated code and its preceding natural language context. CodeBERTScore achieved a higher correlation with human preferences and functional correctness compared to existing metrics.

Another code quality assessment, AICodeReview, explored the integration of AI within code review processes [9]. The study aimed to make code review more effective by developing an IntelliJ IDEA plugin powered by GPT-3.5. The findings suggested that AI-assisted reviews could identify potential issues and improvements in addition to human code review.

The studies by Paul et al. [10] and Yetistiren et al. [6] underscore the advancements and challenges in evaluating AI-generated code. While tools like ChatGPT and GitHub Copilot have shown promise in generating functional code, they show the need for evaluation frameworks and metrics to ensure the reliability and quality of AI-assisted code generation.

III. Methodology

To evaluate the agentic reasoning capabilities of GPT-4.1-based models, we selected 12 algorithmic tasks from the Codewars platform. We evenly distributed the tasks across four Kyu levels: 8 Kyu, 6 Kyu, 3 Kyu, and 1 Kyu. These levels represent increasing complexity, with 8 Kyu being entry-level problems, 6 Kyu being junior-level problems, 3 Kyu medior-level problem and 1 Kyu representing expert-level algorithmic challenges. All chosen tasks were selected to be solvable in the C# programming language and were prompted using Copilot extension within the JetBrains Rider IDE using C# 12.0. The prompt provided to the model for each task followed a consistent format: "*Your task is to implement method for the following task: *task description**".

By such command, we expect the agent to implement method body for the required functionality by itself. The human is only approving or undoing the suggestion in IDE. After each response, the generated code was pasted into the Codewars web interface and executed against its built-in test suite. If the initial solution failed, the test failure messages were copied and provided as feedback in subsequent prompts in the IDE chat window. This process was repeated up to a maximum of eight iterations per task. If the task passed within these constraints, it was marked as successful, otherwise, it was marked as failed.

We used both GitHub Copilot and Windsurf AI during initial testing, but due to consistent performance parity, GitHub Copilot was selected for the primary evaluation in this study. We

Table 1
Performance of GPT-4.1 Agentic Model on Codewars Tasks

Kyu	Task Name	Satisfaction	Success	Retries	Notes
8	Binary to Decimal	91%	Success	0	Solved on first run
8	Count Pos./Sum Neg.	87%	Success	0	Solved on first run
8	Even or Odd	91%	Success	0	Solved on first run
6	IP Validation	85%	Success	0	Elegant 15-line solution
6	Mutual Recursion	79%	Success	0	Solved Hofstadter efficiently
6	Split Strings	89%	Success	0	Clean implementation
3	Tic-Tac-Toe AI	94%	Failed	8	Final logic failed 1/9 tests
3	Egg Drop Problem	89%	Success	2	Recurrence-based DP succeeded
3	Alphabetic Anagrams	96%	Success	0	Permutation-based solution
1	7x7 Skyscrapers	96%	Failed	8	Null results / timeouts
1	Three-Pass Compiler	93%	Success	1	Operand order fix worked
1	Become Immortal	92%	Failed	8	Faulty XOR grid aggregation

focused exclusively on correctness, which was measured by test suite success. No timing or performance metrics were recorded, although execution timeouts and excessive runtime were noted in failure cases. Codewars platform has execution limit of 12000ms, which is specifically measured in performance algorithmic problems.

The goal of the study was to simulate a practical software development scenario where developers iterate on failing code using test feedback and model-assisted debugging. With the enterprise versions of the agentic tools, this is very popular in the software industry nowadays. By doing so, we could examine of how well current agentic LLMs adapt to errors and converge to correct solutions across increasing levels of difficulty.

IV. Results

Table 1 summarizes the performance of the GPT-4.1 agentic model on the selected tasks. Each task entry includes the Kyu level, task name, Codewars community satisfaction rate, number of retries needed, and final result.

The results indicate that GPT-4.1-based agents consistently solved 8 Kyu and 6 Kyu tasks on the first attempt. In low-complexity tasks, the agent was able to come up with correct solutions on the first attempt, most probably by relying on memorized patterns and straightforward syntax. These tasks involved simple iteration, conditions, and did not required any significant reasoning.

Performance degraded at intermediate complexity range (3 Kyu), with one failure and another requiring retries. The successful solution of the Fabergé Egg Drop task after retries demonstrates that the model has capacity for mathematical reasoning when it was guided through error correction. The Tic-Tac-Toe AI task showed a weakness in long-horizon planning as despite numerous retries, the agent could not reliably generate an unbeatable strategy within time constraints. Even with structured error feedback, the model struggled with recursive logic that requires deep planning.

At the highest level of difficulty (1 Kyu), model performance deteriorated further as only one task was successfully solved. Tasks such as the 7×7 Skyscrapers puzzle and the Become Immortal problem showed limitations in scalability and optimization under defined constraints. In both cases, the agent initiated brute-force or overly complex strategies that consumed excessive resources which caused the execution timeout. Other returned incorrect results despite corrections. The Three-Pass Compiler task was successful on second attempt, which indicates that structural transformation is easier for the model than numerical optimization.

Figure 1 illustrates the percentage success rates across different levels of algorithmic katas. The junior-level tasks at 8th and 6th Kyu achieved a 100% success rate, while the medior-level 3rd Kyu tasks saw a reduced success rate of 67%. The senior-level 1st Kyu tasks had the lowest performance, with a success rate of only 33%.

When the problem complexity increases, the reasoning capability and execution planning decreases. The agent was especially challenged by constraint satisfaction problems and large-scale mathematical aggregations. It often failed to adjust strategies adequately within retry limits. To better understand the nature of the failures encountered in this study, we categorized unsuccessful or inefficient solutions into several failure types based on observed patterns during task execution:

1) **Execution Timeouts**

Example: 7×7 Skyscrapers, Fabergé Egg Drop (early versions)
Cause: Excessive recursion depth, lack of pruning, or inefficient search algorithms.

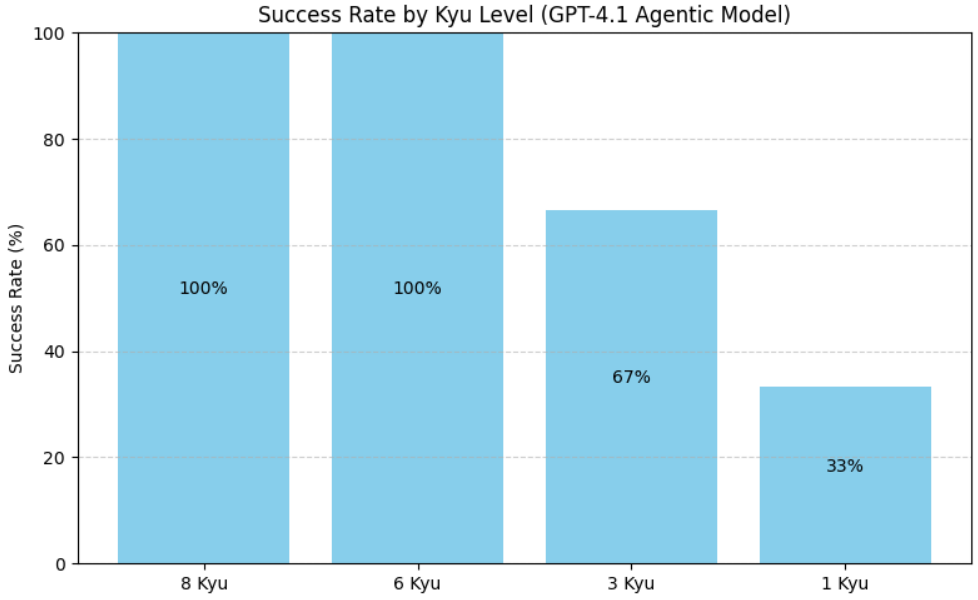


Fig. 1 Success rate for GPT 4.1 Agentic model for different kyu level katas

2) **Incorrect Logic**
   Example: Become Immortal
   Cause: Misapplication of mathematical rules or hallucinated optimizations that didn't respect task constraints. The logic was already-invalid on first run.
3) **Partial Success**
   Example:Tic-Tac-Toe AI
   Cause: Incomplete edge-case handling or inconsistent application of game rules. It failed 1 of 9 tests after applying a rule-based strategy.
4) **Structural Confusion**
   Example: Three-Pass Compiler
   Cause: Misinterpretation of operand evaluation order and confusion over low-level execution semantics.

These categories help to showcase the boundaries of current GPT-4.1 reasoning. Most recoverable cases were from failure types that produced meaningful and isolated error messages. In contrast, tasks that failed due to execution time or logical hallucination often resulted in a prolonged and unproductive retry cycle.

In general, retries helped only when the model had a comprehensible error message to act on. It struggled with execution timeout error messages, which missed a conceptual understanding of algorithmic inefficiency or test case coverage gaps. The agent often continued to refine incorrect logic paths rather than rethinking its strategy.

While GPT-4.1-based agentic models demonstrate good reasoning capabilities, they still lack the meta-reasoning needed for robust performance on complex algorithmic tasks. For that reason, the current agentic LLM models cannot fully replace experienced software developers. Instead, they serve as valuable tools for accelerating the development of repetitive or routine coding tasks.

V. Future work

In this study we did not explore prompt variation, but it would be a strong candidate for future experimentation. The fixed prompt design used here serves as a useful baseline, but likely not show the full model potential under optimized prompting conditions. Error messages from Codewars were sometimes minimal, which may have constrained the model's corrective responses. Future studies might investigate whether transforming error messages into structured debugging statements could enhance the agent's effectiveness. Future research should explore the impact of enhanced prompt engineering techniques. These include better prompting strategies and automatically inserted test cases.

VI. Conclusion

This study provides an assessment of GPT-4.1 agentic reasoning performance across a spectrum of algorithmic challenges from the Codewars platform. We evaluated the performance through

different task difficulties, from 8 Kyu to 1 Kyu. Such evaluation revealed a decline in success rate as complexity increased. The agent reliably and efficiently handled beginner and intermediate tasks with no retries. On the other hand, advanced challenges exposed big limitations in recursive reasoning, constraint handling, and abstraction.

This work highlights both the agentic model's strengths in incremental correction and its weakness in more complex strategic planning. We identified execution timeouts and incorrect logic as dominant contributors to failure. Our findings support the conclusion that current GPT-4.1-based agentic models have good short-term reasoning and reactive adaptation but lack the global planning capabilities required to consistently succeed on expert-level algorithmic problems. Because of that, the current agentic LLM models cannot fully replace experienced software developers but serve as valuable tools for the development of repetitive or routine coding tasks.

References

[1] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. de Oliveira Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, A. Ray, R. Puri, G. Krueger, M. Petrov, H. Khlaaf, G. Sastry, P. Mishkin, B. Chan, S. Gray, N. Ryder, M. Pavlov, A. Power, L. Kaiser, M. Bavarian, C. Winter, P. Tillet, F. P. Such, D. Cummings, M. Plappert, F. Chantzis, E. Barnes, A. Herbert-Voss, W. H. Guss, A. Nichol, A. Paino, N. Tezak, J. Tang, I. Babuschkin, S. Balaji, S. Jain, W. Saunders, C. Hesse, A. N. Carr, J. Leike, J. Achiam, V. Misra, E. Morikawa, A. Radford, M. Knight, M. Brundage, M. Murati, K. Mayer, P. Welinder, B. McGrew, D. Amodei, S. McCandlish, I. Sutskever, and W. Zaremba, "Evaluating large language models trained on code," 2021. [Online]. Available: https://arxiv.org/abs/2107.03374

[2] J. Austin, A. Odena, M. Nye, M. Bosma, H. Michalewski, D. Dohan, E. Jiang, C. Cai, M. Terry, Q. Le, and C. Sutton, "Program synthesis with large language models," 2021. [Online]. Available: https://arxiv.org/abs/2108.07732

[3] Z. Zhang, L. Wen, S. Zhang, D. Chen, and Y. Jiang, "Evaluating gpt's programming capability through codewars' katas," *Lecture Notes in Computer Science*, p. 17–26, 2024.

[4] OpenAI, A. Hurst, and et al., "Gpt-4o system card," 2024. [Online]. Available: https://arxiv.org/abs/2410.21276

[5] OpenAI, "Introducing GPT-4.1 in the API," https://openai.com/index/gpt-4-1/, 2025, accessed: 2025-06-02.

[6] B. Yetiştiren, I. Özsoy, M. Ayerdem, and E. Tüzün, "Evaluating the code quality of ai-assisted code generation tools: An empirical study on github copilot, amazon codewhisperer, and chatgpt," 2023. [Online]. Available: https://arxiv.org/abs/2304.10778

[7] J. Liu, C. S. Xia, Y. Wang, and L. Zhang, "Is your code generated by chatgpt really correct? rigorous evaluation of large language models for code generation," in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, ser. NIPS '23. Red Hook, NY, USA: Curran Associates Inc., 2023.

[8] S. Zhou, U. Alon, S. Agarwal, and G. Neubig, "CodeBERTScore: Evaluating code generation with pretrained models of code," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, H. Bouamor, J. Pino, and K. Bali, Eds. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 13 921–13 937.

[9] Y. Almeida, D. Albuquerque, E. D. Filho, F. Muniz, K. de Farias Santos, M. Perkusich, H. Almeida, and A. Perkusich, "Aicodereview: Advancing code quality with ai-enhanced reviews," *SoftwareX*, vol. 26, p. 101677, May 2024.

[10] D. G. Paul, H. Zhu, and I. Bayley, "Benchmarks and metrics for evaluations of code generation: A critical review," 2024. [Online]. Available: https://arxiv.org/abs/2406.12655

Šifrovací mód AES-XTS pre šifrovanie obsahu diskov

[1]Kamil BERECKÝ, [2]Miloš DRUTAROVSKÝ

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

[1]kamil.berecky@student.tuke.sk, [2]milos.drutarovsky@tuke.sk

Abstrakt – **Článok sa zaoberá problematikou šifrovania diskových zariadení pomocou režimu AES-XTS, ktorý je špeciálne navrhnutý pre ochranu údajov v blokovo orientovaných úložných zariadeniach. Pozornosť je upriamená na dve implementácie režimu XTS v programovacom jazyku C zahŕňajúce riešenie využívajúce knižnicu OpenSSL, ktorá podporuje hardvérovú akceleráciu AES a riešenie založené na minimalistickej knižnici µAES určenej pre vstavané systémy. Porovnanie oboch implementácií prebehlo v rôznych podmienkach vrátane testovania s aktívnou a vyprázdnenou vyrovnávacou pamäťou a pri rôznom počte vlákien. Namerané rýchlostné rozdiely poukazujú na nevyhnutnosť optimalizácie kryptografických aplikácií vzhľadom na charakteristiku použitej hardvérovej architektúry.**

Kľúčové slová – **AES-XTS, šifrovanie diskov, knižnica OpenSSL, knižnica µAES**

I. Úvod

V súčasnej digitálnej ére je bezpečnosť údajov uložených na diskových zariadeniach kritickou požiadavkou pre organizácie aj jednotlivcov. S rastúcim objemom citlivých informácií, ako sú osobné údaje, zdravotné záznamy, finančné dokumenty a dôverné podnikové dáta, sa šifrovanie úložísk stalo nevyhnutným nástrojom na ochranu pred neoprávneným prístupom. Neoprávnené získanie týchto údajov môže viesť k vážnym následkom vrátane straty súkromia, finančných škôd a poškodenia reputácie.

AES [1] patrí medzi najpoužívanejšie symetrické šifrovacie algoritmy na svete vďaka svojej vysokej úrovni bezpečnosti, výkonnosti a širokej hardvérovej podpore. Pri šifrovaní celých diskových zariadení sa však objavujú špecifické požiadavky, ktoré tradičné operačné režimy AES nemusia optimálne spĺňať. Tieto požiadavky viedli k vývoju špecializovaných šifrovacích módov určených priamo pre diskové aplikácie.

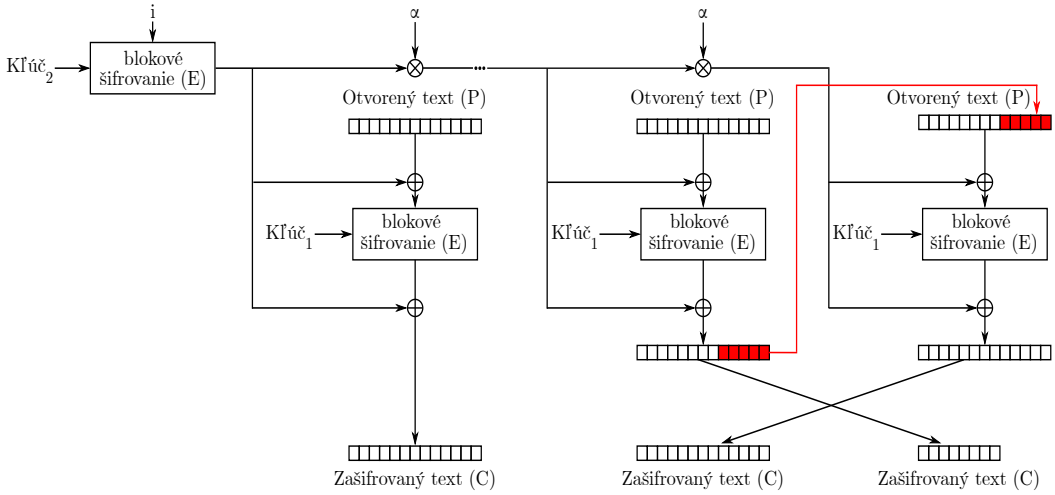
Režim XTS [2] je operačný režim blokovej šifry navrhnutý tak, aby zabezpečil odolnosť voči určitým typom útokov, ktoré sú typické pre šifrovanie diskov. Tento mód kombinuje výhody šifry AES s ďalšími pridanými mechanizmami, ktoré zabezpečujú, že rovnaké dátové bloky na disku budú zašifrované odlišne, čím sa zabraňuje vzniku vzorov, ktoré by mohli byť zneužité útočníkmi.

Napriek širokému prijatiu AES-XTS v open-source aj komerčných aplikáciách, ako sú VeraCrypt [3], BitLocker [4] a FileVault [5], existujú otázky týkajúce sa optimálnej implementácie pre rôzne hardvérové platformy. Výber vhodnej kryptografickej knižnice môže významne ovplyvniť výkon celej aplikácie, čo je často kritické pre reálne nasadenie a využitie.

Cieľom článku je analýza a porovnanie dvoch odlišných implementačných prístupov k režimu XTS v programovacom jazyku C. Prvá implementácia využíva knižnicu OpenSSL [6] s podporou hardvérovej akcelerácie, zatiaľ čo druhá je založená na minimalistickej knižnici µAES [7] určenej pre vstavané systémy. Výsledky tejto práce poskytujú praktické poznatky pre vývojárov pri výbere vhodnej implementácie pre konkrétne hárdverové platformy.

II. Základný opis režimu XTS

AES-XTS je šifrovací režim špeciálne navrhnutý na ochranu údajov v blokovo orientovaných úložných zariadeniach a je rozšírene používaný v pevných diskoch (HDD), SSD diskoch a pamäťových flash kartách. Tento štandard bol opísaný IEEE v roku 2007 a aktualizovaný



Obr. 1 Proces šifrovania jedného bloku dát v režime XTS [9].

v roku 2018 v publikácii IEEE Std 1619™-2018 [2], pričom jeho používanie bolo taktiež schválené americkým Národným inštitútom pre štandardy a technológie (NIST) v roku 2010 v publikácii SP 800-38E [8].

A. Princíp činnosti

Režim XTS využíva AES ako blokovú šifru, ale zahŕňa dve fázy šifrovania AES s rôznymi kľúčmi. V prvej fáze sú vypočítané blokové úpravy z počiatočnej blokovej úpravy (zvyčajne číslo sektora, ktorý šifrujeme), ktorá je zašifrovaná prostredníctvom šifry AES s kľúčom pre úpravu bloku, po ktorom nasleduje násobenie číslom 2 v konečnom poli GF (2^{128}). Druhá fáza je určená na šifrovanie samotných údajov, kde sa každý vstupný a výstupný blok kombinuje s príslušnou blokovou úpravou pomocou $\oplus$ (XOR) operácie pred a po šifrovaní AES. Schéma režimu XTS je znázornená na Obr. 1.

Pre štandardné šifrovanie úplných blokov sa každý blok otvoreného textu P_i najprv kombinuje s blokovou úpravou T_i pomocou $\oplus$ operácie, výsledok sa zašifruje pomocou AES s kľúčom K_1 a nakoniec sa opäť aplikuje $\oplus$ s tou istou blokovou úpravou:

$$C_i = E_{K_1}(P_i \oplus T_i) \oplus T_i$$

B. Blokové úpravy

Dôležitou súčasťou režimu XTS sú blokové úpravy (z angl. tweaks). Bloková úprava je hodnota, ktorá sa používa na modifikáciu šifrovacieho procesu tak, aby rovnaké bloky dát šifrované s tým istým kľúčom ale na rozdielnych pozíciách na disku vytvorili jedinečný šifrovaný výstup. Využitie týchto blokových úprav efektívne bráni rôznym typom útokov založených na analýze vzorov v šifrovaných dátach.

Proces generovania blokových úprav začína počiatočnou blokovou úpravou, ktorá je typicky odvodená od čísla sektora. Táto hodnota sa zašifruje pomocou šifry AES s kľúčom K_2 (určeným pre šifrovanie blokových úprav) a pre každý ďalší blok v sektore sa výsledok násobí primitívnym elementom α (zvyč. číslo 2) v konečnom poli GF(2^{128}). Násobenie dvomi je de facto bitový posun o jednu pozíciu vľavo. Tento bitový posun je ale v konečnom poli problémom pri číslach, ktoré majú MSB (Most Significant Bit) 1. Pre tento prípad je v štandarde definovaný ireducibilný polynóm, pomocou ktorého sa zabezpečuje zotrvanie v rámci konečného poľa:

$$x^{128} + x^7 + x^2 + x + 1$$

C. Kradnutie šifrovaného textu

Pri šifrovaní diskov pomocou XTS sa používa technika nazývaná kradnutie šifrovaného textu (v angl. ciphertext stealing), ktorá je kľúčová pre prácu s dátami, ktorých veľkosť nie je násobkom veľkosti bloku. Namiesto pridávania výplne k poslednému neúplnému

bloku sa „ukradnú“ bajty z predposledného už zašifrovaného bloku. Tieto bajty sa použijú na doplnenie posledného neúplného bloku do plnej veľkosti. Takto doplnený posledný blok sa zašifruje a stane sa novým predposledným blokom. Pôvodný predposledný blok, z ktorého sa odobrali bajty, sa stane posledným blokom (skrátený o počet ukradnutých bajtov). Táto technika (viď. Obr. 1) zabezpečuje správne fungovanie XTS režimu aj pre dáta, ktorých veľkosť nie je presným násobkom veľkosti bloku, a to bez akéhokoľvek nárastu objemu dát.

D. Bezpečnostné vlastnosti

AES-XTS poskytuje niekoľko významných bezpečnostných výhod oproti iným režimom blokových šifier. Režim zabezpečuje difúziu medzi blokmi – malá zmena v otvorenom texte sa prejaví zmenou na viacerých miestach šifrovaného textu, čo sťažuje kryptoanalýzu.
Je taktiež odolný voči útokom na preusporiadanie blokov – použitie blokových úprav odvodených od čísla sektora zabraňuje útočníkovi presúvať šifrované bloky na rôzne pozície.

Napriek svojím silným bezpečnostným vlastnostiam má AES-XTS určité limity, najmä absenciu autentizácie. Režim XTS poskytuje iba dôvernosť, nie však autentizáciu ani integritu dát. Útočník môže modifikovať šifrované dáta, čo môže viesť k predvídateľným zmenám v dešifrovaných údajoch.

III. Implementácie s využitím kryptografických knižníc

V rámci praktickej časti boli vytvorené dve implementácie režimu XTS v programovacom jazyku C, každá zameraná na iný typ hardvérovej platformy.

A. Implementácia s využitím knižnice OpenSSL

Prvá implementácia využíva knižnicu OpenSSL [6] (ver. 3.5.0), ktorá je široko používaná a dobre zdokumentovaná. Na kompiláciu je potrebný prekladač `gcc` (testované s ver. 14.2.0) a príkaz `make` (u nás ver. 4.3). Kompilácia je podporovaná na platformách Windows a Linux. OpenSSL poskytuje hardvérovú akceleráciu šifry AES prostredníctvom inštrukcií AES-NI [10] dostupných na moderných procesoroch, čo výrazne urýchľuje šifrovacie operácie. Pre bezpečné odvodenie šifrovacích kľúčov z používateľského hesla a náhodnej soli sa využíva funkcia na odvodenie kľúča (v angl. Key Derivation Function – KDF) Argon2id, dostupný priamo v knižnici OpenSSL (od ver. 3.2).

Na prácu so zariadeniami sa používajú nízkoúrovňové funkcie operačného systému – na Linuxe funkcie `open()`, `read()`, `write()` a na Windows ich ekvivalenty z Win32 API, ako `CreateFile()`, `ReadFile()`, `WriteFile()` a `CloseHandle()`. Podrobná dokumentácia k projektu je dostupná online v repozitári na GitHube[1] a v bakalárskej práci [11].

B. Implementácia s využitím knižnice µAES

Druhá implementácia je navrhnutá pre vstavané systémy s obmedzenými zdrojmi a využíva minimalistickú knižnicu µAES [7] (ver. 1.9). Knižnica µAES nevyužíva hardvérovú akceleráciu, keďže je primárne určená pre mikrokontroléry a vstavané zariadenia, kde sú dôležitejšie nízke výpočtové a pamäťové nároky a s tým tiež súvisiaca energetická efektívnosť než absolútny výkon.

Okrem knižnice µAES pre AES operácie tento projekt využíva aj knižnicu BLAKE3 [12] ako náhradu za KDF (Key Derivation Function) na bezpečné odvodenie šifrovacích kľúčov z používateľského hesla a náhodnej soli. Implementácia podporuje AES-XTS s 256-bitovými kľúčmi a je navrhnutá pre jednoduchú kompiláciu na systémoch Linux a Windows.

V oboch projektoch je nastavená veľkosť vstupno/výstupného buffra na 4 MB pre minimalizáciu I/O diskových operácií [13] a bezpečnostné mechanizmy ako bezpečné vymazávanie kľúčov z pamäte po použití a upozornenie používateľa na slabé heslo. Podobne ako pre OpenSSL projekt, aj pre túto implemetáciu je dostupná podrobná dokumentácia v repozitári na GitHube[2] a v bakalárskej práci [11].

IV. Experimentálne výsledky

A. Testovacie prostredie

Všetky testy boli vykonané v prostredí Windows Subsystem for Linux (Ubuntu 24.10) na počítači s procesorom AMD Ryzen 7 5800X3D (3,4 GHz, boost 4,5 GHz, 8 jadier/16

[1] https://github.com/XomByik/aes_xts_c
[2] https://github.com/XomByik/maes_xts_c

vlákien), 32 GB DDR4 pamäte a SSD Samsung 980 1 TB. Pre konzistenciu výsledkov sa používalo rovnaké heslo „password“ a virtuálna testovacia partícia s veľkosťou 512 MB.

Merania boli vykonané pomocou štandardného príkazu `time` v Linuxe s presnosťou v milisekundách. Pre každú operáciu sa vykonalo viacero meraní na minimalizáciu vplyvu náhodných výkyvov.

B. Vplyv vyrovnávacej pamäte

Testy boli vykonané v dvoch režimoch – s aktívnou a vyprázdnenou vyrovnávacou pamäťou (`sudo sh -c "sync && echo 3 > /proc/sys/vm/drop_caches"` [14]). Rozdiely medzi testami s aktívnou a vyprázdnenou cache poukazujú na vplyv I/O (vstupno/výstupých) operácií na celkový výkon šifrovania. V prípade aktívnej cache sú dáta už načítané v operačnej pamäti, čím sa minimalizuje čakanie na pomalšie diskové operácie. Naopak, pri vyprázdnenej cache je potrebné všetky dáta načítať z disku, čo zvyšuje celkový čas spracovania dát. Rozdiely pri medzi aktívnou a vyprázdnenou cache boli videteľné pri OpenSSL implementácii, kde je samotné šifrovanie a dešifrovanie veľmi rýchle vďaka hardvérovej akcelerácii, a tak I/O operácie tvoria významnejšiu časť celkového času. Výsledky testov pre implementácie OpenSSL (Tabuľka 1) a µAES (Tabuľka 2) sú zhrnuté v Tabuľke 3.

Tabuľka 1
Výsledky testov režimu AES-XTS-256 s využitím OpenSSL projektu
Pozn.: Veľkosť partície - 512 MB

Číslo testu	**Aktívna cache**		**Vyprázdnená cache**	
	Šifrovanie	**Dešifrovanie**	**Šifrovanie**	**Dešifrovanie**
1	1,096 s	1,029 s	1,594 s	1,841 s
2	1,327 s	0,949 s	1,526 s	1,548 s
3	1,317 s	0,979 s	1,479 s	1,542 s
4	1,127 s	0,938 s	1,559 s	1,501 s
5	1,348 s	0,979 s	1,462 s	1,443 s
6	1,357 s	0,959 s	1,467 s	1,434 s
7	1,327 s	0,958 s	1,498 s	1,480 s
8	1,357 s	1,009 s	1,511 s	1,511 s
9	1,317 s	0,969 s	1,527 s	1,539 s
10	1,377 s	0,979 s	1,511 s	1,491 s
Priemer	1,295 s	0,975 s	1,513 s	1,533 s
Rýchlosť	395,4 MB/s	525,1 MB/s	338,4 MB/s	334,0 MB/s

Tabuľka 2
Výsledky testov režimu AES-XTS-256 s využitím µAES projektu
Pozn.: Veľkosť partície - 512 MB

Číslo testu	**Aktívna cache**		**Vyprázdená cache**	
	Šifrovanie	**Dešifrovanie**	**Šifrovanie**	**Dešifrovanie**
1	87,719 s	142,979 s	88,678 s	141,887 s
2	91,357 s	141,699 s	88,021 s	141,563 s
3	88,397 s	142,049 s	88,723 s	140,541 s
4	87,627 s	141,339 s	87,768 s	141,033 s
5	88,347 s	141,059 s	88,910 s	141,998 s
Priemer	88,689 s	141,825 s	88,420 s	141,404 s
Rýchlosť	5,8 MB/s	3,6 MB/s	5,8 MB/s	3,6 MB/s

Tabuľka 3
Súhrnné porovnanie rýchlostí implementácií režimu XTS
Pozn.: Veľkosť partície - 512 MB

Implementácia	**Operácia**	**Aktívna cache Rýchlosť**	**Vyprázdnená cache Rýchlosť**
OpenSSL	Šifrovanie	395,4 MB/s	338,4 MB/s
	Dešifrovanie	525,1 MB/s	334,0 MB/s
µAES	Šifrovanie	5,8 MB/s	5,8 MB/s
	Dešifrovanie	3,6 MB/s	3,6 MB/s

C. Efektivita paralelizácie

Keďže režim XTS podporuje paralelizáciu [15], pre ďalšiu optimalizáciu výkonu sme experimentovali s paralelizáciou pomocou knižnice OpenMP. Pre OpenSSL projekt sme implementovali paralelné spracovanie sektorov v bloku pomocou direktívy `#pragma omp parallel for`, ktorá rozdeľuje spracovanie jednotlivých sektorov medzi dostupné vlákna procesora.

Ako môžeme vidieť v Tabuľke 4, paralelizácia v našom prípade nepriniesla želaný efekt zvýšenia rýchlosti. Je to spôsobené kombináciou nasledujúcich faktorov:

- hardvérová akcelerácia AES-NI [10] už poskytuje veľmi vysokú rýchlosť šifrovania aj pri jednom vlákne,
- rýchlosť diskových operácií predstavuje úzke hrdlo (v angl. bottleneck), ktorý limituje maximálnu rýchlosť bez ohľadu na počet použitých vlákien,
- réžia spojená so správou vlákien a synchronizáciou môže v niektorých prípadoch prevážiť výhody paralelného spracovania.

Toto potvrdzuje Amdahlov zákon [16], ktorý hovorí, že celkové zrýchlenie je limitované neparalelizovateľnou časťou programu, ktorou sú v tomto prípade práve diskové I/O operácie.

Počas práce s knižnicou µAES sme tiež skúšali implementovať paralelizáciu pomocou knižnice OpenMP, avšak narazali sme na problém pri šifrovaní dát, kde sa bloky preusporiadali a uložili na disk v nesprávnom poradí a tak ich nebolo možné pri dešifrovaní korektne rozšifrovať.

Tabuľka 4
Výsledky paralelného spracovania AES-XTS-256 s OpenSSL
Pozn.: Veľkosť partície - 512 MB

Test	**2 vlákna**		**4 vlákna**		**8 vlákien**	
	Šifrovanie	**Dešifrovanie**	**Šifrovanie**	**Dešifrovanie**	**Šifrovanie**	**Dešifrovanie**
1	1,780 s	1,788 s	1,909 s	1,834 s	1,725 s	1,722 s
2	1,730 s	1,668 s	1,738 s	1,703 s	1,688 s	1,749 s
3	1,689 s	1,474 s	1,792 s	1,729 s	1,812 s	1,786 s
4	1,553 s	1,735 s	1,769 s	1,790 s	1,697 s	1,744 s
5	1,629 s	1,693 s	1,845 s	1,770 s	1,752 s	1,802 s
6	1,618 s	1,734 s	1,829 s	1,722 s	1,703 s	1,755 s
7	1,701 s	1,782 s	1,781 s	1,688 s	1,695 s	1,831 s
8	1,633 s	1,654 s	1,870 s	1,791 s	1,681 s	1,747 s
9	1,601 s	1,652 s	1,835 s	1,717 s	1,773 s	1,815 s
10	1,554 s	1,679 s	1,797 s	1,714 s	1,744 s	1,799 s
Priemer	1,659 s	1,696 s	1,807 s	1,746 s	1,719 s	1,775 s
Rýchlosť	308,6 MB/s	301,9 MB/s	283,3 MB/s	293,2 MB/s	297,8 MB/s	288,4 MB/s

D. Analýza výstupu I/O monitora

Na monitorovanie I/O operácií počas testov sme vytvorili bash skript využívajúci príkaz `vmstat`, ktorý zapisoval údaje o CPU I/O wait, rýchlosti zápisu a rýchlosti čítania súborov. Analýza výstupov z I/O monitora odhalila zaujímavý vzor pri použití rôzneho počtu vlákien. Z týchto logov sme vypočítali nasledujúce priemerné hodnoty:

- **priemerná rýchlosť čítania:** približne 800 MB/s (jednovláknové: 795 MB/s, viacvláknové: 812 MB/s),
- **priemerná rýchlosť zápisu:** približne 900 MB/s (jednovláknové: 887 MB/s, viacvláknové: 913 MB/s),
- **CPU I/O wait:** 2-6% pri jednovláknovom spracovaní, pri viacvláknovom spracovaní až 7% s krátkodobými špičkami 25-31%.

Zvýšené hodnoty `CPU I/O wait` pri viacvláknovom spracovaní jasne ukazujú, že procesor častejšie čaká na dokončenie diskových operácií. Súčasne priebeh I/O operácií bol menej konzistentný pri viacvláknovom spracovaní, s výraznými špičkami a poklesmi v rýchlosti čítania a zápisu, čo naznačuje neefektívne využitie diskového podsystému a možnú fragmentáciu operácií.

Zaznamenali sme niekoľko priamych korelácií medzi vyššími hodnotami `CPU I/O wait` a pomalšími časmi spracovania. Napríklad, počas testu č. 5 pri 4 vláknach (viď. Tabuľka 4), kedy proces trval 1,845 sekúnd, bol zaznamenaný I/O wait na úrovni 5%, zatiaľ čo pri iných testoch s I/O wait okolo 3% boli časy spracovania kratšie.

V našom testovacom prostredí sme používali NVMe SSD disk[3], ktorý má teoretické rýchlosti čítania a zápisu v rádoch tisícok MB/s. Napriek tomu, reálne namerané hodnoty boli na nižšej úrovni. Pre overenie skutočných maximálnych rýchlostí disku v našom testovacom prostredí sme využili aj nástroj `fio` (Flexible I/O Tester) [17], ktorý je štandardom pre benchmarking diskových systémov v Linuxe.

Nástroj `fio` sme nakonfigurovali s nasledujúcimi parametrami:

- `-size=1G` - veľkosť testovacieho súboru,
- `-bs=4M` - veľkosť jedného I/O bloku,
- `-ioengine=sync` - synchrónne I/O operácie,
- `-direct=1` - priamy prístup k disku (neaktívna cache),
- `-rw=readwrite` - striedavé operácie čítania a zápisu.

Výsledky testov potvrdili, že reálne maximálne rýchlosti disku v prostredí WSL (Windows Subsystem for Linux) sa pohybujú okolo 800-950 MB/s pre čítanie a 850-950 MB/s pre zápis, čo zodpovedá hodnotám nameraným pomocou I/O monitora počas našich šifrovacích testov. Nižšie rýchlosti v porovnaní s teoretickými možnosťami NVMe SSD sú spôsobené najmä:

- virtualizačnou vrstvou WSL, ktorá pridáva dodatočnú réžiu pri diskových operáciách [18],
- použitím synchrónnych I/O operácií pre zabezpečenie integrity dát,
- natívnymi obmedzeniami implementácie linuxového jadra v prostredí WSL,
- fragmentáciou I/O žiadostí pri paralelnom spracovaní.

Toto potvrdzuje, že rýchlosť našej OpenSSL implementácie je primárne limitovaná diskovou priepustnosťou, nie výkonom procesora. Hardvérová akcelerácia AES-NI [10] poskytuje dostatočný výpočtový výkon, ale celkový proces je obmedzený rýchlosťou I/O operácií.

V. Záver

Realizované experimenty jasne demonštrujú výrazný rozdiel vo výkone medzi implementáciami. OpenSSL implementácia je približne 68-krát rýchlejšia pri šifrovaní a 146-krát rýchlejšia pri dešifrovaní v porovnaní s µAES implementáciou. Tento extrémny rozdiel je spôsobený kombináciou faktorov: hardvérovou akceleráciou AES-NI [10], optimalizáciou knižnice OpenSSL pre x86 architektúru a rozdielnym zameraním knižníc. Paralelizácia v našich testovacích scénároch nepriniesla lineárne zrýchlenie, pretože celková rýchlosť je limitovaná neparalelizovateľnými časťami programu, ktorými sú v tomto prípade diskové I/O operácie.

Pre desktopové a serverové aplikácie, kde je vysoká rýchlosť kritickým faktorom, je jednoznačne vhodnejšie použiť OpenSSL implementáciu. Pre mikrokontroléry s obmedzenými zdrojmi, kde sú dôležitejšie malá veľkosť kódu a nízka pamäťová náročnosť, je lepšie použiť µAES implementáciu.

Výsledky práce poukazujú na dôležitosť výberu vhodnej kryptografickej knižnice pre konkrétnu hardvérovú platformu a zdôrazňujú význam hardvérovej akcelerácie pre rýchlosť kryptografických operácií v moderných systémoch.

Poďakovanie

Táto práca bola podporená EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 17R05-04-V01-00004 - „Kompetenčné a vzdelávacie centrum kybernetickej a informačnej bezpečnosti TUKE“.

Literatúra

[1] NIST, “Advanced Encryption Standard (AES),” Gaithersburg, MD, USA, Jan. 2023, published 2001, updated 2023. [Online]. Available: https://doi.org/10.6028/NIST.FIPS.197-upd1
[2] IEEE Computer Society, “IEEE Standard for Cryptographic Protection of Data on Block-Oriented Storage Devices,” Dec. 2018, online; accessed October 20, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/8637988
[3] IDRIX, “VeraCrypt Documentation,” May 2025, online; accessed May 8, 2025. [Online]. Available: https://www.veracrypt.eu/en/Documentation.html
[4] Microsoft, “BitLocker Drive Encryption,” May 2025, online; accessed May 8, 2025. [Online]. Available: https://support.microsoft.com/en-us/windows/bitlocker-drive-encryption-76b92ac9-1040-48d6-9f5f-d14b3c5fa178
[5] Apple, “Intro to FileVault,” May 2024, online; accessed May 8, 2025. [Online]. Available: https://support.apple.com/guide/deployment/intro-to-filevault-dep82064ec40/web
[6] OpenSSL Software Foundation, “OpenSSL Library,” May 2025, online; accessed May 8, 2025. [Online]. Available: https://openssl-library.org/

[3]https://semiconductor.samsung.com/consumer-storage/internal-ssd/980/

[7] polfosol, "µAES," Dec. 2024, online; accessed May 8, 2025. [Online]. Available: https://github.com/polfosol/micro-AES
[8] M. Dworkin, "Recommendation for Block Cipher Modes of Operation: The XTS-AES Mode for Confidentiality on Storage Devices," Jan. 2010, online; accessed October 16, 2024. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-38E
[9] Prispievatelia Wikipédie, "Schema of an XTS encryption," May 2016, online; accessed October 20, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Disk_encryption_theory#/media/File:XTS_mode_encryption.svg
[10] ——, "AES instruction set," Apr. 2025, online; accessed May 8, 2025. [Online]. Available: https://en.wikipedia.org/wiki/AES_instruction_set
[11] K. Berecký, "Šifrovací mód AES-XTS pre šifrovanie obsahu diskov," Bakalárska práca, Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, May 2025.
[12] J. O'Connor, "BLAKE3," Apr. 2025, online; accessed May 8, 2025. [Online]. Available: https://github.com/BLAKE3-team/BLAKE3
[13] Srineet, "How Tuning the Buffer Size Led to Much Improved Performance," May 2020, online; accessed May 8, 2025. [Online]. Available: https://vox.veritas.com/blog/backup_exec/how-tuning-the-buffer-size-led-to-much-improved-performance/879902
[14] R. van Riel and P. W. Morreale, "Documentation for /proc/sys/vm/," 2008, online; accessed May 8, 2025. [Online]. Available: https://docs.kernel.org/admin-guide/sysctl/vm.html#drop-caches
[15] M. Alomari, K. Samsudin, and A. R. Ramli, "Implementation of a Parallel XTS Encryption Mode of Operation," *Indian Journal of Science and Technology*, Nov. 2014. [Online]. Available: https://sciresol.s3.us-east-2.amazonaws.com/IJST/Articles/2014/Issue-11/Article13.pdf
[16] Prispievatelia Wikipédie, "Amdahl's law," May 2025, online; accessed May 8, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Amdahl%27s_law
[17] J. Axboe, "Fio - Flexible I/O tester," 2023, online; accessed May 8, 2025. [Online]. Available: https://fio.readthedocs.io/en/latest/
[18] Microsoft, "Windows Subsystem for Linux: Performance Considerations," 2023, online; accessed May 8, 2025. [Online]. Available: https://docs.microsoft.com/en-us/windows/wsl/compare-versions

Nové trendy vo využití magnetických kvapalín v elektrotechnike

[1]*Ján KAŇUCH,* [2]*Peter GIROVSKÝ,* [3]*Nikolas KONDÁŠ*
[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]nikolas.kondas@student.tuke.sk

***Abstrakt* — V tomto príspevku sú prezentované nové trendy vo využívaní magnetických kvapalín v elektrotechnike. Magnetické kvapaliny, ako sú ferrofluidy a magnetoreologické kvapaliny, vykazujú unikátne vlastnosti, ktoré umožňujú ich aplikáciu v chladení, tlmení vibrácií, senzorike, aktuátoroch a bezkontaktných tesniacich systémoch. Je tu popísaný historický vývoj, fyzikálne vlastnosti, typy a zloženie, ako aj na konkrétne oblasti priemyselného a elektrotechnického využitia magnetických kvapalín. Súčasťou príspevku je tiež analýza moderných aplikácií v oblasti mechatroniky, medicíny, energetiky a elektromobility.**

***Kľúčové slová* — Magnetická kvapalina, ferrofluid, magnetoreologická kvapalina**

I. Úvod

Technologický pokrok v elektrotechnike, mechatronike a automatizácii prináša rastúce nároky na spoľahlivosť, účinnosť a zmenšovanie rozmerov zariadení. S tým súvisí aj potreba nových riešení, ktoré umožnia zlepšiť chladenie, tesnenie alebo presné riadenie pohybu bez zložitého mechanického zásahu. Medzi materiály, ktoré v tomto smere majú veľký potenciál, patria magnetické kvapaliny, teda látky, ktoré kombinujú vlastnosti kvapaliny s reakciou na magnetické pole.

Magnetické kvapaliny obsahujú veľmi malé častice kovov, napríklad oxid železitý, ktoré sú rovnomerne rozptýlené vo vode, v oleji alebo v inom nosnom médiu. Keď sa na tieto kvapaliny pôsobí magnetickým poľom, dokážu zmeniť svoje správanie, teda napríklad zmeniť polohu, tuhnúť alebo sa sústrediť v určitom mieste. Tieto vlastnosti sa dnes čoraz viac využívajú v technických aplikáciách.

V elektrotechnike môžu magnetické kvapaliny pomáhať pri chladení výkonných súčiastok, tlmení vibrácií, tvorbe bezdotykových tesnení alebo aj v pohonoch a snímačoch. Ich výhodou je, že reagujú rýchlo a dajú sa presne ovládať. Tým otvárajú možnosti pre tvorbu moderných a kompaktných riešení. Využitie týchto kvapalín dnes nachádzame nielen v laboratóriách, ale už aj v praxi, napr. v počítačoch, elektromobiloch, lekárskych prístrojoch, či riadiacich systémoch.

Aj napriek ich potenciálu však zostáva veľa otázok nezodpovedaných. Je potrebné ďalej skúmať, ako tieto kvapaliny pôsobia pri dlhodobom používaní na okolie, ako ich zloženie ovplyvňuje ich účinok a ako ich čo najlepšie využiť v rôznych zariadeniach. Práve preto sa magnetické kvapaliny stali predmetom výskumu a rozvoja v posledných rokoch.

II. História a základné charakteristiky magnetických kvapalín

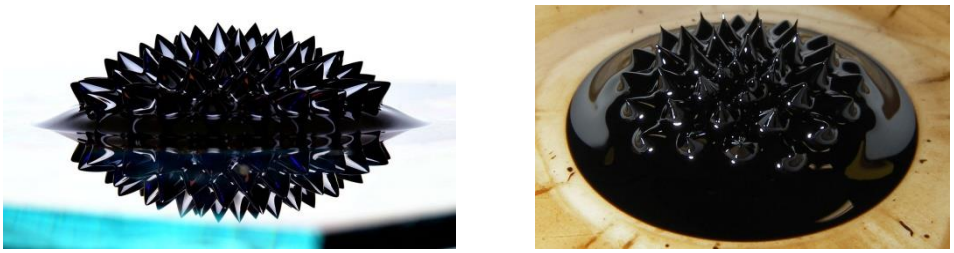
Prvé zmienky o magnetických kvapalinách siahajú až do roku 1938, keď Elmore opísal ich prípravu. Hoci boli tieto kvapaliny pôvodne určené na štúdium magnetických doménových štruktúr v tuhých látkach, ich intenzívny výskum začal až v 60. rokoch 20. storočia.

V roku 1963 pracovali vedci v NASA Lewis Research Center v Clevelande na vývoji a testovaní rôznych dizajnov spaľovacích komôr a vstrekovačov pre motory s kvapalným vodíkom. Ich cieľom bolo vyriešiť problémy s nestabilitou spaľovania, ktoré by mohli ohroziť úspešné vyslanie astronautov na Mesiac. Stephen Papell sa počas svojej práce v NASA zameriaval na kvapalné paliva. V tom čase nebolo jasné, ako by sa palivo v stave beztiaže mohlo dostať do spaľovacej komory počas opätovného naštartovania motora. Papell navrhol pridať do raketového paliva magnetický prach, čím by ho bolo možné pomocou silného magnetu usmerniť do spaľovacej komory. Na Obr.1 Stephen Papell demonštruje účinky magnetického poľa na magnetickú pohonnú látku zmiešanú s vodou.

Papellov návrh bol síce funkčný, no nikdy sa široko nevyužíval. Riešenie spočívalo v pridávaní rozdrvených častíc oxidu železa do paliva, čo by však podľa výskumníka Manfreda Ehresmanna z Inštitútu pre vesmírne systémy na Univerzite v Stuttgarte ovplyvnilo efektivitu rakiet. Napriek tomu získal Papell v roku 1965 patent na magnetickú kvapalinu, ktorá mala zaujímavé vlastnosti nielen v prostredí nulovej gravitácie [1].

Obr. 1 Výskumník Stephen Papell [1].

Ferrofluidy sa mimo magnetického poľa správajú ako bežné kvapaliny, no keď sa nádoba s ferrofluidom priblíži k silnému magnetu, povrch tekutiny získava pevný vzhľad s ostrými hrotmi, ktoré kopírujú línie magnetického poľa (Obr. 2). Tento jav je známy ako Rosensweigova nestabilita, pomenovaná po Ronaldovi Rosensweigovi, americkom chemikovi, ktorý veľkú časť svojej kariéry zasvätil štúdiu správania sa magnetických a ferrofluidných kvapalín.



Obr. 2 Magnetická kvapalina v silnom magnetickom poli [2].

Tieto jedinečné vlastnosti ferrofluidov ich predurčujú na atraktívne vedecké demonštrácie a umelecké inštalácie, aj keď sa nikdy neujali ako prísada do raketových palív. Za posledných 40 rokov sa ferrofluidy využívajú najmä v kvapalinových tesneniach pre rotujúce časti, napríklad v pevných diskoch, kde sú stabilizované magnetmi a umožňujú plynulý, vysokorýchlostný chod s minimálnymi vibráciami a hlukom.

A. Vlastnosti magnetických kvapalín

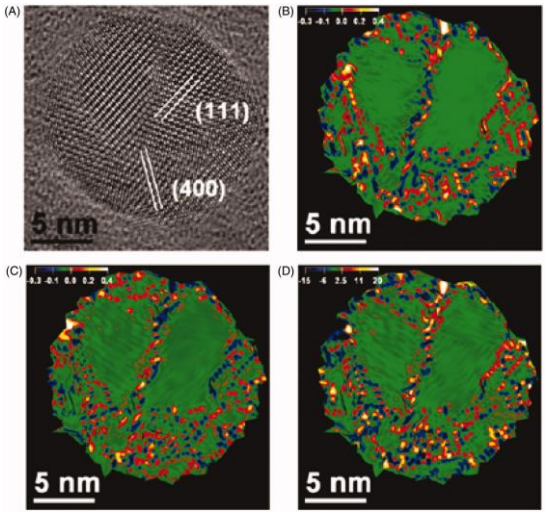
Magnetické vlastnosti - Magnetické kvapaliny reagujú na vonkajšie magnetické pole usporiadaním častíc pozdĺž magnetických indukčných čiar. Tento jav spôsobuje charakteristickú povrchovú štruktúru, známu ako Rosensweigova nestabilita. Saturácia magnetizácie týchto kvapalín závisí od tvaru a veľkosti častíc; napríklad častice v tvare kocky vykazujú vyššiu magnetizáciu ako guľové častice rovnakého objemu.

Reologické vlastnosti - Viskozita magnetických kvapalín sa mení v závislosti od intenzity aplikovaného magnetického poľa, čo umožňuje ich použitie v adaptívnych tlmičoch a ventiloch. Stabilitu týchto kvapalín zabezpečuje chemická úprava častíc, ktorá minimalizuje ich vzájomné zhlukovanie.

Teplotné vlastnosti - Magnetické kvapaliny vykazujú vysokú stabilitu v širokom rozsahu teplôt, čo ich robí atraktívnymi pre technické aplikácie, ako sú chladiace systémy alebo hypertermia. Schopnosť generovať teplo závisí od veľkosti častíc, ich distribúcie a intenzity aplikovaného magnetického poľa.

Tvar a veľkosť častíc -. Tvar častíc významne ovplyvňuje ich vlastnosti. Kvapaliny obsahujúce častice s komplexnými štruktúrami, napríklad „kvetové“ tvary, dosahujú vyššiu účinnosť pri generovaní tepla. Ideálny rozsah veľkosti častíc pre väčšinu aplikácií je 10 až 50

nm [3]. Na Obr. 3 je zobrazený výstup z vysokorozlišovacej transmisnej elektrónovej mikroskopie (HRTEM), ktorá zachytáva kryštalografické roviny nanočastice magnetickej kvapaliny. Obrázky poukazujú na štruktúru a orientáciu rovín (111) a (400), ako aj na analýzu morfológie a rozloženia fáz v rámci jednej častice. Tieto údaje sú dôležité pre pochopenie magnetických vlastností a stability nanočastíc vo ferrofluidných aplikáciách.



Obr. 3 Vysokorozlišovacia transmisná elektrónová mikroskopia (HRTEM) zobrazujúca kryštalografické roviny nanočastice [3].

B. Typy magnetických kvapalín a ich zloženie

Magnetické kvapaliny, známe aj ako ferrofluidy, sú špeciálne navrhnuté suspenzie magnetických nanočastíc (najčastejšie oxid železitý – magnetit alebo hematit) dispergovaných v nosnej kvapaline, ako je voda alebo olej. Na zabránenie aglomerácie častíc sa pridávajú povrchovo aktívne látky (surfaktanty), ktoré vytvárajú stabilizačný film okolo jednotlivých častíc a tým zaisťujú dlhodobú koloidnú stabilitu sústavy.

Rozlišujeme niekoľko základných typov magnetických kvapalín podľa ich zloženia a vlastností:

• Klasické ferrofluidy: Tieto kvapaliny sa skladajú z feromagnetických alebo paramagnetických nanočastíc suspendovaných v nosnom médiu. Charakteristickým znakom je, že sú magneticky aktívne iba v prítomnosti vonkajšieho magnetického poľa.

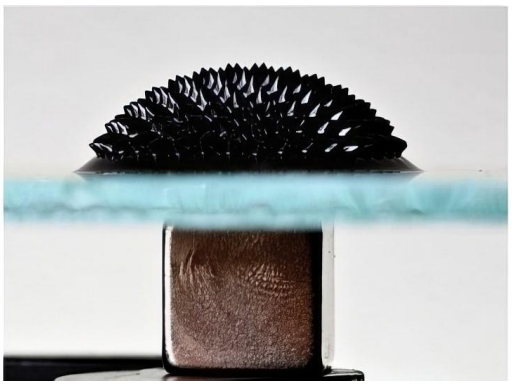
• Iónové magnetické kvapaliny: Tieto sú tvorené na báze iónových kvapalín, ktoré majú schopnosť reagovať na magnetické pole. Príkladom je zlúčenina na báze pyrolu a chloridu železitého, vyvinutá v Japonsku v roku 2004.

• Magnetoreologické kvapaliny: Na rozdiel od klasických ferrofluidov, tieto kvapaliny menia svoju viskozitu v prítomnosti magnetického poľa, čím sa stávajú vhodné pre aplikácie v adaptívnych tlmičoch a brzdách.

• Trvalomagnetizované kvapaliny: Novým typom sú tzv. permanentné magnetické kvapaliny, ktoré si dokážu zachovať magnetické vlastnosti aj po odstránení vonkajšieho poľa. Tento typ bol objavený náhodne pri experimentoch s 3D tlačou kvapalín obsahujúcich oxidy železa, olej a vodu. V dôsledku interakcie medzi magnetickými nanočasticami a surfaktantom vzniká stabilné usporiadanie častíc, ktoré zabezpečuje trvalú magnetizáciu.

Zloženie týchto kvapalín je navrhnuté tak, aby zabezpečovalo stabilitu, reaktivitu na magnetické pole a vhodné fyzikálno-chemické vlastnosti pre špecifické priemyselné, biomedicínske či technologické aplikácie [2].

Na Obr. 4 je zachytený jav, pri ktorom magnetická kvapalina umiestnená na skle vytvára výrazné hrotovité štruktúry. Tieto štruktúry kopírujú priebeh magnetických siločiar, ktoré vznikajú pôsobením magnetu umiestneného pod sklom. Tento efekt demonštruje vysokú citlivosť ferrofluidu na externé magnetické pole a jeho schopnosť tvoriť stabilné geometrické vzory v závislosti od intenzity a rozloženia poľa.



Obr. 4 Magnetická kvapalina - štruktúry kopírujú magnetické siločiary magnetu umiestneného pod sklom [2].

C. Typy nosných kvapalín a ich vplyv na vlastnosti magnetických kvapalín

Nosná kvapalina je jednou z kľúčových zložiek magnetických kvapalín a významne ovplyvňuje ich výsledné fyzikálne vlastnosti, ako sú viskozita, stabilita, magnetická odozva, životnosť a aplikovateľnosť. V magnetoreologických aj feromagnetických kvapalinách sa ako nosič používajú buď nepolárne oleje, polárne kvapaliny (napr. voda), alebo syntetické kvapaliny, ako napríklad silikónové oleje.

V prípade magnetoreologických (MR) kvapalín sa preferujú viskózne oleje s vyššou hustotou, ktoré pomáhajú udržiavať pevné magnetické častice (napr. železný prášok) rovnomerne rozptýlené. Použitie rôznych typov kvapalín ovplyvňuje výslednú viskozitu v stave „off“ aj „on“. Napríklad MR kvapaliny pripravené so silikónovým olejom s viskozitou 100 cP vykazovali pri pokojovom stave viskozitu približne 1,9 P, zatiaľ čo pri vyšších šmykových rýchlostiach ich viskozita výrazne klesala (tzv. shear thinning efekt).

Na druhej strane, ferrofluidy, ktoré využívajú ultrajemné nanočastice (2–10 nm), často používajú vodu alebo olej ako nosné médium. Vzhľadom na ich koloidnú povahu musí byť nosná kvapalina chemicky kompatibilná s použitým surfaktantom, ktorý stabilizuje častice. Dôležitá je aj polarita kvapaliny – napríklad voda je vhodná pre elektrostatickú stabilizáciu, zatiaľ čo oleje sú stabilnejšie vo vysokoteplotnom prostredí.

Nosné kvapaliny zároveň ovplyvňujú magnetickú odozvu. Experimenty ukazujú, že rovnaké magnetické častice môžu pri rôznych kvapalinách vykazovať rozdielne hodnoty viskozity a prahového napätia. Okrem toho je potrebné zohľadniť aj koróznu odolnosť, oxidačnú stabilitu a schopnosť kvapaliny fungovať v rozšírenom teplotnom rozsahu.

Pre aplikácie s vysokými požiadavkami na životnosť a konzistenciu sa preto čoraz častejšie používajú silikónové oleje, ktoré sú chemicky inertné a tepelne stabilné. Naopak, v dočasných alebo lacnejších aplikáciách postačujú aj minerálne oleje alebo voda [4]. Tabuľka 1 zobrazuje hlavné rozdiely vo veľkosti častíc, hustote, nosnej kvapaline, viskozite, potrebnej intenzite poľa a spôsobe aktivácie. Tieto parametre významne ovplyvňujú fyzikálne správanie kvapalín a ich využiteľnosť v technických aplikáciách.

Tabuľka 1
Porovnanie vlastností magnetoreologických kvapalín (MR), elektroreologických kvapalín (ER) a ferrofluidov.

Vlastnosť	***MR kvapaliny***	***ER kvapaliny***	***Ferrofluidy***
Častice (materiál)	Železo, ferity a pod.	Zeolity, polyméry, SiO_2, $BaTiO_3$	Ferity, železo, kobalt, keramika
Veľkosť častíc	0,1 – 10 µm	0,1 – 10 µm	2 – 10 nm
Nosná kvapalina	Nepolárne oleje, polárne kvapaliny, voda	Oleje	Oleje, voda
Hustota (g/cm³)	3 – 5	1 – 2	1 – 2
Viskozita (mP)	100 – 1000	50 – 1000	2 – 500
Požadované pole	~ 3 kV/mm	3 kV/mm	~ 1 kV/mm

D. Stabilizácia magnetických kvapalín

Magnetické kvapaliny (ferrofluidy) si vyžadujú stabilizačné opatrenia, aby zostali dlhodobo homogénne a účinné. Stabilita sa dosahuje kombináciou chemickej úpravy povrchu nanočastíc a pôsobenia fyzikálnych síl v kvapaline. Každá magnetická častica, najčastejšie magnetit (Fe_3O_4), je obalená vrstvou surfaktantu – povrchovo aktívnej látky, ktorá zabraňuje ich zhlukovaniu.

Tento ochranný obal vytvára medzi časticami odpudivé sily, čím sa predchádza aglomerácii a usadzovaniu častíc na dne nádoby.

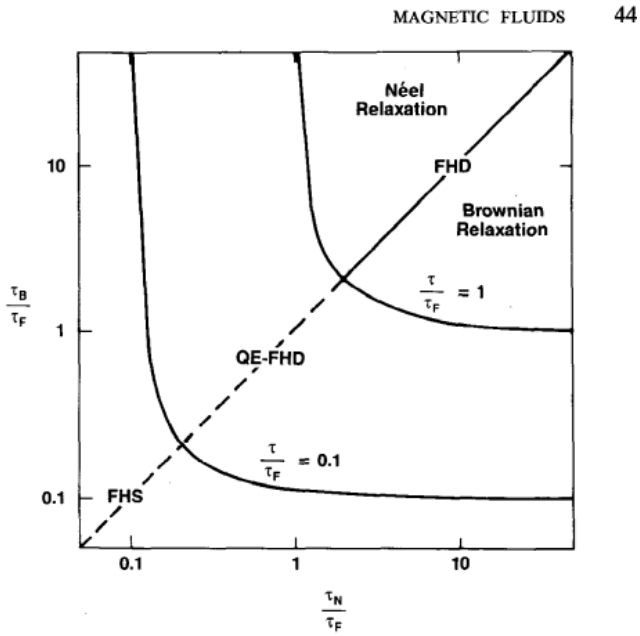
Ďalším kľúčovým faktorom stability je Brownov pohyb – náhodný tepelný pohyb častíc, ktorý udržiava ich rovnomerné rozloženie v nosnej kvapaline, ako je olej alebo voda. Vďaka tomu ostáva kvapalina stabilná aj v prípade dlhodobého skladovania alebo pri pôsobení vonkajšieho magnetického poľa.

Z hľadiska magnetických vlastností hrá dôležitú úlohu tzv. relaxačný mechanizmus, ktorý popisuje, ako rýchlo sa magnetizácia kvapaliny prispôsobuje zmenám vonkajšieho poľa. Existujú dva základné typy:

- Brownovská relaxácia – založená na fyzickej rotácii celej častice v kvapaline; dominuje pri väčších časticiach.
- Néelovská relaxácia – prebieha vo vnútri častice zmenou orientácie jej magnetického momentu; prevláda pri veľmi malých časticiach.

Pre častice magnetitu je hranica medzi týmito mechanizmami približne pri veľkosti 10 nanometrov. Výber správneho typu stabilizácie a častíc ovplyvňuje nielen funkčnosť kvapaliny, ale aj jej použiteľnosť v konkrétnych technických aplikáciách – napríklad v chladiacich systémoch, senzoroch či aktuátoroch. Stabilita je preto základnou podmienkou pre spoľahlivé a opakovateľné správanie magnetickej kvapaliny v praxi [5].

Na Obr. 5 je znázornené približné rozdelenie oblastí toku magnetickej kvapaliny podľa dominantného relaxačného mechanizmu. Graf rozlišuje medzi Brownovskou a Néelovskou relaxáciou, ktoré určujú správanie častíc v magnetickom poli v závislosti od časových mierok. Tento prehľad umožňuje porozumieť prechodom medzi ferrohydrostatickou, kvázirovnovážnou a dynamickou oblasťou toku, čo je kľúčové pri návrhu použitia magnetických kvapalín pre konkrétne aplikácie.



Obr. 5 Približné rozdelenie oblastí toku magnetickej kvapaliny na základe dominantného relaxačného mechanizmu – Brownovského alebo Néelovského [5].

III. Aplikácie magnetických kvapalín v rôznych priemyselných odvetviach

Magnetické kvapaliny patria medzi materiály s unikátnymi fyzikálno-chemickými vlastnosťami, vďaka ktorým nachádzajú široké uplatnenie v rôznych priemyselných oblastiach. Ich schopnosť dynamicky reagovať na vonkajšie magnetické pole a zároveň si zachovať tekutý charakter umožňuje ich využitie v technických aj biomedicínskych aplikáciách.

A. Použitie v medicíne: Magnetická hypertermia, cielená doprava liečiv a diagnostické aplikácie

Magnetické kvapaliny nachádzajú široké uplatnenie v medicíne na presné a efektívne využitie v rôznych terapeutických a diagnostických aplikáciách. Ich schopnosť byť riadené externým magnetickým poľom umožňuje cielené dodávanie liečiv, terapiu rakovinových ochorení, separáciu biologických vzoriek a inovatívne diagnostické metódy.

Na Obr. 6 sú znázornené rôzne vzorky magnetických kvapalín s rozdielnym farebným zafarbením, ktoré poukazujú na ich variabilné zloženie a koncentráciu nanočastíc. Je to

výskumný pokrok v oblasti biomedicínskeho využitia a vo vývoji magnetických kvapalín s prispôsobenými vlastnosťami pre ďalšie špecifické medicínske aplikácie – napríklad na cielenú dopravu liečiv, či diagnostické účely.

Obr. 6 Vzorky magnetických kvapalín s rôznym zložením [6].

Magneticky riadené častice sa využívajú ako nosiče liečiv, čo umožňuje ich presnú lokalizáciu v požadovanej oblasti tela. Táto metóda je obzvlášť užitočná pri chemoterapii, kde sa lieky môžu koncentrovať v nádorových tkanivách, čím sa minimalizujú systémové vedľajšie účinky. Klinické testy s využitím epirubicínu, viazaného na magnetické častice, ukázali úspešné doručenie liečiva do nádorov u približne polovice pacientov.

Okrem dopravy liečiv sa magnetické kvapaliny využívajú aj v hypertermii – terapii, ktorá využíva teplo na ničenie rakovinových buniek. Aplikáciou striedavého magnetického poľa na ferrokvapaliny v nádorových tkanivách dochádza k lokálnemu zvýšeniu teploty, čo vedie k zničeniu rakovinových buniek bez poškodenia okolitých zdravých tkanív.

V oblasti biotechnológie sa magnetické kvapaliny využívajú na separáciu a purifikáciu biomolekúl. Magneticky značené častice umožňujú efektívnu izoláciu mRNA, čo je dôležité pre analýzu génovej expresie. Táto metóda je rýchlejšia a efektívnejšia než tradičné separačné techniky, pričom eliminuje potrebu organických rozpúšťadiel a zjednodušuje purifikačné kroky.

Magnetické kvapaliny nachádzajú široké uplatnenie v moderných diagnostických technikách, najmä v oblasti zobrazovacích metód a analytických testov. Ich unikátne magnetické vlastnosti umožňujú presnejšiu detekciu patologických zmien v organizme a zlepšujú efektivitu existujúcich diagnostických postupov.

Jednou z najvýznamnejších oblastí využitia magnetických kvapalín je magnetická rezonancia (MRI), kde slúžia ako kontrastné činidlá. Nanočastice feromagnetických materiálov, ako je oxid železitý (Fe_3O_4), sa používajú na zvýšenie kontrastu medzi zdravým a patologickým tkanivom. Tento efekt je spôsobený ich schopnosťou ovplyvňovať lokálne magnetické pole, čím sa zvyšuje kvalita obrazov získaných MRI.

Klinické štúdie ukázali, že magnetické kvapaliny môžu výrazne zlepšiť detekciu nádorov a zápalových procesov. Výskum sa v súčasnosti zameriava na vývoj biokompatibilných nanočastíc, ktoré umožnia cielené zobrazovanie špecifických typov rakovinových buniek.

Magnetické kvapaliny sa využívajú aj pri imunomagnetických testoch, ktoré slúžia na rýchlu detekciu mikroorganizmov a biomarkerov v krvi, moči alebo iných telesných tekutinách. Magneticky značené protilátky sa viažu na špecifické antigény, čím umožňujú presnejšiu a citlivejšiu diagnostiku ochorení, ako sú bakteriálne a vírusové infekcie.

Podobný princíp sa uplatňuje v biosenzoroch, kde sú magnetické častice integrované do analytických systémov na detekciu biomolekúl. Tieto metódy majú potenciál na vývoj neinvazívnych testov, ktoré by mohli nahradiť tradičné diagnostické prístupy.

Magnetické kvapaliny sú využívané aj v separačných technikách na izoláciu buniek, nukleových kyselín a proteínov. Napríklad magneticky značené častice sa používajú na separáciu rakovinových buniek z krvného obehu, čo umožňuje skoršiu diagnostiku metastáz a monitorovanie efektivity liečby [6].

B. Aplikácie v automobilovom priemysle

Magnetické kvapaliny, najmä magnetoreologické (MRF), nachádzajú široké uplatnenie v automobilovom priemysle vďaka svojej schopnosti meniť viskozitu pod vplyvom magnetického poľa. Táto vlastnosť umožňuje vývoj adaptívnych systémov tlmenia vibrácií, inteligentných brzdových systémov, spojok a ďalších komponentov zlepšujúcich výkon a bezpečnosť vozidiel [7]. MR tlmiče sa používajú v moderných vozidlách ako súčasť adaptívnych podvozkových systémov. Tradičné tlmiče využívajú hydraulické kvapaliny, avšak pri použití

MR kvapaliny môže byť úroveň tlmenia dynamicky regulovaná podľa jazdných podmienok.

MR tlmič pozostáva z piestu s elektromagnetickou cievkou, ktorý je ponorený v magnetoreologickej kvapaline. Pri aktivácii elektromagnetu sa zmení usporiadanie častíc v kvapaline, čím sa zvýši odpor tlmiča. Tento systém umožňuje takmer okamžitú zmenu tuhosti podvozka v reálnom čase. Na Obr. 7 je znázornená konštrukcia MR tlmiča, ktorý využíva magnetoreologickú kvapalinu (MR fluid) na tlmenie vibrácií.

Obr. 7 Schéma MR tlmiča s detailmi konštrukčných prvkov – MR kvapalina je uložená v oblasti piestu [7].

MR spojky fungujú na podobnom princípe. MR spojky umožňujú plynulé a presne riadené prenosy krútiaceho momentu medzi motorom a prevodovkou. Tieto spojky sú výhodné pre hybridné a elektrické vozidlá, pretože umožňujú presné nastavenie preklzu a efektívnejší prenos výkonu.

MR kvapaliny sa využívajú aj v inteligentných brzdových systémoch (MRF Brakes). Tieto brzdy obsahujú MR kvapalinu umiestnenú priamo v brzde, kde pri aktivácii elektromagnetického poľa sa viskozita kvapaliny zmení, čím sa zvýši brzdný účinok.

C. Aplikácie v mechatronike

Magnetické kvapaliny nachádzajú široké uplatnenie aj v oblasti mechatroniky, kde sa využívajú na tlmenie vibrácií, precízne riadenie pohybov a vývoj inteligentných senzorov a aktuátorov. Ich schopnosť meniť fyzikálne vlastnosti v prítomnosti magnetického poľa umožňuje vývoj adaptívnych systémov, ktoré reagujú na zmeny prostredia v reálnom čase.

Jednou z kľúčových oblastí využitia magnetických kvapalín v mechatronike je tlmenie vibrácií. Magnetické kvapaliny sa používajú v polohovacích a stabilizačných systémoch, kde dokážu efektívne tlmiť mechanické oscilácie. Mechanizmus tlmenia je založený na schopnosti magnetických nanočastíc vytvárať reťazcovú štruktúru v prítomnosti magnetického poľa, čím sa mení viskozita kvapaliny. Pri správnom nastavení magnetického poľa je možné dynamicky regulovať tuhosť tlmiaceho systému, čím sa dosahuje účinné tlmenie vibrácií v širokom frekvenčnom spektre. Príklady využitia:

- Tlmenie vibrácií v precíznych obrábacích strojoch, kde sa vyžaduje eliminácia mikrovibrácií pre vyššiu presnosť výroby.
- Tlmenie vibrácií v robotických ramenách, čo umožňuje hladší a stabilnejší pohyb manipulátorov.
- Odpruženie sedadiel vo vozidlách a lietadlách, kde sa magnetické kvapaliny používajú na aktívne tlmenie nárazov.

Magnetické kvapaliny sa využívajú aj na presné riadenie pohybov v rôznych mechatronických systémoch. Vďaka ich rýchlej odozve na magnetické pole je možné využiť ich na riadenie mechanických pohybov v aktuátoroch a mikropozicionovacích systémoch. Príklady využitia:

- Mikropozicionovacie systémy v optike, kde sa magnetické kvapaliny využívajú na jemné nastavovanie zrkadiel a šošoviek v laserových prístrojoch.
- Presné polohovanie komponentov v elektronickej výrobe, kde je potrebná vysoká presnosť a spoľahlivosť.
- Inteligentné spojky a prevodovky, ktoré umožňujú plynulé riadenie krútiaceho momentu v robotických systémoch.

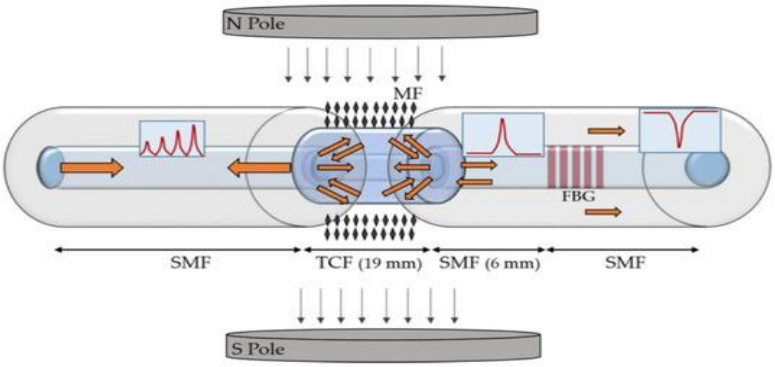
Magnetické kvapaliny sa v posledných rokoch uplatňujú aj v oblasti senzoriky a aktuátorov, najmä v kombinácii s optickými vláknami a mikromechanickými systémami (MEMS). Senzory založené na magnetických kvapalinách sú schopné presne detegovať zmeny magnetického poľa, čo umožňuje ich využitie v rôznych aplikáciách:

• Optické vláknové senzory magnetického poľa, ktoré sú imúnne voči elektromagnetickému rušeniu a nachádzajú uplatnenie v priemyselnej diagnostike.

• Biomedicínske senzory, ktoré využívajú magnetické kvapaliny na presnú detekciu biomolekúl a buniek.

• Akcelerometre a gyroskopy, kde magnetické kvapaliny slúžia ako detekčný prvok na meranie pohybu a vibrácií.

Na Obr. 8. je zobrazený optický senzor založený na TCF-FBG, ktorý využíva magnetickú kvapalinu na detekciu zmien magnetického poľa. Tento typ senzora je vhodný na presné merania aj v prostredí s elektromagnetickým rušením [8].



Obr. 8 Schéma optického senzora založeného na TCF-FBG [8].

Magnetické kvapaliny sa využívajú aj v aktuátoroch, kde umožňujú presnú a rýchlu reguláciu pohybu. Súčasné výskumy ukazujú, že magnetické kvapaliny môžu byť použité na vývoj nových typov lineárnych a rotačných aktuátorov, ktoré majú vysokú odozvu a sú vhodné pre presné riadenie robotických systémov.

Magnetické kvapaliny zohrávajú čoraz významnejšiu úlohu v nových smeroch robotiky, najmä v oblasti mäkkých a kvapalných robotických systémov, kde sa kladie dôraz na prispôsobivosť a schopnosť operovať v extrémnych podmienkach. Na rozdiel od tradičných tuhých robotov umožňujú tzv. inteligentné kvapalné systémy (Smart Fluid Systems – SFS) robotom meniť tvar, reagovať na podnety z okolia a autonómne sa prispôsobovať prostrediu.

Tieto systémy môžu fungovať ako aktuátory – zabezpečujú pohyb prostredníctvom vnútornej zmeny objemu či tlaku. Napríklad kvapalina uzavretá v pružnej membráne dokáže generovať tlak na okolité prostredie, čím vzniká pohybový impulz. Rovnako dôležitá je aj ich funkcia ako senzorov, ktoré reagujú na zmeny teploty, tlaku, magnetického poľa či dokonca chemické látky.

Magnetické kvapaliny majú pozoruhodné vlastnosti, ktoré umožňujú ich využitie v oblasti optiky. Jednou z kľúčových vlastností je ich optická izotropnosť, ktorá sa však vplyvom vonkajšieho magnetického poľa mení – materiál môže vykazovať indukovanú dvojlomnosť. Tento jav umožňuje ich použitie v rôznych magneto-optických zariadeniach. Ferrofluidy môžu meniť svoj tvar v závislosti od pôsobenia magnetického poľa. Táto schopnosť sa využíva napríklad pri vytváraní variabilných šošoviek, kde sa zakrivenie povrchu kvapaliny mení a tým aj jej optická mohutnosť. Takéto riešenie je vhodné v systémoch automatického zaostrovania, v inteligentných kamerách alebo mikroskopoch.

IV. VYUŽITIE MAGNETICKÝCH KVAPALÍN V ELEKTROTECHNIKE

Magnetické kvapaliny sa stali významnou súčasťou moderných elektrotechnických systémov, kde ich jedinečné vlastnosti umožňujú zlepšiť chladiace mechanizmy, ochrániť citlivé súčiastky a zvýšiť presnosť senzorových a aktuátorových systémov. Vďaka ich schopnosti reagovať na magnetické pole si zachovávajú vysokú flexibilitu pri aplikáciách v dynamických prostrediach.

A. Transformátory a indukčné zariadenia

V oblasti energetiky, najmä pri transformátoroch a indukčných zariadeniach, sú magnetické kvapaliny perspektívnym médiom na zlepšenie tepelných a elektrických vlastností izolačných olejov. V posledných rokoch sa testuje nahradenie alebo doplnenie klasického transformátorového oleja magnetickými nanočasticami (najčastejšie oxidu železa), čím vznikajú tzv. magnetické nanokvapaliny (magnetic nanofluids). Jedným z hlavných dôvodov tohto prístupu je zvýšenie tepelnej vodivosti oleja. Pridaním magnetických častíc sa tepelný tok z

cievok a jadra transformátora do chladiaceho média zlepšuje, čo vedie k efektívnejšiemu odvodu tepla a zníženiu prevádzkovej teploty. Tento jav je výraznejší, ak sa v systéme aplikuje vonkajšie magnetické pole – výsledná vodivosť sa totiž stáva anizotropná, teda závislá od smeru poľa.

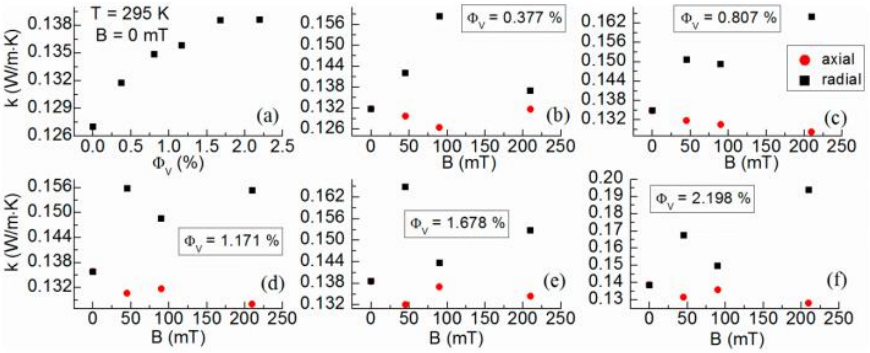
Okrem tepelných parametrov ovplyvňuje prítomnosť magnetických nanočastíc aj dielektrické vlastnosti kvapaliny. Výskumy potvrdzujú, že pridané častice zvyšujú elektrickú pevnosť oleja, čo znižuje riziko elektrického prierazu. Tento efekt súvisí so spomalením vývoja elektrických výbojov (streamerov), ktoré sa za bežných okolností šíria veľmi rýchlo.

Ďalším významným javom je tzv. magnetoviskózny efekt, teda zmena viskozity kvapaliny pod vplyvom pôsobiaceho magnetického poľa. Viskozita rastie so silou poľa aj s koncentráciou častíc. Hoci vyššia viskozita môže znížiť prúdenie, v určitých prípadoch môže byť tento jav výhodný – napríklad pre dynamické riadenie chladenia alebo pri obmedzení vibrácií.

Dôležitým aspektom použiteľnosti magnetických kvapalín v praxi je ich dlhodobá stabilita. Pre zabezpečenie rovnomerného rozptýlenia častíc sa používajú povrchovo aktívne látky (napr. kyselina olejová), ktoré vytvárajú ochrannú vrstvu a bránia zhlukovaniu. Stabilné magnetické nanokvapaliny vykazovali v štúdiách zachovanie vlastností aj po viac ako 24 mesiacoch, čo je kľúčové pre aplikáciu v trvalo zaťažených zariadeniach ako sú transformátory.

Celkovo magnetické kvapaliny otvárajú nové možnosti pre zvýšenie efektivity a životnosti transformátorových olejov, a to najmä vďaka kombinácii vylepšených tepelných, viskozitných a izolačných vlastností [9].

Na Obr. 9 sú znázornené výsledky merania tepelnej vodivosti magnetickej kvapaliny v závislosti od intenzity magnetického poľa pri rôznych koncentráciách magnetických nanočastíc. Merania boli realizované v axiálnom a radiálnom smere, aby sa porovnali rozdiely v smere pôsobenia magnetického poľa.



Obr. 9 Závislosti tepelnej vodivosti magnetickej kvapaliny od intenzity magnetického poľa pri rôznych koncentráciách nanočastíc [9].

B. Chladenie a tepelné manažovanie elektronických súčiastok

S rastúcimi nárokmi na výkon a zmenšovaním rozmerov elektronických súčiastok narastá aj potreba efektívneho chladenia. Tradičné chladiace kvapaliny, ako je voda alebo etylénglykol, často nepostačujú pri vysokom tepelnom zaťažení. Z tohto dôvodu sa v posledných rokoch do popredia dostávajú nanokvapaliny, medzi ktoré patria aj magnetické kvapaliny obsahujúce nanočastice kovov alebo oxidov kovov.

Tieto kvapaliny vykazujú vyššiu tepelnú vodivosť, zlepšenú viskozitnú stabilitu a umožňujú adaptívne tepelné riadenie, najmä v prípade magnetických kvapalín, ktorých správanie možno ovplyvniť vonkajším magnetickým poľom. Nanokvapaliny nachádzajú uplatnenie najmä v nasledujúcich oblastiach:

- Mikrokanálové chladiace systémy – kompaktné riešenia na odvod tepla z vysokovýkonných čipov.
- Tepelné rúrky (heat pipes) a termosyfóny – pasívne chladiace systémy, kde magnetické kvapaliny zlepšujú efektivitu prestupu tepla.
- Termoelektrickéchladiče(TEC) – aktívne systémy, kde nanokvapaliny znižujú tepelný odpor.

Hlavné výhody nanokvapalín využívaných na chladenie elektroniky zahŕňajú:

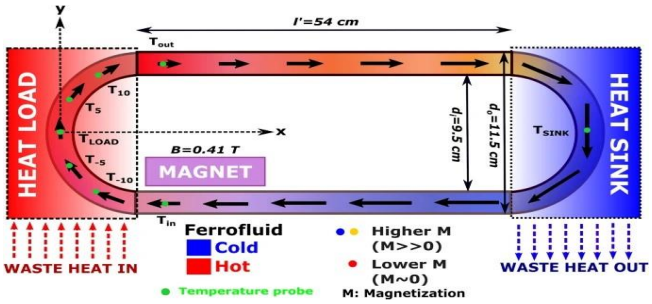
- zvýšenie koeficientu prestupu tepla až o 70 % v porovnaní so základnými kvapalinami,
- možnosť lokálne riadeného prúdenia pomocou magnetického poľa pri magnetických kvapalinách,
- zníženie tepelného odporu medzi čipom a chladiacim médiom,
- zvýšenie spoľahlivosti a životnosti elektronických súčiastok.

Nanokvapaliny na báze magnetických kvapalín sú dnes skúmané aj pre nasadenie v:

- chladení CPU a GPU v počítačoch a serveroch,
- výkonových transformátoroch a elektrických rozvádzačoch,
- elektronike v elektromobiloch a autonómnych zariadeniach.

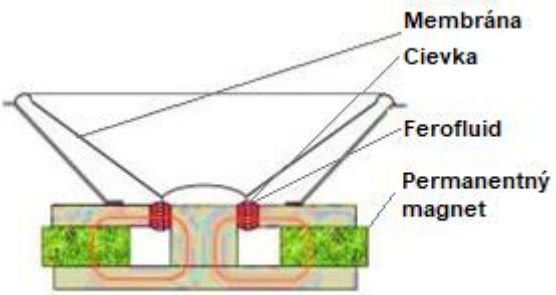
Ich vývoj predstavuje významný krok smerom k novej generácii chladiacich technológií, ktoré budú kompaktné, efektívne a adaptívne [10].

Na Obr. 10 je znázornená schéma tepelného manažmentu pomocou magnetickej kvapaliny. Táto technológia využíva cirkuláciu ferrofluidu medzi zdrojom tepla a chladiacim článkom. Tento princíp zvyšuje efektivitu odvodu tepla, čím sa zvyšuje účinnosť chladenia elektronických komponentov.



Obr. 10 Schéma tepelného manažovania pomocou magnetickej kvapaliny [10].

V reproduktoroch sa magnetická kvapalina často používa ako chladiace médium umiestnené v magnetickej medzere, kde sa nachádza cievka. Ferrofluid je držaný na mieste pôsobením magnetického poľa, a keďže jeho tepelná vodivosť je niekoľkonásobne vyššia ako vodivosť vzduchu, dokáže efektívne odvádzať teplo vznikajúce pri dlhodobom zaťažení reproduktora. Tým sa zvyšuje jeho výkon a predlžuje životnosť súčiastky [9]. Na Obr. 11 je zobrazená schéma reproduktora využívajúceho magnetickú kvapalinu. Táto kvapalina zlepšuje tepelné vlastnosti systému a zároveň tlmí vibrácie membrány. Vďaka tomu sa zvyšuje účinnosť chladenia a kvalita zvuku.



Obr. 11 Reproduktor s magnetickou kvapalinou [9].

C. Tesnenie a ochrana elektromechanických systémov

Magnetické kvapaliny zohrávajú dôležitú úlohu v oblasti tesnenia a ochrany elektromechanických systémov. Ich schopnosť vytvárať dynamické tesnenia v prítomnosti magnetického poľa, ako aj zabezpečovať izoláciu a ochranu citlivých elektronických komponentov, umožňuje ich široké využitie v rôznych priemyselných aplikáciách. Medzi kľúčové oblasti ich využitia patrí tesnenie ložísk a rotujúcich častí elektromotorov, ochrana pevných diskov (HDD), použitie v hermetických systémoch a vákuových komorách, ako aj v tesniacich mechanizmoch potrubí a ventilov.

Magnetické kvapaliny môžu byť použité v „ad hoc ventiloch“, ktoré fungujú na princípe riadenia magnetického poľa. Experimentálne testy ukázali, že takéto systémy dokážu spoľahlivo zadržať kvapaliny a plyny ešte predtým, ako môže dôjsť k úniku a potenciálnej havárii v priemyselných prevádzkach. Tento princíp je možné aplikovať v mikroventiloch s priemerom menším ako 1 mm, ale aj v makroskopických ventiloch s vnútorným priemerom nad 6 mm.

Štúdie ukázali, že magnetické kvapaliny môžu byť efektívne využité v inteligentných tesniacich systémoch, ktoré dokážu automaticky regulovať prietok kvapalín na základe externého magnetického poľa. Takéto systémy môžu byť implementované napríklad v laboratórnych analytických prístrojoch, chladiacich systémoch a v procesoch chemickej výroby.

V oblasti elektromechanických zariadení sa magnetické kvapaliny využívajú predovšetkým na tesnenie ložísk a rotujúcich častí elektromotorov, kde zabezpečujú ochranu pred prachom, vlhkosťou a mechanickým opotrebením. Ložiská, ktoré sú vystavené vysokým otáčkam a mechanickému namáhaniu, vyžadujú spoľahlivú ochranu, aby sa zabránilo ich predčasnému opotrebovaniu a zníženiu účinnosti zariadenia.

Magnetické kvapaliny sa v tomto prípade aplikujú ako dynamické tesnenie, ktoré vďaka prítomnosti magnetického poľa zostáva na svojom mieste a vytvára takmer dokonalé tesnenie medzi pohyblivými a pevnými časťami. Tento princíp tesnenia umožňuje:

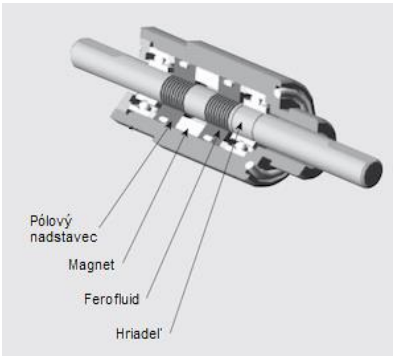
- Minimalizáciu trenia medzi tesniacimi plochami, čím sa predlžuje životnosť ložísk.
- Zníženie energetických strát, ktoré vznikajú v dôsledku mechanického opotrebenia.
- Zvýšenie ochrany pred vonkajšími kontaminantmi, ako sú prach, vlhkosť a iné nečistoty.

Využitie magnetických kvapalín v ložiskách elektromotorov je obzvlášť výhodné v extrémnych prevádzkových podmienkach, ako sú vysoké teploty, podtlakové prostredie alebo vysoká prašnosť.

V oblasti výpočtovej techniky sa magnetické kvapaliny úspešne využívajú na ochranu pevných diskov (HDD), kde zohrávajú kľúčovú úlohu v tesneniach hriadeľov rotujúcich platní. Pevné disky sú citlivé na prachové častice a iné nečistoty, ktoré by mohli poškodiť ich vnútorné mechanizmy, a preto sa magnetické kvapaliny využívajú na utesnenie diskového mechanizmu.

Štúdie preukázali, že správna kombinácia magnetickej sily a koncentrácie nanočastíc v magnetickej kvapaline dokáže výrazne zvýšiť účinnosť tesnenia a zabezpečiť dlhodobú ochranu citlivých elektronických komponentov.

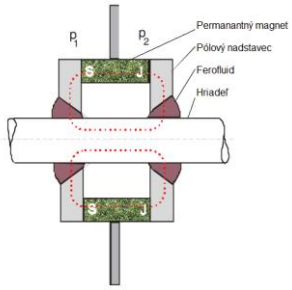
Na Obr. 12 je znázornený princíp ferrofluidného tesnenia v ložisku elektromotora. Magnetické pole udržiava magnetickú kvapalinu v stabilnej polohe medzi pólovými nástavcami, čím vytvára účinnú tesniacu bariéru. Táto bariéra bráni úniku kvapalín a zabraňuje kontaminácii ložiska, čo zabezpečuje spoľahlivú a dlhodobú ochranu vnútorných komponentov.



Obr. 12 Ferrofluidné tesnenie v ložisku elektromotora [13].

Okrem toho sa magnetické kvapaliny môžu využiť aj na tesnenie optických mechanizmov v CD/DVD mechanikách, mikropohonoch a rôznych presných zariadeniach, ktoré vyžadujú hermetické uzavretie.

Feromagnetické kvapaliny sa využívajú aj v tzv. viacvrstvových magnetických tesneniach, ktoré sú schopné zabezpečiť hermetické uzavretie rotačných hriadeľov bez potreby mechanického kontaktu (Obr.13).



Obr. 13 Viacvrstvové magnetické tesnenie hriadeľa [14].

Tieto tesnenia pracujú na princípe viacerých susediacich magnetických pólov, ktoré vytvárajú sériu bariér z magnetickej kvapaliny. Takéto usporiadanie umožňuje udržať tlakový rozdiel až do 10 atmosfér a zároveň výrazne predlžuje životnosť systému vďaka absencii

mechanického opotrebovania. Ďalšou výhodou je schopnosť prevádzky v extrémnych podmienkach, teda od nízkych teplôt (–100 °C) až po vysoké teploty (200 °C). Magnetické tesnenia sú taktiež odolné voči prachu, korozívnym plynom a biologicky agresívnym látkam, čo ich predurčuje na použitie v laboratórnych, farmaceutických či vojenských aplikáciách [11].

D. Magnetické kvapaliny v elektromotoroch a pohonných systémoch

V súčasnom vývoji elektrických pohonov a elektromotorov sa čoraz častejšie zvažuje využitie magnetických kvapalín. Integrácia týchto kvapalín predstavuje inovatívny prístup, ktorého cieľom je zlepšenie funkčnosti, spoľahlivosti a účinnosti týchto systémov. Využitie magnetických kvapalín umožňuje nielen efektívnejší teplotný manažment, ale aj adaptívne riadenie pohybu, zníženie vibrácií a predĺženie životnosti súčiastok bez použitia tradičných mechanických komponentov.

Na rozdiel od pasívnych chladiacich olejov či mazív sú magnetické kvapaliny schopné reagovať na zmenu magnetického poľa. Vďaka tomu sú atraktívnym materiálom pre vývoj inteligentných a energeticky efektívnych pohonných jednotiek, najmä v oblasti robotiky, mechatroniky či elektromobility.

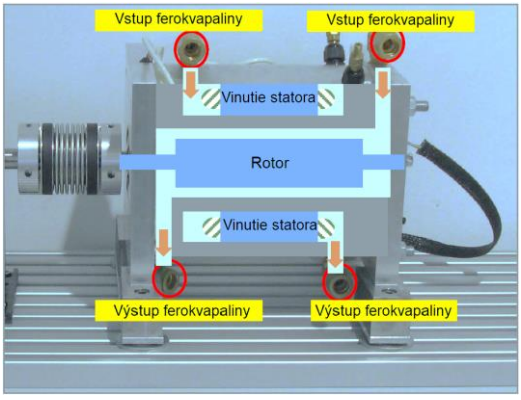
Medzi typické aplikácie využitia magnetických kvapalín patrí ich integrácia do synchrónnych elektromotorov s PM, ktoré sa bežne využívajú v priemyselnej technike, dopravných systémoch a presných pohonoch. Ich hlavnou nevýhodou je vznik intenzívneho tepla v oblasti vinutia a magnetického jadra, čo môže viesť k strate účinnosti a rýchlejšiemu starnutiu izolácií.

Štúdia prezentovaná v rámci Ilmenauského kolokvia [12] preukázala, že vyplnenie vzduchovej medzery medzi rotorom a statorom ferrofluidom znižuje tepelné zaťaženie motora a zároveň vedie k zvýšeniu momentu elektromagnetického poľa. Táto zmena je spôsobená lepším vedením tepla z aktívnych častí vinutia a stabilizáciou magnetického poľa vďaka suspenzii magnetických nanočastíc.

Okrem chladenia má použitie ferrofluidu aj ďalší praktický prínos, umožňuje znižovať spotrebu permanentných magnetov v motore. V praxi to znamená, že je možné použiť menšie množstvo drahých a ekologicky náročných materiálov (napr. NdFeB), pričom výstupný výkon ostáva zachovaný alebo dokonca rastie. To má veľký význam pre budúci vývoj elektromotorov v automobilovom priemysle alebo pohonoch v domácnostiach.

Výsledky meraní ukázali napríklad, že motor chladený ferrofluidom dosiahol pri rovnakom výkone o 10–15 °C nižšiu prevádzkovú teplotu ako referenčný motor. Tým sa výrazne predĺžila jeho životnosť a znížilo riziko poruchy[12].

Na Obr. 14 je znázornené prúdenie ferokvapaliny v experimentálnom elektromotore. Ferokvapalina cirkuluje cez vinutie statora a vzduchové medzery pomocou externého čerpadla. Vďaka tomu dochádza k efektívnejšiemu odvodu tepla z najviac zaťažených častí motora, čo vedie k zníženiu prevádzkovej teploty a zvýšeniu jeho životnosti.



Obr. 14 Tok ferokvapaliny v experimentálnom synchrónnom elektromotore [12].

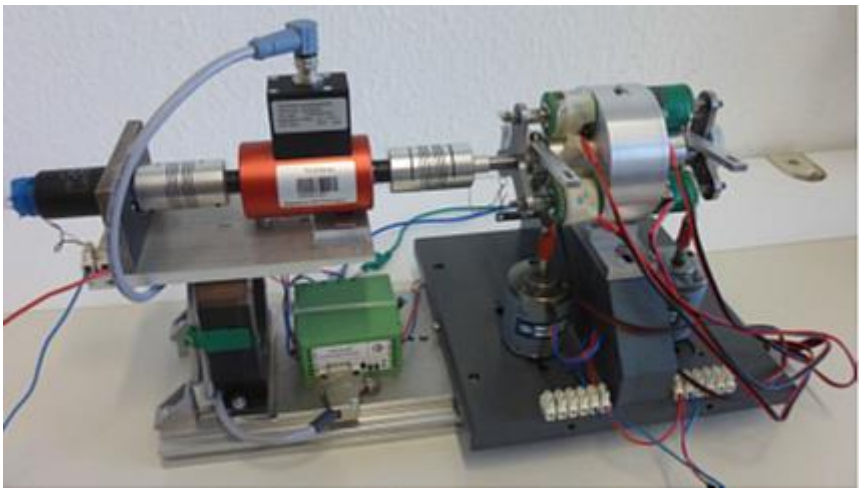
V oblasti aktuátorových systémov a adaptívneho riadenia sa čoraz viac uplatňuje nový prístup využívajúci magnetoreologické kvapaliny (MRF). Tieto kvapaliny menia svoju viskozitu v závislosti od intenzity magnetického poľa – v pokojovom stave sú tekuté, no pri aplikovaní magnetického poľa sa ich častice zoradia a vytvoria štruktúru, ktorá kladie odpor pohybu.

Pre experiment bol zostrojený aktuátor s dvoma zubovými diskami, medzi ktorými sa nachádzala MRF. Magnetické pole sa generovalo prostredníctvom cievok a intenzita bola riadená elektrickým prúdom. Takýto systém umožnil:

- plynulé riadenie krútiaceho momentu bez potreby mechanických prevodoviek,
- zmenu smeru otáčania,
- rýchlu odozvu systému bez mechanického oneskorenia.

V praxi by sa takéto aktuátory mohli využiť ako servopohon pre robotické ramená, presné prenosy sily, alebo ako bezúdržbová alternatíva ku klasickým spojkám a diferenciálom.

Na Obr. 15 je zobrazený experimentálny prototyp aktuátora využívajúceho magneto-reologickú kvapalinu (MRF). Kvapalina je umiestnená medzi otočnými diskami, pričom jej viskozita sa mení v závislosti od pôsobenia magnetického poľa. Zmenou viskozity sa riadi prenos krútiaceho momentu, čím sa umožňuje bezmechanická regulácia pohybu v rôznych aplikáciách mechatroniky a robotiky.



Obr. 15 Experimentálne meranie s MRF aktuátorom [12].

Na základe experimentálnych výsledkov možno konštatovať, že magnetické kvapaliny (najmä ferrofluidy a MRF) predstavujú viacúčelové médium, ktoré môže zlepšiť účinnosť, stabilitu a funkčnú adaptabilitu elektromotorov a pohonných systémov. Výhody použitia magnetických kvapalín spočívajú v tomto:

- znížení tepelných strát a zlepšení chladenia,
- možnosti adaptívne riadiť pohyb a odpor bez mechanického kontaktu,
- zjednodušení konštrukcie a znížení poruchovosti.

Ich využitie sa preto čoraz viac presadzuje nielen v priemysle, ale aj v medicínskej technike, vozidlách a spotrebičoch s vysokými nárokmi na presnosť a tichý chod.

V. ZÁVER

Aj keď sú magnetické kvapaliny známe už niekoľko desaťročí, ich vývoj a využitie stále nie sú úplne preskúmané ani technicky vyčerpané. Výskum v tejto oblasti ukazuje, že ide o materiály s veľkým potenciálom, ktoré v budúcnosti môžu zohrávať kľúčovú úlohu pri navrhovaní nových elektrotechnických zariadení a technológií.

Magnetické kvapaliny umožnili vznik viacerých inovatívnych riešení – od chladiacich systémov cez tlmiče vibrácií až po bezkontaktné tesnenia a snímače. Mnohé z týchto aplikácií sú stále vo vývoji a nepredstavujú zatiaľ prelomové objavy. Napriek tomu sa očakáva, že ich jedinečné vlastnosti, ako napríklad možnosť dynamického ovládania viskozity či stabilita v rôznych podmienkach, budú čoraz viac využívané v nových technických konceptoch.

Zvláštny význam má interdisciplinárny výskum magnetických kvapalín. Na ich vývoji a optimalizácii spolupracujú odborníci z elektrotechniky, fyziky, chémie, strojárstva, ale aj z medicíny či biotechnológií. Takáto spolupráca podporuje tvorbu nových zariadení, ktoré by bez magnetických kvapalín nebolo možné realizovať. V blízkej budúcnosti sa očakáva:

- vývoj nových typov magnetických kvapalín so zvýšenou stabilitou, vyššou teplotnou odolnosťou a nižšou viskozitou,
- širšie využitie v elektromobilite, pri výrobe inteligentných senzorov a pohonov,
- uplatnenie v adaptívnych systémoch riadenia vibrácií (napr. v automobiloch alebo stavebných konštrukciách).

Z hľadiska efektivity sa ukázalo, že magnetické kvapaliny dokážu zlepšiť chladenie, znížiť mechanické trenie, zvýšiť presnosť pohonov a umožniť návrh bezkontaktných systémov. Ich výhody sa prejavujú aj v oblasti spoľahlivosti, adaptability a znižovania ekologickej záťaže.

Napriek tomu však existujú aj obmedzenia, ako je náročnejšia výroba, vyššie náklady alebo obmedzená stabilita pri extrémnych podmienkach.

POĎAKOVANIE

Táto práca bola podporená projektom vedeckej agentúry KEGA podľa zmluvy číslo: 032TUKE-4/2024.

LITERATÚRA

[1] Zagorski, N., „The rise of ferrofluids“, Chemistry World. 2021. [online]. [Dátum: 23. November. 2024]. Dostupné na internete: < https://www.chemistryworld.com/features/the-rise-offerrofluids/4013756.article >.
[2] Petrescu, N., Petrescu, F. I. T., „Permanent Magnetic Fluids“. American Journal of Engineering and Applied Sciences. 2019, Vol. 12, No. 3, pp. 402–412.
[3] Salas, G., Veintemillas-Verdaguer, S., Morales, M., „Relationship between physico-chemical properties of magnetic fluids and their heating capacity“. International Journal of Hyperthermia. 2013,Vol. 29, No. 8, pp. 768–776.
[4] Genç, S., „Synthesis and Properties of Magnetorheological Fluids“, 2022 (PhD thesis). University of Pittsburgh. Dostupné na internete: < https://www.proquest.com/openview/49cd3be0d4f0166debbed92e2907f160/1?pq-origsite=gscholar&cbl=18750&diss=y >.
[5] Rosensweig, R. E., „Magnetic Fluids.“, Annual Review of Fluid Mechanics, 2012, Vol. 19, pp. 437–463.
[6] Namchand, C., Pande, P., „Application of magnetic fluids in medicine and biotechnology“. Indian Journal of Pure Applied Physics. 2001, Vol. 39, No. 10, pp. 683–686.
[7] Verma, P.K., Singh, J.R., Nandwana, H., Apurva, S., „Automotive Applications of Magnetorheological Fluids“. Department of Mechanical Engineering, Gurukul Institute of Engineeringand Technology. 2018, pp. 252–263.
[8] Alberto, N., Domingues, M.F., Marques, C., Andre, P., Antunes, P., „Optical Fiber Magnetic Field Sensors Based on Magnetic Fluid“. A Review Sensors. 2018, Vol. 18, No. 12, pp. 4325-5330.
[9] Azizi, A., Mohseni, H., „Magnetic Field Effect on Thermal, Dielectric, and Viscous Properties of a Transformer Oil-Based Magnetic Nanofluid“. Energies, 2019, Vol. 22, No. 12, pp. 4532-4537.
[10] Moita, A.S., Moreira, A.M., Pereira, J.P., „Nanofluids for the Next Generation Thermal Management of Electronics“. A Review. Symmetry. 2021, Vol. 13, No. 8, p. 1362.
[11] Emmerson, J.O., Shateri, A., Xie, J., Use of magnetic fluids in process system for pipe isolations“. Heliyon, 2024, Vol. 10, p. 35221.
[12] Steinmeier, R., Becker, F., Günther, L., Lysenko, V., Minchenya, V., Zeidis, I., Zimmermann, K. „Approaches to the Application of Magnetic Fluids in Electromechanical Drive Systems.“, 59th Ilmenau Scientific Colloquium, TU Ilmenau, 2017. Dostupné online:< https://www.researchgate.net/publication/321362252 >.
[13] Ferrofluid Sealing Technology. Dostupné online:< https://tymagnets.com/ferrofluid-sealing-technology/>.
[14] Mayer, D., „Future of Electrotechnics: Ferrofluids“. Advances in Electrical and Electronic Engineering. 2008, Vol. 7, No. 1,2, pp. 9-14.

Návrh simulátora jednoduchých logických obvodov

[1]Leo BUČEK, [2]Branislav MADOŠ

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

[1]leo.bucek@student.tuke.sk, [2]branislav.mados@tuke.sk

***Abstract* – Tento článok prezentuje návrh a implementáciu webovej aplikácie určenej na simuláciu kombinačných logických obvodov. Aplikácia umožňuje intuitívne vytváranie obvodov, ich úpravu a vizuálnu reprezentáciu simulácie realizácie logických operácií v reálnom čase. Návrh používateľského rozhrania je v súlade s požiadavkou na jeho jednoduchosť, s využitím funkcionalít ako ťahaj a pusti a notifikačného systému. Logické obvody sú tvorené pomocou prepojení medzi vstupmi a výstupmi jednotlivých komponentov. Aplikácia poskytuje farebnú vizualizáciu signálov pre lepšie pochopenie funkčnosti obvodu. Systém podporuje export obvodu do formátu JSON pre zdieľanie alebo ukladanie obvodu a export do VHDL kódu, ktorý umožňuje ďalšiu prácu s kódom v iných programoch.**

***Keywords* – digitálna logika, logické obvody, VHDL, vizuálna simulácia, webová aplikácia**

I. Úvod

Logické obvody tvoria základ moderných elektronických zariadení, od jednoduchých logických systémov až po komplexné počítačové systémy. Pochopenie fungovania logických obvodov je kľúčové nie len pre študentov informatiky a elektrotechniky, ale aj pre ich rast na úroveň odborníkov v oblasti návrhu hardvéru. Článok sa zaoberá kombinačnými logickými obvodmi, kde výstup obvodu závisí výlučne od aktuálneho vstupu obvodu a nie je ovplyvnený históriou týchto vstupov [1]. Kombinačné logické obvody možno formalizovať rôznymi spôsobmi, medzi ktoré patrí algebraický zápis, pravdivostná tabuľka, zhustená pravdivostná tabulka, množinový zápis alebo Karnaughova mapa [2].

Existuje viacero spôsobov minimalizácie Boolovských funkcií. Delia sa na systematické a heuristické. Medzi systematické minimalizácie patria Harvardské minimalizažné diagramy [3], Marquand-Veitchove diagramy [4] alebo Karnaughové mapy [5], grafické nástroje určené na systematickú minimalizáciu [6] a metóda Quine-McCluskey [7] [8]. Medzi heuristické minimalizácie patrí logický minimalizátor Espresso [9], Espresso-MV [10], Espresso-Exact a Espresso Signature [11], BOOlean Minimizer (BOOM) [12], FC-Min [13] a taktiež BOOM-II [14].

Tento článok prezentuje, v rámci tejto práce navrhnutý, webový simulátor kombinačných logických obvodov, ktorý umožňuje používateľom vytvárať, testovať a simulovať logické obvody v reálnom čase, v prostredí webového prehliadača, bez potreby inštalácie špecifického softvéru. Využitie webových technológií zabezpečuje dostupnosť nástroja na rôznych platformách a umožňuje jednoduché zdieľanie vytvorených obvodov. Hlavným cieľom práce bolo vytvoriť jednoduchý nástroj na tvorbu a simuláciu kombinačných obvodov s intuitívnym používateľským rozhraním, ktoré zjednodušuje prácu s logickými obvodmi a poskytuje vizuálnu reprezentáciu ich fungovania. Dôraz bol kladený na simuláciu v reálnom čase, kde je možné sledovať reakcie obvodu na zmeny vstupných signálov. Systém podporuje logické členy pre návrh logických obvodov ako AND, OR, NOT, NAND, NOR, XOR alebo XNOR.

II. Simulátor logických obvodov

Táto sekcia článku sa zaoberá popisom architektúry, používateľského rozhrania, dostupných komponentov logického systému a ich prepojení, spolu s popisom simulačného algoritmu.

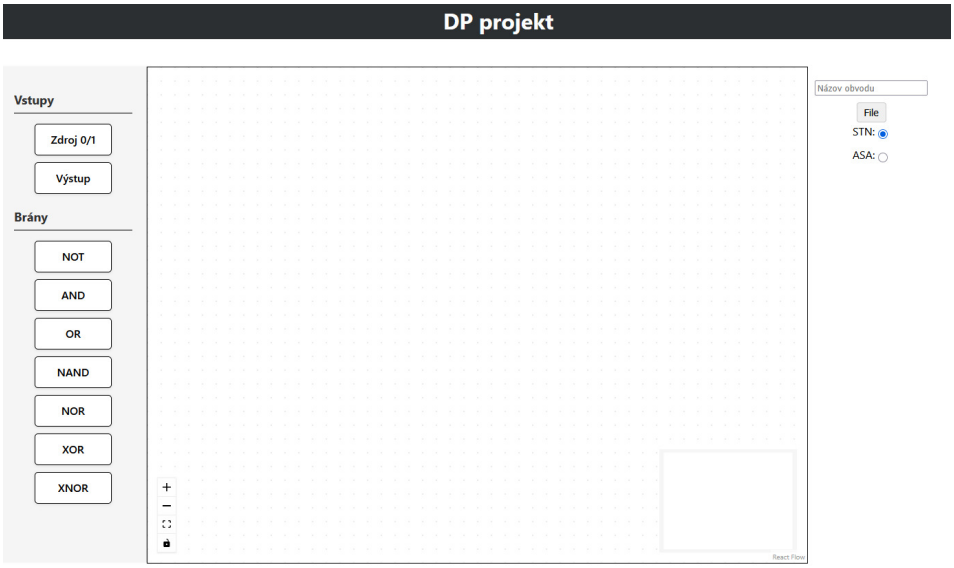
A. Architektúra systému

Aplikácia je implementovaná ako single-page application (SPA) pomocou rámca React. Kľúčové knižnice použité v implementácii sú ReactFlow, React-DnD a štandardné webové technológie ako HTML, CSS a JavaScript. Architektúra bola navrhnutá vo vrstvách:

- Prezentačná vrstva - zodpovedná za zobrazenie používateľského rozhrania
- Aplikačná vrstva - obsahuje logiku aplikácie a simulačný algoritmus
- Dátová vrstva - zodpovedná za správu dát a serializáciu/deserializáciu obvodov

B. Používateľské rozhranie

Používateľské rozhranie bolo navrhnuté s dôrazom na jednoduchosť a intuitívnosť. Hlavné komponenty zahŕňajú toolbox (paleta komponentov), canvas (pracovná plocha), notifikačný systém a pravé menu s možnosťami exportu a importu.



Obr. 1: Hlavné okno navrhnutej aplikácie.

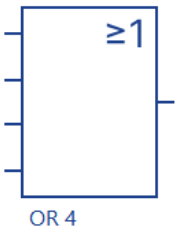
Na Obr. 1 je možné vidieť hlavné okno aplikácie, ktoré obsahuje už niektoré vyššie uvedené komponenty, ako toolbox na ľavej strane obrazovky, odkiaľ sa vyberajú logické členy a pomocou funkcionality ťahaj a pusti sa umiestňujú na canvas. Canvas predstavuje strednú časť okna, kam sa umiestňujú logické členy pri zostavovaní logických obvodov. Na pravej strane okna je umiestnené pravé menu, slúžiace pre prepnutie normy schématických značiek, pomenovanie a export obvodu.

C. Logické komponenty

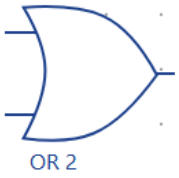
Logické komponenty tvoria základné stavebné prvky pre simuláciu. Každý typ logického člena je implementovaný ako samostatný React komponent so svojou vlastnou vizuálnou reprezentáciou (Obr. 2) a funkcionalitou. Implementácia zahŕňa spomenuté logické hradlá, zdrojový signál pre definovanie počiatočného signálu a výstupný prvok slúžiaci na zobrazenie výstupnej hodnoty logického obvodu.

Každé hradlo implementuje vstupné body na svojej ľavej strane a výstupný bod na pravej strane pomocou ReactFlow. Počet vstupných bodov je dynamicky generovaný na základe špecifikovaného počtu vstupov pre daný člen. Systém podporuje 2 až 8 vstupové hradlá s výnimkou hradla NOT, ktoré má vždy len jeden vstupný bod. Vstupné body sú rovnomerne rozmiestnené po hrane hradla pomocou výpočtu. Každý vstupný bod je limitovaný iba na jeden vstup. Výstupný bod je na pravej strane hradla v strede a slúži ako výstupný terminál pre výslednú hodnotu. Počet výstupov z výstupného bodu nie je limitovaný.

Obr. 3a a 3b zobrazujú vstupné body pomocou elíps čiernej farby. Sú umiestnené na vstupy logických hradiel a v programe sú pre používateľa neviditeľné, kým neumiestni kurzor myši na konkrétny vstupný bod.

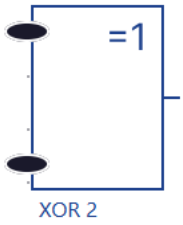


(a) Štvorvstupové OR podľa normy STN.

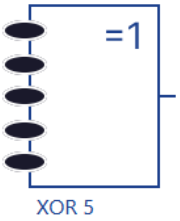


(b) Dvojvstupové OR podľa normy ASA.

Obr. 2: Vizualizácie symbolov hradiel podľa noriem STN a ASA.

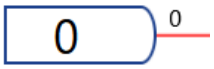


(a) Dvojvstupové XOR s viditeľnými vstupnými bodmi.



(b) Päťvstupové XOR s viditeľnými vstupnými bodmi.

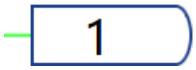
Obr. 3: Vizualizácie symbolov hradiel s viditeľnými vstupnými bodmi.



Obr. 4: Zdroj signálu s hodnotou 0.

Zdroj signálu reprezentuje vstup do logického obvodu, resp. do logických členov, a umožňuje definovať počiatočné vstupné hodnoty signálov. Kľúčovou funkcionalitou v rámci simulácie správania obvodu je prepínanie hodnoty signálnu pomocou kliknutia myši na jeho zdroj (Obr. 4).

Výstupný prvok predstavuje koncový bod obvodu a slúži na zobrazenie výstupnej hodnoty logického obvodu. Na výstupný bod sa môže napojiť iba jedna signálova cesta. Na rozdiel od zdroja signálu, výstupný prvok neposkytuje interaktívnu funkcionalitu pre zmenu stavu, keďže jeho hodnota je výlučne daná hodnotou pripojeného signálu. Jeho úlohou je zobrazenie hodnoty signálu v čitateľnej forme pre používateľa (Obr. 5).



Obr. 5: Výstup signálu s hodnotou 1.

D. Prepojenie medzi komponentmi

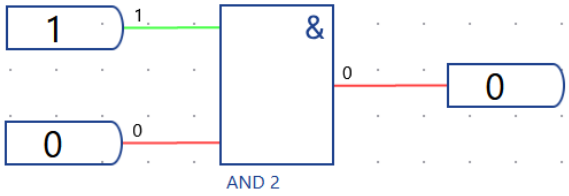
Prepojenie medzi komponentmi je realizované pomocou signálových ciest. Dôležitým aspektom je vizualizácia hodnoty signálu pomocou farieb.

- Zelená farba indikuje logickú hodnotu 1
- Červená farba indikuje logickú hodnotu 0
- Šedá farba indikuje nedefinovaný stav

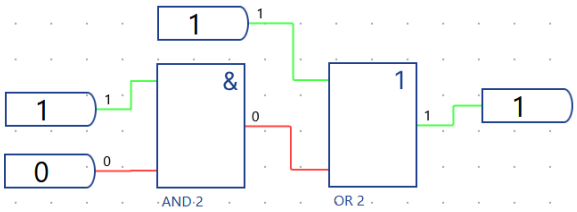
Okrem farebnej identifikácie je zobrazená aj hodnota signálu ako textový štítok pri výstupe logického člena alebo pri výstupe signálu z jeho zdroja, nad signálovou čiarou. Prepojenie medzi komponentmi vzniká natiahnutím čiary pomocou myši od výstupného bodu jedného komponentnu až po vstupný bod druhého komponentu.

Obr. 6 zobrazuje jednoduchý obvod so signálovými cestami, so správnou farbou podľa hodnoty, ktorú signálová cesta prenáša a taktiež je možné vidieť hodnotu signálu na začiatku signálovej cesty. Pri manipulácii s polohou logického komponentu v rámci canvas, ak má komponent vytvorené prepojenia, tieto sú dynamicky aktualizované, takže sa spoj vždy prispôsobí novej pozícii a zachová správne napojenie.

Pokiaľ pripojenie nie je priame, signálová cesta nájde svoj stred a v danom bode sa zlomí a vytvára tak lomené čiary (Obr. 7). Pri vytváraní spojení sú robené validácie ako prevencia samo-pripojení, kde sa kontroluje, či výstup nie je pripojený na svoj vlastný vstup. Pri vytváraní



Obr. 6: Jednoduchý obvod ukazujúci prepojenia komponentov obvodu.

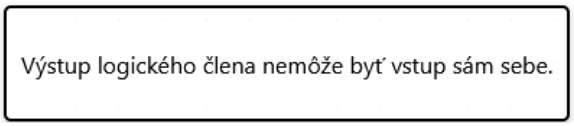


Obr. 7: Zobrazenie lomených čiar v logickom obvode.

nového pripojenia sa overuje limitácia počtu pripojení, či vstupný bod už má nejaké pripojenie alebo nemá.

Pokiaľ má, pripojenie sa nevytvorí. Veľmi dôležitou validáciou je aj algoritmus pre detekciu slučiek. Pri každej snahe o vytvorenie nového prepojenia sa spustí algoritmus na detekciu slučiek. Algoritmus deteguje potenciálny cyklus ešte pred vytvorením prepojenia a pokiaľ algoritmus daný cyklus deteguje, prepojenie nevytvorí. Algoritmus detekcie cyklov funguje na princípe prehľadávania grafu do hĺbky (Depth First Search, DFS).

Tieto validácie pri detekcii chyby poskytnú používateľovi spätnú vazbu pomocou notifikačného systému. Notifikačný systém zobrazí používateľovi notifikáciu v strede používateľského rozhrania ako reakciu na nejakú udalosť. Napríklad pri validácii obvodu alebo pri exporte respektíve importe obvodu. Každá notifikácia je reprezentovaná vizualne s odlišným textom správy podľa danej udalosti, na ktorú reaguje, takže poskytuje používateľovi relevantné informácie (Obr. 8).



Obr. 8: Príklad notifikácie.

E. Simulačný algoritmus

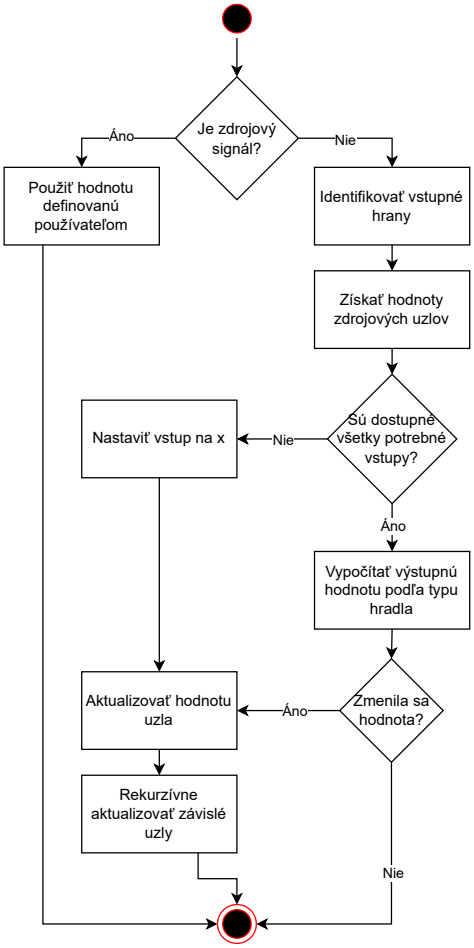
Jadrom aplikácie je výpočet propagácie signálov v obvode. Realizuje výpočet hodnôt signálov v obvode na základe aktuálnych vstupov a logických operácií implementovaných logických členov. Základ je implementovať logické operácie pre jednotlivé logické hradlá.

- **AND** (logický súčin) – výstup je 1, ak sú všetky vstupy 1, inak 0.
- **OR** (logický súčet) – výstup je 1, ak aspoň jeden vstup je 1, inak 0.
- **NOT** (logická negácia) – výstup je inverzná hodnota vstupu.
- **NAND** (negovaný AND) – výstup je 0, ak sú všetky vstupy 1, inak 1.
- **NOR** (negovaný OR) – výstup je 0, ak aspoň jeden vstup je 1, inak 1.
- **XOR** (exkluzívny OR) – výstup je 1, ak je na vstupe iba jedna 1, inak 0.
- **XNOR** (negovaný XOR) – výstup je 0, ak je na vstupe iba jedna 1, inak 1.

Propagácia signálov v obvode je hlavnou úlohou simulačného algoritmu, ktorá zabezpečuje, že zmeny hodnôt na zdrojoch obvodu sa správne šíria cez napojené logické brány. Algoritmus funguje na princípe, že pre každý logický člen sa v obvode identifikujú jeho vstupné signálové cesty a z nich sa získajú hodnoty signálov. Pomocou týchto hodnôt a typu hradla sa vypočíta nová hodnota pomocou funkcie pre výpočet logických operácií. Ak sa hodnota hradla zmenila, aktualizuje sa stav jeho výstupného signálu. Proces sa rekurzívne opakuje pre všetky ostatné logické členy, závislé na tomto hradle. Pre zdroje sa výpočet nevykonáva, keďže ich hodnota je definovaná používateľom.

Tento prístup zabezpečuje, že hodnoty sa propagujú správne cez celý obvod, so zachovanou závislosťou medzi komponentmi. Rekurzívny algoritmus umožňuje efektívne spracovanie aj komplexných obvodov s viacerými logickými hradlami. Implementácia spravuje aj nedefinované stavy.

Ak nie sú k dispozícii všetky požadované vstupy pre daný uzol, výsledná hodnota je nastavená na ×, čo predstavuje nedefinovaný stav. Tento prístup zabezpečuje korektné správanie simulácie aj v prípade nekompletných obvodov a poskytuje používateľovi jasnú vizuálnu indikáciu týchto nedefinovaných stavov.



Obr. 9: Algoritmus pre propagáciu hodnôt signálov.

Po výpočte hodnôt je nevyhnutné aktualizovať vizuálnu reprezentáciu hodnôt v používateľskom rozhraní. Algoritmus (Obr. 9) iteruje, cez všetky signálove cesty v obvode, identifikuje svoj zdrojový logický komponent a jeho aktuálnu hodnotu. Na základe hodnoty signálu sa nastaví hodnota pre príslušnú farbu signálovej cesty a aktualizuje sa štýl signálovej cesty podľa nastavenej hodnoty.

F. Export a Import

Funkcionalita **exportu** obvodov umožňuje používateľom uložiť vytvorený obvod do súboru, ktorý môže byť následne uchovaný, zdieľaný alebo použitý pre offline zálohovanie práce. Proces zahŕňa získanie aktuálneho stavu uzlov a hrán, serializáciu obvodu, konvertovanie serializovaného objektu do formátu JSON. Ďalej sa vytvorí Blob objekt obsahujúci JSON dáta, URL pre Blob, realizuje sa vytvorenie a simulácia kliku na neviditeľný element, čím sa iniciuje stiahnutie súboru vo formáte JSON a následne sa realizuje vymazanie vytvorených referencií na URL. Proces končí poskytnutím notifikácie používateľovi o výsledku operácie.

Import obvodu umožňuje používateľom načítať predtým exportovaný obvod a pokračovať v práci s ním. Proces importu je iniciovaný používateľom, keď pomocou input elementu vyberie súbor. Obsah súboru sa prečíta a parsuje z formatu JSON. Následne prebieha deserializácia a aktualizuje sa stav aplikácie s importovanými logickými členmi. Proces končí poskytnutím notifikácie používateľovi o výsledku operácie.

Export do VHDL Pokročilou funkcionalitou implementovanou v systéme je export logických obvodov do formátu VHDL (VHSIC Hardware Description Language), ktorý predstavuje štandardný jazyk pre popis a modelovanie digitálnych systémov. Táto funkcionalita ma niekoľko krokov, medzi ktoré patrí analýza štruktúry obvodu, generovanie entity, generovanie architektúry, identifikácia logických operácií, spracovanie výstupov a export do súboru. Analýza štruktúry zahŕňa

identifikáciu zdrojov, výstupov a logických hradiel. Generovanie entity spočíva vo vygenerovaní VHDL entity kde sa definuje port a deklarujú sa všetky vstupy a výstupy. Pri generovaní architektúry prebieha deklarácia signálov, ktoré reprezentujú výstupy z logických hradiel. Identifikujú sa prepojenia medzi komponentmi na priradenie signálov vo VHDL. Spracovanie výstupov je generovanie priradenia výstupného portu na správny logický komponent na základe prepojení v obvode. Posledný krok je export do súboru. Konvertuje sa vygenerovaný VHDL kód do formátu textového súboru a implementuje sa rovnaký postup pre stiahnutie tohto súboru používateľom ako pri exportu do súboru typu JSON. Výsledný VHDL kód je štruktúrovaný v súlade so štandardnými konvenciami VHDL, čo zabezpečuje jeho kompatibilitu s bežnými nástrojmi pre simuláciu hardvéru. Obvod je reprezentovaný ako entita s definovaným portom a jeho funkcionalita je implementovaná v architektúre prostredníctvom priradení signálov a logických operácií.

III. Testovanie a validácia

Aplikácia bola podrobená rôznym typom testovania pre overenie jej funkčnosti a použiteľnosti.

A. Jednotkové a integračné testy

Boli implementované testy pre kľúčové funkcie aplikácie:

- Testovanie logických operácií
- Testovanie detekcie cyklov
- Testovanie propagácie signálov
- Testovanie zmeny farby signálových ciest
- Testovanie serializácie a deserializácie
- Testovanie exportu a importu
- Testovanie exportu do VHDL kódu

B. Výkonnostné testovanie

Testovanie výkonu potvrdilo efektívnosť simulačného algoritmu (Tab. 1):

Počet hradiel	Počet komponentov	Počet testovaní	Priemerný čas
10	22	100	0,06 ms
20	34	100	0,14 ms
50	71	100	0,21 ms

Tab. 1: Výsledky testovania logických obvodov.

C. Používateľské testovanie

Používateľské testovanie s 10 účastníkmi potvrdilo intuitívnosť a použiteľnosť aplikácie. Priemerné SUS skóre dosiahlo hodnotu 79,5, čo na SUS škále značí, že používateľské rozhranie je dobré.

IV. Záver

Implementovaný webový simulátor logických obvodov poskytuje nástroj pre vytváranie, testovanie a simuláciu logických obvodov v prostredí webového prehliadača. Aplikácia umožňuje intuitívne vytváranie obvodov pomocou techniky ťahaj a pusti, poskytuje jasnú vizualizáciu signálov a podporuje export a import obvodov. Medzi možné vylepšenia do budúcnosti patrí vylepšená implementácia signálových ciest, možnosť negovať konkrétny vstup priamo na logickom hradle, implementácia zložených komponentov alebo optimalizácia pre mobilné zariadenia.

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] B. Madoš, *Princípy počitáčového inžinierstva*. KPI FEI TUKE, 2025.
[2] ——, "Booleovské funkcie," navštívené 25.4.2024. [Online]. Available: https://ppi.madosonline.sk/index.php?pid=51
[3] H. H. Aitken, *Synthesis of electronic computing and control circuits*. Cambridge, Massachusetts: Harvard University Press, 1951.
[4] E. W. Veitch, "A chart method for simplifying truth functions," in *Proceedings of the 1952 ACM Annual Meeting*. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery (ACM), 1952, pp. 127–133.
[5] M. Karnaugh, "The map method for synthesis of combinational logic circuits," *Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics*, vol. 72, no. 5, pp. 593–599, 1953.
[6] A. Svoboda, "Graficko-mechanické pomůcky užívané při analyse a synthese kontaktových obvodů," in *Stroje na zpracování informací, Symphosium IV on information processing machines*. Prague: Czechoslovak Academy of Sciences, Research Institute of Mathematical Machines, 1959, pp. 9–21, in Czech.
[7] M. V. Quine, "The problem of simplifying truth functions," *American Mathematical Monthly*, vol. 59, no. 8, pp. 521–531, 1952.
[8] E. J. McCluskey, "Minimization of boolean functions," *The Bell System Technical Journal*, vol. 35, no. 5, pp. 1417–1444, 1956.
[9] R. K. Brayton, G. D. Hachtel, C. T. McMullen, and A. L. Sangiovanni-Vincentelli, *Logic Minimization Algorithms for VLSI Synthesis*, 1st ed. Kluwer Academic Publishers, 2000, 9th printing.
[10] R. L. Rudell, "Multiple-valued logic minimization for pla synthesis," University of California, Berkeley, Memorandum UCB/ERL M86-65, Jun. 1986.
[11] P. McGeer, J. V. Sanghavi, R. K. Brayton, and A. L. Sangiovanni-Vincentelli, "Espresso-signature: A new exact minimizer for logic functions," in *Proceedings of the DAC'93*, 1996, pp. 432–440.
[12] P. Fišer and J. Hlavička, "Boom, a heuristic boolean minimizer," *Computing and Informatics*, vol. 22, pp. 19–51, 2003.
[13] P. Fišer and H. Kubátová, "Two-level boolean minimizer boom-ii," in *Proc. 6th International Workshop on Boolean Problems (IWSBP'04)*, Freiberg, Germany, 2004, pp. 221–228.
[14] ——, "Flexible two-level boolean minimizer boom ii and its applications," in *Proc. 9th Euromicro Conference on Digital Systems Design (DSD'06)*, Cavtat, Croatia, 2006, pp. 369–376.

Efektívne fotovoltické aplikácie a úloha zvyšovacieho meniča

[1]Daniel GORDAN, [2]Marek PÁSTOR, [3]Tadeáš KMECIK, [4]Tomáš BASARIK

[1, 2, 3, 4]Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

[1]daniel.gordan@tuke.sk, [2]marek.pastor@tuke.sk, [3]tadeas.kmecik@tuke.sk, [4]tomas.basarik@tuke.sk

***Abstrakt* — Článok opisuje topológiu prekladaného zvyšovacieho meniča (IBC), ktorý je možné použiť vo fotovoltických meničoch pre širokú oblasť aplikácií. Simuláciou v programe PLECS sú porovnané straty IBC s jednoduchým zvyšovacím meničom (boost).**

***Kľúčové slová* — fotovoltika, SiC, PLECS, IBC, boost, zelená energia**

I. Úvod

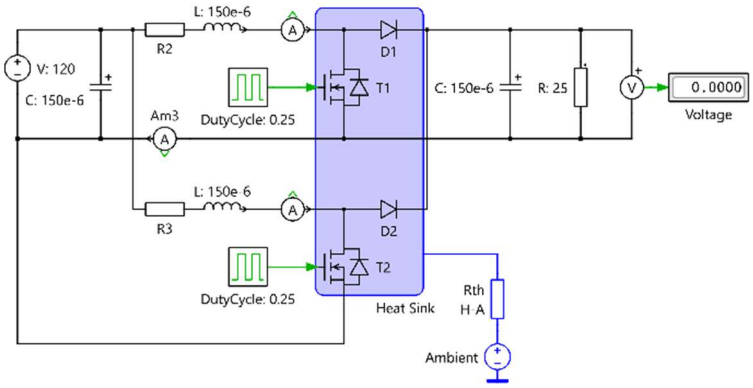
Takmer viac ako desaťročie sme svedkami masívnej implementácie obnoviteľných zdrojov elektrickej energie, ktoré pomáhajú pri snahe o dosiahnutie uhlíkovej neutrality, teda postupnej transformácií od fosílnych palív k zelenej energii. Implementácia obnoviteľných zdrojov vyžadovala a stále vyžaduje výskum v oblasti výkonovej elektroniky, hľadanie nových topológií a samozrejme snahu o dosiahnutie vyššej účinnosti pri už známych topológiách, pretože bez výkonovej elektroniky by neexistovala zelená transformácia [1,2].

Veľkou výhodou fotovoltickej energie je, že je rozšírená, cenovo dostupná a legislatívne nenáročná aj pre bežných užívateľov. V súčasnosti existujú viaceré uplatnenia fotovoltických meničov, najčastejšie sa používa koncepcia, kedy je menič pripojený k sieti, prípadne naviac nabíja záložnú batériu. Okrem toho existujú aj meniče pre ohrev teplej úžitkovej vody (TUV), či na priamy pohon klimatizácie, čím odpadá nutnosť pripojenia meniča do siete. Hlavnou podmienkou pre dosiahnutie vysokej účinnosti celého fotovoltického systému je nutnosť dosiahnutia maximálneho bodu výkonu pri daných svetelných podmienkach (MPP) [3]. MPP nie je možné dosiahnuť bez súčinnosti hardvéru a softvéru. Hardvér, teda menič, musí byť svojou topológiou schopný umožniť riadiacemu mikrokontroléru dostatočnú regulačnú voľnosť, aby MPPT algoritmus naozaj dosiahol maximálny bod výkonu pri daných svetelných podmienkach. Okrem toho musí menič pracovať v širokom rozsahu vstupu a výstupu, aby bolo možné pripojiť rozdielne množstvo panelov. Z tohto dôvodu je v meničoch na vstupe často implementovaný zvyšovací člen [4], ktorý zvýši napätie, pretože bez tohto zvýšenia napätia, by mikroprocesor v regulačnej slučke nedokázal zaťažiť panel dostatočne, teda by nedokázal dosiahnuť maximálny bod výkonu pri danej záťaži, ktorá sa môže meniť [5,6].

V tomto článku je uvedená topológia IBC [7-9], ktorá dokáže zvýšiť napätie a je porovnaná s klasickým zvyšovacím meničom.

II. Zvyšovací stupeň

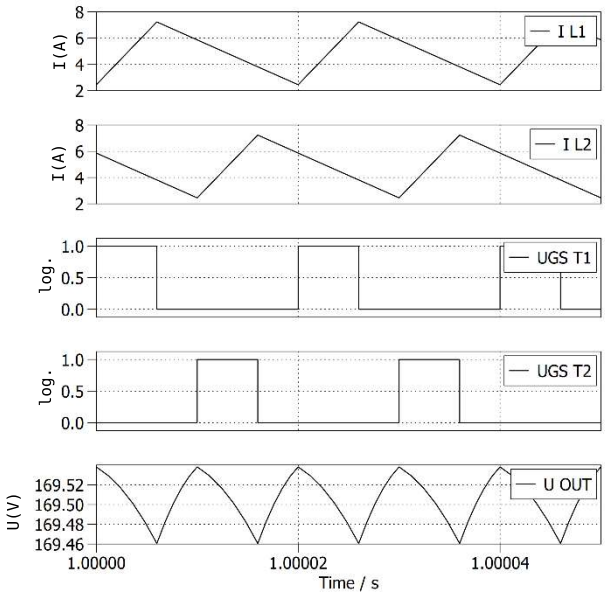
IBC sa skladá z minimálne dvoch klasických zvyšovacích meničov, schéma meniča v programe PLECS je na Obr. 1. Zdroj vstupného napätia 120 V reprezentuje trojicu sériovo zapojených panelov, nasledujú dva totožné stupne zvyšovacieho meniča, jeden je zložený z tranzistora T_1 a diódy D_1 a druhý z T_2 a D_2. Rezistory R_2 a R_3 predstavujú odpor cievky, ktorá je totožná v oboch stupňoch. Výstup oboch stupňov je zapojený paralelne do výstupného kondenzátora. Záťaž má odpor 25 Ω, do ktorej pri výstupnom napätí meniča tečie maximálny prúd panelov. Parametre meniča sú detailne uvedené v Tab. 1. Výhodou tohto riešenia oproti klasickému je, že efektívna hodnota vstupného prúdu sa rozdelí na dva stupne, takže súčiastky sú menej namáhané, na diódach je menší úbytok a straty RI^2 klesnú. Spínací algoritmus tranzistorov je uvedený na Obr. 2.



Obr. 1 Schéma IBC meniča v programe PLECS

Tabuľka 1
Parametre meniča

Parameter	Hodnota
Napájacie napätie	120 V
Výstupné napätie	170 V
Výstupný výkon	1150 W
Spínacia frekvencia	50 kHz
Indukčnosť L	150 μH
Odpor indukčnosti R_2, R_3	50 mΩ
Odpor záťaže R	25 Ω
SiC MOSFET T_1, T_2	IMZA65R027M1H
SiC dióda D_1, D_2	C3D16065D1



Obr. 2 Spínací algoritmus a prúd v indukčnostiach zvyšovacieho stupňa

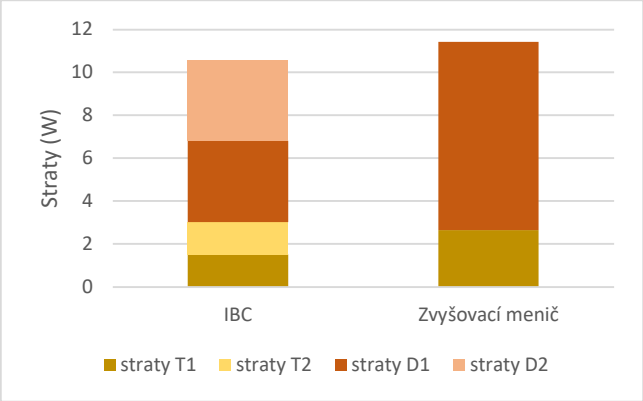
Na Obr. 2 sú uvedené základné priebehy potrebné pre vysvetlenie činnosti IBC meniča. Napätie U_{OUT} a prúdy I_{L1} a I_{L2} sú uvedené v základných jednotkách, napätia $U_{GS\ T1}$ a $U_{GS\ T2}$ predstavujú logické signály, log. 1 pre zapnutý tranzistor a log. 0 pre vypnutý tranzistor.

Po zapnutí tranzistora, Obr. 2, začne prúd v indukčnosti narastať, až do doby, keď tranzistor vypneme. Následne sa energia indukčnosti odovzdá cez diódu do výstupného kondenzátora, čo sa prejaví nárastom napätia U_{OUT}. Funkčnosť meniča je teda analogická zo zvyšovacím meničom, rozdiel spočíva v druhej vetve, kde sú impulzy tranzistora T_2 posunuté o pol periódu. Výsledkom je zmenšenie efektívnej hodnoty prúdu v celej vetve na polovicu a zväčšenie frekvencie zvlnenia vo výstupnom kondenzátore na dvojnásobok. V nasledujúcej podkapitole sú uvedené výhody IBC meniča oproti klasickému zvyšovaciemu stupňu na základe analýzy strát a efektívnych hodnôt prúdov.

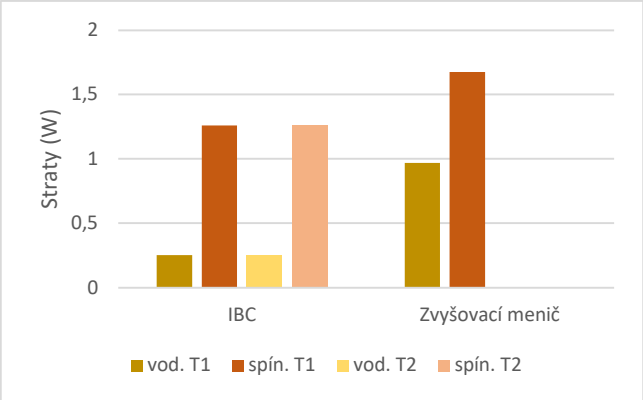
III. Analýza strát meniča

A. Straty na polovodičoch

Straty na polovodičoch simulované v programe PLECS sú uvedené na Obr. 3, pri maximálnom výkone meniča. Modely súčiastok sú prevzaté z oficiálnych stránok výrobcov. V grafe sú uvedené celkové straty polovodičov IBC meniča v porovnaní so zvyšovacím meničom, pričom pri simulácií zvyšovacieho meniča bola odpojená druhá vetva IBC meniča, teda T_2 a D_2. Z grafu vyplýva, že IBC menič dosiahol nižšie celkové straty. Na Obr. 4 sú znázornené spínacie a vodivostné straty. Vodivostné straty, ktoré sú dané od RI^2, klesli v IBC meniči výrazne, avšak na celkových stratách, Obr. 3, sa to až tak výrazne neprejavilo kvôli pomerne vysokým spínacím stratám, ktorým nepomohlo pripojenie ďalšieho tranzistora do obvodu. Výsledkom toho je, že pre optimalizáciu meniča by mohlo byť vhodné zmeniť SiC tranzistor za iný, ktorý by síce mal vyššie R_{DSon}, ale menšiu výstupnú kapacitu, ktorá by mala pozitívny vplyv na redukciu spínacích strát. Po takejto optimalizácií by bolo možné vidieť väčší rozdiel medzi celkovými stratami IBC a zvyšovacieho meniča.



Obr. 3 Celkové straty na polovodičov v IBC a zvyšovacom meniči



Obr. 4 Rozloženie strát na polovodičoch v IBC a zvyšovacom meniči

B. Pasívne komponenty

Z pasívnych komponentov zvyšovacieho meniča je potrebné klásť dôraz najmä indukčnosť, ktorá je kľúčovým komponentom. Je dôležite, aby si pri maximálnom prúde zachovala minimálnu požadovanú indukčnosť a zvládla efektívnu hodnotu prúdu, ktorá spôsobuje oteplenie vinutia a magnetického obvodu.

Druhým veľmi dôležitým prvkom je výstupný kondenzátor, ktorého hodnota definuje výstupné zvlnenie, avšak väčším problémom je, že pracuje so skokmi prúdu, ktoré sa prejavia vysokou efektívnou hodnotou. Táto efektívna hodnota prúdu spôsobuje straty na ekvivalentnom sériovom odpore a spôsobuje oteplenie kondenzátora, ktoré má vplyv na celkovú životnosť. Z tohto dôvodu má každý elektrolytický kondenzátor danú hodnotu uvedenú v katalógovom liste, platí, že čím je menšia efektívna hodnota prúdu kondenzátora, tým je aj výsledný kondenzátor menší, lacnejší.

V Tab. 2 sú uvedené efektívne a maximálne hodnoty prúdov v indukčnostiach, straty na indukčnosti a efektívna hodnota prúdu vo výstupnom kondenzátore. Z údajov vyplýva, že pomocou IBC topológie je možné dosiahnuť výrazne nižšie straty v magnetických obvodoch a zmenšiť nároky na počet, prípadne veľkosť výstupných kondenzátorov.

Tabuľka 2
Parametre veličín pasívnych komponentov

Parameter	Boost	IBC
Efektívna hodnota prúdu v indukčnosti $I_{L(ef)}$	9,78 A	5,03 A
Maximálna hodnota prúdu v indukčnosti $I_{L(max)}$	12,07 A	7,25 A
Celkové straty v indukčnosti	4,76 W	2,53 W
Efektívna hodnota prúdu výstupného kondenzátora I_{C_OUT}	4,39 A	2,49 A

IV. ZHODNOTENIE

V článku je uvedené porovnanie klasického zvyšovacieho meniča a IBC, ktorého výhodou je zníženie celkových strát polovodičov, zníženie strát v indukčnosti spolu s maximálnym prúdom a v neposlednom rade zníženie efektívnej hodnoty prúdu vo výstupnom kondenzátore, čím klesnú rozmery a cena výstupného filtra. Na druhej strane nevýhodou IBC je väčšie množstvo polovodičových komponentov, čo sa z hľadiska ceny môže prejaviť negatívne.

Dôležitý je aj návrh celého riešenia, pri ktorom je potrebné splniť nasledovné podmienky:

1. Aplikovať IBC v aplikácií, ktorá pracuje najmä s väčším prúdom, pretože práve tam sa prejaví výhoda redukcie strát RI^2.
2. Dôsledne vybrať spínacie tranzistory, pretože snaha o výber tranzistora s nízkym odporom v zopnutom stave, môže koniec koncov podstatne zvýšiť spínacie straty, čo pri sa pri IBC s viacerými stupňami prejaví rapídnym zvýšením celkových strát, ktoré budú násobne väčšie ako vodivostné straty.

Výsledná cena výkonovej časti nemusí byť väčšia z nasledujúcich dôvodov:

1. Aj keď je potrebné použiť dve indukčnosti, tak pracujú s menšou efektívnou a maximálnou hodnotou prúdu, čo môže znamenať podobnú cenu v porovnaní s jednou väčšou indukčnosťou.
2. Síce potrebujeme dva tranzistory, avšak môžeme vybrať tranzistor s väčším R_{DSon}, takže cena môže byť podobná
3. Tým, že je frekvencia zvlnenia vo výstupnom kondenzátora dvojnásobná, kondenzátor bude menší a lacnejší
4. Menšie straty sa prejavia zníženými nárokmi na chladenie.

V ďalšom pokračovaní je potrebné nájsť vhodnú metodiku pre výber ideálnych tranzistorov a diód pre IBC topológiu s cieľom minimalizovať celkové straty. Situáciu komplikuje fakt, že zvyšovací člen pre fotovoltiku by mal vedieť pracovať vo viacerých rozsahoch vstupu a výstupu, ktoré sa líšia od danej aplikácie. Z tohto dôvodu je možné preskúmať aj vplyv zapojenia viacerých zvyšovacích členov, prípadne hľadať ďalšie možnosti ich riadenia.

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] D. Yang, X. Wang and J. Kang, "SWOT Analysis of the Development of Green Energy Industry in China : Taking solar energy industry as an example," 2018 2nd International Conference on Green Energy and Applications (ICGEA), Singapore, 2018, pp. 103-107, doi: 10.1109/ICGEA.2018.8356320.
[2] M. Sithan and L. L. Lai, "Application of green technologies in developing countries — Reduced carbon emission and conservation of energy," 2011 IEEE Power and Energy Society General Meeting, Detroit, MI, USA, 2011, pp. 1-7, doi: 10.1109/PES.2011.6039665.
[3] T. Esram and P. L. Chapman, "Comparison of Photovoltaic Array Maximum Power Point Tracking Techniques," in IEEE Transactions on Energy Conversion, vol. 22, no. 2, pp. 439-449, June 2007, doi: 10.1109/TEC.2006.874230.
[4] C. C. C. dos Santos, C. F. Moraes, J. P. da Costa, C. M. O. Stein, E. G. Carati and R. Cardoso, "Photovoltaic Boost Converter Control Operating in the MPPT and LPPT Modes," 2019 IEEE 15th Brazilian Power Electronics Conference and 5th IEEE Southern Power Electronics Conference (COBEP/SPEC), Santos, Brazil, 2019, pp. 1-6, doi: 10.1109/COBEP/SPEC44138.2019.9065604.
[5] J. -M. Kwon, B. -H. Kwon and K. -H. Nam, "Three-Phase Photovoltaic System With Three-Level Boosting MPPT Control," in IEEE Transactions on Power Electronics, vol. 23, no. 5, pp. 2319-2327, Sept. 2008, doi: 10.1109/TPEL.2008.2001906.
[6] Infineon, "6 kW HERIC reference design user guide," 2025, Reference Design User Guide, Infineon
[7] Ronald Crews, "AN-1820 LM5032 Interleaved Boost Converter," 2013, Application Report, Texas Instruments
[8] Texas Instruments, "1.6-kW, Bidirectional Micro Inverter Based on GaN Reference Design," 2023, Reference Design, Texas Instruments
[9] Texas Instruments, "1.6-kW, Bidirectional Micro Inverter Based on GaN Reference Design," 2024, Design Guide, Texas Instruments

Verifikácia RSA certifikátov v IoT uzle s podporou TLS protokolu

[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 **– Článok sa zaoberá využitím digitálnych podpisov založených na kryptografickom algoritme RSA v rámci protokolu TLS 1.3 pre IoT zariadenia. Sú vysvetlené princípy fungovania algoritmu RSA, vrátane jeho výplňových schém PKCS a RSA-PSS, so zameraním na ich implementáciu v zdrojovo obmedzených vstavaných systémoch. Opisuje sa testovanie overovania digitálnych podpisov RSA pomocou knižníc CycloneCrypto, WolfSSL, MbedTLS a TinyTls na vývojovej doske ESP32-C6-DevKitC-1-N8, ktorá je typickým príkladom IoT uzla. Prezentujú sa experimentálne namerané časy overovania digitálnych podpisov, ktoré poskytujú prehľad o výkonnosti jednotlivých knižníc v prostredí vstavaných systémov, a zdôrazňujú výzvy spojené s efektívnym zabezpečením komunikácie v IoT sieťach s obmedzenými výpočtovými zdrojmi.**

Kľúčové slová **– RSA, TLS 1.3, ESP32-C6, PKCS, PSS, Kryptografické knižnice, Digitálny podpis**

I. Úvod

Pre zabezpečenie sieťovej komunikácie medzi dvoma alebo viacerými stranami je nevyhnutné spoľahlivo overiť identitu zúčastnenej strany. Overenie totožnosti zaručuje, že zúčastnená strana komunikácie je skutočne tou, za ktorú sa vydáva. V prípade, že overenie je neúspešné, bezpečná komunikácia nebude inicializovaná a neprebehne. V súčasnosti je bežná internetová komunikácia realizovaná prostredníctvom spojenia HTTPS (Hypertext Transfer Protocol Secure), ktorého jadro tvorí protokol TLS 1.3 (Transport Layer Security, version 1.3) [1]. Protokol TLS 1.3 predstavuje aktuálne najnovšiu verziu protokolu TLS, ktorá podporuje podstatne menšiu sadu kryptografických algoritmov v porovnaní s jeho predchádzajúcimi verziami. Pri zabezpečenej komunikácii s využitím protokolu TLS 1.3 je možné overiť identitu každej zo zúčastnených strán sieťovej komunikácie. Pri bežných internetových spojeniach však zvyčajne klient overuje iba vzdialený webový server, zatiaľ čo server neoveruje indentitu klienta. Pre preukázanie svojej identity, resp. dôveryhodnosti, je potrebné, aby webový server počas nadviazania komunikácie poskytol digitálny certifikát, ktorý obsahuje verejný kľúč a jedinečný digitálny podpis [2] viazaný na jeho špecifiká, akými sú napr. meno webového servera, krajina, organizácia a podobne. Štandardne je overenie tvorené troma zreťazenými certifikátmi, pričom dva certifikáty (certifikát webového servera a stredný – intermediate certifikát) sú odoslané serverom počas inicializácie komunikácie. Koreňové certifikáty typicky vydávajú svetovo uznávané certifikačné autority, ktoré zabezpečujú nielen bezpečnosť, ale aj širokú kompatibilitu, aby ich digitálne certifikáty podporovalo veľké množstvo zariadení. Práve z tohto dôvodu sa pre koreňové certifikáty najčastejšie využíva digitálny podpis na báze RSA (Rivest–Shamir–Adleman) [3]. Tento algoritmus predstavuje jeden z najstarších, ale stále kryptograficky bezpečných algoritmov. V protokole TLS 1.3 sa algoritmus RSA využíva výhradne na vytváranie a overovanie digitálnych podpisov. Algoritmus RSA je založený na matematických operáciách s veľkými číslami, čo predstavuje významné výpočtové nároky pre IoT (The Internet of Things) zariadenia s obmedzenými zdrojmi. Článok analyzuje využitie algoritmu RSA na overovanie digitálnych podpisov zo strany klienta v rámci protokolu TLS 1.3, so zameraním na jeho použitie vo vstavaných zariadeniach. Analýza sa primárne zameriava na časový výkon kryptografických operácií algoritmu RSA [4]. Článok zároveň poukazuje na rôzne kryptografické knižnice, určené najmä pre vstavané zariadenia, a ich možnosť efektívne využívať overovanie digitálneho podpisu RSA.

II. Základná charakteristika algoritmu RSA

Kryptografický algoritmus RSA [4] je jedným z najstarších algoritmov asymetrickej kryptografie, ktorý umožňuje šifrovanie, výmenu kľúčov a vytváranie digitálnych podpisov. V protokole TLS 1.3 [1] sa používa výlučne na digitálne podpisy. RSA zvyčajne pracuje s kľúčmi veľkosti 2048, 3072 alebo 4096 bitov, pričom certifikačné autority vydávajúce koreňové certifikáty bežne podporujú iba 2048-bitové a 4096-bitové kľúče. Generovanie kľúčov pre RSA algoritmus pri digitálnom podpise využíva prvočísla a modulárnu aritmetiku podľa nasledujúcich krokov:

1) **Výber prvočísel:** Na začiatku sa vyberú dve veľké náhodné prvočísla p a q.
2) **Výpočet modulu:** Následne sa vypočíta ich súčin $n = p \cdot q$, ktorý slúži ako základ pre generovanie kľúčov.
3) **Výpočet Eulerovej funkcie:** Ďalším krokom je výpočet Eulerovej funkcie $\varphi(n) = (p-1) \cdot (q-1)$.
4) **Výber verejného exponentu:** Pre verejný exponent sa vo väčšine prípadov volí hodnota $e = 65537$, čo je piate Fermatovo číslo [5]. Dôležité je, aby platilo, že $\gcd(e, \varphi(n)) = 1$, teda e a $\varphi(n)$ musia byť nesúdeliteľné.
5) **Kontrola deliteľnosti:** Ak najväčší spoločný deliteľ $\gcd(e, \varphi(n)) \neq 1$, je potrebné zvoliť nové prvočísla a algoritmus RSA sa musí opakovať od začiatku.
6) **Výpočet súkromného kľúča:** Na konci sa vypočíta súkromný kľúč d, ktorý je inverzný k verejnému exponentu e modulo $\varphi(n)$, a to podľa vzorca $d \equiv e^{-1} \mod \varphi(n)$.

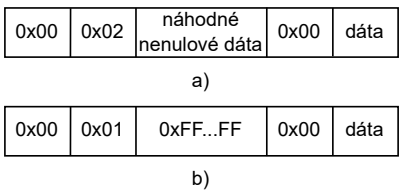
Týmto spôsobom sa generujú kľúče pre RSA algoritmus, kde verejný kľúč je tvorený hodnotami (e, n) a súkromný kľúč je hodnota d.

III. Overenie digitálneho podpisu na báze RSA

V protokole TLS 1.3 je zásadné, aby server a klient boli schopní pomocou algoritmu RSA [4] podpisovať a overovať certifikáty. Článok sa zameriava na proces overovania certifikátov, keďže webové servery zvyčajne nevyžadujú od klientov digitálny certifikát. Výpočet šifrovania a dešifrovania využíva rovnakú štruktúru založenú na modulárnom umocňovaní, no s rôznymi parametrami. Otvorený text sa označuje symbolom M, zašifrovaný text symbolom C. Šifrovanie sa vykoná podľa rovnice $C = M^e \mod n$. Dešifrovanie sa vykoná podľa rovnice $M = C^d \mod n$.

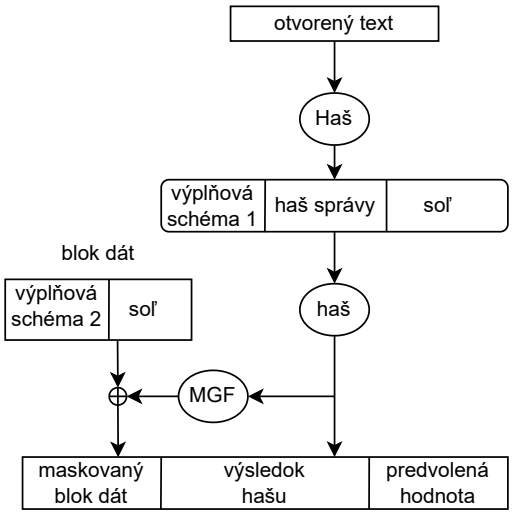
A. Potreba využitia výplňovej schémy

Výplňová schéma OAEP (Optimal Asymmetric Encryption Padding) [6] je dôležitou súčasťou asymetrickej šifry RSA, pretože poskytuje ochranu proti rôznym útokom, ako sú útoky na rozlíšenie šifrovaných dát, útoky s čiastočným odhalením správy a ďalšie. V digitálnych certifikátoch používaných v protokole TLS 1.3 môže algoritmus RSA využiť dve rôzne výplňové schémy. Prvá schéma je PKCS 1.5 (Public Key Cryptography Standards version 1.5) [7], ktorá je jednoduchá na implementáciu, pretože využíva hašovacie funkcie a deterministické dáta. Avšak pri čistom asymetrickom šifrovaní, výplňová schéma PKCS 1.5 používa náhodné dáta, nie deterministické. Schéma PKCS 1.5 pre šifrovanie a digitálny podpis je znázornená na Obr. 1. Druhá schéma, RSA-PSS (RSA – Probabilistic Signature Scheme) [8], je zložitejšia, pretože využíva náhodnú hodnotu (soľ), operáciu MGF (Mask Generation Function)[1] a má komplexnejšiu štruktúru, jej schéma je znázornená na Obr. 2.



Obr. 1: Štruktúra výplňovej schémy PKCS pre RSA: (a) šifrovanie s náhodnými dátami a (b) deterministický digitálny podpis.

[1] https://bitcoinwiki.org/wiki/mask-generation-function



Obr. 2: Štruktúra výplňovej schémy PSS pre RSA šifrovanie a digitálny podpis využívajúca náhodné dáta.

B. Bezpečnosť výplňových schém

Rozdiel medzi týmito dvoma schémami spočíva v tom, že PKCS 1.5 je ľahšie implementovateľná, ale obsahuje známe zraniteľnosti, ktoré môžu predstavovať bezpečnostné riziko. Na druhej strane RSA-PSS poskytuje silnejšiu ochranu, aj keď je jej implementácia náročnejšia. V protokole TLS 1.3 sú používané kryptografické algoritmy na digitálne podpisy serverových certifikátov, ktoré môžu využívať výplňové schémy PKCS 1.5 alebo RSA-PSS. Tieto schémy sú kombinované s hašovacími funkciami, ako sú SHA256 [9] a SHA384 [10]. V kryptografii existujú aj modernejšie a bezpečnejšie verzie výplňovej schémy PKCS[2], ale tie sa v protokole TLS 1.3 nepoužívajú. Kľúče používané v týchto procesoch sú buď 2048-bitové, alebo 4096-bitové. Tieto techniky tvoria základ pre vytváranie a overovanie digitálnych podpisov na báze RSA. Bez ohľadu na použité hašovacie funkcie alebo výplňovú schému, dĺžka kľúča je v praxi buď 2048 bitov, alebo 4096 bitov. Tabuľka 1 uvádza zoznam 19 vybraných webových serverov a ich príslušných zreťazených certifikátov používaných v protokole TLS 1.3, vrátane typov a veľkostí kľúčov pre serverový, stredný (intermediate) a koreňový certifikát. Veľkú popularitu predstavujú certifikáty na báze eliptických kriviek SECP (Standards for Efficient Cryptography Prime) 256R1 a 384R1.

Tabuľka 1: TLS 1.3 Certifikáty serverov.

Server	**Koreňový certifikát**	**Stredný certifikát**	**Certifikát servera**
google.com	SECP384R1	SECP256R1	SECP256R1
facebook.com	RSA 2048 PKCS	RSA 2048 PKCS	SECP256R1
instagram.com	RSA 2048 PKCS	RSA 2048 PKCS	SECP256R1
cnn.com	RSA 2048 PKCS	RSA 2048 PKCS	RSA 2048 PSS
youtube.com	SECP384R1	SECP256R1	SECP256R1
twitter.com	RSA 4096 PKCS	SECP384R1	SECP256R1
netflix.com	RSA 2048 PKCS	SECP256R1	SECP256R1
malwarebytes.com	RSA 4096 PKCS	SECP384R1	SECP256R1
espressif.com	RSA 2048 PKCS	RSA 2048 PKCS	RSA 2048 PSS
hbo.com	RSA 2048 PKCS	RSA 2048 PKCS	RSA 2048 PSS
bbc.com	RSA 2048 PKCS	RSA 2048 PKCS	RSA 2048 PSS
proton.me	RSA 4096 PKCS	RSA 2048 PKCS	RSA 4096 PSS
tuke.sk	RSA 4096 PKCS	RSA 2048 PKCS	RSA 4096 PSS
upjs.sk	RSA 4096 PKCS	RSA 4096 PKCS	RSA 2048 PSS
refresher.sk	SECP384R1	SECP256R1	SECP256R1
csfd.cz	RSA 4096 PKCS	RSA 2048 PKCS	RSA 2048 PSS
alza.sk	SECP384R1	SECP256R1	SECP256R1
eset.sk	RSA 4096 PKCS	SECP384R1	SECP256R1
datart.sk	SECP384R1	SECP384R1	SECP256R1

[2]https://www.techtarget.com/searchsecurity/definition/Public-Key-Cryptography-Standards

IV. Overenie digitálnych podpisov v IoT zariadeniach

IoT predstavuje vstavané zariadenia, ktoré sú často vybavené rôznymi senzormi slúžiacimi na špecifické účely v priemysle alebo domácnostiach [11]. Napríklad ide o zariadenia pre zavlažovacie systémy, biometrické snímače, dochádzkové systémy a podobne. Pri zabezpečovaní IoT zariadení je nevyhnutné aplikovať bezpečné protokoly, ako je TLS 1.3, a to nielen pri komunikácii so vzdialenými servermi na internete, ale aj so zariadeniami v lokálnej sieti.

Vstavané systémy majú obmedzené zdroje, čo spôsobuje problémy s implementáciou bezpečných protokolov alebo samotných šifrovacích algoritmov. Napríklad RSA, hoci poskytuje vysokú úroveň bezpečnosti, je výpočtovo náročnejší v porovnaní s modernými algoritmami na digitálny podpis založenými na eliptických krivkách (ECDSA - Elliptic Curve Digital Signature Algorithm) [12]. V Tabuľke 2 je znázornené porovnanie ekvivalentnej kryptografickej bezpečnosti medzi algoritmami RSA a ECDSA [13]. Pri implementácii týchto výpočtovo náročnejších algoritmov je nevyhnutné používať kryptografické knižnice optimalizované pre vstavané zariadenia, ktoré poskytujú efektívne a prispôsobené implementácie týchto algoritmov.

Tabuľka 2: Ekvivalentná kryptografická bezpečnosť RSA a ECDSA uvedená v bitoch [13]

Dĺžka kľúča RSA	**Dĺžka kľúča ECDSA**	**Ekvivalentná bezpečnosť**
2048	224	112
3072	256	128
7680	384	192

Medzi reálne problémy, ktoré môžu nastať pri implementácii náročných šifrovacích protokolov v IoT zariadeniach, patria:

- IoT zariadenia majú obmedzený výpočtový výkon, pamäť flash a RAM (Read Access Memory).
- Kvôli obmedzeným zdrojom IoT zariadenia používajú zjednodušené alebo neúplné implementácie kryptografických funkcií, čo môže viesť k potencionálnym zraniteľnostiam.
- IoT zariadenia často používajú rôzne protokoly a kryptografické algoritmy, ktoré môžu byť v súčastnej dobe zastaralé a veľa krát už nepodporované inými sieťovými zariadeniami.

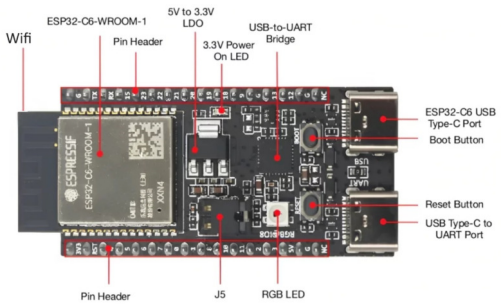
V. Vstavaná platforma na báze zariadenia ESP32-C6

Overenie digitálneho podpisu na báze RSA v TLS 1.3 protokole bolo vykonané na vývojovej doske ESP32-C6-DevKitC-1-N8 [14]. ESP32-C6 je ultra-nízkoenergetický systém na čipe (SoC) s jednojadrovým procesorom typu RISC-V (Reduced Instruction Set Computer - Version 5) [15]. Vývojová doska ESP32-C6-DevKitC-1-N8[3], ktorá je znázornená na Obr. 3, využíva SoC[4] a poskytuje vysoký výkon pri nízkej spotrebe energie.

Kľúčové parametre:

- **Procesor**: RISC-V jednojadrový procesor (32-bitový) s frekvenciou až 160 MHz,
- **RAM**: 512 kB HP SRAM a 16 kB LP SRAM,
- **Wi-Fi**: Podpora Wi-Fi 6 v pásme 2,4 GHz,
- **Bluetooth**: Podpora Bluetooth 5,
- **Ostatné**: Podpora IEEE 802.15.4 (Zigbee 3.0 a Thread 1.3), USB Type-C port.

RISC-V je open-source inštrukčná sada (ISA), ktorá určuje spôsob, ako procesor spracováva príkazy pre prácu s hardvérom a softvérom. Tento štandard sa stal populárny v počítačovom priemysle vďaka svojej flexibilite a otvorenosti pre vývoj.



Obr. 3: Vývojová doska ESP32-C6-DevKitC-1-N8.

[3]https://www.espressif.com/en/products/socs/esp32-c6
[4]https://www.ansys.com/blog/what-is-system-on-a-chip

VI. Kryptografické knižnice pre zariadenie ESP32-C6

Existuje viacero kryptografických knižníc pre vstavané zariadenia. Knižnice, ktoré poskytujú jednoduchú integráciu s prostredím ESP-IDF, určeným pre vývoj, testovanie a ladenie aplikácií pre ESP32-C6, sú:

- **CycloneCrypto (v2.5.0)**[5] – je kryptografická sada nástrojov navrhnutá na použitie vo vstavaných systémoch. Táto kryptografická knižnica má výhodu vo väčšej prehľadnosti a poskytuje optimalizované implementácie kódov pre rôzne typy procesorov. Nakoľko je však menej známa, má obmedzenú dokumentáciu.
- **MbedTLS (v3.1)**[6] – ide o kryptografickú knižnicu priamo implementovanú v prostredí ESP-IDF. Jej využitie je v danom prostredí veľmi jednoduché, čo zahŕňa efektívnu manipuláciu cez grafické prostredie menuconfig. MbedTLS má predvolene zapnutú hardvérovú akceleráciu pre zariadenie ESP32-C6.
- **WolfSSL (v5.8.0)**[7] – predstavuje robustnú kryptografickú knižnicu, ktorá má náročnejšiu implementáciu do prostredia ESP-IDF, taktiež zložitejšiu možnosť konfigurácie. Výhoda tejto knižnice spočíva v podpore širokej škály algoritmov a efektívneho, rýchleho kódu pre zariadenie ESP32-C6.
- **TinyTls (v1.0.0)**[8] – ide o oklieštenú knižnicu, ktorá podporuje protokol TLS a niektoré jeho kryptografické algoritmy. Táto knižnica má pomalšie a menej efektívne algoritmy, ale na základe jej kompaktnosti sú veľmi ľahko čitateľné a portovateľné.

A. Otestovanie overenia digitálneho podpisu RSA pre CycloneCrypto, MbedTLS, WolfSSL

Na zariadení ESP32-C6 bol testovaný algoritmus overovania digitálneho podpisu na báze RSA s použitím 2048-bitových a 4096-bitových kľúčov. Výsledky meraní pre všetky knižnice, okrem TinyTls, sú zhrnuté v Tabuľke 3. Pri meraní doby overenia podpisu pomocou knižnice TinyTls sa ukázalo, že modulárne umocňovanie, ktoré tvorí základnú operáciu algoritmu RSA, má rôzne trvanie v závislosti od vstupných dát. Z tohto dôvodu sú výsledky pre knižnicu TinyTls uvedené samostatne v Tabuľke 4. V rámci implementácie algoritmu RSA v knižnici TinyTls sme nahradili funkciu násobenia dvoch čísel efektívnejšou implementáciou z knižnice MonoCypher[9]. Overenie digitálneho podpisu sa po nahradení funkcie násobenia zrýchlilo o 44 %.

Na základe nameraných časov poskytuje kryptografická knižnica WolfSSL najrýchlejšie operácie pri overovaní RSA podpisu, a to napriek tomu, že pri meraní nepoužívala hardvérovú akceleráciu. Táto skutočnosť poukazuje na efektivitu softvérovej implementácie. Zároveň však platí, že pri overovaní digitálneho podpisu nezohráva hardvérová akcelerácia až takú kľúčovú úlohu pri dosahovaní vysokého výkonu, ako je to napríklad pri podpisovaní s využitím algoritmu RSA. Pri vytváraní digitálneho podpisu sa pracuje s niekoľkonásobne väčšími exponentmi, kde sú jednotlivé rozdiely oveľa viditeľnejšie.

Knižnica MbedTLS je ako jediná z testovaných knižníc nakonfigurovaná s hardvérovou akceleráciou, ktorá je dostupná na zariadení ESP32-C6 pre RSA do veľkosti kľúča 3072 bitov.

[5] https://www.oryx-embedded.com/products/CycloneCRYPTO.html
[6] https://www.trustedfirmware.org/projects/mbed-tls/
[7] https://www.wolfssl.com/
[8] https://github.com/Anthony-Mai/TinyTls
[9] https://monocypher.org/

Tabuľka 3: Časové porovnanie overenia digitálneho podpisu na báze RSA pre knižnice CycloneCrypto, MbedTLS, WolfSSL (v milisekundách, operácie vykonané v konštantnom čase, frekvencia procesora 160 MHz).

RSA	Optimalizácia	CycloneCrypto	MbedTLS	WolfSSL
		PKCS v1.5		
2048	-O2	54,18	27,12	15,08
2048	-Os	83,49	27,76	19,16
4096	-O2	207,08	101,09	56,98
4096	-Os	319,85	103,93	73,34
		PSS		
2048	-O2	54,57	27,33	15,43
2048	-Os	83,88	27,96	19,49
4096	-O2	207,78	101,45	57,72
4096	-Os	320,58	104,29	74,04
		Bez výplňovej schémy		
2048	-O2	53,97	27,02	14,91
2048	-Os	83,25	27,61	18,99
4096	-O2	206,71	101,04	56,83
4096	-Os	319,43	103,84	73,21

Tabuľka 4: Časové porovnanie overenia digitálneho podpisu na báze RSA pre TinyTLS (v milisekundách, operácie nevykonané v konštantnom čase, frekvencia procesora 160 MHz).

RSA	Optimalizácia	TinyTLS	Bez výplňovej schémy
		PKCS v1.5	
2048	-O2	124,69	123,56
2048	-Os	131,65	132,81
4096	-O2	196,97	196,37
4096	-Os	209,47	210,12
		PSS	
2048	-O2	120,99	120,09
2048	-Os	128,02	129,07
4096	-O2	190,65	189,68
4096	-Os	202,36	203,83

B. Implementácia algoritmu RSA z knižnice TinyTls

Kryptografická knižnica TinyTls poskytuje najpomalšiu operáciu pre algoritmus RSA a zároveň časovo závislú od vstupných dát. Napriek tomu bola táto implementácia využitá v aplikácii Minimalistický TLS 1.3 klient [16], pretože poskytuje jednoduchú portovateľnosť a efektívny prehľad jednotlivých funkcií, ktoré sa pri overovaní digitálneho podpisu využívajú. Z hľadiska praktického využitia nadviazania TLS 1.3 spojenia nie je doba trvania použitého overenia digitálneho podpisu obmedzujúca a skutočnosť, že pri overení certifikátu nechránime žiadny tajný kľúč, znamená, že časová závislosť nepredstavuje bezpečnostné riziko časového útoku postrannými kanálmi [17]. Testovanie RSA operácií (2048-bit a 4096-bit, s výplňovými schémami PKCS#1.5 a PSS) na platforme `ESP32` pomocou knižnice `TinyTLS` ukázalo nízku pamäťovú spotrebu RAM. Pri optimalizácii pre veľkosť (`-Os`) dosahujú tieto operácie využitie zásobníka (stacku) v rozsahu od 8,9 do 9,4 kB bez akéhokoľvek použitia haldy (heapu). Pri optimalizácii pre rýchlosť (`-O2`) sa nároky mierne znižujú na 8,9–9,3 kB stacku, opäť bez potreby využitia heapu. Uvedené výsledky potvrdzujú, že implementácia `TinyTLS` je mimoriadne efektívna pre vstavané systémy s obmedzenou pamäťou.

VII. Záver

V článku boli opísané metódy overenia digitálnych podpisov založených na algoritme RSA, s osobitným zreteľom na ich využitie vo vstavaných zariadeniach. Boli predstavené princípy fungovania algoritmu RSA a jeho uplatnenie v protokole TLS 1.3. V kontexte digitálneho podpisu RSA sa objasnili výplňové schémy PKCS v1.5 a RSA-PSS spolu s ich praktickým významom pre kryptografickú bezpečnosť. Osobitná pozornosť bola venovaná časovej analýze overenia digitálneho podpisu RSA pomocou kryptografických knižníc CycloneCrypto, WolfSSL, MbedTLS a TinyTLS na vývojovej doske ESP32-C6-DevKitC-1-N8. Analýza ukázala, že knižnica WolfSSL dosahuje najvyššiu rýchlosť pri operáciách s algoritmom RSA.

Napriek tomu bola v aplikácii Minimalistického TLS 1.3 klienta [16] použitá implementácia z knižnice TinyTls (s funkciou násobenia prevzatou z knižnice MonoCypher) pre jej jednoduchú portovateľnosť a prehľadnosť kódu. Pri reálnom internetovom pripojení by overovanie digitálnych podpisov s použitím knižnice TinyTLS pre certifikáty RSA 4096 PSS, RSA 2048 PKCS a RSA 4096 PKCS nepresiahlo 37 % celkového času potrebného na nadviazanie zabezpečeného spojenia. Výsledky tejto analýzy majú osobitný význam pre oblasť IoT, kde obmedzené výpočtové zdroje zohrávajú kľúčovú úlohu pri zabezpečovaní bezpečnej TLS 1.3 komunikácie.

VIII. Poďakovanie

Táto práca bola podporená Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky v rámci projektu KEGA 041TUKE-4/2025 „Laboratórium na výskum a vývoj pokročilých IoT senzorových sietí a širokopásmových senzorov“.

Literatúra

[1] N. Kumari and A. K. Mohapatra, “A comprehensive and critical analysis of TLS 1.3,” *Journal of Information and Optimization Sciences*, vol. 43, no. 4, pp. 689–703, 2022. [Online]. Available: doi.org/10.1080/02522667.2021.1968579
[2] W. Stallings, “Digital Signature Algorithms,” *Cryptologia*, vol. 37, no. 4, pp. 311–327, 2013. [Online]. Available: doi.org/10.1080/01611194.2013.797044
[3] M. Ding, “Analyzing the Research and Application of the RSA Cryptosystem in the Context of Digital Signatures,” *TCSISR*, vol. 5, 2024. [Online]. Available: doi.org/10.62051/j2p6g984
[4] C. Paar, J. Pelzl, and T. Güneysu, *The RSA Cryptosystem*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2024, pp. 205–240. [Online]. Available: doi.org/10.1007/978-3-662-69007-9_7
[5] Department of Mathematical and Statistical Sciences, University of Alberta, “Fermat Numbers,” 2012. [Online]. Available: https://www.math.ualberta.ca/~isaac/math324/s12/fermat_numbers.pdf
[6] A. Boldyreva, H. Imai, and K. Kobara, “How to Strengthen the Security of RSA-OAEP,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5876–5886, 2010. [Online]. Available: doi.org/10.1109/TIT.2010.2070330
[7] B. Kaliski, “PKCS 1: RSA Encryption Version 1.5,” RSA Laboratories East, Request for Comments 2313, Mar. 1998. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc2313
[8] S. A. Kakvi, “On the Security of RSA-PSS in the Wild,” Cryptology ePrint Archive, Paper 2019/1268, 2019. [Online]. Available: doi.org/10.1145/3338500.3360333
[9] S. R. Prasanna and B. S. Premananda, “Performance Analysis of MD5 and SHA-256 Algorithms to Maintain Data Integrity,” in *2021 International Conference on Recent Trends on Electronics, Information, Communication and Technology (RTEICT)*. IEEE, 2021, pp. 246–250. [Online]. Available: doi.org/10.1109/RTEICT52294.2021.9573660
[10] I. Cicek, “FPGA Hardware Implementation of a SHA384 Accelerator for Internet of Things Applications,” *European Journal of Science and Technology*, Jun. 2021. [Online]. Available: doi.org/10.31590/ejosat.951580
[11] P. Nguyen, A. Goknil, G. Erdogan, S. Tokas, N. Ferry, and T. T. T. Tran, “Advances in secure iot data sharing,” *Foundations and Trends in Privacy and Security*, vol. 7, no. 1, pp. 1–73, 2024. [Online]. Available: doi.org/10.1561/3300000042
[12] J. Bauer, R. Staudemeyer, A. Fragkiadakis, and H. Pöhls, “ECDSA on Things: IoT Integrity Protection in Practise,” in *Lecture Notes in Computer Science*, vol. 9977. Springer, Dec. 2016. [Online]. Available: doi.org/10.1007/978-3-319-50011-9_1
[13] M. Savari, M. Montazerolzohour, and Y. E. Thiam, “Comparison of ECC and RSA algorithm in multipurpose smart card application,” in *2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec)*. IEEE, 2012, pp. 49–53. [Online]. Available: doi.org/10.1109/CyberSec.2012.6246121
[14] Espressif Systems, “ESP32-C6 Datasheet,” Espressif Systems, Tech. Rep., Apr. 2023. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32-c6_datasheet_en.pdf
[15] V. S. Chakravarthi and S. R. Koteshwar, “RISC-V-Based Processor Systems,” in *SOC-Based Solutions in Emerging Application Domains*. Springer, Cham, 2025. [Online]. Available: doi.org/10.1007/978-3-031-85044-8_10
[16] J. Pajkos, E. Kupcova, M. Pleva, and M. Drutarovský, “ESP32 Microcontroller based Lightweight TLS 1.3 Client for IoT Applications,” in *2021 International Conference on Recent Trends on Electronics, Information, Communication and Technology (RTEICT)*. Czech Republic: IEEE, May 2025, pp. 1–6. [Online]. Available: doi.org/10.1109/RADIOELEKTRONIKA65656.2025.11008381
[17] F. Koeune, *Timing Attack*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019, pp. 1–3. [Online]. Available: doi.org/10.1007/978-3-642-27739-9_138-2

Digitálne podpisovanie občianskym preukazom a bezpečnostné riziko rodného čísla

[1]*Karolína POLAČKOVÁ,* [2]*Miloš DRUTAROVSKÝ*

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

[1]karolina.polackova@student.tuke.sk, [2]milos.drutarovsky@tuke.sk

Abstrakt **– Článok sa zaoberá bezpečnostnou analýzou elektronického podpisovania pomocou slovenského občianskeho preukazu s čipom (SVK eID) so zameraním na ochranu osobných údajov. Opisuje formát elektronického podpisu PAdES, pri ktorom sa do PDF dokumentu vkladá kvalifikovaný certifikát vo formáte X.509. Ukazuje, že tento certifikát obsahuje viaceré osobné údaje vrátane rodného čísla podpisujúcej osoby, ktoré sa tak môže stať prístupným každému, kto má k dokumentu prístup. V článku sú ukázané spôsoby extrahovania týchto údajov pomocou bežne dostupných nástrojov, ako sú Python, OpenSSL alebo Adobe Acrobat Reader. Analýza sa ďalej venuje právnym a bezpečnostným dôsledkom, najmä v súvislosti s GDPR, ochranou osobných údajov a rizikom zneužitia identity.**

Kľúčové slová **– PAdES, SVK eID, elektronický podpis, občiansky preukaz, rodné číslo, certifikát, bezpečnosť**

I. Úvod

V digitálnej dobe, keď sa elektronická komunikácia stáva štandardom vo verejnej správe aj v súkromnom sektore, zohráva digitálna autentizácia dokumentov kľúčovú úlohu pri zabezpečení právnej záväznosti a dôveryhodnosti. Elektronické podpisovanie pomocou slovenského občianskeho preukazu s čipom (SVK eID) poskytuje občanom možnosť vytvárať kvalifikovaný elektronický podpis, ktorý je z právneho hľadiska ekvivalentom vlastnoručného podpisu. Táto technológia umožňuje bezpečné podpisovanie dokumentov na diaľku a zároveň spoľahlivo overuje totožnosť podpisujúceho.

V pozadí tohto procesu však prebiehajú technické úkony, ktoré môžu mať nečakané dôsledky z pohľadu ochrany súkromia. Konkrétne ide o to, že digitálny podpis vo formáte PAdES vkladá do podpísaného dokumentu kvalifikovaný certifikát podpisujúcej osoby vo formáte X.509, ktorý obsahuje osobné údaje podpisujúcej osoby. Tieto údaje, ktoré tvoria elektronickú identitu osoby, zahŕňajú nielen meno a priezvisko, ale aj celé rodné číslo [1], t.j. údaj, ktorý sa považuje za citlivý a podlieha prísnej právnej ochrane podľa európskeho nariadenia GDPR a slovenského zákona o ochrane osobných údajov [2].

Tento aspekt digitálneho podpisovania pomocou SVK eID je všeobecne známy, je však často prehliadaný, najmä zo strany bežných používateľov, ktorí si neuvedomujú, že podpísaním dokumentu sprístupňujú svoje rodné číslo každému, kto má k dokumentu prístup. Článok preto analyzuje technický mechanizmus podpisovania prostredníctvom SVK eID, demonštruje spôsoby extrakcie osobných údajov z certifikátu pomocou nástrojov ako OpenSSL a Adobe Acrobat Reader a v neposlednom rade sa venuje aj právnym a bezpečnostným rizikám, ktoré z tohto faktu vyplývajú.

Cieľom tohto článku je upozorniť na potrebu zodpovedného prístupu k spracúvaniu osobných údajov v kontexte elektronického podpisovania a poukázať na možnosti zníženia rizík pri zachovaní funkčnosti a právnej platnosti digitálneho podpisu.

II. Technické pozadie podpisovania so SVK eID

SVK eID [1] je karta, ktorá obsahuje kryptografický čip na uchovávanie privátneho kľúča. Pri podpise dokumentu sa z obsahu vypočíta hash, resp. urobí sa z dokumentu jeho odtlačok,

ktorý sa šifruje privátnym kľúčom v karte. Výsledkom je elektronický podpis, ktorý sa uloží spolu s certifikátom používateľa vo formáte X.509 [3], ktorý obsahuje identifikačné údaje podpisovateľa. Certifikát používateľa je kľúčovou súčasťou podpisu a umožňuje tretím stranám overiť pravosť podpisu.

Na Slovensku sa najčastejšie používa algoritmus RSA s veľkosťou kľúča 3072 bitov, hashovacia funkcia SHA-256 a podpisový algoritmus RSASSA-PKCS1-v1_5 [4]. Certifikát používateľa vydáva kvalifikovaná certifikačná autorita, napríklad Disig a.s., a jeho platnosť je typicky 5 rokov.

Porovnanie PAdES, CAdES a XAdES formátov

Existujú tri hlavné formáty elektronického podpisu, ktoré možno využívať so SVK eID:

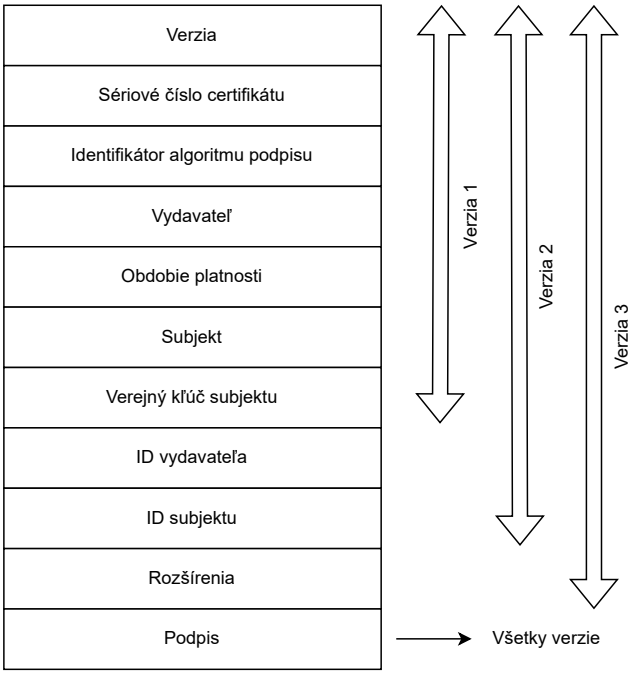
- **PAdES [5]** – pre PDF dokumenty, certifikát je vložený priamo v PDF.
- **CAdES [6]** – podpis vo formáte ASN.1, používa kontajnery `.p7s`, `.asice`.
- **XAdES [7]** – podpis určený pre XML, využívajúci `XML Signature`.

Vo všetkých troch formátoch sa vkladá certifikát podpisujúcej osoby, ktorý obsahuje aj jej rodné číslo. Rozdiely sú najmä v spôsobe jeho uloženia a v možnostiach extrakcie týchto údajov z dokumentu. Hoci sa tento problém týka všetkých formátov (PAdES, CAdES aj XAdES), v tomto článku sa ďalej zameriavame výlučne na formát PAdES.

III. Digitálny certifikát a osobné údaje

Pri elektronickom podpise pomocou SVK eID sa používa kvalifikovaný certifikát, vydaný podľa nariadenia eIDAS [8]. Tento certifikát je súčasťou výsledného PDF dokumentu vo formáte PAdES, pričom ide o certifikát typu X.509, ktorý prepája verejný kľúč s elektronickou identitou podpisujúcej osoby.

Certifikát obsahuje osobné údaje ako meno, priezvisko, mesto bydliska, ale taktiež aj celé rodné číslo v čitateľnej podobe. Štruktúru certifikátu je vidieť na obrázku 1, kde je zobrazená položka `Subject`, ktorá tieto informácie obsahuje. Ich súčasťou môžu byť aj ďalšie voliteľné rozšírenia, ktoré dopĺňajú údaje o subjekte.

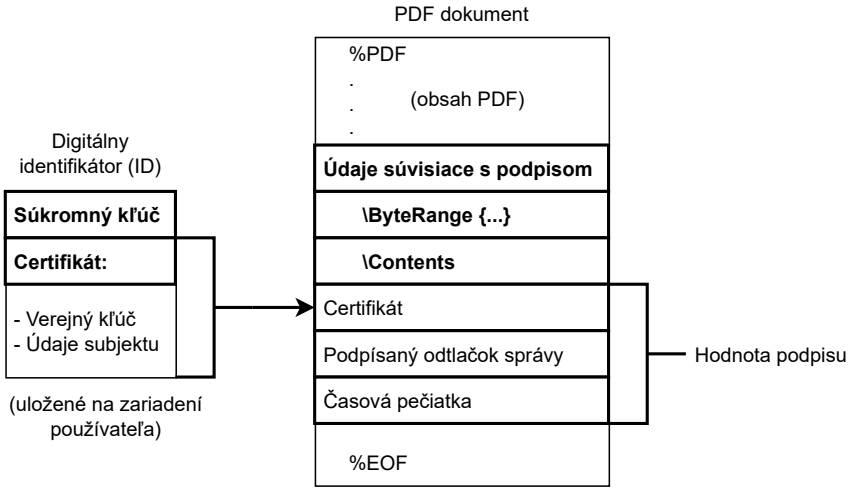


Obr. 1: Štruktúra X.509 certifikátu [9]

Keďže tento certifikát je verejne dostupný každému, kto si otvorí podpísaný dokument, obsah osobných údajov v ňom predstavuje potenciálne riziko z pohľadu ochrany súkromia. V nasledujúcich častiach preto analyzujeme, ako možno tieto údaje získať pomocou nástrojov ako OpenSSL alebo Adobe Acrobat Reader a aké právne, či bezpečnostné dôsledky z toho vyplývajú.

IV. Digitálny podpis vo formáte PAdES a riziká úniku osobných údajov

Pri elektronickom podpisovaní PDF dokumentov pomocou SVK eID sa používa formát PAdES, ktorý vkladá elektronický podpis priamo do štruktúry PDF. Tento podpis obsahuje certifikát vo formáte X.509, v ktorom sa nachádzajú osobné údaje, vrátane už spomínaného rodného čísla podpisujúcej osoby. Tieto údaje sú súčasťou podpisového objektu a je možné ich extrahovať pomocou nástrojov ako Python [10], pdfsig [11], OpenSSL [12] a Adobe Acrobat Reader [13]. Štruktúru podpísaného PDF dokumentu je možné vidieť na obrázku 2.



Obr. 2: Štruktúra podpísaného PDF dokumentu [14]

Obrázok 2 zobrazuje, ako sa elektronický podpis vkladá do PDF dokumentu. Ľavá časť obrázku znázorňuje SVK eID kartu, ktorá obsahuje digitálny identifikátor s certifikátom a súkromným kľúčom. Súkromný kľúč kartu nikdy neopúšťa, takže pri podpisovaní sa použije priamo v karte na vytvorenie podpisu z odtlačku dokumentu. Do PDF sa následne vloží samotný podpis spolu s časťou certifikátu, nie však súkromný kľúč.

V pravej časti obrázku je znázornená štruktúra podpísaného PDF dokumentu, v ktorom sa nachádza tzv. podpisová časť. Tá obsahuje napríklad:

- pole `ByteRange` so zoznamom podpísaných bajtov,
- pole `Contents`, v ktorom je zakódovaný podpis v štandarde PKCS#7 [15] spolu s certifikátom.

Keďže podpisová časť obsahuje celý certifikát, všetky osobné údaje, vrátane rodného čísla, sú prístupné komukoľvek, kto má k dokumentu prístup. Tieto údaje sú totiž súčasťou štruktúry `SignedData`, ktorá je čitateľná.

Na overenie podpisu možno použiť nástroje ako napríklad OpenSSL, kde celý postup overenia je ukázaný v [16].

Podpis sa dá overiť pomocou príkazu:

```
openssl cms -verify \
  -binary \
  -in signature.p7s \
  -inform DER \
  -content signed_data.bin \
  -CAfile fullchain.pem \
  -purpose any
```

Príkaz overí, či podpis zodpovedá podpísanému obsahu a či bol certifikát vydaný dôveryhodnou certifikačnou autoritou.

Výplňová schéma podpisu

Overenie výplňovej schémy (ang. paddingu) v podpise je kľúčové pre bezpečnostnú analýzu. Pri SVK eID kartách sa používa algoritmus RSA, kde sa podpis generuje šifrovaním hash hodnoty (odtlačku dokumentu) pomocou súkromného kľúča. Na overenie typu výplne je potrebné dešifrovať podpis verejným kľúčom a analyzovať výslednú binárnu štruktúru.

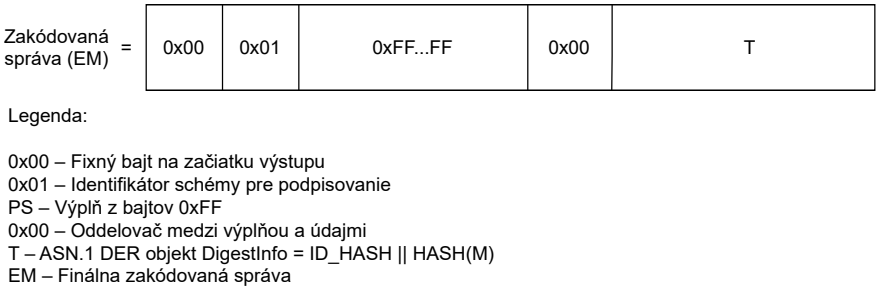
Postup pozostáva z troch krokov:

- Extrakcie samotného podpisu zo súboru `.p7s` pomocou skriptu, ktorý načíta štruktúru CMS (Cryptographic Message Syntax) a z nej získa pole `signatureValue`.
- Získania verejného kľúča z certifikátu, kde sa najprv extrahuje certifikát z podpísaného súboru a následne sa z neho exportuje verejný kľúč do súboru `public_key.pem`.
- Dešifrovania podpisu verejným kľúčom, čím sa zistí štruktúra podpisového bloku. Pri PKCS#1 v1.5 [4] začína sekvenciou bajtov `0x00 0x01` nasledovanou výplňou a ASN.1 štruktúrou s identifikátorom hashovacej funkcie a hodnotou hash.

Ukážka výstupu z dešifrovaného bloku:

```
Raw RSA decrypted block:
00 01 FF FF FF ... 00 30 31 30 ...
```

Tento dešifrovaný blok zodpovedá štandardnej výplňovej schéme PKCS#1 v1.5, ktorá sa bežne používa aj pri podpisovaní pomocou SVK eID. Schéma štruktúry podpisového bloku je znázornená na obrázku 3.



Obr. 3: Formát podpisu podľa PKCS#1 v1.5 s paddingom a ASN.1 štruktúrou `DigestInfo`, ktorý je použitý aj v SVK eID [17], [18]

V. Získanie rodného čísla z digitálneho podpisu

V predchádzajúcich častiach sme sa zaoberali procesom elektronického podpisovania pomocou SVK eID, pričom sme ukázali, že rodné číslo podpisujúcej osoby je súčasťou certifikátu vo formáte X.509. V tejto sekcii sa zameriame na konkrétne kroky, ako tieto údaje, vrátane rodného čísla, možno extrahovať z podpísaných dokumentov pomocou nástrojov ako Python, OpenSSL a Adobe Acrobat Reader.

A. Získanie rodného čísla pomocou OpenSSL

Certifikát, ktorý je súčasťou digitálneho podpisu, obsahuje rodné číslo podpisujúcej osoby. Tento certifikát je možné extrahovať a analyzovať pomocou nástroja OpenSSL (verzia 3.5.0) [12]. Následujúce kroky opisujú, ako získať tieto informácie.

1) **Extrahovanie certifikátu zo súboru podpísaného dokumentu:**

```
openssl pkcs7 \
  -in podpis.p7s \
  -inform DER \
  -print_certs \
  -out certs.pem
```

Tento príkaz extrahuje certifikát zo súboru podpísaného dokumentu (v tomto prípade vo formáte `p7s`) a uloží ho do súboru `certs.pem`.

2) **Zobrazenie certifikátu a nájdenie rodného čísla:**

```
openssl x509 \
  -in certs.pem \
  -text \
  -noout
```

Tento príkaz zobrazí celý obsah certifikátu v textovej podobe. Vo výstupe bude položka `Subject`, kde bude uvedené rodné číslo. Napríklad:

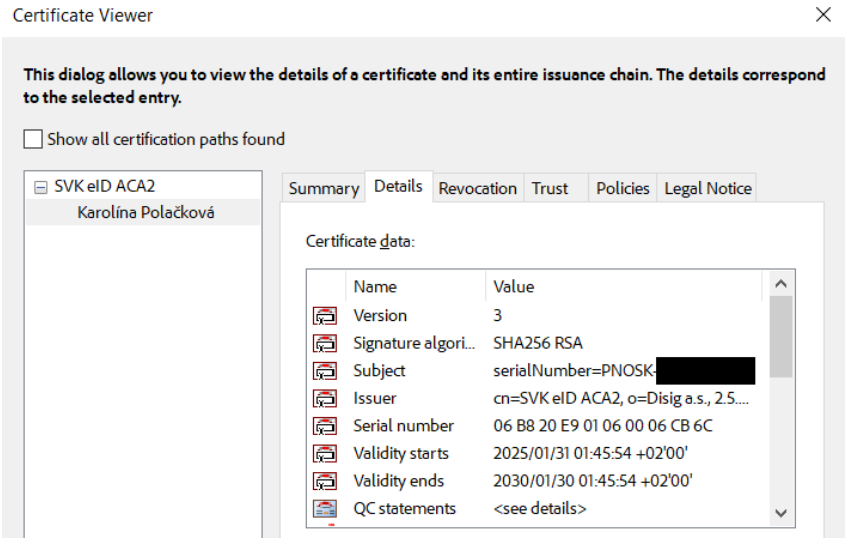
```
subject=CN = Karolina Polackova,
         GN = Karolina,
         SN = Polackova,
         L  = Presov,
         C  = SK,
         serialNumber = PNOSK-rodneCislo
```

B. Získanie rodného čísla pomocou Adobe Acrobat Reader

Rodné číslo je taktiež možné získať z certifikátu prostredníctvom aplikácie Adobe Acrobat Reader (verzia 25.001.20474) [13], ktorá umožňuje zobraziť certifikát použitý na podpísanie dokumentu. Tento proces prebieha nasledovne:

1) **Otvorenie podpísaného PDF dokumentu:**
   Dokument sa otvorí v aplikácii Adobe Acrobat Reader.
2) **Zobrazenie podpis:**
   V dokumente sa klikne na tlačidlo *Signature Panel*, čím sa otvorí panel s informáciami o podpise.
3) **Zobrazenie detailov certifikátu:**
   V paneli *Signatures* sa zobrazí zoznam podpisov, pričom prvý záznam má tvar `Rev. 1: Signed by [Meno osoby]`. Na tento záznam sa klikne pravým tlačidlom myši a zvolí sa možnosť *Show Signature Properties*.
   V novom okne *Signature Properties* sa v spodnej časti nachádza sekcia *Signer Info*, kde je tlačidlo *Show Signer's Certificate*. Po jeho kliknutí sa otvorí nástroj *Certificate Viewer*, v ktorom sa prejde na kartu *Details*.
   Na tejto karte je zobrazená štruktúra certifikátu vrátane jednotlivých polí a OID identifikátorov. Medzi nimi sa nachádza aj atribút `serialNumber`, ktorý obsahuje rodné číslo vo formáte `RRMMDD/XXXX`.

Tento údaj je priamo viditeľný aj na obrázku 4 nižšie a nevyžaduje žiadne špeciálne oprávnenie, takže každý, kto má prístup k dokumentu, má zároveň prístup aj k tomuto osobnému údaju. Na uvedenom obrázku je rodné číslo z bezpečnostných dôvodov začiernené.



Obr. 4: Zobrazenie certifikátu v Adobe Acrobat Reader (rodné číslo začiernené)

Tento postup umožňuje zobraziť certifikát a získať rodné číslo, ktoré je priamo a verejne viditeľné ako súčasť certifikátu podpisujúcej osoby.

VI. Bezpečnostné riziká a ochrana osobných údajov

Ako sme ukázali v predchádzajúcich častiach, kvalifikovaný certifikát je pevnou súčasťou každého elektronicky podpísaného PDF dokumentu. Tento certifikát obsahuje viacero osobných údajov, vrátane celého rodného čísla podpisujúcej osoby, ktoré je uložené v poli `serialNumber`. Tieto údaje sú verejne dostupné každému, kto si otvorí dokument – bez nutnosti akéhokoľvek prihlasovania či overenia identity.

Takéto sprístupnenie citlivého identifikátora, akým je rodné číslo, predstavuje vážne riziko jeho zneužitia. Rodné číslo umožňuje jednoznačnú identifikáciu osoby a spolu s ďalšími údajmi (napr. meno, priezvisko, mesto) môže byť zneužité na rôzne podvody, neoprávnený prístup k službám alebo vytváranie falošných identít.

Z nášho pohľadu je ochrana súkromia dôležitá a osobné údaje by sa mali spracúvať len v nevyhnutnom rozsahu. Zahrnutie celého rodného čísla do verejne dostupného certifikátu podľa nás nezodpovedá princípu minimalizácie údajov. V kontexte nariadenia GDPR a zákona č. 18/2018 Z.z. [19] o ochrane osobných údajov by mal byť rozsah spracúvaných údajov primeraný konkrétnemu účelu, v tomto prípade overeniu elektronického podpisu. Rodné číslo na tento účel podľa nášho názoru nie je nevyhnutné použiť.

Príklady zneužitia v praxi

Rodné číslo môže byť v nesprávnych rukách zneužité napríklad na:

- **identitné podvody** – napr. vybavenie pôžičiek alebo leasingu na cudzie meno,
- **cielený phishing** – útočník môže vytvárať personalizované správy, ktoré zvyšujú dôveryhodnosť podvodu,
- **falošné zmluvy alebo registrácie** – využitie údajov na podpisovanie zmlúv bez vedomia dotknutej osoby.

Odporúčania

Aby sa znížilo riziko zneužitia osobných údajov, odporúčame:

- **odstrániť rodné číslo** z verejne dostupných certifikátov, alebo ho aspoň **nahradiť pseudonymom alebo jednorazovým identifikátorom**,
- **zvážiť úpravu legislatívneho a technického rámca** vydávania kvalifikovaných certifikátov na Slovensku, a to s dôrazom na ochranu súkromia,
- **zvýšiť povedomie používateľov** o tom, aké údaje sa v certifikátoch nachádzajú a aké sú riziká ich verejného zverejnenia.

VII. Záver

Z analýzy vyplýva, že súčasný systém elektronického podpisovania pomocou SVK eID, hoci technicky vyspelý a právne záväzný, má vážny nedostatok z pohľadu ochrany osobných údajov. Automatické vkladanie celého rodného čísla do kvalifikovaného certifikátu, ktorý je následne súčasťou každého podpísaného dokumentu vo formátoch PAdES, XAdES aj CAdES, predstavuje významné riziko. Ide o informáciu, ktorá má vysokú hodnotu pre potenciálnych útočníkov a môže byť zneužitá pri identitných podvodoch, phishingových útokoch alebo neoprávnenom profilovaní osôb.

Zverejnenie rodného čísla bez akejkoľvek autentizácie zo strany príjemcu dokumentu predstavuje porušenie zásad minimalizácie a nevyhnutnosti spracúvania údajov podľa nariadenia GDPR a slovenského zákona o ochrane osobných údajov. Navyše, väčšina používateľov si túto skutočnosť ani neuvedomuje, čo zvyšuje riziko nevedomého ohrozenia ich súkromia.

Jedným z odporúčaných riešení je, aby sa Slovenská republika ako členský štát EÚ zaoberala možnosťou technickej a legislatívnej úpravy procesu vydávania kvalifikovaných certifikátov. Ako potenciálne riešenia vnímame napríklad pseudonymizáciu rodného čísla, jeho nahradenie iným identifikátorom alebo poskytnutie možnosti výberu rozsahu údajov, ktoré budú v certifikáte zahrnuté. Z nášho pohľadu je rovnako dôležité zvyšovať informovanosť používateľov SVK eID o tom, aké údaje podpisovaním zverejňujú.

Ochrana digitálnej identity by mala byť rovnako dôležitá ako zabezpečenie právnej platnosti dokumentov. Inak môže dôjsť k oslabeniu dôvery občanov voči systému elektronickej komunikácie.

VIII. Poďakovanie

Práca bola podporená EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 17R05-04-V01-00004 – „Kompetenčné a vzdelávacie centrum kybernetickej a informačnej bezpečnosti TUKE“.

Literatúra

[1] Úrad podpredsedu vlády SR pre investície a informatizáciu, “eID karta,” https://www.slovensko.sk/sk/eid/_eid-karta/, 2022.

[2] V. Milošovič, “Čo patrí medzi osobné údaje?” November 2021, accessed: 2025-06-03. [Online]. Available: https://www.securion.sk/co-patri-medzi-osobne-udaje/

[3] D. Cooper, S. Santesson, and S. Farrell, “Internet X.509 public key infrastructure certificate and certificate revocation list (crl) profile,” Internet Engineering Task Force (IETF), Request for Comments 5280, May 2008. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc5280

[4] J. Jonsson and B. Kaliski, “PKCS #1: RSA cryptography specifications version 2.2,” Internet Engineering Task Force (IETF), Request for Comments 8017, November 2016. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8017

[5] E. T. S. I. (ETSI), “Electronic signatures and infrastructures (ESI); PAdES digital signatures; part 1: Building blocks and pades baseline signatures,” European Telecommunications Standards Institute (ETSI), Technical Specification, February 2016. [Online]. Available: https://www.etsi.org/deliver/etsi_en/319100_319199/31914201/01.01.00_30/en_31914201v010100v.pdf

[6] ——, “Electronic signatures and infrastructures (ESI); CAdES digital signatures; part 1: Building blocks and cades baseline signatures,” European Telecommunications Standards Institute (ETSI), Technical Specification, June 2023. [Online]. Available: https://www.etsi.org/deliver/etsi_en/319100_319199/31912201/01.03.01_60/en_31912201v010301p.pdf

[7] ——, "Electronic signatures and infrastructures (ESI); XAdES digital signatures; part 1: Building blocks and xades baseline signatures," European Telecommunications Standards Institute (ETSI), Technical Specification, February 2022. [Online]. Available: https://www.etsi.org/deliver/etsi_en/319100_319199/31913201/01.02.01_60/en_31913201v010201p.pdf
[8] European Commission, "eIDAS regulation," May 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation
[9] Microsoft Corporation, "X.509 public key certificates," https://learn.microsoft.com/en-us/windows/win32/seccertenroll/about-x-509-public-key-certificates, 2021.
[10] P. S. Foundation, "Python programming language – official website," https://www.python.org/, 2025, accessed: 2025-06-03.
[11] Ubuntu Manual Pages, "pdfsig - portable document format (PDF) digital signatures tool," https://manpages.ubuntu.com/manpages/lunar/man1/pdfsig.1.html, 2019, accessed: 2025-06-03.
[12] O. S. Foundation, "Openssl," https://www.openssl.org/, 2025.
[13] Adobe, "Čo môžem robiť s aplikáciou acrobat reader?" https://helpx.adobe.com/sk/reader/using/do-reader.html, 2022, accessed: 2025-06-03.
[14] A. S. Incorporated, "Digital signatures in PDF," https://pubhtml5.com/ymbo/bzab/Acrobat_DigitalSignatures_in_PDF, accessed: 2025-06-03.
[15] B. Kaliski, "PKCS #7: Cryptographic message syntax version 1.5," RSA Laboratories, Request for Comments 2315, March 1998. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc2315
[16] K. Polačková, "Digitálne podpisy na báze algoritmov RSA, DSA, ECDSA a EdDSA," Bakalárska práca, Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, máj 2025.
[17] Stack Exchange User, "Why PS does differ between PKCS#1 v1.5 padding for signature and for encryption?" https://crypto.stackexchange.com/questions/61178/why-ps-does-differ-between-pkcs1-v1-5-padding-for-signature-and-for-encryption, 2018.
[18] F. Valsorda, "Bleichenbacher 06 signature forgery in python RSA," https://words.filippo.io/bleichenbacher-06-signature-forgery-in-python-rsa/, 2016.
[19] Národná rada Slovenskej republiky, "Zákon č. 18/2018 z.z. o ochrane osobných údajov a o zmene a doplnení niektorých zákonov," Zbierka zákonov Slovenskej republiky, 2018. [Online]. Available: https://www.slov-lex.sk/ezbierky/pravne-predpisy/SK/ZZ/2018/18/20240701

Platform for sharing projects

[1]*Veronika KALEJOVÁ,* [2]*Miroslav MURIN,* [3]*Ondrej KAINZ,*
[4]*Miroslav MICHALKO*

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

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

Abstract **– This thesis focuses on the design and implementation of an application that enables multiple users to collaborate on the same software project in real time. The tool is primarily intended for computer science students, who often work on team projects that require collaborative editing of source code. In the early stages, we analyzed existing solutions on the market and identified several limitations, such as high usage costs and limited support for programming languages. For the frontend, we used a technology called Flutter, while the backend was implemented using Node.js with the Express.js framework. Data is stored in a relational MySQL database. The final solution provides an environment for collaborative coding, emphasizing a simple design and fast real-time updates. The application was successfully implemented and tested, with only minor adjustments from the original design. The results of the implementation and testing confirmed the usability of the proposed solution and its potential for future development.**

Keywords **– coding in real time, pair programming, assigning tasks, compiling code, Flutter app**

I. Introduction

Direct collaboration is a key aspect of student teamwork at universities. However, many computer science students find team-based work challenging, largely due to the lack of effective tools for real-time collaborative coding. While students often use Git to work on shared projects, it does not offer the same experience as direct, real-time access to the code. Existing tools provide limited support in this area. Our goal is to develop an effective solution that enables real-time collaboration in coding projects.

Modern software development increasingly relies on real-time communication and seamless integration of tools. While professionals often work with advanced platforms that combine code sharing, task management, and messaging, students are frequently limited to using a combination of disconnected tools that are not tailored to their specific needs. This fragmentation complicates the development process and increases the cognitive load, especially for beginners who are still learning to navigate team dynamics and development environments.

By creating a unified platform that integrates all essential components—such as real-time code editing, project organization, task tracking, and team communication—we aim to provide students with an accessible and practical environment that supports their learning and collaboration. The goal is not only to enhance productivity, but also to foster a better understanding of modern development workflows through an intuitive and supportive toolset.

II. Background and Implementation Tools

When designing a platform that supports team collaboration and real-time code editing, it is essential to understand both the theoretical background of collaborative work and the technologies that enable such functionality. This chapter provides an overview of the key principles behind effective teamwork, explores existing tools used for collaborative programming, and describes the technologies chosen for the development of the proposed solution. Understanding these foundations is crucial for identifying gaps in current solutions and justifying the architectural and technological decisions made during implementation.

A. *Effective team work*

Effective teamwork plays a crucial role in solving complex tasks, especially in academic environments where students gain practical experience through project-based work. To ensure successful collaboration, several prerequisites must be met, such as clear communication, transparent task distribution, the ability to track progress, and feedback. Modern platforms that allow multiple users to work in parallel on different parts of a project and support real-time collaboration form the essential foundation for such teamwork.

Key collaborative methods that can be supported by the proposed platform include live code review and pair programming. Live code review enables immediate feedback and enhances the quality of the resulting code, which is valuable not only during development but also in interview scenarios. There are many methods of live code review, with the most common being so-called tool-assisted reviews, which involve the use of supporting tools, and over-the-shoulder reviews, where the reviewer physically observes the author explaining the code [1]. However, the effectiveness of this method depends on the availability of appropriate software.

Pair programming is commonly used in software development and IT education [2]. It is based on the cooperation of two developers assuming the roles of driver and navigator [3]. This form of collaboration requires tools that support real-time code sharing, which is often limited or missing in many existing solutions.

B. *Existing solutions*

Solutions and tools for team collaboration in software development often provide useful features, but their main problem is fragmentation. Users have to use several different apps for communication, task management, and coding. Collaborative programming tools, such as JetBrains Code With Me, Visual Studio Live Share, and Google Colab, enable real-time code editing and shared work environments, but they often lack integrated features for task management and communication within a single platform.

Task management platforms like Trello provide visual tools for organizing work through Kanban boards [4], as shown in Figure 1. They improve team planning but are not specifically designed for programming. In one of the studies, students confirmed that their work was more effective and their perception of the application was positive, as it supported their collaboration on projects [5].

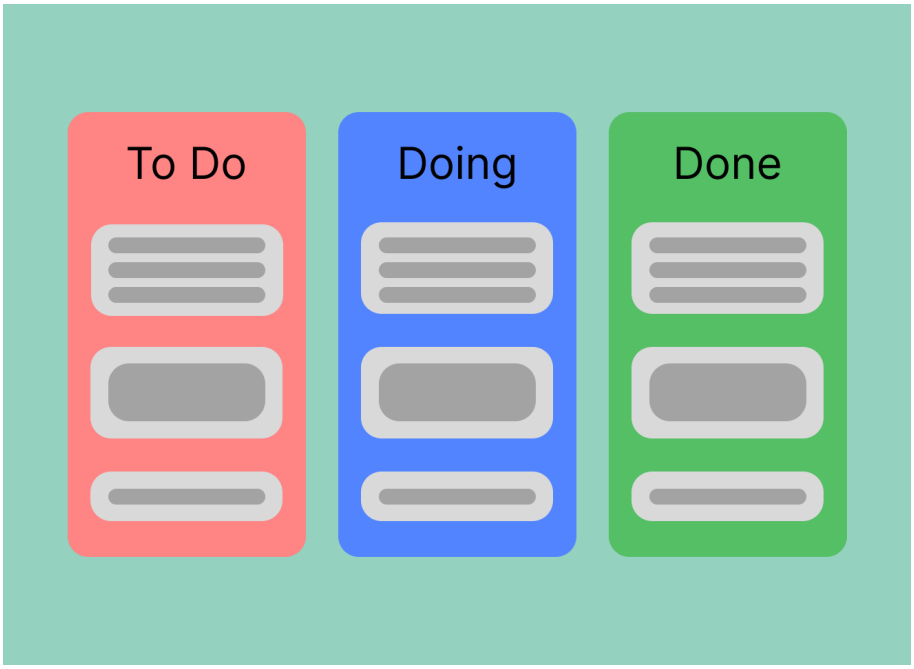


Fig. 1 Kanban Board

For effective team collaboration, it would be ideal to integrate collaborative programming and task management features into a single unified solution.

C. *Technologies Applied*

While designing the platform for project sharing, we selected technologies based on their effectiveness and suitability. The frontend is built using Flutter, a framework developed by Google.

It enables the creation of cross-platform applications through the Dart programming language. Dart is an object-oriented language, and Flutter is essentially a library built with Dart [6]. This programming language has a syntax similar to that of Java [7]. Flutter has found applications across various industries. For example, BMW uses it in their **BMW Connected** app, which allows users to control vehicle functions via a mobile device [8].

Dart supports both JIT and AOT compilation. The AOT compiler translates Dart into efficient native code, making Flutter fast, which benefits both developers and users. The JIT compiler enables the hot reloading feature, allowing Flutter to refresh the application in less than one second during development [6].

The backend is implemented using Node.js, a widely-used JavaScript runtime environment. Node.js focuses on simplicity and efficiency, making it an ideal choice for developers who want to quickly build and deploy modern web applications [9]. In Node.js, a web server is managed by a single JavaScript function that handles all incoming client requests. Depending on the request type, it responds with an HTML page or JSON data. Express simplifies this process by providing useful tools and allowing the logic to be divided into smaller parts [10].

III. Designing solution

Before starting the actual implementation of the platform, it was necessary to carefully design its structure, features, and user interface to ensure that the final product would be intuitive, efficient, and meet the needs of its users. The design process focused on creating a balanced solution that combines functionality with simplicity, taking into account the typical workflow of student teams. This chapter describes the proposed user interface, system architecture, and conceptual model, which together form the basis for the subsequent development phase.

A. Designing user interface

The user interface, shown in Figure 2, is going to be designed with a focus on simplicity and intuitive controls, ensuring quick access to the application's main features. Navigation between the three main modules (projects, chats and tasks) will be handled through a side menu. This menu will be hidden by default, but users will be able to display it when needed.

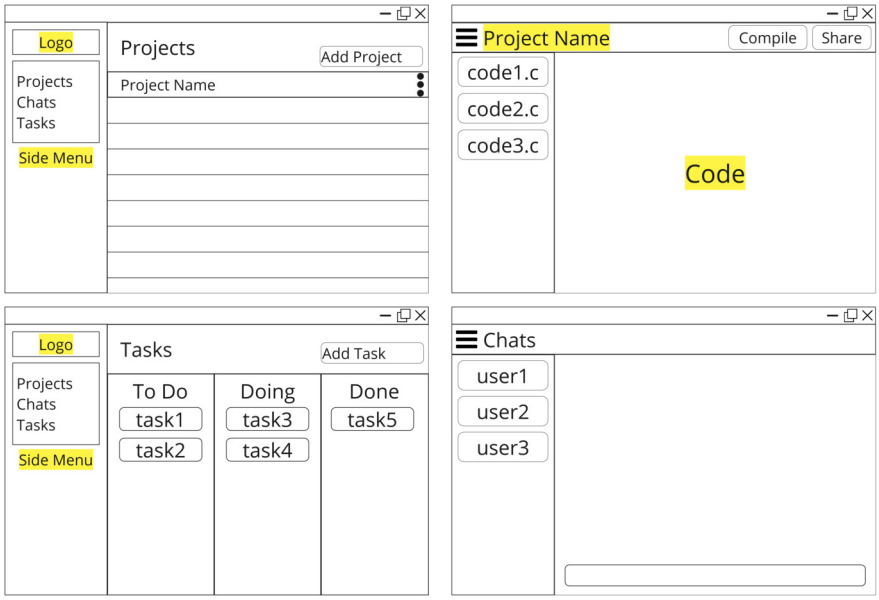


Fig. 2 User interface

In the Projects section, users will be able to create new projects, edit existing ones, share them with team members, and open them in the integrated development environment (IDE). The IDE is going to include a file list and an integrated code editor for writing and editing code in the C programming language. It will also allow users to compile the code and view any errors that occur. The Chats module will be used for communication among team members and will support group conversations. The Tasks section will feature a Kanban system that is going to allow users to visually monitor the progress of their work. Tasks will be organized into three columns based on their current status, and users will be able to move them between columns using a drag-and-drop function.

B. System architecture

The architecture of the application will be based on the client-server model. The client side will be developed using the Flutter framework and will provide the user interface. The server will be implemented in JavaScript frameworks Node.js and Express.js, and will handle client requests, operate a REST API, and manage access to a MySQL database, which will store data about users, projects, tasks, and messages. For real-time communication, such as collaborative programming or chatting, the application will use Socket.IO technology. Code compilation in the C language will be performed on the server using the GCC compiler, which will return output or errors to the users.

The frontend of the application will be organized into a folder structure with separate files for each screen, implemented in the Dart programming language. These screens will interact and communicate with the backend and the database. The application will include files for login, registration, project display, and file manipulation. Each project and its related data will be dynamically fetched from the database. The application structure will also include a main file called `main.dart` and two main directories, one for the application screens and another for smaller widgets that will support specific functionalities.

MySQL database will be used as the central storage for tasks, messages, projects, and files. Projects will represent a key part of the application, so it will be important to store them efficiently along with their structure, including different types of files such as source code and documentation.

C. Conceptual model

The main component of the solution will be the project, which will serve as the central point for managing and organizing student activities. Each student will be able to manage multiple projects, with each project having an assigned creator responsible for its coordination and maintenance. Projects will either be private or shared with other students as needed. They will be structured to support multiple files, making it easy to manage source code, documentation, and other related content. Shared projects will include a group chat, integrated directly into the project, to facilitate effective communication and collaboration among team members. In addition to the group chat for shared projects, the system will also support individual communication between students, enabling them to send direct messages to their collaborators.

Each project will also include a task management module, allowing students to create, assign, and track tasks with specific deadlines and responsible users. Project files will be editable and shareable. Every student will have their own account containing personal information.

IV. Implementing solution

After completing the design phase, the next step was to transform the proposed concepts into a fully functional application. The implementation was divided into two main parts — the server side and the client side — each developed using technologies chosen for their flexibility, scalability, and suitability for real-time applications. This chapter provides a detailed description of how the individual components of the platform were implemented, what changes were made during development, and how specific technical challenges were addressed to ensure the system operates reliably and efficiently.

A. Server side

The implementation of the server-side part of the application was designed to efficiently process user requests and enable seamless real-time collaboration. For this purpose, a MySQL database is used to store and retrieve data related to users, projects, and other entities. To handle communication with the database, we used the `mysql2` library. For storing file content, we used the LONGBLOB data type, which allows for secure storage of large amounts of binary data, including source code and documents.

Each project is stored in the database as a separate record containing an identifier, name, and a reference to its creator. Files associated with projects are stored in a separate table, where each record includes a file identifier, project identifier, file name, type, actual content in LONGBLOB format, and information about the user who created the file.

Additional tables were created in the database, such as the users, messages, and tasks tables. The users table contains a unique identifier, name, email address, and a hashed password. The tasks table is linked to specific projects and users, and each task can be in various stages of completion. Tasks also have a deadline assigned.

Real-time communication in the application was implemented using Socket.IO, a tool enabling bidirectional communication between the client and the server. When a user opens a file, the

client is assigned to a specific room, and any changes made to the document are immediately broadcasted to other connected users through the `sendTextUpdate()` function.

The compilation of C code was performed on the server using Node.js together with the GCC compiler. After receiving source code from the client, the code is saved in a temporary file and compiled via the `gcc` command. In case of successful compilation, the user is informed that no errors were detected; otherwise, compilation errors are captured and returned to the user. Communication between client and server occurred through HTTP POST requests, with responses returned as JSON, which the Flutter client processed and displayed.

During user registration, passwords were encrypted with bcrypt using a salt before storage in the database to enhance security. For authentication, JSON Web Tokens (JWT) were utilized. After successful login or registration, a signed token containing user identification data is generated and sent to the client. This token is required for authenticated requests and had a one-hour expiration time to improve security.

B. Client side

The client side of the application was developed using Flutter and includes nine screens, two of which (related to chat) were eventually not used. The `login.dart` and `register.dart` screens are stateless widgets that handle user login and registration through the `login()` and `register()` functions found in `authService.dart`. Stateful widgets include screens like `projects.dart`, which appears after login and shows a list of projects, and `project.dart`, which serves as the main development environment. This screen includes features such as loading file contents and tasks, creating and editing these items, compiling C files, sharing the project, and opening the group chat. It is divided into four main parts: a control panel, a list of project files, a code or documentation editor, and a task list. The group chat uses the `fetchMessages()` and `sendMessage()` functions. Messages are visually aligned depending on whether they were sent by the current user or other team members. Another widget, called `tasks.dart`, is responsible for fetching, editing, and deleting tasks.

C. Implementation Adjustments

During the development of the application, several changes were made. The most significant adjustment was the redesign of the chat system, originally intended to support one-on-one conversations, it was replaced with group chats linked to specific projects, which better supports team collaboration. From a technical perspective, we didn't use Socket.IO for messaging and instead focused on a more simple approach using a MySQL database for message storage, which simplified implementation and ensured that messages remain accessible even after re-logging into the system.

Task management also underwent changes. Instead of the originally planned Kanban board with drag-and-drop functionality, we implemented a simpler mechanism using control arrows for moving tasks between phases. Additionally, a button for editing and deleting tasks was added. Later, task management functionality was integrated directly into the project, as shown in Figure 3, editor, with a sidebar on the right displaying all tasks assigned to the current project. Users can edit, delete, or update the task status directly while working with the code, without having to switch screens.

Other modifications included simplifying the file system—projects now work only with individual files, without support for directory structures. The user interface was enriched with intuitive controls for renaming and deleting files directly within the application, with confirmation prompts to prevent accidental data loss.

V. Conclusion

The main goal of this thesis was to create an application that helps computer science students work together more effectively. Its key feature is the ability to write code together in real time, which we made possible using Socket.IO. Besides the code editor, the app also includes a built-in C code compiler, so users can quickly test their programs. The group chat helps team members stay in touch without needing other communication tools. Tasks are managed using a Kanban board, where users can assign tasks, change their status, and view them directly in the project workspace.

During development, we looked at existing tools to design a solution that fits real user needs. As we built the app, we made several changes to improve how it works and make it more stable. Feedback from users during testing showed us what could be improved, giving us useful ideas for future updates. These improvements will help the platform support teamwork and project sharing even better.

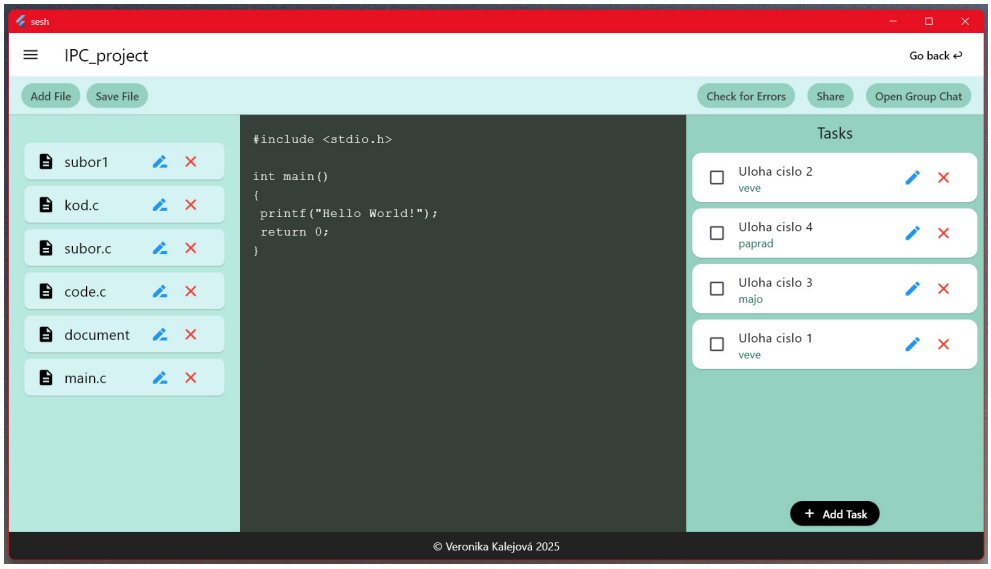


Fig. 3 Project page

A. *Testing the solution*

Testing of the platform involved 12 students with varying technical skills and experience levels, providing diverse perspectives and feedback. To gather user opinions, a standardized System Usability Scale (SUS) questionnaire was used alongside additional open-ended questions aimed at identifying weaknesses, improvement suggestions, and any missing or unclear features. Participants also rated the application overall using a 1-to-5 star scale.

Key feedback highlighted the need to improve the user interface by enlarging chat bubbles, adjusting color contrast, and making control elements clearer and more consistent. Functional suggestions included enabling message sending with the Enter key, automatic login after registration, visual highlighting of overdue tasks, warnings when leaving unsaved files, an interactive deadline calendar, notifications for newly assigned tasks, showing compilation results within the app interface, and drag-and-drop task management. Additional ideas involved integrating GitLab, enhancing real-time chat synchronization, adding emojis and multimedia attachments in chat, and offering dark mode, language selection, and code syntax highlighting options.

Overall, users praised the platform's simplicity, and collaborative features such as instant synchronization and Kanban task management. All participants stated they would recommend the app to peers, reflected in a high average SUS score of 80.42 (grade "A"), indicating strong usability and user satisfaction. This feedback confirms that the application effectively meets the needs of its target audience and provides valuable tools for collaborative project work.

B. *Future Development*

In the next phase of development, it would be beneficial to add several features that make the system easier to use and more useful. One important feature is the ability to create folders inside projects and work with different types of files. This would make the platform more similar to popular tools like GitHub or GitLab, which organize projects in folders.

Right now, the editor does not have common features found in many code editors, such as syntax highlighting, automatic indentation, or support for the Tab key. Without these, writing code can be more difficult, especially in languages like Python where indentation matters. Another important improvement would be to allow users to run their code directly inside the app, not just compile it. Adding a simple terminal would help users test their programs in real time. Supporting more programming languages, especially Python, would also make the app useful for more users.

Finally, it would be helpful to see what other users are doing in the document at the same time. For example, colored cursors could show where each team member is working. This would help users follow their teammates more easily.

VI. Summary

The primary objective was to create a unified platform that supports team collaboration, reduces tool fragmentation, and allows users to focus on project development. The development environment offers two main functionalities: editing C source code and creating technical documentation.

Users can write and modify content in real time, with changes instantly visible to all team members. Additionally, the platform enables users to compile code and verify correctness through displayed error messages. Task management is integrated via a clear Kanban board or a project-specific task list, helping users plan their work, track progress, and assign tasks efficiently among team members. A group chat feature further enhances communication, allowing quick interaction without switching to external tools. By combining these components into one cohesive system, the application significantly simplifies collaboration among students or colleagues.

One of the most important recommendations for further development is deploying the application on a remote server. Currently, the app functions only on a single device, which limits real-world use. During testing, running two separate instances on one device was necessary to simulate real-time collaboration, but this is impractical outside of controlled environments. Other suggested improvements include integrating an advanced source code editor with syntax highlighting and extending code processing capabilities to support program execution within the app through an integrated terminal. At present, the system only supports compiling C code.

Despite some adjustments during development, the project successfully met its goals, producing a functional platform for project sharing that supports team collaboration and offers tools for effective group programming. The development process also provided valuable experience working with Flutter and related technologies, detailed further in this work.

References

[1] M. Jureczko, Kajda, and P. Górecki, "Code review effectiveness: an empirical study on selected factors influence," *IET Software*, vol. 14, no. 7, pp. 794–805, 2020. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-sen.2020.0134

[2] I. Graßl and G. Fraser, "Equitable student collaboration in pair programming," in *2024 IEEE/ACM 46th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*, 2024, pp. 274–285.

[3] ——, "The abc of pair programming: Gender-dependent attitude, behavior and code of young learners," in *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*, 2023, pp. 115–127.

[4] C. Pereira, A. Santos, L. Machado, and L. Zaina, "How developers feel about tools: an investigation on software startup professionals experience with virtual kanban boards," in *2022 IEEE/ACM 15th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*, 2022, pp. 1–10.

[5] P. Garcia-Diaz, J. Garcia-Gomez, J. Redoli-Granados, and D. de la Mata-Moya, "Case study: the use of trello for collaborative work in laboratory practice on engineering subjects," in *2021 1st Conference on Online Teaching for Mobile Education (OT4ME)*, 2021, pp. 31–34.

[6] E. Windmill, 2020.

[7] S. Sharma, S. Khare, V. Unival, and S. Verma, "Hybrid development in flutter and its widgits," in *2022 International Conference on Cyber Resilience (ICCR)*, 2022, pp. 1–4.

[8] S. Zindl, "Flutter on windows desktop: a use case based study," p. 58, 2021. [Online]. Available: http://elib.uni-stuttgart.de/handle/11682/11740

[9] P. Teixeira, *Professional Node.js: Building Javascript Based Scalable Software*, ser. ITPro collection. Wiley, 2012. [Online]. Available: https://books.google.sk/books?id=ZH6bpbcrlvYC

[10] E. Hahn, *Express in Action: Writing, building, and testing Node. js applications*. Simon and Schuster, 2016.

Autentizovaná výmena symetrických kľúčov s doprednou ochranou

[1]*Jozef KOVALČÍN,* [2]*Miloš DRUTAROVSKÝ*

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

[1]jozef.kovalcin@student.tuke.sk, [2]milos.drutarovsky@tuke.sk

Abstrakt – **Súčasné protokoly na výmenu kľúčov, založené na asymetrickej kryptografii ako Diffie-Hellman, čelia existenčnej hrozbe zo strany kvantových počítačov implementujúcich Shorov algoritmus. Existujúce post-kvantové alternatívy (lattice-based, code-based, hash-based) síce poskytujú kvantovú odolnosť, ale za cenu významne zvýšených výpočtových nárokov a veľkostí kľúčov, čo ich činí nepraktickými pre zariadenia s obmedzenými zdrojmi.**

Článok prezentuje základnú implementáciu protokolu SAKE (Symmetric-key Authenticated Key Exchange), ktorý dosahuje vzájomnú autentizáciu a doprednú ochranu bez asymetrických kryptografických stavebných blokov, pričom využíva iba symetrické primitíva a hašovacie funkcie vhodné pre vstavané zariadenia s obmedzenými výpočtovými zdrojmi. Implementácia využíva kryptografickú knižnicu Monocypher a demonštruje praktickú aplikáciu v systéme na zabezpečený prenos súborov s automatickou rotáciou kľúčov a mechanizmami resynchronizácie. Experimentálne vyhodnotenie prenosu súborov medzi PC počítačmi poukazuje na porovnateľný výkon s tradičnými Diffie-Hellman riešeniami pri zachovaní post-kvantovej bezpečnosti.

Kľúčové slová – **SAKE protokol, symetrická kryptografia, postkvantová bezpečnosť, dopredná ochrana, synchronizácia kľúčov**

I. Úvod

Rastúca hrozba kvantových počítačov pre súčasné kryptografické systémy výrazne ovplyvňuje výskum post-kvantovo bezpečných alternatív. Diffie-Hellmanov protokol [1], publikovaný v roku 1976, revolučne zmenil moderný prístup k bezpečnej komunikácii umožnením výmeny kryptografických kľúčov cez nezabezpečený kanál bez potreby vopred zdieľaných tajomstiev [2]. Jeho široké praktické využitie v protokoloch ako TLS/SSL [3], IPSec [4], SSH a VPN riešeniach urobilo z neho základný stavebný blok internetovej bezpečnosti. Protokol sa spolieha na výpočtovú náročnosť problému diskrétneho logaritmu v konečných poliach alebo eliptických krivkách, čo zabezpečuje jeho odolnosť voči útokom s využitím klasických počítačov.

Shorov algoritmus [5] dokáže vyriešiť problém diskrétneho logaritmu a faktorizácie celých čísel v polynomiálnom čase na kvantovom počítači, čím sa stávajú všetky protokoly založené na týchto problémoch (RSA, DH, ECDH) zraniteľné. NIST [6] v roku 2016 inicioval proces štandardizácie post-kvantovej kryptografie, ktorý viedol k štandardizácii algoritmov ako Kyber (lattice-based), Dilithium (lattice-based) a SPHINCS+ (hash-based). Tieto riešenia však prinášajú značné praktické výzvy v podobe výrazne zvýšených veľkostí kľúčov a výpočtovej komplexnosti. Napríklad lattice-based schémy ako Kyber-768 vyžaduje verejné kľúče veľkosti 1,2 kB a privátne kľúče 2,4 kB, čo je 18-násobne (verejný) a 75-násobne (privátny) väčšie ako ECC-256, zatiaľ čo code-based schémy ako Classic McEliece majú veľkosti kľúčov až 1,3 MB. Lattice operácie sú navyše 5-10× pomalšie ako ECC operácie, čo spolu so zvýšenou energetickou náročnosťou predstavuje kritické obmedzenia pre IoT zariadenia s batériovým napájaním.

Zatiaľ čo asymetrická kryptografia čelí potenciálnym zraniteľnostiam zo Shorovho algoritmu [5], symetrické kryptografické stavebné bloky zostávajú prirodzene odolné voči známym kvantovým útokom s vhodnými veľkosťami kľúčov. Groverov algoritmus [7] poskytuje iba kvadratické zrýchlenie pre útoky na symetrické šifry. Symetrické protokoly na

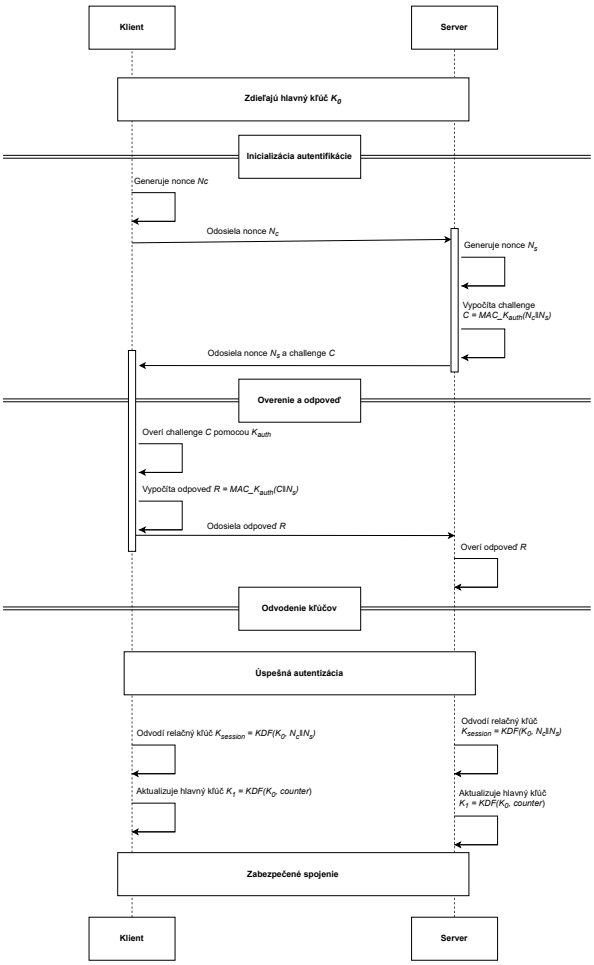
výmenu kľúčov predstavujú alternatívny prístup, ktorý sa spolieha výlučne na hašovacie funkcie a symetrické šifry. Protokol SAKE (Symmetric-key Authenticated Key Exchange) [8] predstavuje inovatívny prístup k dosiahnutiu doprednej ochrany a vzájomnej autentizácie bez asymetrických stavebných blokov, pričom využíva len hašovacie funkcie a symetrické primitíva, čím poskytuje prirodzenú post-kvantovú bezpečnosť bez výpočtových nákladov spojených s asymetrickými operáciami.

Základnou výzvou v symetrických protokoloch je *problém synchronizácie*. Keď strany aktualizujú svoje zdieľané kľúče na dosiahnutie doprednej ochrany, zlyhania siete môžu spôsobiť desynchronizáciu, čím zabránia ďalšej komunikácii. SAKE to rieši elegantnou technikou resynchronizácie pomocou duálnych reťazcov kľúčov, ktoré ohraničujú synchronizačnú medzeru a umožňujú automatickú obnovu spojenia.

Článok je organizovaný nasledovne: Sekcia 2 opisuje formálny model protokolu SAKE. Sekcia 3 poskytuje detaily implementácie. Sekcia 4 sa zameriava na bezpečnosť a obmedzenia protokolu. Sekcia 5 prezentuje experimentálne testovanie a výsledky výkonnosti.

II. Protokol SAKE, základné princípy a opis

Protokol SAKE umožňuje dvom stranám s vopred zdieľanými symetrickými kľúčmi vykonať bezpečnú autentizáciu a vytvoriť nové relačné kľúče s doprednou ochranou [9]. Protokol výmeny kľúčov poskytuje doprednú ochranu, ak odhalenie dlhodobých tajných kľúčov neohrozí bezpečnosť predošlých relácií. Hlavnou inováciou je riešenie problému synchronizácie pomocou systému evolúcie kľúčov, ktorý dokáže automaticky obnoviť synchronizáciu aj po výpadkoch siete. Protokol využíva dva typy hlavných kľúčov, ktoré sa nezávisle vyvíjajú. Prvým je odvodzovací kľúč K používaný na výpočet relačných kľúčov pre šifrovanie dát, druhým je autentizačný kľúč K' používaný na autentizáciu správ a sledovanie synchronizácie. Oba kľúče sa vyvíjajú pomocou jednosmernej funkcie $K_{j+1} = \text{BLAKE2b}(K_j \| j \| \text{tag})$, čím sa zabezpečuje dopredná ochrana. Základný princíp protokolu je znázornený v algoritme 1 a jeho zjednodušený priebeh je ilustrovaný na obrázku 1.



Obr. 1 Vzájomná autentizácia v implementácii protokolu SAKE

Algoritmus 1 Fázy protokolu SAKE

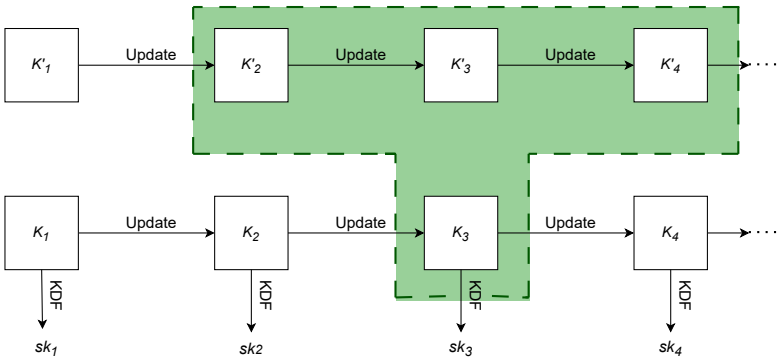
1: Klient a server zdieľajú hlavný kľúč K_0
2: Klient generuje nonce N_c a odosiela ho serveru
3: Server generuje nonce N_s a challenge $C = \mathrm{MAC}_{K_{auth}}(N_c \| N_s)$
4: Server odosiela nonce N_s a challenge C klientovi
5: Klient overí challenge a vypočíta odpoveď $R = \mathrm{MAC}_{K_{auth}}(C \| N_s)$
6: Klient odosiela odpoveď R serveru
7: Server overí odpoveď R
8: Obaja odvodzujú reláčny kľúč $K_{session} = \mathrm{KDF}(K_0, N_c \| N_s)$
9: Obaja aktualizujú hlavný kľúč $K_1 = \mathrm{KDF}(K_0, \mathrm{counter})$

Riešenie problému synchronizácie je založené na nerovnomernom udržiavaní autentizačných kľúčov medzi účastníkmi protokolu. Iniciátor udržiava trojicu autentizačných kľúčov K'_{j-1}, K'_j, K'_{j+1} reprezentujúcich predchádzajúci, aktuálny a nasledujúci kľúč, zatiaľ čo respondent má iba aktuálny autentizačný kľúč K'_j. Táto nerovnosť umožňuje iniciátorovi detegovať synchronizačnú medzeru pomocou synchronizačného parametra $\delta_{AB} \in \{-1, 0, 1\}$ [8], ktorý označuje vzájomnú pozíciu strán v evolúcii kľúčov.

Obrázok 2 ilustruje kľúčovú inováciu protokolu SAKE – asymetrické udržiavanie autentizačných kľúčov medzi iniciátorom a respondentom.

Protokol SAKE pozostáva z maximálne piatich správ organizovaných do troch logických fáz [8]. V prvej fáze si iniciátor A a respondent B vymieňajú náhodné čísla (nonces). Iniciátor odošle svoj identifikátor a náhodné číslo r_A, na čo respondent odpovedá vlastným náhodným číslom r_B spolu s autentizačným kódom τ_B.

Druhá fáza je venovaná detekcii synchronizácie, kde iniciátor overí autentizačný kód respondera pomocou všetkých troch kľúčov a na základe toho určí synchronizačnú medzeru.



Obr. 2 Architektúra reťazca kľúčov SAKE s evolúciou autentizačných kľúčov K'_{j-1}, K'_j, K'_{j+1}

Následne odošle synchronizačný bit ϵ a vlastný autentizačný kód τ_A.

Tretia fáza zahŕňa potvrdenie a dokončenie protokolu. Respondent odošle potvrdenie τ'_B a v prípade potreby (ak $\epsilon = 1$) iniciátor odošle dodatočné potvrdenie τ'_A. Po úspešnom dokončení si obe strany odvodia rovnaký relačný kľúč pomocou funkcie na odvodenie kľúča (KDF), $sk = \mathrm{KDF}(K, r_A \| r_B)$.

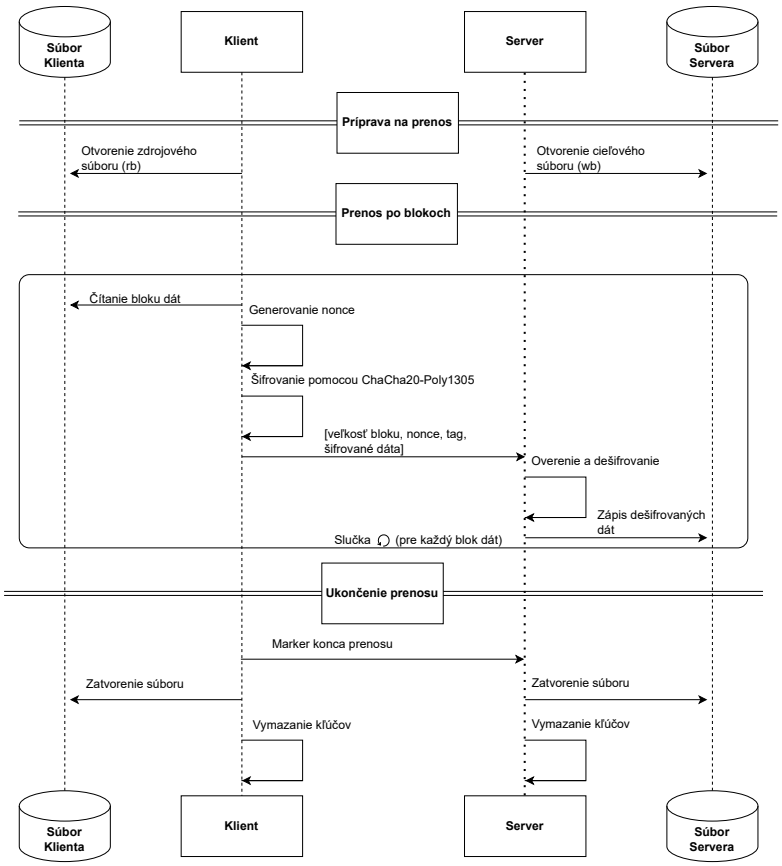
Mechanizmus resynchronizácie funguje tak, že iniciátor overí autentizačný kód od respondera postupne pomocou všetkých troch kľúčov K'_{j-1}, K'_j, K'_{j+1}. Na základe toho, ktorý kľúč úspešne overí správu, určí synchronizačnú medzeru δ_{AB}. Synchronizačný bit ϵ následne informuje respondenta, ako má aktualizovať svoje kľúče, čím sa zabezpečí, že po protokole sú obe strany garantovane synchronizované. Tento mechanizmus je kľúčovým prínosom protokolu SAKE, pretože dokáže automaticky obnoviť synchronizáciu aj po sieťových výpadkoch alebo iných poruchách komunikácie.

III. Testovacia implementácia na overenie protokolu

Implementácia protokolu SAKE je navrhnutá v jazyku C s podporou pre operačné systémy Windows a Linux. Využíva kryptografickú knižnicu Monocypher [10] 4.0.2, ktorá poskytuje moderné symetrické kryptografické stavebné bloky s dôrazom na bezpečnosť a výkonnosť. Kompletná implementácia vrátane zdrojového kódu, dokumentácie a užívateľského manuálu je dostupná v GitHub repozitári [11]. Detailný opis implementácie, vrátane architektúry systému, algoritmov a testovacích scenárov, je k dispozícii v [12].

Základom implementácie je hašovacia funkcia BLAKE2b [13], ktorá slúži na evolúciu kľúčov s 256-bitovou výstupnou dĺžkou, odvodzovanie autentizačných kľúčov a implementáciu challenge-response protokolu. Pre šifrovanie súborov je použitá symetrická AEAD [14] šifra ChaCha20-Poly1305 [15], ktorá poskytuje integrovanú autentizáciu a šifrovanie v jednom kroku [16]. Odvodzovanie kľúčov z hesla je implementované pomocou funkcie Argon2i [17] s konfigurovateľnými parametrami pamäte a počtu iterácií. Generovanie bezpečných náhodných čísel je riešené platformovo špecificky – BCrypt API pre Windows a syscall getrandom pre Linux.

Pre prenosy veľkých súborov implementácia obsahuje mechanizmus automatickej rotácie kľúčov, ktorý poskytuje dodatočnú bezpečnosť obmedzením množstva dát šifrovaných jedným kľúčom. Proces rotácie je iniciovaný klientom každých n blokov dát, pričom klient najprv signalizuje serveru začiatok rotácie kľúčov. Server potvrdzuje svoju pripravenosť a následne si obe strany vymieňajú nové nonce hodnoty. Z týchto nových hodnôt sa odvodí nový relačný kľúč z aktuálneho hlavného kľúča. Validácia synchronizácie prebieha pomocou výmeny validačných kódov medzi stranami, čo zabezpečuje, že obe strany majú rovnaký odvodený dočasný kľúč. Po úspešnom potvrdení validácie pokračuje prenos súborov s novým relačným kľúčom, čím sa minimalizuje riziko kompromitácie väčšieho množstva dát v prípade úspešného útoku.



Obr. 3 Proces šifrovaného prenosu súboru v klient-server aplikácii (zjednodušená verzia)

Obrázok 3 znázorňuje aplikáciu na bezpečný prenosu súboru s využitím protokolu SAKE. Proces začína počiatočnou autentizáciou a vytvorením relačného kľúča, pokračuje šifrovaným prenosom dát s periodickou rotáciou kľúčov pre zvýšenú bezpečnosť a končí sa úspešným dokončením prenosu s verifikáciou integrity.

Implementácia zahŕňa niekoľko dôležitých bezpečnostných mechanizmov zameraných na praktické aspekty nasadenia. Všetky funkcie porovnávajúce citlivé údaje sú implementované v konštantnom čase, čím sa zabraňuje časovým útokom (timing attacks). Citlivé dáta sú bezpečne vymazané z pamäte pomocou `volatile` prístupu, ktorý zabezpečuje, že kompilátor neoptimalizuje tieto operácie.

Validácia všetkých sieťových vstupov je implementovaná s timeout mechanizmami, ktoré zabezpečujú odolnosť voči distribuovaným útokom odmietnutia služby (DDOS). Ošetrenie spracovania chýb zahŕňa automatické opakovanie (retry mechanism) pre sieťové operácie pri výpadkoch komunikácie. Oddelenie domén pri odvodzovaní kľúčov je dosiahnuté pomocou špecifických kódov (tagov), čo zabezpečuje, že kľúče odvodené pre rôzne účely sú kryptograficky nezávislé.

Architektúra implementácie umožňuje jednoduchú integráciu do existujúcich systémov a poskytuje jasné API pre základné operácie protokolu SAKE vrátane inicializácie, vykonania protokolu a správy kľúčov.

IV. Bezpečnosť, odolnosť voči útokom a obmedzenia

Protokol SAKE poskytuje kľúčovú vlastnosť ohraničenej synchronizácie, čo znamená, že ak sa počiatočná synchronizácia medzi stranami líši maximálne o jeden krok evolúcie kľúčov, potom každá úspešná komunikačná relácia vždy vedie k úplnej synchronizácii oboch strán.

Táto vlastnosť funguje v troch základných scenároch. V prvom prípade, keď sú strany už synchronizované, obe používajú rovnaký aktuálny dočasný kľúč pre MAC operácie a po úspešnej relácii zostávajú synchronizované. V druhom prípade, keď je iniciátor o jeden krok vpredu v evolúcii kľúčov, respondent má starší kľúč, ktorý iniciátor stále udržiava vo svojej trojici kľúčov. Protokol deteguje túto situáciu a respondent vykoná aktualizáciu

kľúča, čím sa dosiahne synchronizácia. V treťom prípade, keď je iniciátor o jeden krok vzadu, respondent má novší kľúč, ktorý iniciátor už má pripravený v svojej trojici. Iniciátor vykoná aktualizáciu a dosiahne sa synchronizácia.

SAKE dosahuje doprednú ochranu prostredníctvom jednosmernej evolúcie kľúčov založenej na kryptograficky bezpečnej hašovacej funkcii BLAKE2b. Tento mechanizmus zabezpečuje tri kľúčové vlastnosti bezpečnosti. Predchádzajúce kľúče nemožno vypočítať z aktuálnych kľúčov kvôli jednosmernej vlastnosti evolučnej funkcie, čo znamená, že aj pri kompromitácii súčasného stavu nie je možné obnoviť minulé kľúče. Relačné kľúče sa odvodzujú z vyvinutých hlavných kľúčov namiesto zo statických dlhodobých kľúčov, čím sa zabezpečuje, že každá relácia má jedinečný kryptografický materiál. Autentizačné kľúče sa vyvíjajú nezávisle od odvodzovacích kľúčov, čo poskytuje dodatočnú vrstvu ochrany proti rôznym typom útokov.

Ak protivník kompromituje stranu po nadviazaní spojenia, nemôže vypočítať predchádzajúce relačné kľúče, pretože tieto boli odvodené z predtým vyvinutých hlavných kľúčov, ktoré sú už nedostupné. Relačný kľúč pre každú epochu je vypočítaný z príslušného hlavného kľúča a dočasných nonce hodnôt, pričom inverzia evolučnej funkcie na získanie starších kľúčov nie je výpočtovo realizovateľná. Nonce hodnoty sú navyše dočasné a bezpečne vymazané po skončení relácie.

Spoliehajúc sa výlučne na symetrické stavebné bloky a hašovacie funkcie, SAKE poskytuje prirodzenú odolnosť voči kvantovým útokom. Hašovacie funkcie ako BLAKE2b [13] sú voči Groverovmu algoritmu [7] odolné, s tým, že poskytujú iba kvadratické zrýchlenie, vyžadujúce stále exponenciálny počet operácií pre praktický útok. Symetrické šifry majú podobnú odolnosť voči kvantovým útokom pri vhodných veľkostiach kľúčov. Protokol je úplne imúnny voči Shorovmu algoritmu [5], ktorý narúša protokoly založené na diskrétnych logaritmoch a faktorizácii, pretože nevyužíva žiadne asymetrické kryptografické operácie.

Implementácia demonštruje odolnosť voči bežným kryptografickým útokom. Replay útoky sú znemožnené jedinečnými nonces v každej relácii a nepretržitou evolúciou kľúčov. Man-in-the-Middle útoky sú detegované vďaka požiadavke na vopred zdieľané kľúče a MAC overeniu všetkých správ. Timing útoky sú eliminované použitím funkcií s konštantným časom vykonania. Pasívne odpočúvanie je neefektívne vďaka doprednej ochrane, ktorá chráni minulé komunikácie aj pri budúcej kompromitácii.

Protokol má však určité inherentné obmedzenia vyplývajúce z vlastností symetrických protokolov. Je zraniteľný voči Key Compromise Impersonation (KCI) útokom [18] na rozdiel od Diffie-Hellmanovho protokolu s jednorazovými kľúčmi kvôli závislosti medzi aktualizovanými hlavnými kľúčmi. Nepodporuje paralelné vykonávanie relácií kvôli zdieľaným vyvíjajúcim sa symetrickým kľúčom, čo môže byť limitujúce v niektorých aplikáciách. Škálovateľnosť je obmedzená požiadavkou na vopred zdieľané symetrické kľúče medzi každým párom komunikujúcich strán.

V porovnaní s Diffie-Hellmanom [1] ponúka SAKE významné výhody v podobe postkvantovej bezpečnosti, ľahších výpočtov a vhodnosti pre zariadenia s obmedzenými zdrojmi. Nevýhody zahŕňajú absenciu podpory paralelných relácií, zraniteľnosť voči KCI útokom a požiadavku na vopred zdieľané kľúče, čo môže komplikovať nasadenie v určitých scenároch.

V. Analýza výkonnosti

Testovanie výkonnosti bolo vykonané na nasledujúcej konfigurácii:

- **Hardvér:** Osobný počítač – Intel Core i5-12500H, 16 GB RAM; Osobný počítač – Intel Core i7-8700K, 32 GB RAM; Osobný počítač – AMD Ryzen 7 7800X3D, 32 GB RAM.
- **OS:** Windows 11 24H2, Ubuntu Linux 20.04 LTS, Kali Linux 2025.1a.
- **Kompilátory:** : MinGW-w64 (GCC) 13.2.0 pre Windows, GCC 14.2 pre Linux.
- **Sieť:** Ethernet, 1 Gbps.
- **Testovací súbor:** 4,6 GB obraz (.iso) operačného systému Windows.

Testovanie prebiehalo medzi dvoma samostatnými počítačmi, pričom jeden z nich využíval operačný systém Windows a druhý operačný systém Linux. Takto navrhnuté testovacie prostredie umožnilo overenie multiplatformovej kompatibility implementovaného riešenia. Každé meranie bolo vykonané trikrát za identických podmienok, pričom v tabuľke 1 sú uvedené priemerné hodnoty získané z týchto opakovaných meraní.

Tabuľka 1
Výsledky testu prenosu veľkého súboru (4,6 GB)

Metrika	**DH implementácia**	**SAKE implementácia**
Veľkosť súboru	4,6 GB	
Veľkosť bloku	4096 bajtov	
Frekvencia rotácie kľúčov	Každých 1024 blokov (4 MB)	
Čas prenosu (Server)	343,0 s	343,6 s (+0,2%)
Čas prenosu (Klient)	342,8 s	343,3 s (+0,1%)
CPU čas (Server)	26,6 s	27,6 s (+3,8%)
CPU čas (Klient)	49,5 s	49,8 s (+0,6%)
Využitie CPU (Server)	7,8%	8,0%
Max. pamäť (Server)	68,8 MB	68,8 MB
Max. pamäť (Klient)	68,9 MB	68,9 MB
Priepustnosť	13,7 MB/s	
Stabilita	Stabilné	

Výsledky uvedené v tabuľke 1 demonštrujú na PC platforme porovnateľný výkon implementácie SAKE s tradičným Diffie-Hellman prístupom. Minimálny rozdiel v čase prenosu (0,6 sekundy) a CPU využitia (0,2%) potvrdzuje, že symetrické operácie v SAKE neprispievajú k významnému výkonnostnému zaťaženiu v porovnaní s asymetrickými operáciami DH protokolu. Prakticky identické využitie pamäte (68,8 MB) naznačuje efektívnu implementáciu bez nadmernej spotreby systémových zdrojov.

Dôležité je poznamenať, že tieto výsledky reprezentujú iba vybranú časť testovania implementácie. Kompletná analýza výkonnosti a bezpečnostné testy v rôznych útokových scénároch je uvedená v [12], ktorá obsahuje experimentálne vyhodnotenie a analýzu implementácie.

VI. Záver

Článok prezentuje implementáciu protokolu SAKE využívajúcu kryptografickú knižnicu Monocypher, ktorá demonštruje jeho použiteľnosť na aplikácii pre zabezpečený prenos súborov. Implementácia úspešne dosahuje doprednú ochranu pomocou evolúcie kľúčov založenej na hašovacej funkcii BLAKE2b a poskytuje automatickú rotáciu kľúčov pre dlhodobé prenosy.

Kľúčové vlastnosti implementácie zahŕňajú upravený pôvodný 5-správový protokol na protokol výzvy a odpovede pre autentizáciu, cross-platform podporu pre Windows a Linux, detekciu a ošetrenie chybových stavov vrátane resynchronizácie. Aj keď implementácia využíva upravený autentizačný protokol, zachováva kľúčové bezpečnostné vlastnosti SAKE vrátane doprednej ochrany a post-kvantovej bezpečnosti.

Pri návrhu implementácie sa prihliadalo aj na požiadavky typické pre zariadenia Internetu vecí, ako sú nízke výpočtové nároky, obmedzené pamäťové zdroje či potreba energetickej efektívnosti. Tieto vlastnosti predurčujú riešenie na potenciálne nasadenie aj v takýchto podmienkach, pričom pričom projekt vytvára základ pre následné testovanie a optimalizáciu v týchto podmienkach. Testovaniu na báze IoT zariadení sa budeme venovať v našom ďalšom výskume.

Experimentálne vyhodnotenie na PC platfoprme potvrdilo porovnateľný výkon s tradičnými Diffie-Hellman [1] riešeniami s minimálnym dodatočným zaťažením. Výsledky naznačujú, že SAKE môže poskytovať post-kvantovú bezpečnosť bez kompromisu vo výkonnosti, čo je dôležité pre zariadenia s obmedzenými zdrojmi.

Implementácia však obsahuje určité obmedzenia, ktorých riešenie by mohlo viesť k širšiemu praktickému využitiu symetrických protokolov v post-kvantovom prostredí. Identifikované oblasti pre zlepšenie zahŕňajú: hybridný prístup kombinujúci symetrické protokoly s post-kvantovým vytváraním počiatočných kľúčov, rozšírenie na skupinovú komunikáciu s viacerými účastníkmi, riešenie zraniteľnosti voči KCI útokom pri zachovaní výhod čisto symetrických operácií, a ďalšiu optimalizáciu pre zariadenia s obmedzenými zdrojmi.

Odstránením týchto obmedzení by sa symetrické protokoly mohli stať lákavou alternatívou k post-kvantovým asymetrickým schémam, najmä v scenároch kde sú kľúčové nízke výpočtové nároky a energetická efektívnosť. Výsledky naznačujú, že symetrické protokoly ako SAKE môžu poskytovať post-kvantovú bezpečnosť zariadeniam, najmä v prostrediach s obmedzenými zdrojmi.

Poďakovanie

Táto práca bola podporená EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu c. 09I05-03-V02-00019 – „Kľúčové digitálne technológie v oblasti bezdrôtových širokopásmových odolných prenosových sietí“.

Literatúra

[1] W. Diffie and M. Hellman, “New directions in cryptography,” *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976. [Online]. Dostupné na: https://doi.org/10.1109/TIT.1976.1055638

[2] M. Bellare, A. Desai, E. Jokipii, and P. Rogaway, “A concrete security treatment of symmetric encryption,” in *38th Annual Symposium on Foundations of Computer Science*, 1997, pp. 394–403. [Online]. Dostupné na: https://doi.org/10.1109/SFCS.1997.646128

[3] E. Rescorla, “The transport layer security (tls) protocol version 1.3,” RFC 8446, 2018. [Online]. Dostupné na: https://doi.org/10.17487/RFC8446

[4] S. Kent and K. Seo, “Security architecture for the internet protocol,” RFC 4301, 2005, internet Engineering Task Force (IETF). [Online]. Dostupné na: https://doi.org/10.17487/RFC4301

[5] P. W. Shor, “Algorithms for quantum computation: discrete logarithms and factoring,” in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 124–134. [Online]. Dostupné na: https://doi.org/10.1109/SFCS.1994.365700

[6] NIST, “Post-quantum cryptography standardization,” National Institute of Standards and Technology, 2016–2022. [Online]. Dostupné na: https://csrc.nist.gov/projects/post-quantum-cryptography

[7] L. K. Grover, “A fast quantum mechanical algorithm for database search,” in *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, 1996, pp. 212–219. [Online]. Dostupné na: https://doi.org/10.1145/237814.237866

[8] G. Avoine, S. Canard, and L. Ferreira, “Symmetric-key authenticated key exchange (sake) with perfect forward secrecy,” IACR Cryptology ePrint Archive, 2022. [Online]. Dostupné na: https://eprint.iacr.org/2019/444

[9] M. Bellare and B. S. Yee, “Forward-security in private-key cryptography,” in *Topics in Cryptology – CT-RSA 2003*, 2003, pp. 1–18. [Online]. Dostupné na: https://doi.org/10.1007/3-540-36563-X_1

[10] L. Marsenne, “Monocypher: an easy to use, easy to deploy crypto library,” 2017–2023. [Online]. Dostupné na: https://monocypher.org/

[11] J. Kovalčín, “Sake protocol implementation,” GitHub Repository, 2025. [Online]. Dostupné na: https://github.com/JozefKovalcin/SAKE-project

[12] ——, “Ochrana klient-server komunikácie pomocou prúdovej šifry chacha20,” Bakalárska práca, Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025, pp. 16-59.

[13] J.-P. Aumasson, L. Henzen, W. Meier, and R. C.-W. Phan, “Sha-3 proposal blake,” Submission to NIST, 2008.

[14] P. Rogaway, “Authenticated-encryption with associated-data,” in *Proceedings of the 9th ACM conference on Computer and communications security*, 2002, pp. 98–107. [Online]. Dostupné na: https://doi.org/10.1145/586110.586125

[15] D. J. Bernstein, “The poly1305-aes message-authentication code,” in *Fast Software Encryption (FSE 2005)*, ser. Lecture Notes in Computer Science, vol. 3557. Springer, 2005, pp. 32–49. [Online]. Dostupné na: https://doi.org/10.1007/11502760_3

[16] A. Klein, M. Kuehlewind, and M. Edman, “Chacha20-poly1305 cipher suites for tls 1.2,” RFC 7905, 2016. [Online]. Dostupné na: https://doi.org/10.17487/RFC7905

[17] A. Biryukov, D. Dinu, and D. Khovratovich, “Argon2: new generation of memory-hard functions for password hashing and other applications,” in *IEEE European Symposium on Security and Privacy*, 2016, pp. 292–302. [Online]. Dostupné na: https://doi.org/10.1109/EuroSP.2016.31

[18] S. Blake-Wilson, D. Johnson, and A. Menezes, “Key agreement protocols and their security analysis,” in *6th IMA International Conference on Cryptography and Coding*, 1997, pp. 30–45. [Online]. Dostupné na: https://doi.org/10.1007/BFb0024447

Automated discovery of topology structure in private network environments

[1]*Erika Abigail KATONOVÁ,* [2]*Gabriel STRELA,* [3]*Rastislav PETIJA,* [4]*Peter FECIĽAK*

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

[1]erika.abigail.katonova@tuke.sk, [2]gabriel.strela@student.tuke.sk, [3]rastislav.petija@cnl.sk, [4]peter.fecilak@cnl.sk

***Abstract* — Modern network infrastructures face challenges in topology management due to increasing complexity and the proliferation of connected devices. This paper presents an automated network topology discovery system designed for private network environments using standard network protocols. The proposed solution uses Cisco Discovery Protocol (CDP), Link Layer Discovery Protocol (LLDP), and Secure Shell (SSH) to create comprehensive network maps. Implemented on a cost-effective Raspberry Pi platform, the system employs a modular architecture consisting of network discovery, device discovery, and data processing modules. The solution uses recursive discovery algorithms starting from network gateways to map device interconnections and routing relationships. Experimental validation demonstrates topology mapping across multiple network segments, identifying device types, interface configurations, and routing information. The system generates structured JSON outputs representing both physical and logical network topologies, enabling integration with existing network systems. Results demonstrate discovery of network environments while maintaining low operational costs compared to commercial solutions.**

***Keywords* — automated network mapping, CDP, JSON, LLDP, Raspberry Pi, network topology, SSH**

I. Introduction

In today's digitally connected world, organizations face challenges in managing and securing their network infrastructures. The increase in connected devices has reached large numbers, with Cisco estimating there will be over 75 billion networked devices by 2025, compared to approximately 31 billion in 2020 [1]. This rapid growth, driven by the rise of the Internet of Things (IoT) adoption, cloud computing migration, and the shift to remote work, has significantly increased the complexity of enterprise networks.

As networks evolve, traditional methods of documenting network topologies manually have become outdated and impractical. Network configurations change frequently as organizations adapt to evolving business requirements, making static documentation unreliable. For example, IBM's Security X-Force reported a 37% rise in cyberattacks targeting network infrastructure in 2023 [2]. A These vulnerabilities allow attackers to exploit undetected gaps, often maintaining a foothold in networks for extended periods.

While there are many tools available to address these challenges, current network management solutions often fall short for organizations with limited resources. Commercial platforms like SolarWinds, PRTG, and Cisco Prime Infrastructure offer comprehensive topology discovery capabilities but come with high costs, including licensing fees, specialized hardware, and the need for skilled staff. These requirements make them inaccessible for small and medium-sized enterprises.

On the other hand, open-source alternatives such as Nagios, LibreNMS, and OpenNMS provide cost-effective options but require high technical expertise for setup and maintenance. This complexity presents another barrier, leaving many organizations without the visibility they need to manage their networks effectively.

This gap between expensive commercial solutions and complex open-source tools creates a need for an accessible, affordable network discovery system. Small and medium organizations require solutions that are easy to deploy, manage, and integrate into their workflows, without

sacrificing functionality. This paper therefore introduces an automated network topology discovery system designed to address these challenges in private network environments. The solution uses the set of standard network protocols - Cisco Discovery Protocol (CDP), Link Layer Discovery Protocol (LLDP), and Secure Shell (SSH) - to create comprehensive network topology maps without requiring expensive hardware or advanced technical expertise.

The system is implemented on the Raspberry Pi platform, demonstrating how single-board computers can deliver enterprise-grade network discovery capabilities at a fraction of the cost of traditional tools.

The key contributions of this research include:

- **Affordable discovery platform:** A complete topology discovery solution deployable on hardware under 100€, reducing infrastructure costs by a lot compared to commercial alternatives.
- **Multi-protocol integration strategy:** A combination of CDP, LLDP, and SSH protocols that maximizes device coverage across network environments while providing device information.
- **Recursive discovery algorithm:** An intelligent exploration strategy that maps network topologies by following routing relationships and neighbor advertisements
- **Dual-representation topology model:** Simultaneous generation of physical and logical network views providing insights into both infrastructure and network configurations.
- **Standardized output format:** JSON-based topology representation enabling seamless integration with existing network management workflows and tools

The proposed system offers a practical alternative to expensive commercial solutions and complexity of open-source tools, as well as provides network discovery on an accessible and affordable platform. Experimental results demonstrate system's effectiveness in topology mapping across multi-segment networks including device identification, interface mapping, and routing relationship discovery.

This work contributes to the standardizing of network management technology by proving that sophisticated topology discovery capabilities can be delivered through commodity hardware and standard protocols, making network visibility more accessible.

This paper is structured as follows Section II surveys related work in network topology discovery and analyzes existing solution limitations. Section III details the system architecture and design principles. Section IV evaluates system performance through comprehensive experimental analysis. Section V discusses implications and future research directions, while Section VI concludes and summarizes this work.

II. Related Work and Background

Network topology discovery has been a focus of extensive research, resulting in a variety of techniques for monitoring and protocol-based discovery methods.

A. Protocol-Based Discovery Approaches

Protocol-based discovery methods use the existing network protocols to identify and map devices within a network. Two primary standards for neighbor discovery in enterprise environments are Cisco Discovery Protocol (CDP) and Link Layer Discovery Protocol (LLDP).

CDP, developed by Cisco Systems, operates at the data link layer and provides device information including capabilities, addresses, and interface details [3]. Despite its proprietary nature, CDP has become deployed in enterprise environments due to its comprehensive data sets and effectiveness in topology construction.

LLDP, standardized as IEEE 802.1AB [4], provides vendor-neutral neighbor discovery capabilities. While offering broader vendor support than CDP, LLDP typically provides less detailed device information compared to CDP. Research by Zhang [5] demonstrated that combining both protocols provides a superior coverage in diverse network environments.

B. Active Discovery Techniques

Active discovery methods use techniques such as traceroute, SNMP polling, and ARP scanning to map network topologies. Techniques like those proposed by Bejerano and Rastogi [6] leverage traceroute to determine the paths packets take across networks. These methods are particularly effective for mapping multi-segment networks and understanding routing paths.

SNMP-based approaches provide detailed insights into network devices, including interface states, traffic statistics, and routing information. Donnet and Friedman [7] demonstrated the utility

of SNMP in large-scale internet topology mapping. However, these methods depend on proper SNMP configuration across all devices, which can be a significant limitation in networks with heterogeneous configurations or stringent security policies.

ARP-based discovery focuses on local area networks (LANs) and uses address resolution to identify devices. While effective within smaller, localized environments, ARP scanning struggles to scale to larger networks or those segmented by routers.

C. Commercial and Open-Source Solutions

Commercial network management platforms such as SolarWinds Network Topology Mapper, PRTG Network Monitor, and Cisco Prime Infrastructure offer comprehensive topology discovery capabilities but require significant investments. These solutions provide feature sets including performance monitoring, alerting, and reporting capabilities.

Open-source alternatives including Nagios, Cacti, and LibreNMS provide cost-effective monitoring solutions but often require substantial technical support for deployment and customization. As highlighted by Kumar [8] the trade-offs between feature completeness and deployment complexity in open-source network management tools.

D. Limitations of Existing Approaches

Despite the availability of both commercial and open-source solutions, there are still significant limitations in current network topology discovery and management approaches, such as:

- **Cost barriers:** High licensing and hardware costs associated with commercial solutions make them inaccessible to resource-constrained organizations.
- **Complexity:** Many solutions require dedicated infrastructure and specialized technical expertise for effective deployment and operation.
- **Protocol limitations:** Single-protocol approaches, such as those relying only on CDP, LLDP, or SNMP, may fail to capture all devices in heterogeneous environments. This can result in incomplete topology maps and blind spots.
- **Scalability concerns:** Some methods struggle to scale effectively in large or distributed networks, limiting their usability in modern, complex network environments.

The proposed system directly addresses these limitations by combining a multi-protocol discovery approach with a focus on simplicity, affordability, and extensibility.

III. System Architecture and Design

The proposed network topology discovery system employs a modular architecture designed to provide network mapping capabilities while maintaining flexibility and scalability. By separating concerns into distinct functional layers, the system ensures independent development, testing, and refinement of each component while maintaining seamless integration across the entire platform.

A. Architecture Overview

The system architecture consists of three primary modules, each responsible for a specific set of functions, working together to deliver comprehensive topology discovery capabilities.

The *Network Interface Module* serves as the foundation layer, responsible for basic network interaction and communication management. This module utilizes the *netifaces* library to collect local network configuration data and manages network connections within the resource constraints of the Raspberry Pi platform. It handles network interface detection, IP address resolution, and gateway identification, providing essential networking bases for higher level discovery operations.

The *Device Discovery Module* implements the core discovery logic using a multi-protocol approach. This module organizes the deployment of CDP and LLDP protocols for neighbor discovery, creates SSH connections for device interrogation, and performs ARP scans to detect devices not supporting standard discovery protocols. The module maintains discovery state, tracks processed devices, and implements retry mechanisms to handle temporary network issues or device unavailability.

The *Data Processing Module* handles the complex task of data transformation, validation, and generation of the output topology. This module converts raw discovery data collected from discovery protocols into structured JSON representations suitable for further processing and integration with network management systems. It performs data normalization, relationship mapping, and consistency validation to ensure the completeness of the generated topology

representations. This module ensures that the output is both technically accurate and ready to be used in broader network management workflows.

B. Multi-Protocol Discovery Strategy

The system implements a hierarchical strategy that maximizes network coverage. The strategy operates through a sequential decision tree.

The first step in the hierarchical strategy involves establishing SSH connections to identified network devices using stored credentials. This connection attempt serves as the primary gateway for device interrogation and system gathers rich, detailed insights, enabling precise topology mapping and enhanced device understanding. This method also allows for querying specific configuration data, making it particularly valuable in multi-vendor environments.

When SSH connectivity is successfully established, the system proceeds with *Protocol-Based Discovery* through the authenticated session. During this phase, CDP and LLDP commands are executed remotely on the target device to gather neighbor information and device capabilities. These protocols operate at the data link layer and can identify adjacent network devices along with interface configurations and management addresses. The SSH session also enables collection of device configuration data including routing tables and interface statistics. This approach allows the system to gather both neighbor discovery data and configuration information through a single authenticated session, maximizing information collection efficiency while minimizing network overhead.

When SSH connectivity cannot be established due to credential unavailability, authentication failures, or device limitations, the system transitions to *Fallback Scanning* mechanisms. This phase uses ARP-based network scanning to discover devices not accessible through authenticated protocols. The fallback approach ensures coverage across network environments. This mechanism proves particularly valuable for identifying end-user devices, IoT equipment, and network hardware from vendors that may not support standard discovery protocols or SSH access. The fallback scanning maintains network visibility even in environments with limited credential access or diverse device ecosystems.

This layered approach ensures comprehensive coverage across heterogeneous network environments while adapting to varying levels of device support for discovery protocols. The strategy maximizes information collection by prioritizing high-information protocols while maintaining fallback mechanisms for complete network coverage.

C. Data Model and Representation

The system employs a graph-based data model using the *NetworkX* library to represent network topology relationships. This model naturally captures the interconnections between network devices and enables the application of graph algorithms for network visualization.

The JSON output format provides dual representations of the discovered network topology to address different analytical and operational requirements. The physical topology representation captures actual device connections including interface information, MAC addresses, physical link properties and connectivity characteristics. This representation enables comprehensive understanding of the actual network infrastructure.

The logical topology representation focuses on routing relationships, network segments and logical connections that manage traffic flow across the network infrastructure. This view provides insights into routing behavior and network segmentation. The logical representation includes routing protocol information, administrative distances, next-hop relationships and network reachability, data essential for network optimization and troubleshooting activities.

Both representations maintain referential integrity through device and interface identifiers, enabling correlation between physical infrastructure elements and their logical functions.

IV. Experimental Results and Evaluation

The system was evaluated in a controlled laboratory environment and virtual environment of GNS3 consisting of routers interconnected, and routing made by OSPF protocol and end devices.

A. Discovery performance

The system successfully discovered and mapped the entire test topology. Discovery results included identification of routing devices with complete interface configurations, enabling comprehensive understanding of the network's routing infrastructure. The system found potential host addresses across host network segments, providing visibility into device address allocation patterns. Inter-router connections were mapped through CDP neighbor discovery, revealing the

physical and logical relationships between routing devices. Additionally routing table information was collected including static routes, connected networks, and dynamic OSPF route advertisements, enabling understanding of traffic flows.

B. Protocol effectiveness

Analysis of protocol utilization revealed distinct strengths and coverage patterns across the implemented discovery methods. CDP provided comprehensive information for Cisco devices including interface specifications that enabled device characterization and relationship mapping. LLDP offered broader vendor compatibility across heterogeneous network environments but with more limited information sets compared to CDP's vendor specific extensions. ARP scanning identified devices not supporting SSH protocol, ensuring network coverage even in environments with limited protocol support.

C. Output quality assessment

The generated JSON topology representations captured both physical and logical network aspects. Physical Topology mapped device interfaces, MAC addresses, and direct connectivity relationships. The output included detailed interface information enabling physical network documentation. Logical Topology identified network segments and routing relationships. The logical representation reflected the OSPF routing domain and inter-network connectivity.

D. Resource utilization

System resource consumption during discovery operations demonstrated efficient utilization of the Raspberry Pi's limited hardware capabilities. Peak memory usage reached only 9% of the available 2GB RAM capacity and indicating substantial headroom for larger network environments. Network bandwidth consumption showed minimal impact on production traffic, ensuring that discovery operations do not interfere with normal network operations. Storage requirements for JSON outputs typically remained under 1MB for the tested topologies, indicating efficient data representation and minimal storage overhead for most deployment scenarios.

V. Discussion and Future Work

The system offers several significant advantages over existing commercial solutions. Cost-effectiveness represents a primary benefit, as deployment on commodity hardware significantly reduces infrastructure costs compared to dedicated corporate solutions. Protocol diversity through the multi-protocol approach ensures broad device compatibility across environments, avoiding the coverage gaps that affect single-protocol solutions. Deployment simplicity minimizes configuration requirements and enables installation without extensive technical expertise or specialized training. Integration capability through standard JSON output enables seamless connection with existing network management workflows and tools, avoiding vendor concerns.

However, several limitations must be acknowledged in the current implementation. The system exhibits dependency on device credentials for comprehensive discovery, which may limit coverage in environments with restricted access policies or inconsistent credential management. Scalability constraints may affect performance in very large network environments due to the sequential nature of the discovery process and hardware limitations of the Raspberry Pi platform. Additionally, potential security issues rise from automated credential management, requiring careful consideration of credential storage and access control policies in production deployments.

Several enhancement opportunities could significantly expand the system's capabilities and applicability. Container deployment through Docker containerization would improve deployment flexibility and scalability by enabling consistent deployment across diverse infrastructure environments and enabling integration with modern orchestration platforms. A comprehensive web interface providing graphical topology visualization and management capabilities would enhance usability for administrators who prefer visual network representations over JSON data structures. Database backend implementation would enable storage for historical analysis, trend detection and change tracking over time. Performance optimization through multi-threading and parallel processing techniques could substantially improve discovery performance for larger network environments while maintaining compatibility with deployment platforms.

VI. Conclusion

This paper introduced an automated network topology discovery system designed for private

network environments, leveraging standard protocols to achieve effective and accessible network mapping. By integrating CDP, LLDP, ARP, and SSH, the system provides comprehensive topology discovery capabilities while maintaining cost efficiency through its implementation on Raspberry Pi hardware.

Experimental validation confirmed the system's effectiveness in discovering and mapping topologies across multiple network segments. Key achievements include accurate identification of device types, detailed interface configurations, and precise routing relationships. The modular architecture not only supports the current functionality but also lays the foundation for future enhancements. Additionally, the use of a JSON output format ensures seamless integration with existing network management systems.

This solution addresses a significant gap for affordable network topology discovery tools, offering a practical alternative to expensive commercial platforms. Future work may focus on extending the system's capabilities to enhance scalability, introduce intuitive graphical interfaces, and expand protocol support.

References

[1] Cisco Systems, "Cisco Annual Internet Report (2018–2023) White Paper," Cisco, San Jose, CA, USA, Tech. Rep., 2020.
[2] IBM Security, "X-Force Threat Intelligence Index 2023," IBM Corporation, Armonk, NY, USA, Tech. Rep., 2023.
[3] Cisco Systems, "Cisco Discovery Protocol Version 2," Cisco Systems, San Jose, CA, USA, 2013.
[4] IEEE Computer Society, "IEEE Standard for Local and metropolitan area networks - Station and Media Access Control Connectivity Discovery," IEEE Std 802.1AB-2016, pp. 1-146, Mar. 2016.
[5] L. Zhang, M. Wang, and J. Liu, "Hybrid topology discovery in heterogeneous network environments,'" IEEE Trans. Network Service Manage., vol. 16, no. 3, pp. 1045-1058, Sep. 2019.
[6] Y. Bejerano and R. Rastogi, "Robust monitoring of link delays and faults in IP networks," IEEE/ACM Trans. Networking, vol. 14, no. 5, pp. 1092-1103, Oct. 2006.
[7] B. Donnet and T. Friedman, "Internet topology discovery: A survey," IEEE Commun. Surveys Tuts., vol. 9, no. 4, pp. 56-69, 4th Quart. 2007.
[8] A. Kumar, S. Patel, and R. Sharma, "Comparative analysis of open-source network monitoring tools," Int. J. Network Manage., vol. 30, no. 2, pp. e2089, Mar. 2020.

Classification of Data Flows Based on Network Traffic Characteristics

[1]*Erika Abigail KATONOVÁ,* [2]*Adam FABIŠÍK,* [3]*Rastislav PETIJA,* [4]*Peter FECIĽAK*

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

[1]erika.abigail.katonova@tuke.sk, [2]adam.fabisik@student.tuke.sk, [3]rastislav.petija@cnl.sk, [4]peter.fecilak@cnl.sk

***Abstract* — This paper presents a method for classifying network traffic using a deep learning model based on Long Short-Term Memory (LSTM) neural networks. The aim of the study is to distinguish between various types of network services, such as HTTP, HTTPS, VPN, VoIP, and others, without relying on deep packet inspection (DPI). Instead, the classification relies on statistical and structural characteristics of the traffic flow, such as IP structure, port ranges, traffic intensity, and temporal patterns. Due to the limitations of existing public datasets, a custom dataset was generated simulating realistic traffic in a segmented enterprise network. The model architecture combines three bidirectional LSTM layers with dropout and batch normalization, followed by dense layers and a softmax output. The model was trained on the simulated dataset and tested on unseen data to evaluate generalization. The proposed approach achieved an average classification accuracy of 88–90% on unseen data, outperforming other tested configurations. The results suggest that flow-level traffic analysis using LSTM is a viable and effective solution for scenarios where DPI is infeasible or undesirable.**

***Keywords* — Deep learning, LSTM, Network traffic classification, Simulation, Time series analysis, Packet flow**

I. Introduction

With the rapid growth of network traffic and the increasing diversity of online services, the ability to accurately classify network flows has become essential for maintaining performance, security, and efficient resource allocation. Traditional methods for traffic classification, such as port-based analysis or Deep Packet Inspection (DPI), are either outdated due to protocol obfuscation and encryption or computationally expensive and privacy-invasive.

Recent advances in machine learning and deep learning have opened new possibilities for flow-based classification, which relies on statistical and behavioral features extracted from network flows rather than payload content. These approaches are particularly attractive in environments where DPI is infeasible or undesirable, such as encrypted traffic scenarios or privacy-sensitive networks.

In this study, we propose a flow-level traffic classification model based on a deep learning architecture using Long Short-Term Memory (LSTM) neural networks. Our goal is to distinguish among multiple types of services (e.g., HTTP, HTTPS, VPN, VoIP, etc.) by analyzing structured features like IP and port patterns, packet lengths, flow durations, and temporal behaviors.

Unlike many previous studies that rely on outdated or insufficient public datasets, we address this limitation by generating a custom dataset that simulates realistic network activity across segmented enterprise zones (internal, DMZ, and external). The proposed model is trained and evaluated on this dataset, demonstrating its ability to generalize and achieve high classification accuracy without requiring access to raw packet contents.

II. Foundation of Research

Traditional approaches to network traffic classification have relied heavily on port-based methods or Deep Packet Inspection (DPI). Port-based techniques attempt to identify services based on well-known port numbers (e.g., 80 for HTTP, 443 for HTTPS), but these are increasingly unreliable due to port obfuscation and dynamic allocation. DPI, while more accurate, requires full access to packet payloads and often violates user privacy or is rendered ineffective by encryption [1].

As a result, researchers have increasingly turned to machine learning (ML) techniques that analyze flow-level metadata [2]. These include statistical methods using features like packet size, duration, or interarrival time, as well as supervised classifiers such as decision trees, support vector machines (SVM), and k-nearest neighbors (k-NN). Although these models have shown moderate success, they struggle with complex temporal dependencies and scale poorly with dynamic traffic types [3].

Deep learning methods, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have recently gained attention for their ability to automatically learn patterns in traffic flows [4]. LSTM networks, a form of RNN, are especially well-suited for sequential data, capturing both short- and long-term dependencies. Several studies like [5] and [6] have demonstrated that LSTM-based models outperform traditional ML classifiers in identifying encrypted traffic or multiplexed protocols.

However, many of these studies rely on outdated public datasets (e.g., KDD Cup 1999, CICIDS2017), which may not reflect current network behavior or service diversity. In contrast, our approach uses a custom-generated dataset that better represents modern enterprise segmentation and traffic patterns, improving model robustness and real-world applicability.

III. Dataset and Feature Engineering

To ensure flexibility, realism, and control over the classification task, we generated a custom network traffic dataset instead of relying on existing public datasets such as KDD Cup 1999 [7] or CICIDS2017 [8]. These older datasets often suffer from a lack of contemporary protocol representation, class imbalance, and outdated attack models. Our custom dataset simulates realistic communication flows in a modern enterprise environment by incorporating time-based behavior, subnet segmentation, and protocol-specific properties.

A. Network Segmentation and Flow Simulation

The synthetic dataset simulates communication across three primary network segments [9]:

- Internal (192.168.x.x) – representing employee machines and internal systems.
- DMZ (172.16.x.x) – public-facing services such as web or mail servers.
- External (10.0.x.x) – external clients or internet-based sources.

This segmentation adheres to RFC 1918 standards [10] for private IP address allocation and reflects common enterprise network architectures.

Traffic flows are generated using a Python script that creates individual records representing one session or connection. Each flow includes the following information: source and destination IP address, ports, protocol (TCP/UDP), service label (e.g., HTTP, HTTPS, VPN), hour of activity, packet rate, flow duration, and derived features such as packet size category or protocol type.

Service usage is influenced by time-of-day patterns. For example, HTTP and HTTPS traffic is more frequent during working hours, while VPN or VoIP may have different usage peaks. This is implemented using weighted random selection based on predefined peak intervals, adding temporal variability to the dataset.

B. Feature Extraction and Encoding

Once the raw flows are generated, several preprocessing steps are applied to prepare the data for training:

- **IP Feature Extraction:** Each IP address is parsed to extract its second-to-last and last octets, normalized to [0,1]. The subnet type (internal, DMZ, external, or other) is encoded as a categorical feature.
- **Port Feature Extraction:** Ports are categorized as well-known (<1024), registered (1024 – 49151), or dynamic (≥49152). A binary flag indicates whether the port is standard for the given service (e.g., 443 for HTTPS).
- **Categorical Encoding:** Fields such as *protocol*, *protocol_category*, and *packet_size_category* are encoded using LabelEncoder from scikit-learn.
- **Numerical Normalization:** Continuous features (flow duration, packet rate, traffic intensity, etc.) are normalized to [0,1] using MinMaxScaler.

All extracted features are combined into a single input vector for each flow. This final vector comprises 3 features derived from the source IP, 3 from the destination IP, 5 features from the source port, 5 from the destination port, 6 numerical features (such as duration, rate, and intensity), and 3 encoded categorical variables.

This results in a 21-dimensional feature vector representing each network flow, capturing structural, behavioral, and temporal characteristics essential for classification.

IV. Model Design and Training

To classify the generated network flows based on their statistical and structural features, we designed a deep learning model using Long Short-Term Memory (LSTM) neural networks. LSTM networks are a subclass of Recurrent Neural Networks (RNNs) that are particularly well-suited for processing sequential or time-dependent data, as they are capable of capturing both short-term fluctuations and long-range dependencies in input sequences [11]. In the context of network traffic analysis, this is crucial for learning patterns that evolve over time within each flow.

A. Architecture Overview

The proposed model consists of multiple layers stacked sequentially to form a deep neural network. The architecture was implemented using the Keras API with a TensorFlow backend. The input to the model is a 21-dimensional feature vector for each flow, which is reshaped to be compatible with LSTM layers.

The architecture consists of several key components designed to effectively process and classify network traffic. It begins with an input reshaping layer that transforms the flat input vector into a sequence format suitable for LSTM processing. This is followed by three bidirectional LSTM layers, which enable the model to capture dependencies in both forward and backward directions, enhancing its understanding of contextual relationships within the input features. Batch normalization layers are incorporated after each LSTM block to stabilize and accelerate the training process by normalizing layer inputs. To mitigate overfitting, dropout layers with a rate of 0.5 are applied, randomly disabling neurons during training. The architecture also includes two fully connected dense layers, with 128 and 64 units respectively, to facilitate abstraction and integration of learned features. Finally, a dense output layer with softmax activation is used to generate class probabilities for the eight predefined traffic categories.

B. Loss Function and Optimization

We used the *sparse_categorical_crossentropy* loss function, which is appropriate for multi-class classification problems where the class labels are represented as integers. The model was trained using the Adam optimizer, which is an adaptive gradient method that combines the advantages of RMSProp and momentum. The initial learning rate was set to 0.0005.

To ensure stable and balanced training, several techniques were applied. Class weighting was used to address class imbalance in the dataset by assigning higher importance to underrepresented classes. A validation split was implemented, reserving 20% of the training set for validation to monitor overfitting during the training process. Additionally, gradient clipping with a clip norm of 1.0 was employed to prevent gradient explosion during backpropagation, ensuring a smoother optimization process.

C. Training Parameters

The model was trained for 150 epochs with a batch size of 64 samples and a step count of 200 per epoch. Each training run was initialized with a fixed random seed (e.g., 42) to ensure reproducibility. During training, metrics such as accuracy and loss were monitored on both the training and validation sets. The training history, including performance curves, was stored and visualized for later evaluation.

D. Implementation Details

The entire model was implemented in Python 3.8 using open-source libraries:

- TensorFlow and Keras for model definition and training.
- NumPy and Pandas for data manipulation.
- Scikit-learn for preprocessing (label encoding, scaling).

The trained model was exported in .keras format for later inference. Additionally, the fitted label encoders and scalers were serialized into a .pkl file to ensure consistent preprocessing during prediction on new data.

V. Evaluation and Results

To evaluate the performance of the proposed LSTM-based classification model, we conducted a series of experiments using the synthetic dataset described earlier. The evaluation focused on classification accuracy, generalization to unseen data, and class-wise performance across various traffic types. Additionally, we analyzed training dynamics and model behavior under different random seeds to assess overfitting risks and stability.

A. *Experimental Setup*

The dataset was split into training and testing subsets using an 80/20 ratio. The training set was used to fit the model parameters, while the test set - unseen during training - was used for final evaluation. The random seed was set to 42 for all core experiments to ensure reproducibility. Additional seeds were used to assess model variance across runs.

Model performance was evaluated using several metrics. Accuracy, representing the proportion of correctly classified flows, served as a primary indicator of overall model effectiveness. A confusion matrix was used to visualize the relationship between true and predicted labels for each class, providing detailed insights into classification performance. Additionally, class-wise accuracy was calculated to assess how effectively the model recognized each specific traffic type.

B. *Training Performance*

The model achieved a training accuracy of approximately 90% and validation accuracy of around 88-89%, is shown in Fig. 1, indicating good generalization. The training loss decreased steadily over the course of 150 epochs, while the validation loss plateaued after approximately 100 epochs. No major signs of overfitting were observed for seed 42.

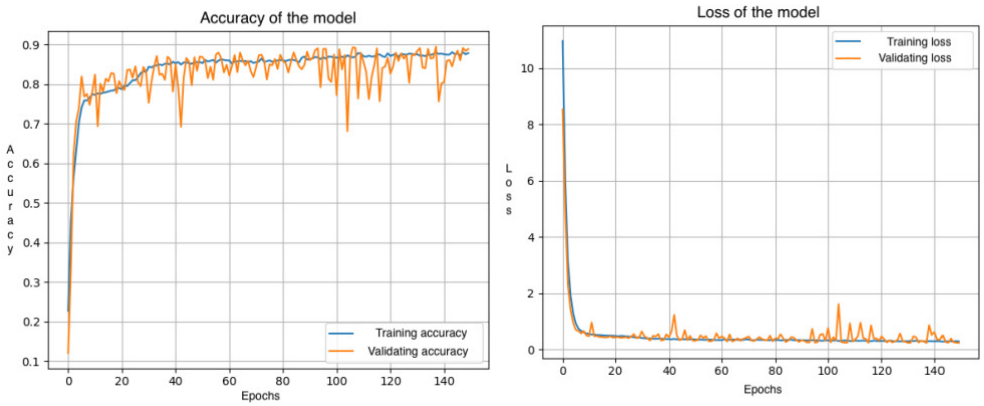


Fig. 1 Training curves: model accuracy (left) and loss (right) across epochs.

Other seed values (e.g., 0, 21, 100) showed stronger signs of overfitting, with training accuracy reaching 89-90% but test accuracy dropping to 50-60%, highlighting the importance of initialization and seed selection.

C. *Confusion Matrix and Class-wise Accuracy*

The confusion matrix for the test set Fig. 2 shows that the model performs well across most classes. The majority of misclassifications occurred between similar services such as SMTP and IMAP, which may exhibit overlapping characteristics.

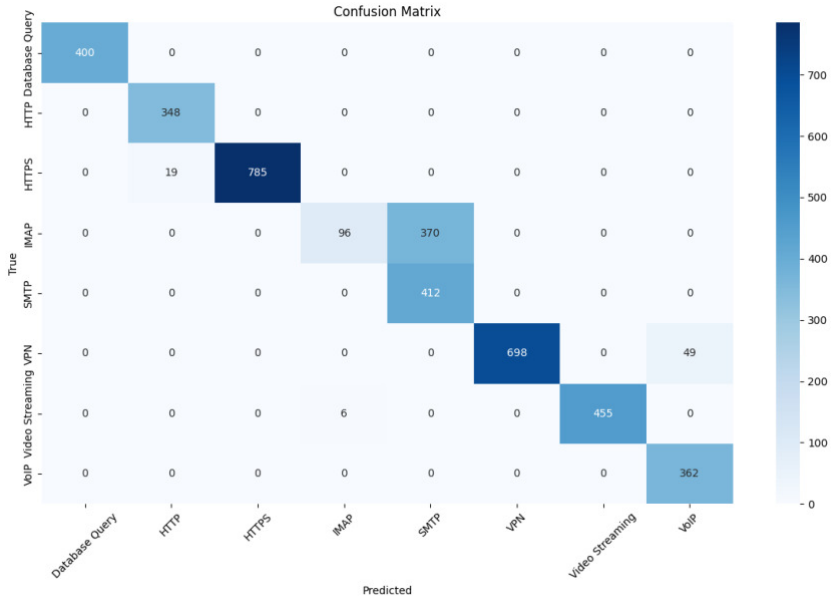


Fig. 2 Confusion matrix on the test set (seed 42).

Table 1 presents the class-wise accuracy scores. Web-based traffic types (HTTP/HTTPS) achieved near-perfect accuracy, while mail-related services and VoIP were less reliable. This is likely due to less distinctive patterns in their flow statistics or higher overlap between classes.

Table 1
Class-wise Accuracy on Test Data (Seed 42)

Traffic type	Accuracy
HTTP	98%
HTTPS	96%
Database Query	93%
Video Streaming	91%
VPN	87%
VoIP	84%
SMTP	79%
IMAP	74%

D. Distribution and Prediction Analysis

To better understand how well the model reproduced traffic distribution, we compared the predicted label distribution with the original distribution in the test set. Visualizations such as pie charts (Fig. 3) and hourly distribution graphs (Fig. 4) demonstrate a strong correlation between real (original) and predicted class frequencies, suggesting good overall alignment.

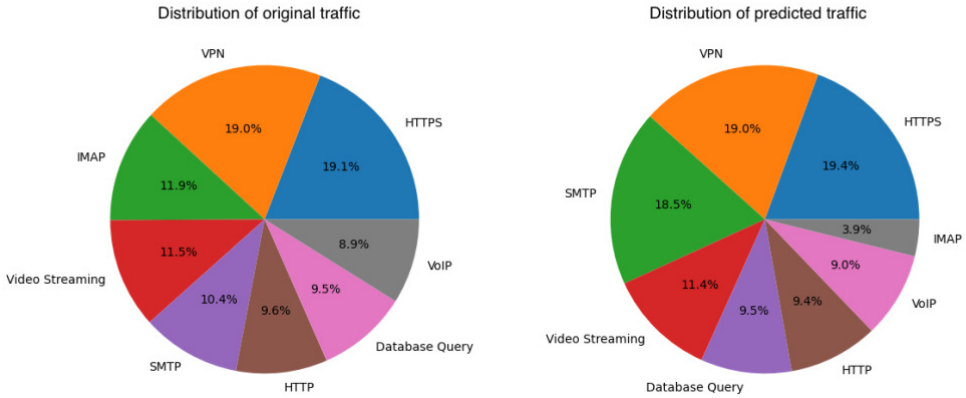


Fig. 3 Pie chart comparison of predicted vs. original label distribution.

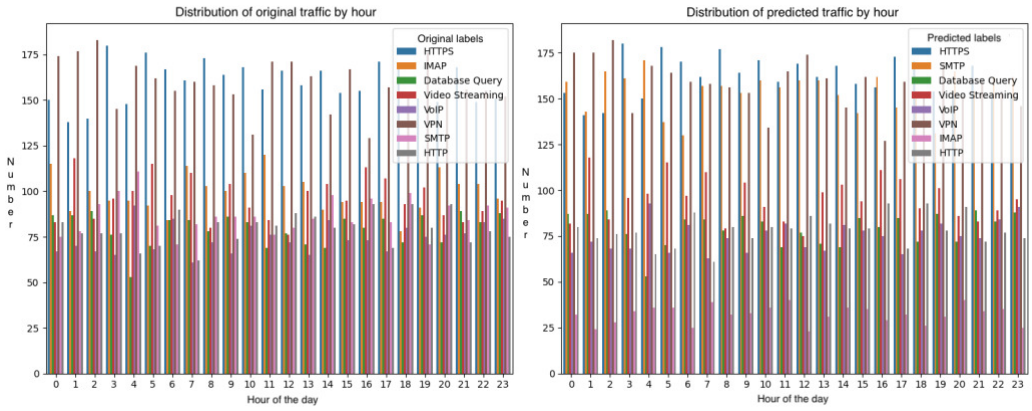


Fig. 4 Prediction accuracy across different hours of the day.

E. Summary of Findings

The model performs robustly on unseen data when initialized with a stable seed (e.g., 42), achieving consistent accuracy across a wide range of traffic classes. It generalizes well to different time intervals and maintains prediction consistency under realistic traffic fluctuations. However, sensitivity to initialization was observed, and services with less distinctive patterns exhibited lower classification reliability.

VI. Conclusion and Future Work

In this paper, we presented a deep learning-based method for network traffic classification that leverages flow-level features and temporal patterns. By combining a custom-generated dataset with a robust LSTM architecture, we demonstrated that accurate traffic classification is achievable without relying on Deep Packet Inspection (DPI) or payload access.

Our model achieved an overall test accuracy of up to 90%, with strong performance across most service classes, particularly web traffic. The use of bidirectional LSTM layers allowed the model to effectively capture time-related dependencies in flow data, while normalization and dropout ensured stable training and good generalization. Additionally, we showed that flow-based traffic analysis can be enhanced with features derived from IP addresses, ports, and protocol categories.

Despite promising results, several limitations remain. First, the dataset was synthetically generated, which, while useful for controlled testing, may not capture the full complexity of real-world traffic. Second, the model showed sensitivity to initialization (random seed), which suggests the need for more robust training strategies or ensemble methods.

Future work will focus on extending the system to operate on real captured traffic, possibly integrating with network monitoring tools or intrusion detection systems. Moreover, incorporating additional contextual features, such as flow timing patterns or metadata from encrypted sessions, may further improve performance. A user-friendly interface or API could also support broader adoption in operational environments.

References

[1] M. S. Sheikh and Y. Peng, “Procedures, Criteria, and Machine Learning Techniques for Network Traffic Classification: A Survey,” 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9791420
[2] F. Shaman, B. Ghita, N. Clarke and A. Alruban, "User Profiling Based on Application-Level Using Network Metadata," 2019 7th International Symposium on Digital Forensics and Security (ISDFS), Barcelos, Portugal, 2019.
[3] S. Jebaraj and M. Joshi, “Port-Based Classification of Network Traffic Analysis using the ColaSoft Capsa Tool for Enhanced Cybersecurity,” 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10420917
[4] P. Dhruv, S. Naskar, “Image Classification Using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN): A Review”, In: Swain, D., Pattnaik, P., Gupta, P. (eds) Machine Learning and Information Processing. Advances in Intelligent Systems and Computing, vol 1101. Springer, Singapore, 2020.
[5] S. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in Proc. IEEE CICSDA, 2009.
[6] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” in Proc. Int. Conf. Inf. Syst. Secur. Privacy (ICISSP), 2018, pp. 108-116.
[7] X. Zeng, "Unmasking Intruders: An In-Depth Analysis of Anomaly Detection Using the KDD Cup 1999 Dataset," 2024 3rd International Conference on Artificial Intelligence and Computer Information Technology (AICIT), Yichang, China, 2024.
[8] Kurniabudi, D. Stiawan, Darmawijoyo, M. Y. Bin Idris, A. M. Bamhdi and R. Budiarto, "CICIDS-2017 Dataset Feature Analysis With Information Gain for Anomaly Detection," in IEEE Access, vol. 8, pp. 132911-132921, 2020.
[9] Kallatsa, Markus, “Strategies for Network Segmentation: A Systematic Literature Review”, 2024.
[10] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear, “Address Allocation for Private Internets,” RFC 1918, 1996. [Online]. Available: https://www.rfc-editor.org/rfc/rfc1918
[11] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

A Predictive System for Network Failure and Congestion Using Time Series Analysis

[1]*Erika Abigail KATONOVÁ,* [2]*Alex SZAFFKO,* [3]*Rastislav PETIJA,* [4]*Peter FECIĽAK*

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

[1]erika.abigail.katonova@tuke.sk, [2]alex.szaffko@student.tuke.sk,
[3]rastislav.petija@cnl.sk, [4]peter.fecilak@cnl.sk

***Abstract* — The stability and reliability of computer networks are critical for both enterprise infrastructure and daily life. With the increasing complexity and scale of modern networks, traditional monitoring approaches often fail to provide timely warnings of impending failures or congestion. This paper presents the design, implementation, and evaluation of a predictive system for network failure and congestion, leveraging time series analysis and machine learning techniques. The system utilizes synthetic and real-world datasets to train and test statistical models, specifically AutoRegressive Integrated Moving Average (ARIMA) for short-term predictions and Seasonal ARIMA (SARIMA) for long-term forecasting. The results demonstrate that ARIMA models can accurately predict short-term CPU and interface utilization, enabling proactive alerts, while SARIMA captures seasonal trends in network traffic. The proposed approach offers a practical solution for early detection of critical network states, with potential for further enhancement through hybrid and multi-metric models.**

***Keywords* — ARIMA, congestion detection, machine learning, network reliability, network monitoring, SARIMA, time series analysis, failure prediction.**

I. INTRODUCTION

The continuous development of network technologies and the growing volume of data transmission have made the prediction of failures and congestion in network environments an increasingly important topic [1]. As the number of connected devices and the reliance on network services rise, ensuring the stability and availability of networks becomes critical. Outages and network unavailability can have direct impacts on business productivity and individual users, ranging from loss of access to essential services to severe failures in critical infrastructures such as hospitals, banks, and industrial enterprises [2]. Even short outages can result in significant financial losses and reputational damage, with major IT infrastructure failures costing organizations hundreds of thousands of euros per hour [3]-[5].

The economic and societal consequences of network failures are far-reaching. For businesses, network outages can halt operations, disrupt supply chains, and erode customer trust. In sectors such as healthcare, finance, and public safety, the consequences can be even more severe, potentially endangering lives or causing widespread service disruptions. The increasing adoption of cloud computing, remote work, and the Internet of Things (IoT) has further heightened the dependency on reliable network infrastructure, making resilience and proactive management a top priority for organizations worldwide.

Traditional network monitoring approaches, such as those based on the Simple Network Management Protocol (SNMP), provide basic status information but often lack the ability to detect and predict problems in a timely manner. These methods are typically reactive, alerting operators only after a failure or congestion event has occurred. In large-scale and complex networks, the sheer volume and diversity of monitoring data can overwhelm conventional tools, limiting their effectiveness in early detection and prevention of failures. Without advanced analytical and predictive capabilities, it is difficult to identify hidden patterns, anticipate emerging issues, and take preventive action before disruptions escalate.

The evolution of network environments has introduced new challenges for monitoring and management. Modern networks are characterized by high levels of dynamism, heterogeneity, and

scale, driven by trends such as virtualization, software-defined networking, and the proliferation of IoT devices. These factors increase the likelihood of unexpected failures and congestion, while also complicating the task of distinguishing between normal fluctuations and genuine anomalies. As a result, there is a growing need for intelligent, data-driven approaches that can provide actionable insights and support proactive decision-making.

Recent advances in machine learning and time series analysis offer promising solutions to these challenges. By leveraging historical data and statistical models, it is possible to forecast future network states, detect early warning signs of impending problems, and enable timely interventions. Predictive analytics can transform network monitoring from a reactive process into a proactive one, reducing downtime, improving service quality, and optimizing resource utilization.

This paper addresses these challenges by proposing a predictive system that leverages time series analysis and machine learning to forecast potential failures and congestion in network environments. The system is designed to process both synthetic and real-world datasets, applying statistical models to generate timely alerts and support proactive network management. The main contributions of this work are:

- The design and implementation of a predictive system for network failure and congestion, utilizing ARIMA and SARIMA models for short- and long-term forecasting.
- A comprehensive evaluation of the system using synthetic and real network data, demonstrating its effectiveness in early detection of critical network states.
- An analysis of the limitations of traditional and machine learning-based approaches, and a discussion of future directions for enhancing predictive network monitoring.

This paper is organized as follows: Section II describes the methodology, including data sources, models, and evaluation metrics. Section III details the system implementation and architecture. Section IV presents the results and evaluation of the predictive models. Section V concludes the paper and outlines directions for future work.

II. Methodology

The methodology for developing the predictive system was based on a combination of synthetic and real-world datasets to ensure both controlled experimentation and practical relevance.

A. Data

Synthetic Data Generation is initial data collection involved gathering SNMP (Simple Network Management Protocol) metrics from a virtual Cisco router using a Python script and the pysnmp library. These data provided valuable insights into the structure and characteristics of network metrics, such as CPU and memory utilization. However, it was determined that data collected from virtual environments did not sufficiently reflect the complexity and variability of production networks. To address this, synthetic time series were manually created to simulate a range of network behaviors:

- Normal operation: Low-variance, stable utilization patterns generated using random walks with small step sizes.
- Sudden spikes: Abrupt increases in utilization, implemented by injecting step changes at random intervals to mimic overload events.
- Gradual increases: Slow, monotonic rises in utilization, simulating congestion buildup.

Each synthetic dataset consisted of 200-300 time points, normalized to the [0, 100] range. Anomalies were injected at known locations to facilitate controlled testing of model sensitivity.

For the *Real Network Data* was the CESNET-TimeSeries24 dataset [6] used for long-term congestion prediction. This dataset contains time-aggregated (10 minutes, 1 hour, 1 day) network metrics, such as total bytes transferred, from a real academic network environment. The dataset spans several months and includes millions of records, providing a realistic basis for evaluating the system's performance in practical scenarios. Preprocessing involved normalization, handling missing values, and resampling to ensure consistent time intervals.

All datasets were split into training and testing subsets, typically using an 80/20 split, based on the paper [7]. For short-term predictions, synthetic data were primarily used, while long-term forecasting relied on the CESNET dataset.

B. Models

Several predictive models were considered and evaluated during the development of the system:

ARIMA (AutoRegressive Integrated Moving Average): Used for short-term prediction of CPU and network interface utilization. ARIMA models are well-suited for time series data with trends and are effective for forecasting over short horizons. The model parameters (p, d, q) were selected based on validation performance and information criteria [8].

SARIMA (Seasonal ARIMA): Applied for long-term prediction of network traffic, particularly when seasonal patterns are present. SARIMA extends ARIMA by incorporating seasonal terms, making it suitable for datasets with periodic fluctuations. The seasonal period was determined based on the data's natural cycles (e.g., weekly seasonality for hourly data) [9].

LSTM (Long Short-Term Memory) **neural networks:** Explored as a deep learning approach for time series prediction. LSTM networks are a type of recurrent neural network (RNN) capable of learning long-term dependencies. Despite their theoretical suitability for time series, LSTM models did not yield satisfactory results in this study due to the limited size and variability of the available datasets, often producing nearly constant predictions [10].

Other models: Linear Regression, Moving Average, and Random Forest classifiers were also tested. These models were found to be less effective for the specific characteristics of the network data, particularly in capturing non-linear and seasonal patterns.

Parameter selection: Model parameters for ARIMA and SARIMA were selected using a combination of grid search and information criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [11]. For each dataset, multiple configurations were evaluated, and the best-performing model was chosen based on validation set performance.

Based on experimental results, ARIMA was selected as the primary model for short-term predictions, while SARIMA was chosen for long-term forecasting due to its ability to model seasonality in network traffic.

C. Evaluation Metrics

The performance of the predictive models was assessed using standard metrics for time series forecasting. The following metrics were used:

- **Root Mean Square Error (RMSE):** Measures the average magnitude of the prediction error, with greater emphasis on larger errors [12].
- **Mean Absolute Error (MAE):** Represents the average absolute difference between predicted and actual values [12].
- **Mean Absolute Percentage Error (MAPE):** Expresses the prediction error as a percentage, facilitating comparison across different scales [13].
- **Normalized Root Mean Square Error (NRMSE):** Used for long-term predictions to enable comparison between datasets of different scales.
- **R-squared (R^2):** Indicates the proportion of variance in the observed data explained by the model, with values closer to 1 signifying better fit [14].

These metrics provided a comprehensive evaluation of model accuracy and enabled objective comparison of different approaches and configurations. RMSE and MAE measure the average magnitude of prediction errors, MAPE expresses errors as a percentage, NRMSE allows for comparison across datasets of different scales, and R^2 quantifies the proportion of variance explained by the model.

III. System Implementation

The predictive system was implemented in Python and designed with modularity and extensibility in mind. The overall architecture consists of several interconnected modules responsible for data ingestion, preprocessing, model training, prediction, evaluation, and alerting. Fig. 1 illustrates the high-level architecture and data flow of the system.

The system functions in two primary modes. In the short-term prediction mode, it uses synthetic time-series data to predict CPU and network interface utilization over a brief horizon, such as 15 minutes into the future. In the long-term prediction mode, it leverages real-world data from the CESNET dataset to forecast aggregated network traffic over extended periods, capturing seasonal trends and patterns.

Data is first ingested and preprocessed, then split into training and testing sets. The appropriate predictive model (ARIMA or SARIMA) is trained on the training data and used to generate forecasts on the test set. The results are evaluated using the selected metrics, and alerts are

generated if critical thresholds are exceeded.

The main components of the system are as follows:

- **Data processing module:** Responsible for loading, cleaning, and preprocessing time series data. This includes normalization, handling missing values, and splitting data into training and testing subsets.
- **Prediction module:** Implements ARIMA for short-term and SARIMA for long-term forecasting. The module allows for dynamic configuration of model parameters (order of ARIMA, seasonal period for SARIMA).
- **Evaluation module:** Computes performance metrics (RMSE, MAE, MAPE, NRMSE, R^2) and detects threshold violations that may indicate potential failures or congestion.
- **Alerting module:** Generates alerts (via HTTP requests) when predicted values exceed predefined critical thresholds, enabling proactive network management.
- **User interface (Streamlit):** An interactive web interface was developed using Streamlit to facilitate testing and visualization during development. The UI allows users to configure model parameters, visualize time series data and predictions, and monitor evaluation metrics in real time. This component was primarily used for experimentation and is not required for production deployment.
- **Deployment (Docker):** The final implementation is packaged as a Docker container, ensuring portability and ease of deployment in different environments.

IV. Results and Evaluation

This chapter evaluates the prediction models across various scenarios. It covers short-term forecasting with ARIMA for CPU utilization and network interface performance, followed by long-term predictions using SARIMA to capture seasonal traffic trends.

A. Short-Term Prediction (ARIMA)

All short-term ARIMA predictions in this study were performed on a synthetic dataset that was manually generated. These artificially created scenarios were specifically designed to fulfill the requirements of the bachelor's thesis assignment and to enable controlled evaluation of the predictive models. The synthetic data allowed for the simulation of a variety of network behaviors, including normal operation, sudden spikes, and gradual increases in load, with anomalies injected at known locations for precise assessment of model performance.

The ARIMA model was employed for short-term prediction of CPU and network interface utilization using these synthetic datasets. The ARIMA model was configured to predict 15 minutes ahead, with parameters optimized for each scenario.

The results demonstrated that ARIMA achieved high accuracy in both stable and dynamic conditions. For CPU utilization, the model attained a Root Mean Square Error (RMSE) as low as 1.63 and a Mean Absolute Percentage Error (MAPE) of 1.87% in stable scenarios. In cases with high CPU events, the RMSE increased to 4.47 and MAPE to 4.16%, but the model maintained its ability to provide timely warnings of potential threshold violations.

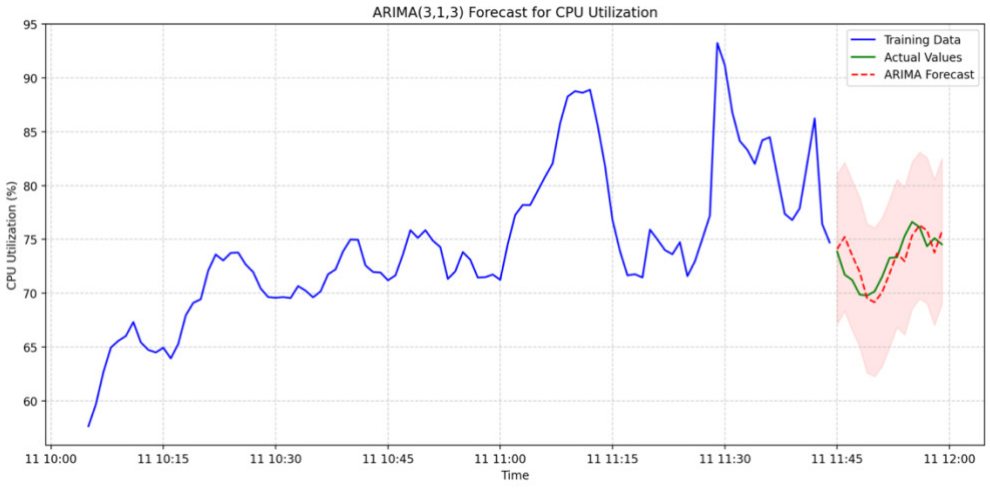


Fig. 1 ARIMA(3,1,3) forecast for CPU utilization. The blue line represents training data, the green line actual values, and the red dashed line the ARIMA forecast with confidence intervals.

Fig. 1 shows an example of ARIMA(3,1,3) forecasting CPU utilization, with the training data, actual values, and forecasted values (with confidence intervals) clearly visualized. The model's ability to follow the actual trend and provide a reasonable confidence interval is evident.

In scenarios with high CPU load events, the model was able to forecast threshold violations, as shown in Fig. 2 and the detailed view in Fig. 3. The red points indicate forecasted high load events above the 90% threshold.

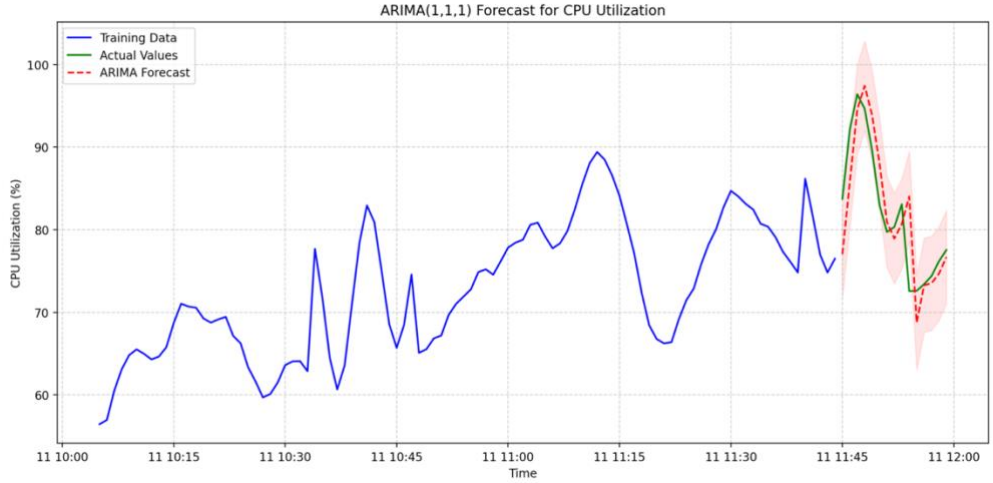


Fig. 2 Short-term ARIMA forecast for CPU utilization with high load events.

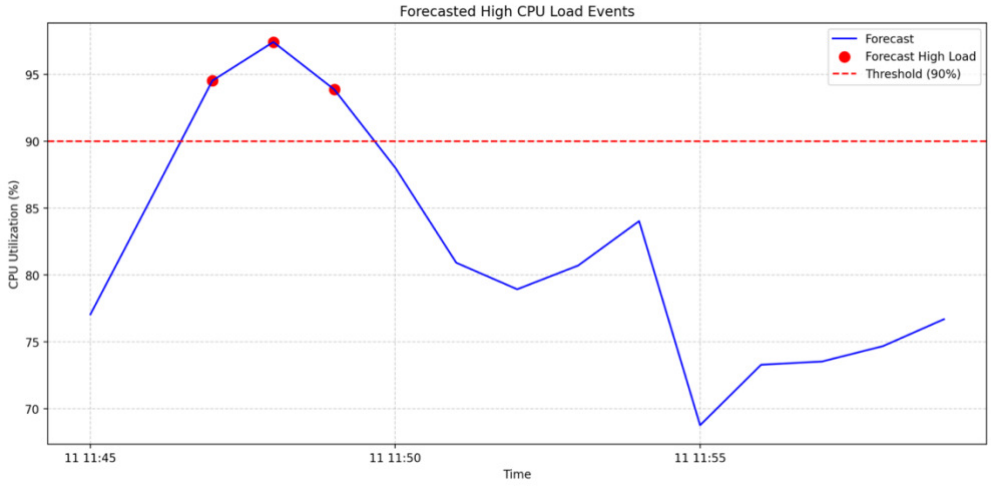


Fig. 3 Detail of ARIMA forecast showing predicted high CPU load events (red points) above the 90% threshold.

B. Short-Term Prediction for Network Interface

For network interface utilization, ARIMA achieved an RMSE of 4.92 and MAPE of 6.13% in stable cases, and an RMSE of 9.37 and MAPE of 11.00% in scenarios with sudden spikes. The model effectively identified periods where utilization exceeded critical thresholds, enabling proactive congestion management. Fig. 4 shows a typical ARIMA(1,0,3) forecast for network utilization, while Fig. 5 and Fig. 6 illustrate the model's performance in scenarios with congestion.

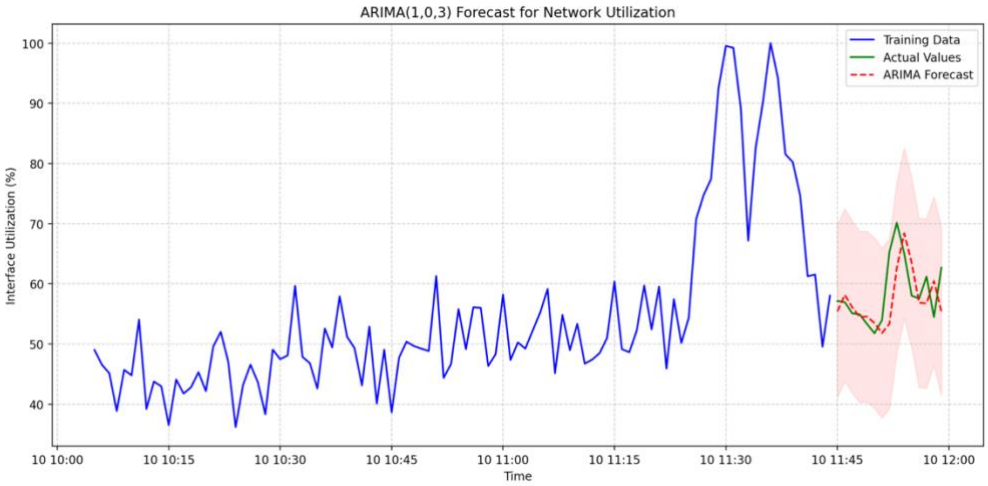


Fig. 4 ARIMA(1,0,3) forecast for network interface utilization.

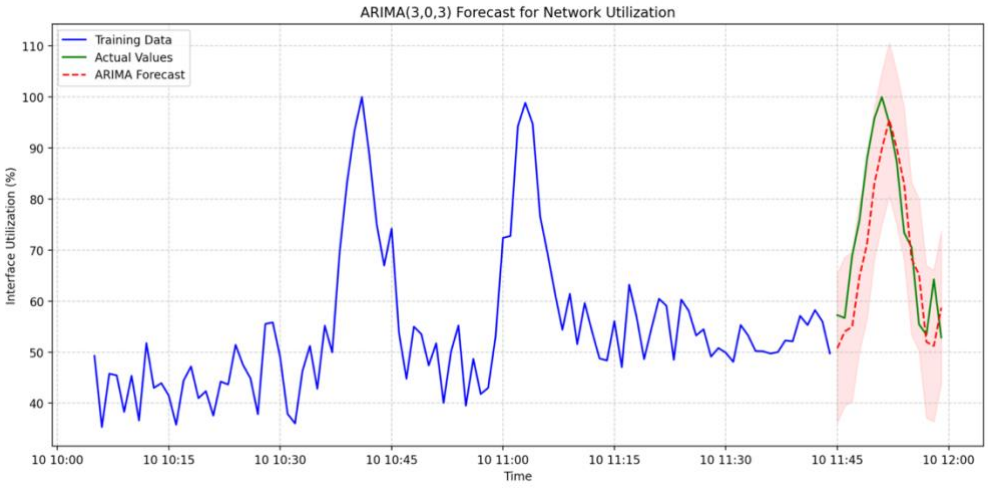


Fig. 5 ARIMA(3,0,3) forecast for network interface utilization with a congestion event.

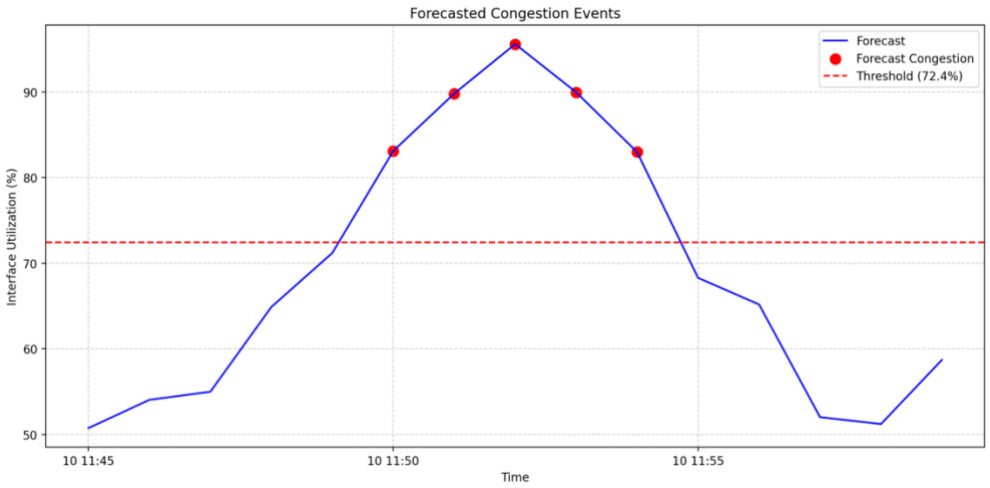


Fig. 6 Detail of ARIMA forecast showing predicted congestion events (red points) above the 95th percentile threshold.

C. Long-Term Prediction (SARIMA)

The SARIMA model was applied to the CESNET-TimeSeries24 dataset for long-term forecasting of aggregated network traffic. The model was optimized for weekly seasonality and trained on one month of historical data. The evaluation focused on the model's ability to capture seasonal trends and provide accurate forecasts over a one-week horizon.

Fig. 7 shows the SARIMA forecast for the total number of bytes transferred, with congestion events highlighted above the threshold.

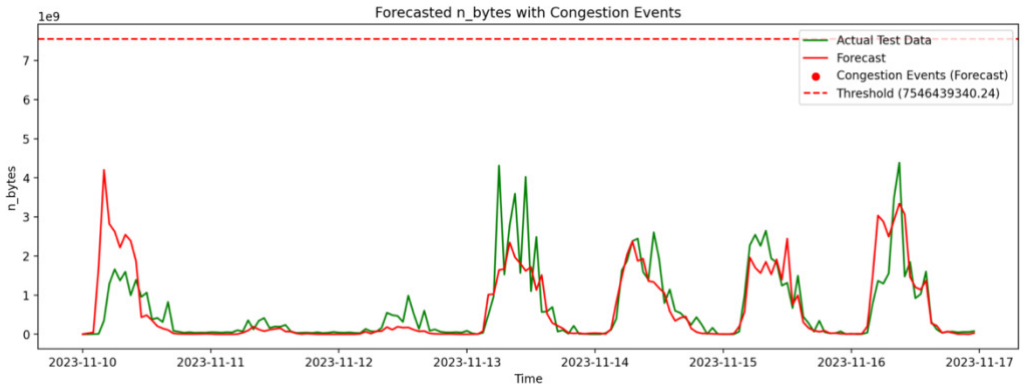


Fig. 7 SARIMA long-term forecast for total bytes transferred (n_bytes) with predicted congestion events.

SARIMA achieved a Normalized Root Mean Square Error (NRMSE) of 0.9684 and R^2 score of 0.5324 on the test set. These results demonstrate that SARIMA can effectively model and predict seasonal patterns in network traffic, although the precision may be limited by the complexity and variability of real-world data. The model's ability to anticipate long-term trends makes it a valuable tool for capacity planning and anomaly detection in large-scale networks.

D. Limitations and Discussion

While the ARIMA and SARIMA models demonstrated effectiveness with the available datasets, several limitations emerged during evaluation. First, data limitations were evident: synthetic datasets, while suitable for controlled experiments, may not fully reflect the complexity of real network environments. Similarly, the CESNET dataset, though representative, is restricted to aggregated metrics and lacks detailed interface-level data. Second, attempts to utilize LSTM neural networks for time-series prediction were unsuccessful due to insufficient data volume and variability. These models frequently produced nearly constant predictions, underscoring the need for larger and more diverse datasets to harness the full potential of deep learning. Third, the current system focuses on predicting individual metrics, such as CPU or interface utilization, rather than integrating multiple metrics for comprehensive anomaly detection. This single-metric approach may limit the system's ability to identify complex failure or congestion scenarios. Lastly, the models were specifically tuned for the given datasets and scenarios, raising concerns about their generalizability. Broader validation across diverse network environments and with additional data sources is required to ensure the system's applicability in varied contexts.

Despite these limitations, the results confirm the feasibility and practical value of time series analysis for proactive network monitoring. The system provides a foundation for future enhancements, such as hybrid approaches that combine statistical and machine learning models, and the integration of multiple metrics for more robust detection of network anomalies.

V. Conclusion and Future Work

This paper presented the design, implementation, and evaluation of a predictive system for network failure and congestion, leveraging time series analysis and machine learning techniques. The system was tested on both synthetic and real-world datasets, demonstrating the effectiveness of ARIMA models for short-term prediction of CPU and interface utilization, and SARIMA models for long-term forecasting of network traffic with seasonal patterns.

The results confirm that statistical time series models can provide accurate and timely predictions, enabling proactive alerts and supporting network operators in preventing outages and congestion. The modular architecture and use of open-source technologies ensure that the system can be easily adapted to different environments and extended with new data sources or models.

Despite the promising results, several challenges remain. The current system is limited by the availability and diversity of training data, particularly for deep learning approaches. The focus on single-metric prediction restricts the ability to detect more complex or multi-faceted network anomalies. Furthermore, the models were validated on specific datasets and scenarios, and broader generalization requires further testing in diverse network environments.

Future work may aim to address these limitations through several key initiatives. First, the integration of multiple metrics and data sources will be prioritized to enable holistic anomaly detection and minimize false positives. Second, hybrid approaches that combine statistical models with machine learning classifiers will be explored to enhance accuracy and robustness. Third,

efforts will be made to acquire and annotate larger, more diverse datasets to support the development and validation of advanced predictive models, including deep learning techniques. Finally, the system will be evaluated in real-world production environments to assess its practical impact and scalability, ensuring its readiness for deployment in complex network scenarios.

By advancing predictive network monitoring, this work contributes to the development of more resilient and reliable network infrastructures, capable of meeting the demands of increasingly complex digital ecosystems.

REFERENCES

[1] A. K. M. Bahalul Haque, M. Oahiduzzaman Mondol Zihad, and M. Rifat Hasan, "5G and Internet of Things - Integration Trends, Opportunities, and Future Research Avenues," Singapore: Springer Nature, Singapore, 2023, pp. 217–245. [Online]. Available: https://doi.org/10.1007/978-981-99-3668-7_11

[2] D. Turner, K. Levchenko, A. C. Snoeren, and S. Savage, "California fault lines: understanding the causes and impact of network failures," SIGCOMM Comput. Commun. Rev., vol. 40, no. 4, p. 315–326, Aug. 2010. [Online]. Available: https://doi.org/10.1145/1851275.1851220

[3] "Summary of the aws service event in the northern virginia (us-east-1) region," Amazon, Tech. Rep., 12 2021, https://aws.amazon.com/message/12721/.

[4] D. Madory, "Facebook's historic outage, explained," 2021, https://www.kentik.com/blog/facebooks-historic-outage-explained/.

[5] S. Mursaleen, "Centurylink outage and everything that went down with it," 2020, https://www.catchpoint.com/blog/centurylink-outage-2020.

[6] J. Koumar, K. Hynek, T. Čejka, and P. Šiška, "Cesnet-timeseries24: Time series dataset for network traffic anomaly detection and forecasting," Scientific Data, vol. 12, no. 1, p. 338, Feb 2025.

[7] Muraina, Ismail. "Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts." 7th international Mardin Artuklu scientific research conference. 2022.

[8] Sirisha, U.M., Belavagi, M.C. and Attigeri, G., "Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison". Ieee Access, 10, 2020.

[9] Dubey, Ashutosh Kumar, et al. "Study and analysis of SARIMA and LSTM in forecasting time series data." Sustainable Energy Technologies and Assessments 47, 2021.

[10] DiPietro, Robert, and Gregory D. Hager. "Deep learning: RNNs and LSTM." Handbook of medical image computing and computer assisted intervention. Academic Press, 2020.

[11] Kasali, Jimoh, and Adediwura Alaba Adeyemi. "Model-data fit using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the sample-size-adjusted BIC." Square: Journal of Mathematics and Mathematics Education 4.1, 2022.

[12] Hodson, Timothy O. "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not." Geoscientific Model Development Discussions 2022, 2022.

[13] Román, S. A. R. L., Sebastian Zelada Collazos, and Jorge Antonio Corzo Chavez. "Model for Reducing Mean Absolute Percentage Error through Smoothing and Time Series Forecasting In a Tourism SME: A Case Study." Journal of Machine Intelligence and Data Science (JMIDS) 5, 2024.

[14] Gao, Jian. "R-Squared (R2)–How much variation is explained?." Research Methods in Medicine & Health Sciences 5.4, 2024.

Dynamické Modelovanie v Prostredí Coach

Zuzana GIBOVÁ

Katedra matematiky a teoretickej informatiky, Fakulta elektrotechniky a informatiky, Technická Univerzita, Košice, Slovensko

zuzana.gibova@tuke.sk

***Abstrakt* — Článok sa zaoberá dynamickým modelovaním fyzikálnych javov v prostredí Coach. Toto prostredie umožňuje pomocou modelu riešiť diferenciálne rovnice, ktorých matematické riešenie je často zložité. Coach ponúka aj ďalšie funkcie, ako je zobrazenie funkčných závislostí modelovaných veličín vo forme grafov a tabuliek, simulovanie modelu pri zmenených podmienkach alebo využitie grafickej derivácie. V článku je uvedený konkrétny príklad dynamického modelovania pohybu odpálenej golfovej loptičky. V závere je ukázané ako je možné pomocou analógie medzi veličinami popisujúcimi dynamický dej a veličinami popisujúcimi elektrický dej, vytvoriť model na prechodové javy v sériovom RC obvode.**

***Kľúčové slová* — diferenciálna rovnica, dynamické modelovanie, prostredie Coach, simulácia**

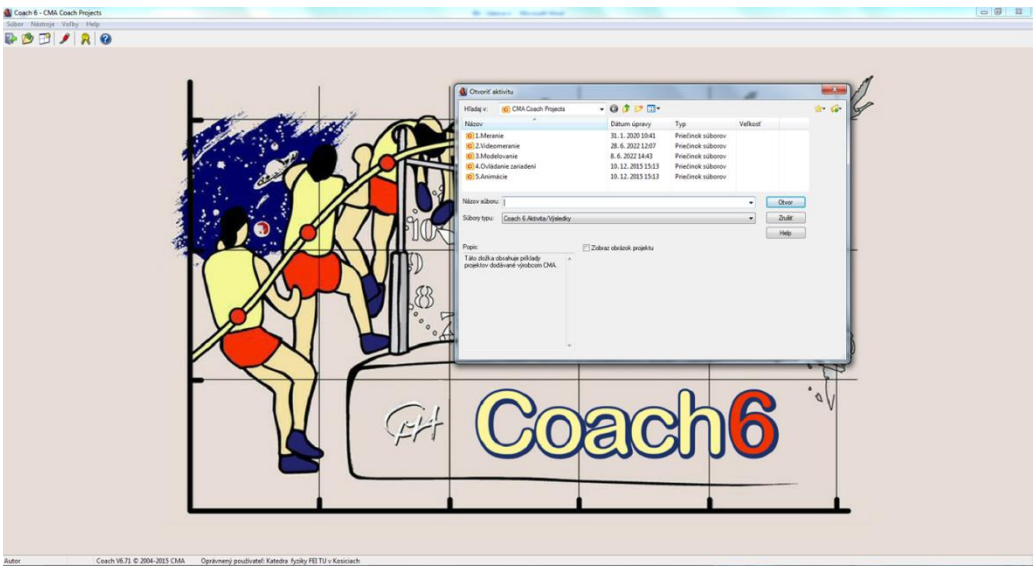
I. Úvod

Dynamické modelovanie umožňuje skúmať javy, ktoré sa menia v čase, napr. pohyb telesa, priebeh prúdu alebo napätia v obvode, kmitanie telies, prechodové javy v RLC obvode a iné. Spomenuté príklady sú z matematického hľadiska často popísané diferenciálnymi rovnicami, ktorých riešenie nie je vždy triviálne. Pri dynamickom modelovaní sa využívajú algebrické metódy približného riešenia týchto rovníc. V článku je spomenutá Eulerova metóda, ktorá sa používa pri riešení diferenciálnych rovníc v prostredí Coach.

II. Modelovanie v Coachi

A. *Prostredie Coach*

Systém Coach [1] je systém počítačom podporovaného laboratória, ktoré spája nástroje pre meranie, ovládanie zariadení počítačom, videomeranie, modelovanie, tvorbu interaktívnych animácií, pokročilé spracovanie dát a analýzu dát (obr. 1).



Obr.1 Hlavná ponuka prostredia Coach.

Tento systém nachádza svoje využitie vo fyzike a v iných prírodovedných predmetoch, napr. ho používajú študenti FEI na laboratórnych cvičeniach z fyziky. Je využívaný na PF UPJŠ, ale je aj súčasťou prírodovedných laboratórií na mnohých základných a stredných školách na Slovensku a v zahraničí.

Modelovanie v Coachi umožňuje používať a vytvárať dynamické modely v grafickom alebo v textovom móde a takto získané výsledky porovnať z vypočítanými alebo nameranými dátami. Vytvorené modely umožňujú predpovedať priebeh jednotlivých dejov a pohybov telies pri znemených počiatočných podmienkach. V článku je popísané modelovanie v textovom móde, pomocou ktorého je skúmaný pohyb telesa pri šikmom vrhu v odporovom prostredí a v prostredí bez odporu (tzv. dynamický model). Súčasne je ukázané ako je možné modelovať aj elektrické deje pomocou analógie k dynamickému deju.

B. Fyzikálne hľadisko

Budeme skúmať pohyb golfovej loptičky hmotnosti m = 46 g, priemeru d = 40 mm, ktorá bola odpálená z trávnika pod uhlom $\alpha = 30°$ s počiatočnou rýchlosťou v_0 = 75 m/s (parametre loptičky zvolené podľa [2]). Pri tejto rýchlosti na loptičku pôsobí vzduch odporovou silou F_o, v dôsledku ktorej sa okolo loptičky tvoria víry tzv. turbulentné prúdenie, ktoré jej pohyb spomaľuje. Odporová sila závisí od rozmerov a rýchlosti loptičky, v prípade turbulentného prúdenia, pre teleso guľového tvaru, je jej veľkosť daná

$$F_o = \frac{1}{2} C\rho S v^2, \tag{1}$$

kde S je plocha prierezu telesa v smere kolmom na smer pohybu, ρ je hustota tekutiny (pre vzduch 1,293 kg/m^3) a C je faktor odporu tekutiny (závisí od tvaru telesa, pre guľu C = 0,37). Okrem toho, na loptičku pôsobí tiažové pole silou F_g, ktorá ho priťahuje k zemi. Veľkosť tiažovej sily je daná

$$F_g = mg, \tag{2}$$

kde m je hmotnosť telesa a g je tiažové zrýchlenie. Výsledná sila F pôsobiaca na loptičku je daná súčtom pôsobiacich síl.

Pre jednoduchosť predpokladajme, že loptička sa pohybuje v rovine XY. V tomto prípade sa výsledná sila $\vec{F}$ pôsobiaca na loptičku rozkladá na dve zložky sily: $\vec{F}_x$ vzhľadom na os X a $\vec{F}_y$ vzhľadom na os Y (veľkosti zložiek sily nazývame súradnice sily a označujeme F_x, F_y). Podobne sa rozkladá na zložky aj odporová sila (jej súradnice sú F_{ox}, F_{oy}). Vzhľadom na os X na loptičku pôsobí len odporová sila. Potom pre x - ovú súradnicu výslednej sily

$$F_x = F_{ox} = -\frac{1}{2} C\rho S {v_x}^2. \tag{3}$$

Vzhľadom na os Y pôsobia na loptičku dve sily, odporová a tiažová sila. Výsledná sila je daná ich súčtom, potom pre y - ovú súradnicu výslednej sily

$$F_y = F_{oy} + F_g = -\frac{1}{2} C\rho S {v_y}^2 - mg. \tag{4}$$

Znamienko mínus v rovniciach (3) a (4) naznačuje, že spomínané sily majú opačný smer ako je pohyb loptičky vzhľadom na daný súradnicový systém a v_x, v_y sú súradnice rýchlosti.

Zaujíma nás ako sa bude loptička pohybovať v poli zemskej tiaže, teda ako sa bude meniť jej rýchlosť v čase, aký bude tvar jej trajektórie, akú maximálnu výšku dosiahne a ako ďaleko doletí na trávnik. Z matematického hľadiska potrebujeme vypočítať súradnice rýchlosti v_x, v_y loptičky a súradnice jej polohy x, y. Pohybové rovnice potrebné pre výpočet, vyplývajú z druhého Newtonovho pohybového zákona [3] v tvare

$$F_x = m\frac{\mathrm{d}v_x}{\mathrm{d}t} = ma_x \tag{5}$$

$$F_y = m\frac{\mathrm{d}v_y}{\mathrm{d}t} = ma_y, \tag{6}$$

kde a_x, a_y sú súradnice zrýchlenia. Pri výpočte sa ľavé strany rovníc (5) a (6) nahradia rovnicami (3) a (4) a riešia sa diferenciálne rovnice. Týmto spôsobom sa vypočítajú súradnice rýchlosti v_x a v_y a následne pomocou nich súradnice polohy x, y. V prípade pohybu loptičky v prostredí bez odporu (C = 0) je výpočet triviálny, na loptičku pôsobí len tiažová sila, teda $F_x = 0, F_y = -mg$ (obr. 2). Dosadením týchto podmienok do rovníc (5) a (6) dostaneme diferenciálne rovnice prvého rádu so separovateľnými premennými, ktoré po separácii sú

$$0 = m\frac{\mathrm{d}v_x}{\mathrm{d}t} \rightarrow \mathrm{d}v_x = 0\,\mathrm{d}t \tag{7}$$

$$-mg = m\frac{\mathrm{d}v_y}{\mathrm{d}t} \rightarrow \mathrm{d}v_y = -g\,\mathrm{d}t. \tag{8}$$

Tieto rovnice sa ďalej upravujú integrovaním oboch strán pomocou určitého integrálu

$$\int_{v_{0x}}^{v_x} \mathrm{d}v_x = \int_0^t 0\,\mathrm{d}t \tag{9}$$

$$\int_{v_{0y}}^{v_y} \mathrm{d}v_y = \int_0^t -g\,\mathrm{d}t, \tag{10}$$

kde v_{0x}, v_{0y} sú súradnice počiatočnej rýchlosti v čase $t = 0$ s a v_x, v_y sú súradnice rýchlosti v čase *t*. Výpočtom pre súradnice rýchlosti

$$v_x - v_{0x} = 0 \rightarrow v_x = v_{0x} = v_0 \cos a, \tag{11}$$

$$v_y - v_{0y} = -gt + 0 \rightarrow v_y = v_{0y} - gt = v_0 \sin a - gt, \tag{12}$$

kde súradnice počiatočnej rýchlosti $v_{0x} = v_0 \cos a$, a $v_{0y} = v_0 \sin a - gt$, sú vyjadrené pomocou uhla α, ktoré je zrejmé z obr. 2. Vzťah medzi rýchlosťou a dráhou je daný

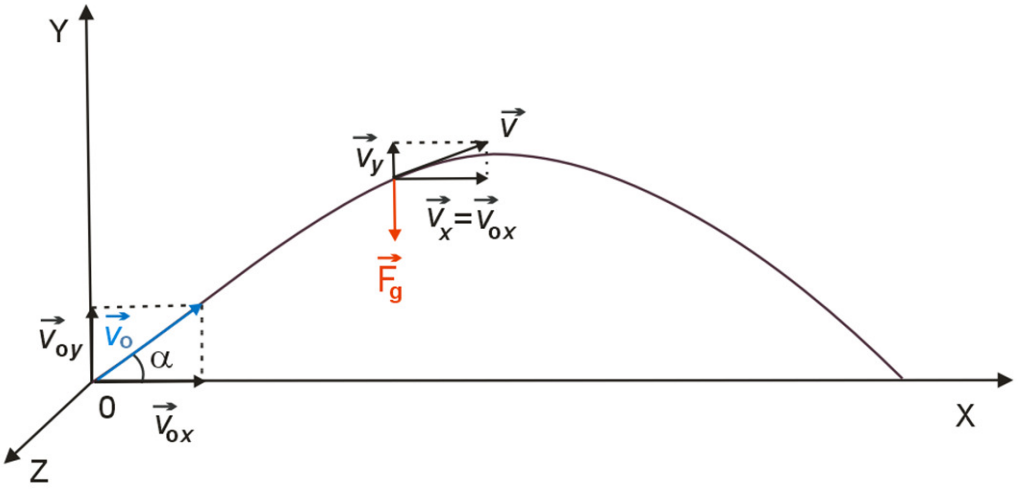
$$s = \int_0^t v\,\mathrm{d}t + s_0, \tag{13}$$

ktorý vieme použiť aj na vyjadrenie súradníc polohy *x*, *y* loptičky

$$x = \int_0^t v_x\,\mathrm{d}t + x_0 = v_0\,t\cos a + x_0 \tag{14}$$

$$y = \int_0^t v_y\,\mathrm{d}t + y_0 = v_0\,t\sin a - \frac{1}{2}gt^2 + y_0, \tag{15}$$

kde x_0, y_0 sú súradnice počiatočného bodu, z ktorého je loptička odpálená [4].



Obr. 2 Trajektória pohybu odpálenej golfovej loptičky pod uhlom α s počiatočnou rýchlosťou v_0 v poli zemskej tiaže bez vplyvu odporovej sily. Na loptičku pôsobí len tiažová sila.

Maximálna výška, ktorú loptička pri svojom pohybe v prostredí bez odporu dosiahne je bod, kde celková rýchlosť loptičky je daná len súradnicou v_x ($v_y = 0$ m/s) a $y = h_{\max}$. Dosadením týchto podmienok do rovníc (12) a (15), riešením sústavy a po úprave je maximálna výška

$$h_{\max} = \frac{v_0^2 \sin^2 \alpha}{2g}. \tag{16}$$

Dolet loptičky (maximálna vzdialenosť, kde loptička dopadne na trávnik) určíme pomocou podmienok $x = d_{\max}, y = 0$ m, ktoré sa dosadia do rovníc (14) a (15). Riešením sústavy a po úprave je dolet

$$d_{\max} = \frac{v_0^2 \sin 2\alpha}{g}. \tag{17}$$

V prípade, že sa loptička pohybuje v odporovom prostredí (C = 0,37) výpočet diferenciálnych rovníc je zložitý, preto je výhodnejšie pri výpočte použiť dynamický model.

C. Dynamický model v prostredí Coach

Pri tvorbe dynamického modelu použijeme Eulerovu metódu približného riešenie diferenciálnych rovníc [5], ktorá vychádza z pohybovej rovnice (5) upravenej na tvar

$$a = \frac{F(t,\ s,\ v)}{m} = a(t,\ s,\ v), \tag{18}$$

ktorá je východiskom výpočtu postupnosti pre rýchlosť, dráhu a čas

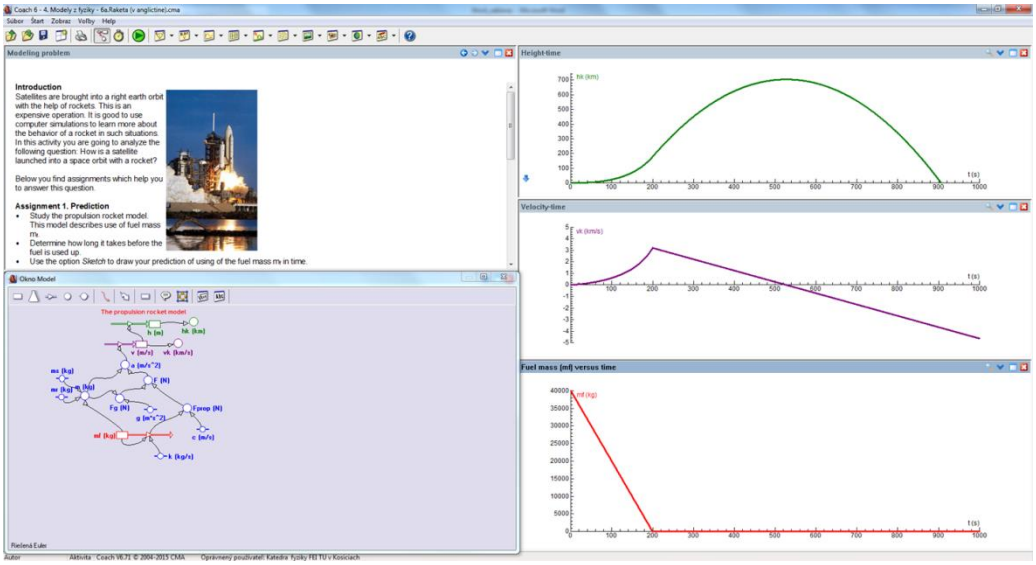
$$v_{i+1} = v_i + a\Delta t, \tag{19}$$

$$s_{i+1} = s_i + v\Delta t, \tag{20}$$

$$t_{i+1} = t_i + \Delta t. \tag{21}$$

V dynamickom modeli sú diferenciálne rovnice nahradené diferenčnými rovnicami, kde pre veľkosť $\Delta v = v_{i+1} - v_i$ a $\Delta s = s_{i+1} - s_i$ je rozhodujúca veľkosť časového kroku $\Delta t = t_{i+1} - t_i$. Hodnota časového kroku ovplyvňuje presnosť numerického riešenia diferenciálnych rovníc touto metódou, preto je vhodné voliť Δt čo najmenšie.

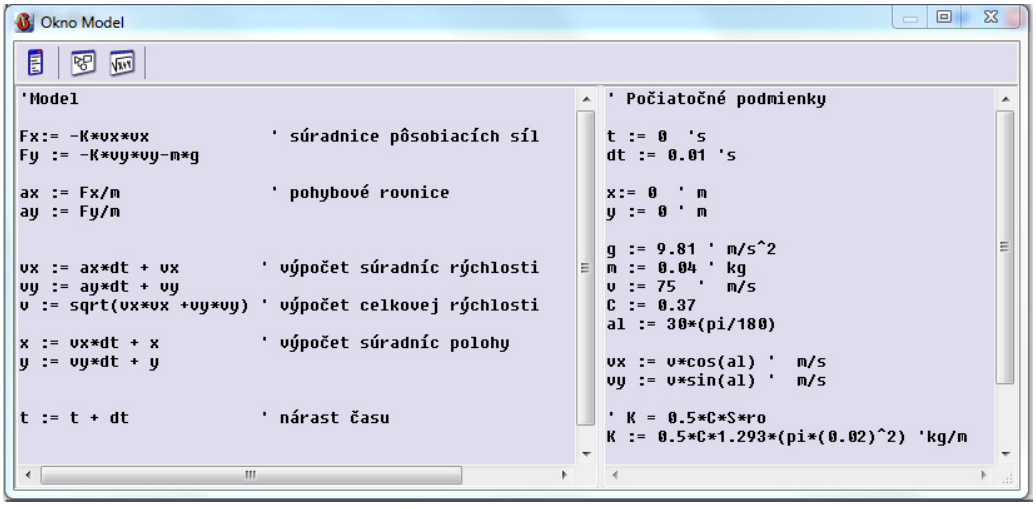
Prostredie Coach ponúka priestor na modelovanie, v ktorom sú už vytvorené vzorové modely alebo sa vytvárajú vlastné projekty s modelmi. Súčasťou každého projektu je modelovacie okno (okno vľavo dole na obr. 3), ktoré slúži na tvorbu modelu. V danom projekte je možné zvoliť ľubovoľný počet okien, do ktorých je možné vložiť grafické závislostí modelovaných veličín, text, obrázky ako aj použiť ďalšie funkcie, ktoré projekt ponúka [6].



Obr. 3 Projekt modelovanie v Coachi. Okno vľavo dole slúži na tvorbu modelu (grafický mód), počet zobrazených okien je voliteľný.

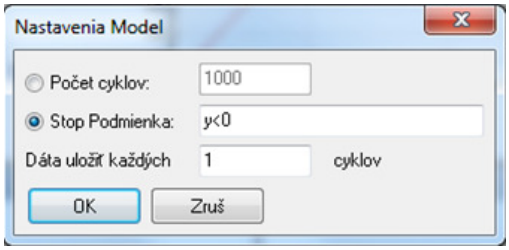
D. Model pohybu golfovej loptičky

Pri vytváraní dynamického modelu potrebujeme poznať pôsobiacu silu na teleso, pohybovú rovnicu, počiatočné podmienky a hodnotu dt. Pri modelovaní pohybu loptičky v odporovom prostredí použijeme vzťahy (3), (4) pre vyjadrenie súradníc pôsobiacej sily (v modeli je pre odporovú silu zavedená konštanta $K = \frac{1}{2}CS\rho$). Pohybové rovnice upravíme na tvar daný rovnicou (18) a zvolíme dt = 0,01 s, $x_0 = 0$ m, $y_0 = 0$ m.



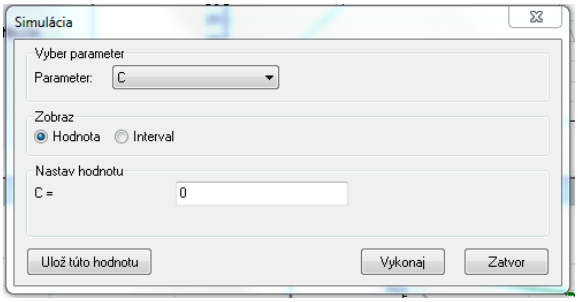
Obr. 4 Dynamický model pohybu golfovej loptičky vytvorený v modelovacom okne Coach (textový mód).

Postupnosť krokov výpočtu bude daná rovnicami (19) – (21). Model vytvárame v ľavej časti modelovacieho okna a počiatočné podmienky, ktoré vyplývajú zo zadania úlohy, uvedieme v pravej časti tohto okna (obr. 4). Podmienku pre ukončenie výpočtov môžeme pridať do modelu, nastaviť v príslušnom okne (obr. 5) alebo zadať počet iterácií (opakovaní výpočtov). Pre získanie potrebných informácií o pohybe, definujeme grafické závislosti súradníc rýchlosti od času, celkovej rýchlosti od času a závislosť y – ovej súradnice od x – ovej, ktorých grafy umiestnime do ďalších okien. Na ich vytvorenie použijeme funkciu *Graf*, ktorú získame stlačením ikony na hornej lište modelovacieho prostredia . Na rovnakej lište sa nachádza aj tlačidlo na spustenie modelu (zelené tlačidlo, obr. 3). Po skončení výpočtov sa v pripravených grafoch zobrazia vypočítané hodnoty veličín, s ktorými môžeme ďalej pracovať.



Obr. 5 Okno pre nastavenie modelu.

Vytvorený model je možné použiť aj v prípade pohybu loptičky v prostredí bez odporu. Buď zmeníme priamo v modeli hodnotu faktora odporu C na nulu alebo použijeme funkciu *Simulácia* (kliknutím v okne Model pravým tlačidlom myši vyvoláme túto funkciu). V okne pre simuláciu vyberieme parameter, ktorý chceme simulovať, zvolíme $C = 0$ a potvrdíme *Vykonaj*.

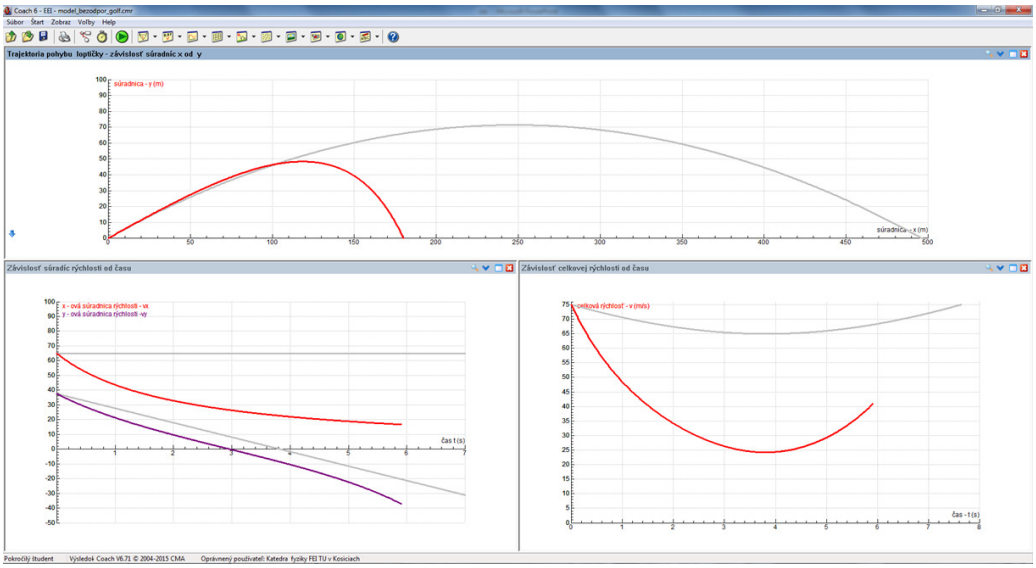


Obr. 6 Okno pre simuláciu zvolených parametrov.

Nové grafické závislosti pre simulovaný parameter sú zobrazené v rovnakých grafoch inou farbou, čo umožňuje vzájomné porovnanie (na obr. 7 sú pre prostredie bez odporu zobrazené sivou farbou, v prostredí s odporom červenou a fialovou farbou).

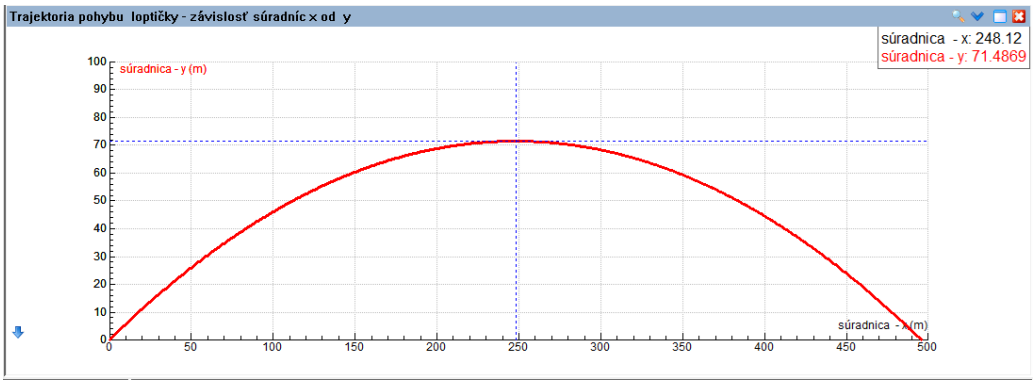
E. Interpretácia výsledkov

Z grafu závislostí súradníc polohy y od x vidieť (obr. 7, graf hore, sivá farba), že trajektóriou pohybu loptičky je parabola, ktorá v prípade pohybu v prostredí bez odporu, je symetrická vzhľadom na os y a svoje maximum dosahuje v strede trajektórie v jej najvyššom bode. V prípade pohybu s odporom, parabola nie je symetrická (červená farba na obr. 7). Hodnotu maximálnej výšky a doletu loptičky určíme pomocou funkcie *Prezerať* (kliknutím v grafe pravým tlačidlom myši), ktorá umožňuje pomocou pohyblivého súradnicového systému určiť hodnotu súradníc v ľubovoľnom bode trajektórie [7]. Požadované hodnoty sú zobrazované v pravom hornom rohu grafu (obr. 8). V prípade pohybu bez odporu, golfová loptička dosahuje po odpale maximálnu výšku 71,49 m a dolet 496,23 m. Výpočtom použitím rovníc (16), (17) vychádzajú porovnateľné hodnoty $h_{max} = 71,67$ m a $d = 496,57$ m. Tieto hodnoty sú v prostredí s odporom $h_{max} = 48,37$ m a $d = 180,08$ m, pri ich výpočte bol použitý vyššie spomenutý model. Pre presnejšie určenie maximálnej výšky je možné použiť aj deriváciu funkcie $y = f(x)$ podľa x (funkcia *Analyzuj / Spracuj* – derivácia) a určiť jej extrém, t.j. $dy/dx = 0$. V grafe sa zobrazí na pravej strane nová os, ktorá zobrazuje hodnoty derivácie dy/dx, krivka derivácie (zelená krivka na obr. 9) a čiara, ktorá naznačuje polohu, kde je $dy/dx = 0$ (čierna priamka idúca cez 0 na osi na pravej strane grafu). V mieste, kde sa pretína zelená krivka s čiernou priamkou, je extrém funkcie $y = f(x)$ (červená krivka na obr. 9).



Obr. 7 Grafické závislosti súradníc polohy *y* od *x* (graf hore), súradníc rýchlosti od času (graf vľavo dole) a celkovej rýchlosti od času (graf vpravo dole) v prostredí bez odporu (sivá farba), v prostredí s odporom (červená, fialová farba).

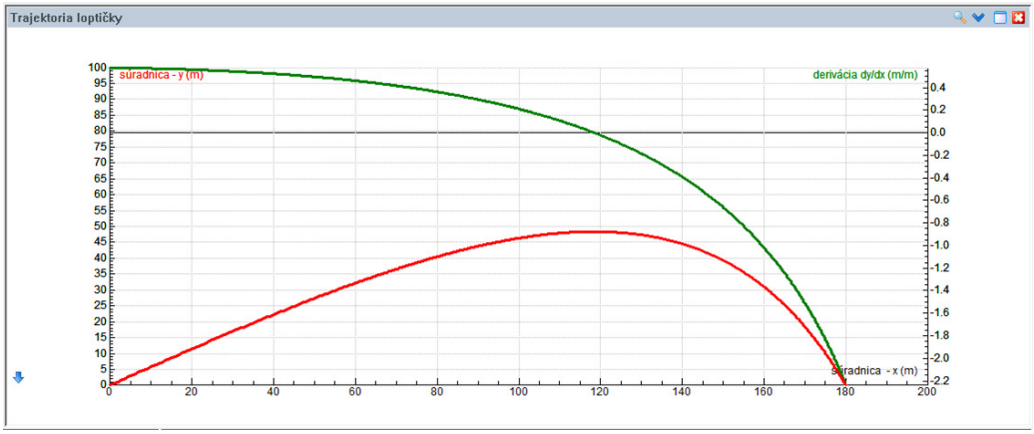
Podľa [2], by mala golfová loptička zvolených parametrov dopadnúť do vzdialenosti menšej ako 256 m. Použitím dynamického modelovania sme zistili, že loptička, ktorá sa pohybuje v odporovom prostredí, kde sa okolo nej tvoria víry, spĺňa túto požiadavku, jej dolet je 180,08 m. V prípade ideálneho prostredia (bez odporu) to neplatí, dolet loptičky je oveľa väčší.



Obr. 8 Využitie funkcie *Prezerať* na určenie hodnoty maximálnej výšky trajektórie pohybu golfovej loptičky.

Z grafických závislosti súradníc rýchlosti od času (obr. 7, graf vpravo dole, sivá farba) vyplýva, že v prostredí bez odporu je *x* - ová súradnica rýchlosti konštantná (sivá priamka rovnobežná s osou *x*), loptička v smere osi *x* vykonáva rovnomerný pohyb. Zatiaľ čo *y* - ová súradnica rýchlosti klesá, v najvyššej výške dosahuje nulovú hodnotu a potom narastá (sivá priamka, ktorá zviera s osou *x* určitý uhol). Jej zmena je ovplyvnená len tiažovou silou, ktorá je konštantná, preto loptička v smere osi *y* vykonáva rovnomerne spomalený pohyb a po dosiahnutí maximálnej výšky sa ďalej pohybuje rovnomerne zrýchleným pohybom. Pohyb golfovej loptičky v odporovom prostredí je v jednotlivých smeroch oveľa zložitejší. Z obr. 7 (graf vpravo dole, červená farba) vidieť, že *x* - ová súradnica rýchlosti klesá exponenciálne s časom a *y* - ová súradnica rýchlosti sa mení podobne ako pri pohybe bez odporu, po krivke, ktorá nie je priamka (graf vpravo dole, fialová farba).

Vplyv rôzneho pôsobenia výslednej sily na loptičku, možno vidieť aj na grafe celkovej rýchlosti od času (obr. 7, vpravo dole, červená farba). Celková rýchlosť loptičky po odpálení klesá z hodnoty 75 m/s na hodnotu 24,22 m/s, ktorú dosahuje v maximálnej výške, a pri klesaní loptičky k zemi, narastá jej rýchlosť na hodnotu 40,82 m/s. Pokles rýchlosti je výraznejší ako jej nárast. V dôsledku odporovej sily konečná hodnota rýchlosti loptičky sa líši od jej počiatočnej hodnoty. Pôsobením odporovej sily na loptičku, dochádza k strate jej mechanickej energie, ktorá sa počas jej pohybu premení na teplo, čo sa prejaví na koncovej rýchlosti loptičky. V prípade nulového odporu je hodnota počiatočnej a koncovej rýchlosti rovnaká (obr. 7, vpravo dole, sivá farba).



Obr. 9 Funkcia *Derivácia* na určenie extrému funkcie, t.j. maxima funkcie $y = f(x)$, ktorá je trajektóriou golfovej loptičky pri pohybe v prostredí s odporom.

F. Ďalšie možnosti

Ako bolo v úvode spomenuté je možné použiť dynamické modelovanie pri štúdiu elektrických dejov (napr. priebeh elektrického prúdu a napätia v obvodoch, kmitanie elektromagnetického oscilátora a iné). Postup pri modelovaní je rovnaký ako bolo ukázané vyššie, kde využijeme analógiu veličín (tabuľka 1).

Tabuľka 1
Analógia medzi veličinami pri dynamickom modelovaní

Mechanický dej	Elektrický dej
sila F	napätie U
prírastok rýchlosti $a\mathrm{d}t$	prírastok prúdu $\mathrm{d}i$
rýchlosť v	prúd i
dráha s (alebo súradnice x, y)	náboj q

Pohybová rovnica (18) sa pri skúmaní daného deja nahradí 2. Kirchhoffovym zákonom pre daný obvod. Prírastok prúdu $\mathrm{d}i$ sa určí na základe rovnice pre elektrický obvod, ďalšie rovnice modelu sú

$$q_{i+1} = q_i + i\mathrm{d}t, \tag{22}$$

$$i_{i+1} = i_i + \mathrm{d}i, \tag{23}$$

$$t_{i+1} = t_i + \mathrm{d}t. \tag{24}$$

Rovnica (19) z modelu pre mechanický dej bola nahradená analogickou rovnicou (23) a rovnica (20) analogickou rovnicou (22) pre elektrický dej.

Napríklad pri skúmaní prechodových javov v sériovom RC obvode pri nabíjaní kondenzátora je model daný rovnicami

$$q = q + i\mathrm{d}t, \tag{25}$$

$$i = \frac{U - \frac{q}{C}}{R}, \tag{26}$$

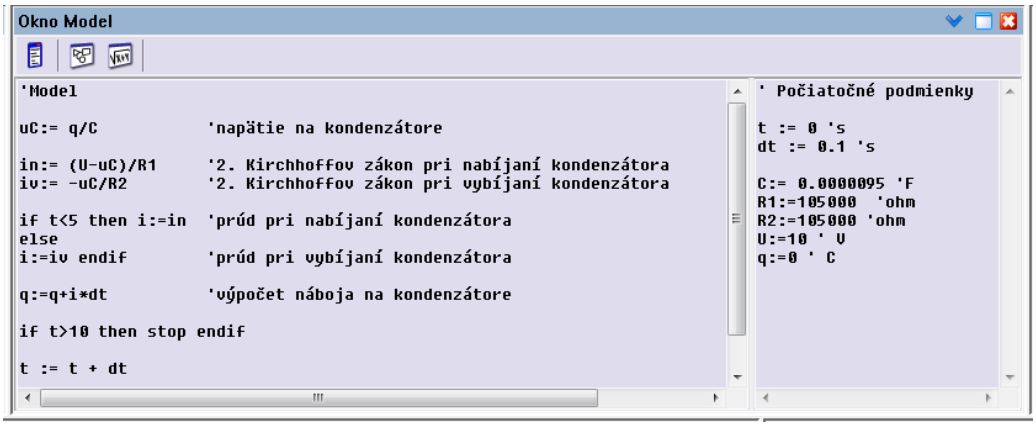
$$t = t + \mathrm{d}t, \tag{27}$$

kde rovnica (26) bola odvodená z 2. Kirchhoffovho zákona pre tento prípad

$$U = u_R + u_C = Ri + \frac{q}{C}. \tag{28}$$

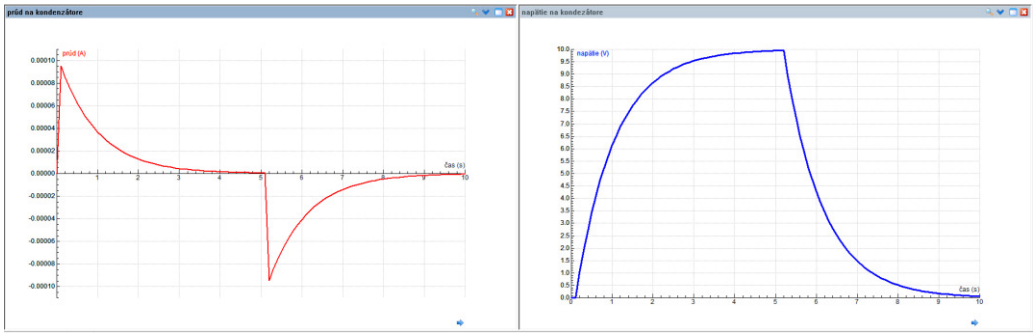
V prípade vybíjania kondenzátora je rovnica pre prúd daná

$$i = \frac{-q}{RC}. \tag{29}$$



Obr. 11 Príklad dynamického modelovania elektrického deja.

Výsledný model zobrazujúci súčasne nabíjanie a vybíjanie kondenzátora je na obr. 11 a grafické priebehy prúdu a napätia od času na obr. 12 [5].



Obr. 12 Priebeh prúdu (červená krivka) a napätia (modrá krivka) pri nabíjaní a vybíjaní kondenzátora v obvode s rezistorom.

III. ZÁVER

Ako bolo v článku ukázané, dynamické modelovanie je vhodnou formou skúmania fyzikálnych procesov, ktoré sa menia v čase. Prostredie Coach využíva pri riešení diferenciálnych rovníc popisujúcich tieto procesy Eulerovu metódu, ktorá poskytuje pomerne presne hodnoty. Ovládanie prostredia a jednotlivých funkcií je intuitívne a tvorba modelov je ľahko pochopiteľná. Prostredie Coach predstavuje ďalší nástroj použitia vo vzdelávaní v oblasti matematiky, informatiky a prírodovedných predmetov.

LITERATÚRA

[1] https://cma-science.nl/home-2 (citované 3.7.2025)
[2] https://golfaky.cz/sk/content/5-golfova-lopticka-opis-technologia-golfovej-lopticky (citované 3.7.2025)
[3] D. Olčák, Z. Gibová, *Mechanický pohyb častice a telesa*. Košice: TU, 2014, s. 26-30.
[4] Z. Gibová, *Dynamika hmotného bodu. E – skriptá.* Košice: TU, 2014. https://zuzana.gibova.website.tuke.sk/index1.htm (citované 3.7.2025)
[5] O. Lepil, L. Richterek, *Dynamické modelovaní*. Olomouc: Repronis, 2007, s. 9-11, 17-19, 43-46.
[6] P. Demkanin, *Počítačom podporované prírodovedné laboratórium*. Bratislava: Knižničné a edičné centrum FMFI UK, 2006. s. 37–39, 81, 82, 88–90.
[7] Z. Gibová, *Fyzikálne meranie*, Košice: TU, 2019, s. 108, 116-120, 123. https://zuzana.gibova.website.tuke.sk/files/Fyzikalne-meranie.pdf (citované 3.7.2025)

Konštrukcia aktívneho balanséra na báze spínaných kondenzátorov

[1]*Daniel MARCIN,* [2]*Milan LACKO*

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

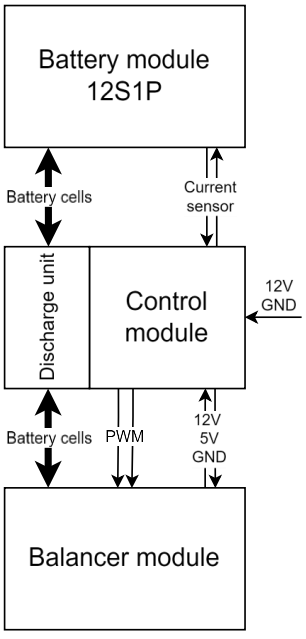
[1]daniel.marcin@tuke.sk, [2]milan.lacko@tuke.sk

***Abstrakt* — Táto práca sa zaoberá aktívnym balansovaním článkov na báze kondenzátorov s cieľom zvýšiť výkon a životnosť batériových systémov pre elektromobily. Pomocou simulácií v prostredí MATLAB Simulink sme analyzovali tri typy balansovacích obvodov: spínané kondenzátory, dvojúrovňové spínané kondenzátory a topológiu s prídavným kondenzátorom. Pre overenie výsledkov simulácií v praxi sme vyvinuli reálny prototyp balanséra, čo si vyžadovalo návrh a výrobu vlastného hardvéru. Ten pozostáva z batériového modulu, riadiaceho modulu a modulu balanséra. Hlavným zameraním tohto článku je návrh a implementácia modulu balanséra.**

***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 [1], [2]. 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. Existujú rôzne topológie zapojení s použitím rôzneho množstva kondenzátorov. Tieto topológie sú jednoduché na riadenie, ale môžu byť časovo zdĺhavé. Výsledky simulácií ukázali význam takéhoto riešenia, preto sme simulácie potrebovali posunúť o krok vpred. Pre overenie výsledkov v praxi bolo preto potrebné navrhnúť vhodný hardvér, čomu sa v stručnosti venujeme v tomto článku [3].



Obr. 1 Batériový modul, ovládací modul, prepojenia medzi modulmi

II. HARDVÉROVÉ POŽIADAVKY A ANALÝZA RIEŠENIA

Na overenie presnosti simulácií sme navrhli a zostrojili hardvér pre modul balanséra. Kvôli modularite, prispôsobiteľnosti a jednoduchým úpravám je hardvér rozdelený do troch samostatných DPS modulov: batériový, riadiaci a balansovací modul. Tieto moduly sú prepojené vhodnou kabelážou. Riadiaci modul má kľúčovú úlohu v monitorovaní v reálnom čase, tvorbe riadených nevyvážeností medzi článkami a zbere údajov počas nabíjania a vybíjania. Tiež ovláda činnosť modulu balanséra. Výhodou tohto návrhu je, že jeden riadiaci modul podporuje viaceré konfigurácie balansovacích obvodov. Bloková schéma zapojenia je znázornená na Obr. 1.

A. Batériový modul

Batériový modul obsahuje držiaky článkov zapojené do série, pričom každý z nich je pripojený ku konektoru určenému na meranie napätia a balansovanie. Tieto prepojenia sú chránené SMD poistkami s menovitým prúdom 3 A. Celý modul je dodatočne istený poistkou s hodnotou 6 A. Do zapojenia je integrovaný prúdový senzor ACS712-05, ktorý meria prúd v rozsahu od -5 A do 5 A pri napájaní 5 V a je pripojený cez konektor k mikrokontroléru. Tento návrh pomocou držiakov zároveň umožňuje jednoduchú výmenu článkov. Použité sú články APR18650M1B na báze LFP od spoločnosti Lithium Werks [4].

B. Ovládací modul

Riadiaci modul pozostáva z dvoch hlavných častí: riadiacej jednotky a vybíjacej jednotky. Riadiaca časť obsahuje mikrokontrolér AVR128DB64 (128 kB flash, 24 MHz), obvody na spracovanie signálov, RTC modul DS1307, pamäť pre záznam údajov, programovací konektor a rozhrania pre externé zariadenia. Mikrokontrolér komunikuje cez I2C a SPI a podporuje 22 diferenciálnych 12-bitových ADC kanálov, ktoré zabezpečujú presné meranie napätia a teploty článkov [5].

Používateľské rozhranie pozostáva zo štyroch tlačidiel, potenciometra a LCD displeja (20x4 znaky s modulom PCF8574). RTC modul zabezpečuje presné meranie času aj počas výpadku napájania vďaka záložnej batérii. Dôležité údaje ako napätie, prúd a časové značky sa ukladajú na microSD kartu cez SPI rozhranie.

Systém je napájaný 12 V zdrojom, pričom pre logické obvody sa napätie znižuje na 5 V pomocou stabilizátora. Vybíjacia jednotka zabezpečuje nastavenie počiatočných podmienok článkov prostredníctvom spínaných rezistorov (3,3 Ω / 5 W typu AX5W-3R3) ovládaných relé (AZ943-1AH-12DE), čím sa dosahuje vybíjací prúd 1C (1,1 A).

Sériové zapojenie 12 článkov spôsobuje celkové napätie 39 V, čo komplikuje presné meranie napätí jednotlivých článkov. Pôvodne boli použité dva multiplexory HEF4067BT.653, no pre nefungujúce zapojenie bol systém vylepšený o meraciu dosku INFINEON TLE9012DQU, ktorá zabezpečuje snímanie napätia každého článku. Namerané údaje následne poskytuje cez UART.

C. Balansovací modul

Modul balanséra (Obr. 2) pozostáva predovšetkým zo spínačov a kondenzátorov a je navrhnutý s dôrazom na flexibilitu a prispôsobivosť pre budúce výskumné aplikácie. Obsahuje viacero topológií aktívneho balansovania článkov založených na kondenzátoroch, ktoré sú integrované na jednej DPS. Pomocou prepojok je možné upraviť konfiguráciu obvodu – buď zmenou typu topológie alebo počtu zapojených kondenzátorov. Pre každý článok sú použité dva unipolárne MOSFET tranzistory typu N (P140LF4QL), ktoré riadia tok prúdu v oboch smeroch. Pripojenie jednotlivých článkov k balansovaciemu obvodu je riešené viacpinovými konektormi, ktoré zabezpečujú pevné a efektívne prepojenie s ostatnými modulmi.

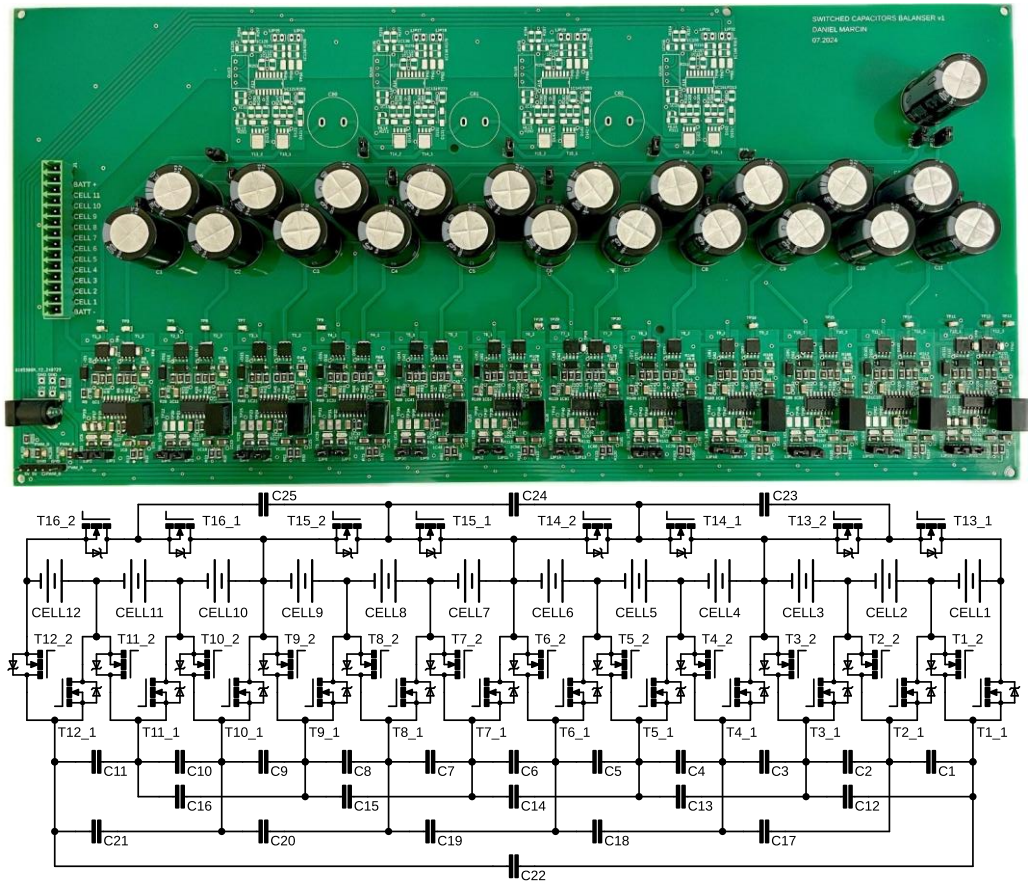
Aby sa zabezpečilo presné meranie počas výskumu balansovania, pomocné obvody nie sú napájané z batériových článkov, ale z dvoch samostatných napájacích zdrojov. MOSFETy sú napájané 12 V zdrojom, zatiaľ čo ovládacie obvody sú napájané z 5 V zdroja. INFINEON meracia doska je napájaná priamo z batériového packu. Samotné balansovanie je riadené dvoma komplementárnymi PWM signálmi (PWM_A a PWM_B) generovanými mikrokontrolérom.

Každý pár MOSFET tranzistorov je riadený izolovaným dvojkanálovým budičom (UCC21330BDR) s nastaviteľnou bezpečnostnou medzerou. Tento budič využíva 5 V napájanie pre logiku vstupu a zároveň slúži ako referenčné napätie pre PWM signály. Na vstupoch PWM aj napájania sú zapojené RC filtre podľa odporúčaní výrobcu. Bezpečnostná medzera približne 500 ns je nastavená odporom R_{DT} s hodnotou 56 kΩ. Výkonová časť je napájaná pomocou izolovaných DC/DC meničov 12 V – 12 V, ktoré zároveň nabíjajú bootstrap kondenzátory. Zapnutie MOSFET-ov je riadené odporom R_{ON} (15 Ω) a vypnutie je zabezpečené Schottkyho diódou a odporom R_{OFF} (3,3 Ω). Celá schéma je replikovaná 12-krát.

Ako bolo uvedené vyššie, pomocou prepojok je možné aktivovať rôzne topológie aktívneho balansovania na báze kondenzátorov. Podporované sú štyri rôzne konfigurácie:

- Spínané kondenzátory (SC)
- Dvojúrovňové spínané kondenzátory (DTSC)
- Dvojúrovňové spínané kondenzátory s pridaným kondenzátorom (DTSCAC)
- Modulárne spínané kondenzátory (MSC)

Tento systém zabezpečuje vysokú modularitu pre rôzne experimentálne usporiadania a zvyšuje výskumné možnosti vďaka jednoduchému prepínaniu medzi jednotlivými topológiami v rámci jednej hardvérovej platformy. Návrh kladie dôraz na flexibilitu, presnosť a efektivitu, čo z neho robí vhodný nástroj pre pokročilé štúdie v oblasti manažmentu batérií a aktívneho balansovania. V závislosti od konkrétnej balansovacej stratégie môžu byť kondenzátory zapojené paralelne cez jeden, dva alebo tri články a preto sú použité kondenzátory s rôznymi napäťovými úrovňami.



Obr. 2 DPS a zapojenie balansovacieho modulu

III. ZÁVER

Táto práca sa zaoberala metódami aktívneho balansovania článkov pomocou kondenzátorov, určenými pre batériové systémy elektrických vozidiel. Na základe predchádzajúcich simulácií a teoretických poznatkov [6] bol navrhnutý a zostrojený hardvérový prototyp určený pre praktické overenie týchto konceptov. Balansér umožňuje prepínanie medzi rôznymi topológiami balansovania: SC, DTSC, DTSCAC a MSC.

Hardvér bol konštruovaný s ohľadom na modularitu a flexibilitu pre budúci výskum. Zavedenie meracej dosky pre napätie zvýšilo presnosť monitorovania a zabezpečilo spoľahlivý zber údajov počas testovania. Vybíjacia jednotka v riadiacom module umožňuje konzistentnú prípravu článkov a zabezpečuje tak opakovateľné testovanie.

Úvodné testy ukázali, že balansér dokáže efektívne vyrovnávať napätie v reálnom čase, čo predstavuje pevný základ pre ďalší výskum. Vďaka modulárnemu návrhu je možné jednoducho, pomocou prepojok prepínať medzi jednotlivými stratégiami balansovania.

Nasledujúcou fázou je vývoj softvéru s cieľom implementovať automatizované balansovanie a hlbšia analýza energetickej efektívnosti jednotlivých topológií. Ďalšie testovanie v rôznych prevádzkových podmienkach prispeje k zlepšeniu výkonnosti, účinnosti a bezpečnosti batérií pre aplikácie v elektromobiloch.

Záverom možno konštatovať, že vyvinutý hardvér predstavuje silnú a flexibilnú platformu pre

rozvoj techník aktívneho balansovania článkov na báze kondenzátorov a tvorí solídny základ pre budúci výskum i praktické nasadenie v systémoch manažmentu batérií elektrických vozidiel.

Poďakovanie

Táto práca bola podporená z projektov APVV-18-0436, VEGA 1/0363/23 a KEGA 059TUKE-4/2024.

Referencie

[1] M. A. Hannan, Md. M. Hoque, A. Hussain, Y. Yusof, and P. J. Ker, 'State-of-the-Art and Energy Management System of Lithium-Ion Batteries in Electric Vehicle Applications: Issues and Recommendations', *IEEE Access*, vol. 6, pp. 19362–19378, 2018, doi: 10.1109/ACCESS.2018.2817655.

[2] P. Nur Halimah, S. Rahardian, and B. A. Budiman, 'Battery Cells for Electric Vehicles', *International Journal of Sustainable Transportation Technology*, vol. 2, no. 2, pp. 54–57, Oct. 2019, doi: 10.31427/IJSTT.2019.2.2.3.

[3] D. Andrea, *Battery Management Systems for Large Lithium-Ion Battery Packs*. Artech House, 2010.

[4] Lithium Werks Inc., '18650 Lithium Ion Power Cell'. Accessed: Nov. 29, 2023. [Online]. Available: https://a123batteries.com/product_images/uploaded_images/APR18650M1B.pdf

[5] Microchip Technology Inc., ' AVR128DB28/32/48/64'. Accessed: Jun. 02, 2025. [Online]. Available: https://ww1.microchip.com/downloads/aemDocuments/documents/MCU08/ProductDocuments/DataSheets/AVR128DB28-32-48-64-DataSheet-DS40002247.pdf

[6] D. Marcin, M. Lacko, D. Bodnár, and L. Pancurák, 'Capacitor-Based Active Cell Balancing for Electric Vehicle Battery Systems: Insights from Simulations', *Power Electronics and Drives*, vol. 9, no. 1, pp. 317–330, Jan. 2024, doi: 10.2478/pead-2024-0020.

Mental Health Support Application Designed for Local Data Processing and User Autonomy

[1]*Simona MICHALIKOVÁ*, [2]*Marek HORVÁTH*

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

[1]simona.michalikova@student.tuke.sk, [2]marek.horvath@tuke.sk

Abstract **– The article describes the development of a mobile application designed to support mental well-being using established methods for stress reduction, formation of healthy daily habits, guided journaling, and visualization of emotional states. The solution prioritizes user privacy and ensures full offline operation, even in critical situations. The design is informed by scientific literature, expert consultations, and feedback from the intended user group. The application responds to persistent challenges including limited access to professional care, social stigma, and the daily pressures affecting mental health by providing accessible and interactive self help tools.**

Keywords **– Mental Health, Mobile Application, React Native, Stress Management, User Privacy**

I. Introduction

Mental health is a key component of overall health and well-being, encompassing emotional, psychological, and social balance that enables individuals to manage everyday stress, make decisions, work productively, and maintain relationships. The *World Health Organization (WHO)* defines mental health as a state in which individuals realize their own potential, cope with normal life stresses, work productively, and contribute to their community [1]. Despite this positive framing, millions of people worldwide face various forms of mental disorders, with one in eight people globally affected and depression alone impacting more than 280 million individuals [2]. The COVID-19 pandemic further intensified this trend, causing a 25% rise in the prevalence of depression and anxiety in its first year. Meta analyses suggest that nearly half the world's population will experience at least one mental disorder by age 75 [3], highlighting that mental health is now a major global public health challenge. Alongside high prevalence, a shortage of mental health professionals such as psychologists and psychiatrists persists worldwide, resulting in long waiting times, limited access to therapy, and unaffordable costs for many people [4], [5]. Stigma associated with mental illness remains a significant barrier, discouraging individuals from seeking help due to fears of judgment or social rejection [6].

Given these persistent obstacles, there is growing interest in the use of technology as a complementary tool for mental health care. Mobile applications and online platforms offer users accessible ways to engage in self monitoring, journaling, relaxation techniques, and psychoeducational activities. These digital solutions are generally accessible, anonymous, low-cost, and allow for a high degree of personalization. Systematic reviews confirm that digital apps can effectively support mental well-being and serve preventive functions [7]. The potential of these technologies also lies in their capacity to collect continuous data on mood and behavior. Devices such as smartphones and wearables can capture real-time data, sometimes identifying early warning signs of mental health deterioration that would be missed by traditional diagnostics [8]. For both individuals and clinicians, such tools represent a practical extension of support. Nevertheless, digital interventions also introduce new challenges, particularly regarding data privacy, informed consent, and the transparency of underlying algorithms [9]. The need to validate their effectiveness and avoid digital exclusion for vulnerable groups is well documented in the literature [10].

This article presents the development of a mobile application for mental health support, focusing on daily mood tracking, structured journaling, meditation playback, breathing exercises, daily challenges, and progress visualization. The solution is designed as a frontend only application

with no user registration or internet connection required, ensuring that all data is processed and stored locally to maximize user privacy. The primary goal is to provide a digital tool addressing the ongoing problem of limited access to mental health care, with design decisions guided by user needs, technical limitations, and ethical principles. The resulting prototype functions as a proof of concept and a starting point for further expansion and research on the effectiveness of digital mental health interventions.

II. Related Work

Recent advances in digital mental health support are reflected both in research and in practice, with growing attention given to mobile technology, artificial intelligence, and practical solutions that respect user privacy and promote accessibility. Many studies confirm that digital applications can provide effective assistance for stress, anxiety, and depression management, particularly when they are simple to use and available without restrictions. *The National Institute of Mental Health* reports that technologies such as mobile apps, SMS services, wearable devices, artificial intelligence tools, and virtual reality platforms have the potential to complement traditional therapy or offer support to individuals who do not have direct access to mental health professionals [11]. These digital tools allow for ongoing monitoring of symptoms and are accessible at any time, which is especially valuable in crisis situations or when professional help is not available.

Various studies also point out both the benefits and the limitations associated with digital solutions for mental health. Naslund and colleagues [12] emphasize the reduction of access barriers, but at the same time highlight concerns about the loss of human contact, the necessity of professional guidance, and the need to protect user privacy. Other authors [13], [14] draw attention to the importance of independent evaluation, transparent operation, and secure handling of sensitive data. Although there is a wide variety of digital tools available, only a small number have undergone rigorous scientific validation. Open questions remain regarding personalization, interactivity, and sustaining user engagement. The use of artificial intelligence, including chatbots and systems that provide recommendations or motivational messages, is increasing. Still, experts agree that such technologies must be developed with attention to ethical standards, clear functioning, and that their role should remain supplementary to professional care [15]. In this context, the chatbot in our application is used strictly for reflective support, never as a substitute for diagnosis or therapy.

Recent research [16] further addresses the importance of secure data storage and the minimization of unnecessary data transfer. Studies focusing on mobile applications and wearable technology demonstrate that continuous collection of behavioral data can help identify early changes in mental state, but often relies on external servers and brings questions about ownership and control over personal information. To respond to these risks, our application deliberately avoids any synchronization with remote storage and relies entirely on local data handling. Effective applications in this area are usually characterized by ease of use, clear psychological foundations, motivational content, and the ability to support emotional regulation. Collaboration between developers, clinical experts, and end users, as well as systematic testing, are repeatedly recommended in the literature [17], [18], [19]. The features implemented in this project, such as mood tracking, structured journaling, daily challenges, relaxation tools, and elements of gamification, are chosen both according to scientific recommendations and direct feedback from the target user group, with the intention of combining proven methods into one accessible and trustworthy digital solution.

III. Methodology

This project combines expert guidance, practical feedback, and an emphasis on privacy at every stage. The aim was to create an independent and interactive tool to support mental well-being through features like mood tracking, guided reflection, breathing exercises, stress management activities, and motivational tasks. The design and development process was shaped by repeated consultations with a clinical psychologist and iterative feedback from target users. Consultations with an expert in emotional self regulation and behavioral change contributed insights into digital versions of proven practices. These included mood journaling, daily affirmations, and ways to keep users engaged over time. An additional chatbot based on large language models was included to provide motivational support in a strictly supportive, not clinical role. All data collection was strictly limited to what users entered directly into the application. No external datasets or analytics were used. Entries such as mood ratings, journal notes, completed tasks, reminders, and chatbot messages are stored locally. Meditation audio files are sourced from openly licensed platforms, and motivational tasks were curated and saved in a structured JSON file. This design provided full control over the app's data structure and ensured compliance with ethical and copyright standards.

From a technical perspective, the application is fully implemented on the frontend with no server-side infrastructure. All processing and storage take place on the user's device. This approach

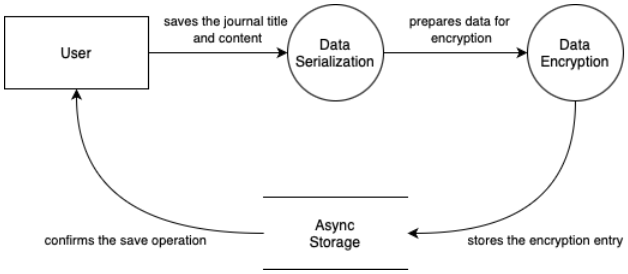


Fig. 1 Data flow diagram of serialization and encryption of sensitive records

was motivated by three main factors. Sensitivity of personal data, the need for offline use, and reduced operational costs. Users do not need to register, log in, or synchronize data with any cloud service. Sensitive data such as journal entries and mood ratings are encrypted with the AES-256 algorithm. Data is decrypted and loaded into memory only as needed, so no unencrypted data is stored on the device outside of runtime. The serialization and encryption process is shown in Figure 1.

To deliver meditation audio, the application streams files from public Google Drive links. This allows updates to the audio library without modifying the application package. Google Drive was chosen for reliability and integration. The separation between local and remote data handling is shown in Figure 2.

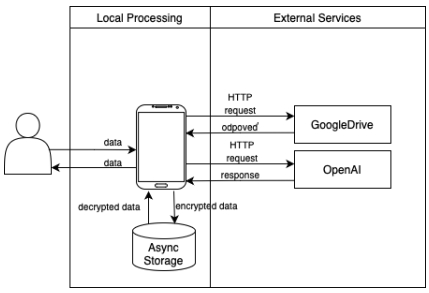


Fig. 2 Local and external data interactions

For user experience, an interactive prototype was evaluated with the System Usability Scale (SUS) by five experts. This prototype covered the main application screens, navigation, transitions, and visual design. The usability score was 80, suggesting a high level of intuitiveness and meeting user expectations. Based on the feedback, improvements were made to navigation, icon clarity, and color selection.

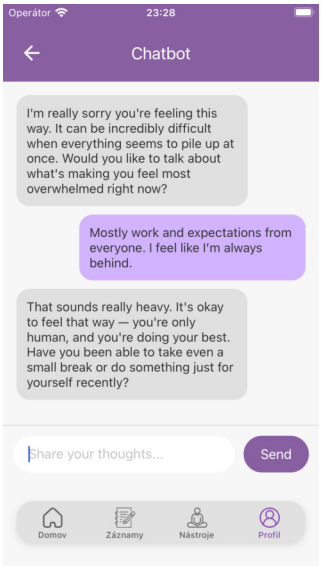


Fig. 3 Chat interface with the AI assistant providing motivational responses

The AI chatbot is integrated via secure HTTPS calls to the OpenAI API. Each call is stateless, so no conversation history is retained. The API key is managed via environment variables using dotenv, never exposed in the public code. This preserves secure, one-time communication in a client-only environment. The chatbot provides short, motivational feedback in response to user input. The chat interface is shown in Figure 3.

The application footprint was measured with simulations. Regular use, such as daily mood logs, weekly journal entries, and challenge completions, keeps monthly data usage below 100 kilobytes. This ensures the app remains lightweight even with long-term use and is suitable for devices with limited storage or connectivity. In summary, the methodology emphasizes simplicity, privacy, and relevance. Expert input, a purely frontend architecture, and encrypted local storage made it possible to develop a maintainable and self contained mobile tool to help users manage emotional well-being. The modular structure and scalability of the solution allow for future updates such as additional languages, content, or optional sync, while still respecting core privacy and offline principles. This framework can be adapted to future wellness applications in mobile environments.

IV. Results

To evaluate the effectiveness and relevance of the proposed mobile application, a two-phase user survey was conducted before and after implementation. The initial survey identified the target group's mental health challenges, coping strategies, and expectations from digital tools, while the second part focused on user experience, perceived usefulness, and emotional impact following app testing. The pre-implementation survey involved participants primarily aged 18–24 (72.7%), with 63.6% identifying as female. Most respondents reported experiencing stress or anxiety either occasionally (36.4%) or several times a week (27.3%). The most common coping strategies included physical activity, talking to close ones, breathing exercises, and journaling. Notably, 63.7% had already used mental health apps, mostly for meditation, breathing, and journaling. Participants also expressed interest in features like daily challenges, reminders, anonymous chat, and mood tracking, which directly influenced the app's design. Figure 4 summarizes the preferred coping strategies, highlighting a clear preference for simple, self-guided practices. These insights informed the inclusion of modules that promote emotional self-regulation and awareness.

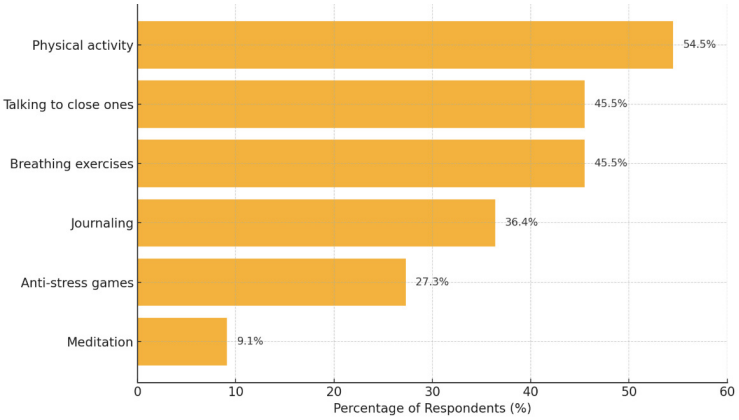


Fig. 4 Preferred coping strategies for stress and anxiety (pre-study survey)

After development, users tested the app over several days and provided feedback. A majority (81.8%) found the app *"very easy"* to use, while 63.6% described its design as *"modern and pleasant"*. The most used functionalities included breathing exercises, mood tracking, daily challenges, and meditation sounds, all of which align with initial user expectations. Table 1 presents average usefulness ratings on a 5-point scale. The highest scoring features were daily challenges, reminders, and breathing exercises, while even less frequently used modules, such as meditation and games, were received positively.

In addition to subjective usefulness, the most frequently used features during testing are illustrated in Figure 5. These usage patterns confirm alignment between user expectations and actual interaction behavior, especially regarding modules focused on self-regulation and motivation.

Users also shared feedback on privacy, gamification, and emotional effects. All participants appreciated the app's offline functionality and lack of registration, contributing to a greater sense of safety and trust. The gamified elements, such as points and trophies, were perceived positively by all respondents, with 36.4% describing them as *"very motivating"*. In terms of emotional impact, 72.7% of users reported feeling *"somewhat calmer"* after using the app, and

Feature	Average Rating (1–5)
Daily challenges / motivational tasks	4.8
Reminders (notifications)	4.8
Breathing exercises with animation	4.6
Mood tracking and visualization	4.5
Help contacts	4.3
Structured journaling and diary	4.2
Chatbot	4.1
Meditation sounds	4.0
Anti-stress games	4.0

Table 1
User ratings of app features after testing (1 = lowest, 5 = highest)

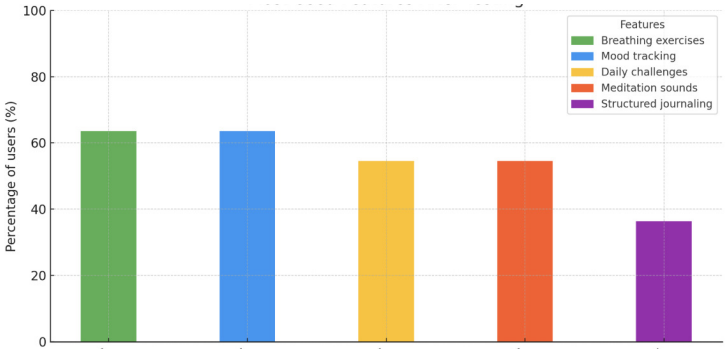


Fig. 5 Most frequently used app features after testing

27.3% felt *"definitely calmer"*. No negative reactions were reported, and all participants expressed willingness to recommend the application to others.

While breathing exercises and mood tracking showed the highest engagement, meditation sounds and daily challenges were also frequently used. Modules like the chatbot and diary, though rated positively, saw lower usage, suggesting potential for improved contextual integration or design adjustments. Features focused on self regulation, such as breathing exercises, mood tracking, and motivational tasks had the greatest perceived impact. The strong reception of structured routines highlights users' preference for consistency in mental wellness habits. From a technical standpoint, diagnostics using *Flipper* confirmed stable performance across devices. *Chatbot API* response times averaged under two seconds, and all main modules functioned reliably during repeated testing. Minor issues, such as occasional animation lag or audio delay under poor connectivity, were observed but did not significantly affect usability. Results confirm that a lightweight, privacy preserving mobile application can effectively support users emotional well-being. The high satisfaction with functionality, coupled with strong emotional responses and minimal technical issues, highlights the potential of low threshold digital tools for everyday mental health support.

V. Discussion

The project results indicate that the implemented mobile application can support users mental well-being by providing accessible tools on personal devices. During development, the process was guided by current knowledge in mental health, user feedback, and a focus on comfort, clarity, and privacy. Testing confirmed technical stability and reliable operation during repeated feature use. Among all functionalities, users most valued breathing exercises, reminders, and daily challenges, which is in line with recommendations from behavioral therapy and stress management literature. Users also reported that regular app use encouraged more effective daily self-reflection. A clear advantage noted by users was the ability to use the application without registration, with all data stored locally. This feature addresses concerns about misuse of sensitive personal information and increases user trust, which is essential for mental health tools. Offline operation is particularly valuable for users in crisis or with unstable internet connections. Feedback highlighted several directions for further development:

- The chatbot feature was used less often, possibly due to its basic conversational scope. Future versions could benefit from more advanced models or offer different modes (such as support, motivation, or factual information) to increase engagement.

- Meditation audio playback depends on internet connectivity. Implementing offline audio downloads or a hybrid storage model would increase reliability.
- Mood tracking and statistics could be enhanced with trend analysis, pattern detection, and personalized recommendations or notifications.
- Gamification elements (points, trophies) could be extended by unlocking new exercises, providing feedback, or offering personalized advice based on achieved milestones.
- The current solution, built on local storage, is efficient and secure for prototypes or personal use. If scaling for a larger user base, future updates should consider secure cloud storage and multidevice synchronization.
- User adaptability is important. Although tested with a limited group, the app features are versatile enough for students and working adults. To generalize further, future work should examine variables like age, cultural background, or digital literacy, and personalize onboarding or content according to user profile.
- Selected modules could serve as integration points for blended care, for example, enabling mood data export or journaling to share with professionals.

From a technical and ethical perspective, the project demonstrates that evidence-based psychological methods can be implemented in a mobile environment without sacrificing privacy or usability. AES-encrypted local storage, minimal permissions, and the absence of mandatory accounts are in line with ethical principles for mental health technology and make the app a potential model for low-risk interventions in privacy sensitive or offline settings positive feedback on the clear structure and exercise variety confirms the benefit of balancing guided activities with space for self-reflection. This balance can help reduce user resistance and encourage engagement, especially among those less familiar with mental health tools. For stronger validation, longitudinal studies would be needed to assess sustained impact and user retention.

The presented solution offers an accessible and practical way to support mental well-being. It demonstrates that even with modest resources, it is possible to develop a mobile app with tangible social value. Future work may include expanding the user base, adding language support, or integrating with healthcare systems or therapeutic services, according to user needs and feedback.

VI. Conclusion

This project confirms that focused attention on accessible digital tools for mental health is both needed and feasible. The experience gained from designing and testing a locally stored mobile application highlights how even simple interventions can lower barriers to care, support user autonomy, and respond flexibly to different needs. Further research and development in this area remain essential, especially given persistent gaps in mental health services and the importance of privacy. Prioritizing user feedback, ethical standards, and adaptability in digital solutions can lead to broader societal impact, better support for prevention, and the creation of scalable tools that are inclusive, secure, and ready for integration into healthcare systems or broader well-being strategies.

Acknowledgment

This work was supported by project KEGA No. 061TUKE-4/2025 Building Bridges between University and High School ICT Education.

References

[1] W. H. Organization, *Promoting Mental Health: Concepts, Emerging Evidence, Practice*. Geneva: World Health Organization, 2004.
[2] ——, "Mental health: strengthening our response," 2022, accessed: 2025-01-26. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response
[3] R. C. e. a. Kessler, "Age of onset and cumulative risk of mental disorders: A cross-national analysis of population surveys from 29 countries," *The Lancet Psychiatry*, vol. 10, no. 7, pp. 522–531, 2023.
[4] W. H. Organization, "Mental health workforce statistics," 2020. [Online]. Available: https://www.who.int
[5] P. S. e. a. Wang, "Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the who world mental health surveys," *The Lancet*, vol. 370, no. 9590, pp. 841–850, 2007.
[6] P. W. Corrigan and A. C. Watson, "The paradox of self-stigma and mental illness," *Clinical Psychology: Science and Practice*, vol. 9, no. 1, pp. 35–53, 2002.
[7] A. Almuqrin, R. Hammoud, I. Terbagou, S. Tognin, and A. Mechelli, "Smartphone apps for mental health: systematic review of the literature and five recommendations for clinical translation," *BMJ Open*, vol. 15, no. 2, p. e093932, 2025.
[8] G. Gillett and K. E. A. Saunders, "Remote monitoring for understanding mechanisms and prediction in psychiatry," *Current Behavioral Neuroscience Reports*, vol. 6, no. 1, pp. 51–56, 2019.
[9] J. Torous and L. W. Roberts, "The ethical use of mobile health technology in clinical psychiatry," *BJPsych Bulletin*, vol. 41, no. 6, pp. 322–326, 2017.
[10] G. Fitzpatrick and G. Ellingsen, "A review of the use of digital technologies in mental health," *Sociology of Health & Illness*, vol. 42, no. 2, pp. 334–348, 2020.

[11] National Institute of Mental Health, "Technology and the future of mental health treatment," 2023, accessed: 2025-05-01. [Online]. Available: https://www.nimh.nih.gov/health/topics/technology-and-the-future-of-mental-health-treatment

[12] J. A. Naslund, K. A. Aschbrenner, L. A. Marsch, and S. J. Bartels, "Opportunities for and tensions surrounding the use of technology-enabled mental health services," *Psychiatric Services*, vol. 71, no. 2, pp. 121–127, 2020. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC6946841/

[13] J. Torous, K. Jän Myrick, N. Rauseo-Ricupero, and J. Firth, "The intersection of technology and mental health: enhancing access, care, and well-being," *Current Psychiatry Reports*, vol. 25, no. 8, pp. 467–475, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37602724/

[14] F. Reer, W. Tang, and T. Quandt, "Technology in mental health treatment: A review of teletherapy, mobile health apps, virtual reality, and artificial intelligence," *International Journal of Adolescent Medicine and Health*, 2023. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/ijamh-2023-0113/html

[15] Y. Parean and R. Yehuda, "Artificial intelligence and mental health: Applications and ethical considerations," *Dialogues in Clinical Neuroscience*, vol. 18, no. 2, pp. 203–210, 2016. [Online]. Available: https://www.tandfonline.com/doi/full/10.31887/DCNS.2016.18.2/parean

[16] A. García-Caballero, J. L. Reyes-Ortiz, F. Archidona-Sánchez, J. C. López, and J. Medina-Quero, "Long-term mobile monitoring for mental health: Challenges and opportunities," *Frontiers in Psychiatry*, vol. 11, p. 523, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyt.2020.00523/full

[17] V. Salgadó, A. García-Caballero, C. Rodríguez-Díaz, and S. Seseña, "Technology and mental health: The role of artificial intelligence," *European Psychiatry*, vol. 64, no. S1, 2023. [Online]. Available: https://www.cambridge.org/core/journals/european-psychiatry/article/technology-and-mental-health-the-role-of-artificial-intelligence/7056DC68CF2F384756708C8F06AC6C9D

[18] T. Li, C. Cheung, and P. Lee, "Use of smartphone apps, social media, and web-based resources to support mental health: A user perspective study," *Intersect: The Stanford Journal of Science, Technology, and Society*, vol. 12, no. 3, 2019. [Online]. Available: https://ojs.stanford.edu/ojs/index.php/intersect/article/download/2859/1636/10215

[19] C. Hollis, C. J. Falconer, J. L. Martin, C. Whittington, S. Stockton, C. Glazebrook, and E. B. Davies, "Technological and digital interventions for mental health and wellbeing: An overview of systematic reviews," *Frontiers in Digital Health*, vol. 3, p. 754337, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdgth.2021.754337/full

Simulácia mild hybridného pohonu

[1]*Erik PODRACKÝ,* [2]*František ĎUROVSKÝ*

[1,2] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach

[1]erik.podracky@student.tuke.sk, [2]frantisek.durovsky@tuke.sk

***Abstrakt* — Jednou z možností, ako znižovať emisie vozidiel so spaľovacími motormi, je využitie niektorej z foriem hybridného pohonu. Článok sa zaoberá najjednoduchšou formou mild hybridného pohonu v konfigurácii P0. Opisuje nové funkcie, ktoré tento typ pohonu prináša, a demonštruje správanie sa takéhoto pohonu vo vybraných jazdných režimoch pomocou simulácií.**

***Kľúčové slová* — AVL Cruise M, jazdný cyklus, mild hybridný pohon, štartér-generátor**

I. Úvod

Aktivity človeka, hlavne v oblasti priemyslu a dopravy, vyžadujú veľa energie, ktorá je dnes získavaná predovšetkým spaľovaním fosílnych palív. To je spojené s produkciou skleníkových plynov, ktoré významne prispievajú k otepľovaniu planéty. Preto v poslednej dobe silnejú snahy o zavádzanie technológií, ktorý by produkovali menej emisií. Snaha o dekarbonizáciu ekonomiky sa týka aj dopravy, ktorá sa na tvorbe skleníkových plynov podieľa významnou mierou.

Snaha o znižovanie množstva skleníkových plynov v doprave sa premieta do legislatívy, ktorá núti výrobcov vozidiel postupne znižovať množstvo emisií produkovaných vozidlami. Zatiaľ posledná zmena stanovila množstvo CO_2 produkovaného flotilou vozidiel vyrábaných u daného výrobcu na hodnotu 95 g CO_2/km. Táto hodnota je už taká prísna, že ju len s problémami dokážu splniť vozidlá nižšej triedy s malým objemom spaľovacieho motora (ICE). Vozidlá strednej a vyššej triedy s väčším objemom ICE tieto limity splniť nedokážu. Nakoľko zákazníci preferujú hlavne väčšie vozidlá, museli výrobcovia hľadať cesty ako vyhovieť legislatíve i zákazníkom.

Optimálnym bezemisným vozidlom je elektromobil (EV). (Bezemisným tu budeme označovať vozidlo, ktoré pri jazde neprodukuje CO_2.) Nevýhodou súčasných EV sú trakčné akumulátory, ktoré majú malú mernú energiu, sú ťažké a drahé. To je hlavný dôvod, prečo z pohľadu zákazníka súčasné EV, napriek svojim nesporným výhodám, nedokážu plne nahradiť vozidlá s ICE.

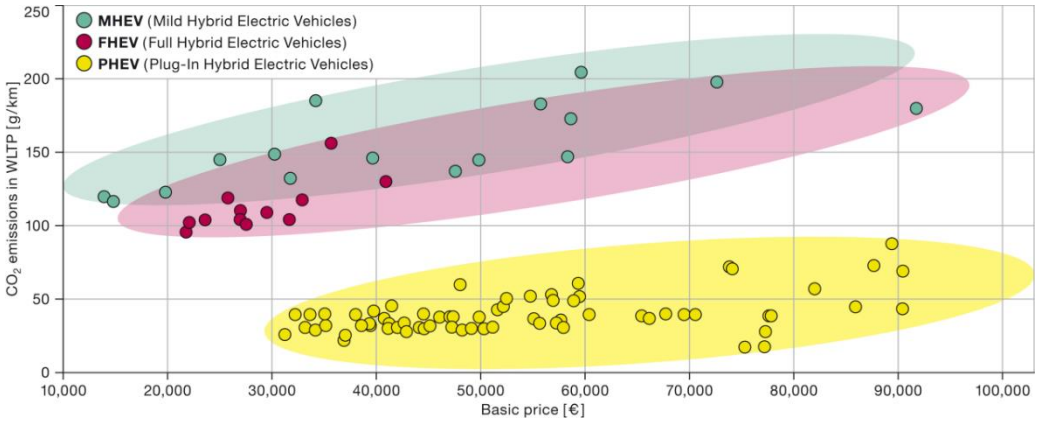
Jedným z riešení tohto problému je hybridný pohon, ktorý spája výhody vozidiel s elektrickým a spaľovacím motorom. Hybridné elektrické vozidlo (HEV) je vybavené ICE, ktorý je doplnený o jeden alebo viac elektromotorov, zásobník elektrickej energie (trakčnú batériu) a príslušné výkonové meniče. Podľa usporiadania pohonného reťazca, výkonu elektromotora a kapacity batérie možno deliť HEV rôznym spôsobom. Pri delení podľa výkonu elektromotora dostaneme tri kategórie vozidiel: mikro hybrid, mild hybrid (MHEV) a full hybrid (FHEV). Základné parametre jednotlivých kategórií aj s možnosťami redukcie CO_2 sú uvedené v Tab.1 [1].

Hoci podľa Tab.1 majú najväčší potenciál na redukciu emisií FHEV, aj tie spĺňajú emisné limity len hranične. Podľa prehľadu hybridných vozidiel na nemeckom trhu [2] súčasné emisné normy spĺňajú len Plug-in-hybridy (P-HEV), ktoré profitujú z možnosti získať časť energie na jazdu z externej elektrickej siete. Nevýhodou P-HEV je však ich vysoká cena (Obr.1).

Tabuľka 1
Charakteristiky hybridných elektrických vozidiel

Typ HEV	Typ batérie	Napätie HV batérie[1)]	Výkon elektromotora	Potenciál redukcie CO_2[2)]
Mikro hybrid	Olovená	12 V	<5 kW	~3%
Mild hybrid	Li-ion	24-48 V	5-20 kW	do 15 %
Full hybrid	Li-ion	200-400 V	20-80 kW	15-30 %

[1)] HV batéria je prídavná batéria určená na ukladanie elektrickej energie na jazdu.
[2)] Potenciál redukcie CO_2 zodpovedá testom na dynamometri, nie reálnej jazde.



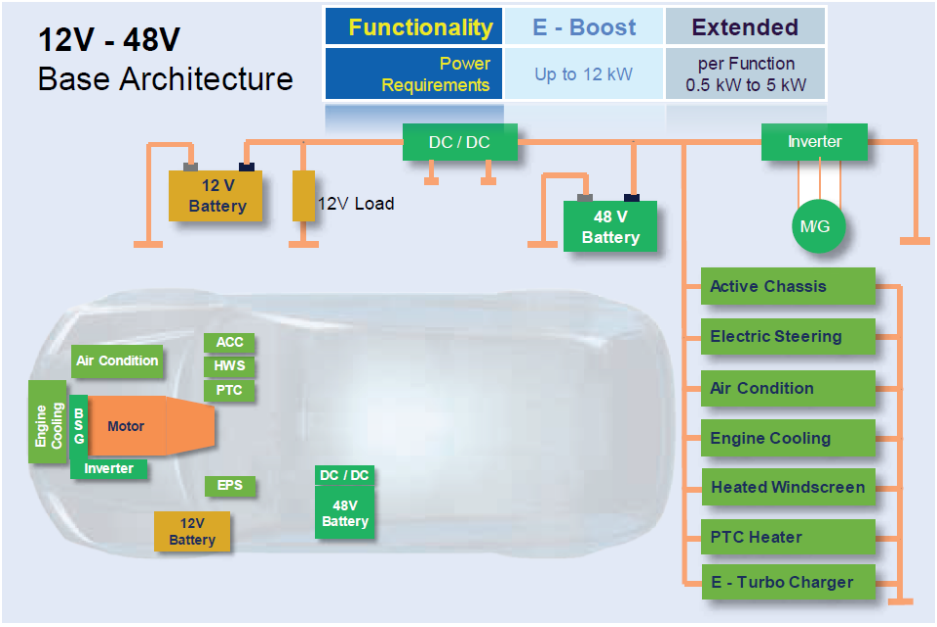
Obr. 1 Prehľad nemeckého trhu s hybridnými vozidlami. Emisie CO_2 sú vyhodnotené podľa jazdného cyklu WLTP (smernica EÚ 2017/1151). Ceny sú z verejne dostupných zdrojov vrátane 19% DPH [2].

Určitým kompromisom medzi cenou vozidla a mierou redukcie CO_2 je využitie mild hybridnej (MH) technológie a duálnej palubnej siete (obr.2). Jej základom je elektromotor (EM), ktorý dokáže pracovať v motorickom alebo generátorickom režime a plní funkciu štartéra aj generátora. Jeho výkon sa pohybuje v rozsahu 5-20 kW. Vzhľadom na vyššie výkony je elektrický systém vozidla doplnený o 48V sieť s vlastným akumulátorom. Do 48 V siete je možné pripojiť aj ďalšie spotrebiče s vyšším výkonom. Na napájanie spotrebičov s napätím 12V sa používa samostatná sieť 12V s vlastným akumulátorom. Energiu pre túto sieť dodáva zo 48 V siete DC/DC menič.

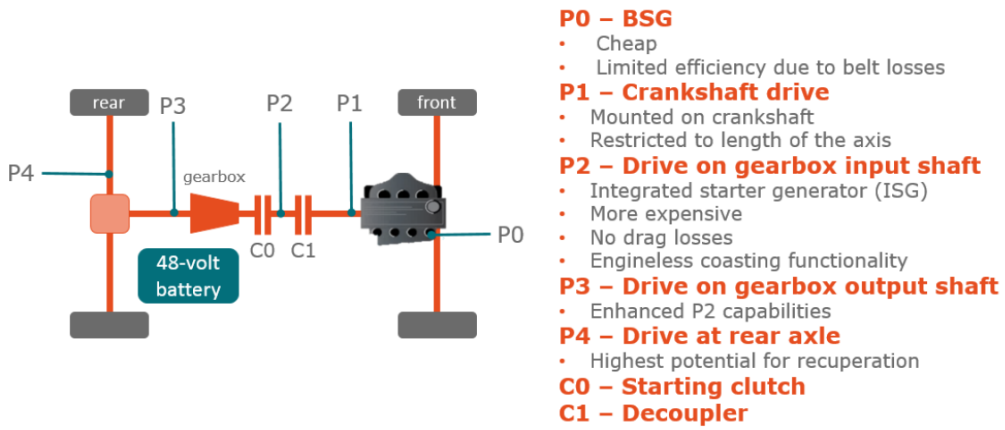
Podľa umiestnenia EM v trakčnom reťazci vozidla sa jednotlivé MH konfigurácie označujú skratkami P0 až P4 (obr.3). Základné vlastnosti pohonu v danej konfigurácii sú uvedené na obr.3 v legende. Vo všeobecnosti platí, že čím je EM bližšie ku kolesám, tým má pohon vyššiu účinnosť a nižšie straty v prevodoch. Na druhej strane, EM pracuje s vyšším momentom, je rozmernejší a ťažší. V konfigurácii P3 a P4 je EM pevne spojený s kolesami, preto nemôže štartovať ICE pri stojacom vozidle, a motor musí byť vybavený samostatným štartérom. V usporiadaní P0 a P1 je EM zase pevne spojený s ICE, takže účinnosť elektromotora je znížená o straty v ICE.

Nakoľko sa jedná o nový trend v pohone vozidiel, bolo v nedávnej dobe publikovaných viacero prehľadových článkov, napr. [1] a štúdií [5]-[7], kde sa analyzoval jednak potenciál jednotlivých konfigurácii na zníženie emisií, zvýšenie účinnosti pohonu vozidla a zlepšenie jeho jazdných vlastností. Ďalej boli robený odhad, ako sa bude vyvíjať cena technológie vďaka postupnej optimalizácii výroby a vývoja komponentov MH pohonu.

Článok je usporiadaný nasledovne. V kapitole 2 sú opísané funkcie, ktoré mild-hybridný pohon poskytuje a v kap.3 je prezentovaný simulačný model MH pohonu s výsledkami simulácií jednotlivých jazdných režimov v programe AVL Cruise M.



Obr. 2 Mild hybridný pohon s duálnou palubnou sieťou 12/48 V [3].



Obr. 3 Konfigurácie mild hybridného pohonu podľa umiestnenia elektromotora [4].

II. MILD HYBRIDNÝ POHON V KONFIRGURÁCII P0

A. Hlavné komponenty mild hybridného pohonu

Konfigurácia P0 obsahuje elektromotor (EM) poháňaný remeňom od kľukového hriadeľa (tzv. Belt Starter Generator – BSG). Remeňom sú zároveň poháňané ďalšie zariadenia, ako je vodné čerpadlo, ventilový rozvod a kompresor klimatizácie, ktoré spolu tvoria skupinu označovanú ako Front End Accessory Drive – FEAD. BSG je umiestnený na mieste pôvodného alternátora a plní zároveň funkciu štartéra. Jeho výkon podľa Tab.1 sa pohybuje v rozsahu 5-20 kW. Prenos takéhoto výkonu cez remeň vyžaduje pevné napnutie remeňa, čo jednak zaťažuje ložiská zariadení v rámci FEAD a zvyšuje trecie straty. Preto je FEAD vybavený adaptívnym napínačom, ktorý zvýši napnutie remeňa len v prípade požiadavky na prenos vyššieho výkonu. Vo funkcii alternátora zvýšené napnutie remeňa nie je nutné. Uvedený výkon (krátkodobý) stačí na naštartovanie zohriateho motora v režime Start-Stop. Na štartovanie studeného motora sa zvykne používať klasický štartér napájaný zo siete 12 V. BSG býva riešený ako synchrónny motor s permanentnými magnetmi a v housingu má integrovaný striedač/usmerňovač.

Vzhľadom na vyššie výkony, s ktorými BSG pracuje, je tento zapojený do siete 48V. K 48 V sieti je možné pripojiť ďalšie spotrebiče vyššieho výkonu, ako sú kompresor pre klimatizáciu, vyhrievanie katalyzátora, elektrické turbo a pod (obr.3). To prináša možnosť redukcie zariadení v rámci FEAD a možnosť ďalších úspor energie a zníženia emisií.

Batéria pre MH pohon býva lítium-iónová s napätím 48V (MHEV batéria). Objavili sa aj MH pohony s batériami 24V a 36V, ale to boli skôr výnimky. MHEV batéria je na rozdiel od batérií pre elektromobily optimalizovaná na vysokú výkonovú hustotu, preto má iné chemické zloženie, napr. lítium-železo-fosfát (LFP). To umožní zmenšiť jej rozmery a cenu bez toho, aby sa redukovali výkonové požiadavky. Kapacita batérie býva cca 0,5 až 1 kWh, s výkonom do 25 kW.

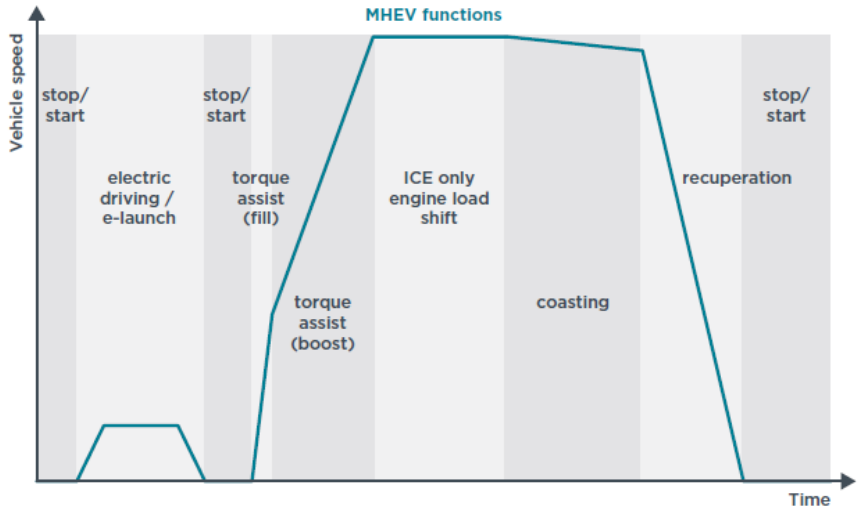
Napájanie spotrebičov v 12V sieti je cez DC/DC menič, ktorého výkon býva cca 3 kW. Aj keď tok výkonu je prednostne zo 48V siete do 12V siete, menič je obojsmerný a v prípade potreby dokáže preniesť výkon aj opačným smerom.

Hoci konfigurácia P0 neposkytuje maximálnu redukciu emisií, má toto usporiadanie viacero výhod. Predovšetkým nevyžaduje zásadné zmeny v usporiadaní motorového priestoru (BSG sa montuje na miesto pôvodného alternátora) a cena za hybridizáciu je prijateľná (cca 1000,- Euro).

B. Doplnkové funkcie mild hybridného pohonu

Doplnenie výkonného elektromotora/generátora a ďalšieho zásobníka energie do pohonného reťazca umožní realizovať nové funkcie a zlepšiť tak jazdné vlastnosti vozidla (obr.4). Jedná sa o nasledujúce funkcie.

Pokročilý štart/stop systém (Advanced Start-Stop system). Voči klasickému štart/stop systému sa ICE nevypína len v čase, keď vozidlo stojí, ale už počas spomaľovania pred zastavením. ICE tak pracuje kratší čas a generuje menej emisií. Vozidlo však musí byť vybavené elektricky poháňanými zariadeniami, ktoré zabezpečia tlak v brzdovom systéme a dostatočné mazanie a chladenie aj v čase, keď je ICE vypnutý. Pri rozjazde je zase možné ICE štartovať neskôr, teda v čase, keď je už vozidlo v pohybe, a motor sa otáča. Rozjazd na malú rýchlosť zabezpečí elektromotor. ICE sa tak štartuje pri vyšších otáčkach motora, čo prináša nižšie emisie a plynulejší štart.



Obr.4 Funkcie mild hybridného pohonu [1]

Jazda nízkou rýchlosťou (Electric launching) – malý výkon BSG je dostatočný na jazdu veľmi nízkou rýchlosťou. Dojazd je limitovaný kapacitou MHEV batérie a rýchlosť možnosťami remeňového prevodu preniesť potrebný výkon.

Prídavný moment (Torque assist) sa uplatní v dvoch jazdných režimoch:

- *pri rozbehu (torque fill)* - nábeh momentu ICE pri nízkych otáčkach je pomalý a kvôli zotrvačnosti mechanických komponentov ICE a obmedzenej rýchlosti prúdenia vzduchu v motore len pomaly narastá. V tejto fáze činnosti ICE tak BSG dokáže krátkodobo zvýšiť moment a pomôcť ICE prekonať slabé miesto v momentovej charakteristike;
- *pri jazde (torque boost)* - umožňuje krátkodobo zvýšiť maximálny moment ICE o moment BSG. Aj keď je trvanie tohto stavu limitované stavom nabitia MHEV batérie, je pre väčšinu jazdných situácií postačujúce.

Posun pracovného bodu ICE (Engine load shift) – pri jazde ustálenou rýchlosťou zvyčajne pracuje ICE v oblasti s nízkym momentom a nízkou účinnosťou. Ak sa v tomto režime aktivuje BSG v generátorickom režime, zvýši sa zaťaženie ICE a jeho pracovný bod sa posunie do oblasti s vyššou účinnosťou spaľovacieho procesu. Vyrobená elektrická energia sa ukladá do batérie.

Plachtenie (Coasting) – tu sa jedná o voľný dobeh vozidla s vypnutým ICE. Prídavný EM sa využíva na pohon príslušenstva cez FEAD. Nakoľko v konfigurácii P0 je BSG pevne spojený s motorom, dá sa tento režim využívať len obmedzene.

Rekuperácia (Recuperation) – pri aktívnom brzdení možno využívať BSG v generátorickom režime. Pri brzdení sa prednostne využíva brzdný moment od BSG a mechanické brzdy kryjú len moment, ktorý je nad možnosti BSG. Okrem generovania elektrickej energie sa takto znižujú aj emisie produkované opotrebením brzdového obloženia.

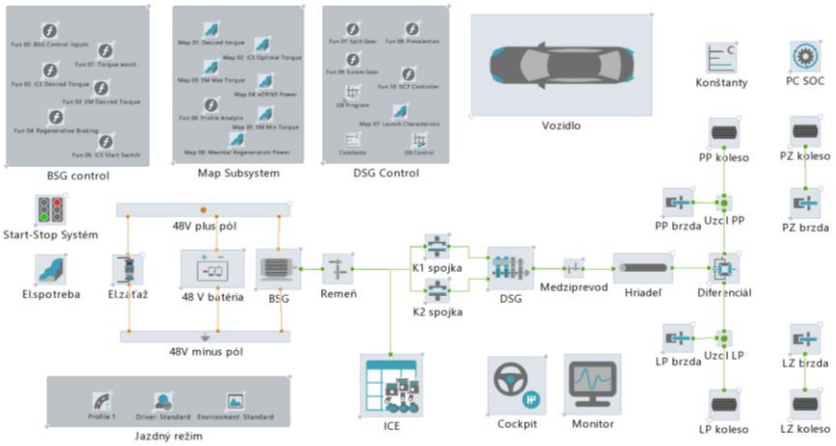
Schopnosť EM pokryť krátkodobo výkonové špice umožní okrem hore uvedených funkcií znížiť výkon ICE (tzv. downsizing). Počas prevádzky potom bude ICE pracovať v pásme s vyššou účinnosťou a nižšími emisiami.

C. Komponenty vozidla pripojené na 48V sieť

Ako vidno na obr.2, k 48V sieti môžu byť pripojené ďalšie spotrebiče s vyšším výkonom, a to aj také, ktoré sú u bežného ICE poháňané mechanicky. Presun pohonu z mechanického na elektrický umožní ich presnejšie riadenie a zapínanie len vtedy, keď je to potrebné. Využitie vyššieho napätia zase zníži straty pri prenose energie od zdroja k spotrebičom. Medzi spotrebiče pripojené k 48V sieti patria napríklad:

Pohon kompresora klimatizácie - elektrický pohon umožní jednak reguláciu otáčok kompresora, ako aj jeho zapínanie nezávisle od toho, či ICE beží alebo nie. Odstránenie remeňového prevodu zase zníži trecie straty ICE a namáhanie ložísk v rámci FEAD.

Elektrické vyhrievanie katalyzátora – časté využívanie štart/stop funkcie hlavne v meste spôsobí, že katalyzátor nie je dostatočne vyhriaty od výfukových plynov, takže jeho schopnosť redukcie emisií je obmedzená. Elektrické vyhrievanie tento problém odstráni. Pri dostatočnej kapacite akumulátora je možné katalyzátor vyhriať na pracovnú teplotu ešte pred prvým studeným štartom ICE.



Obr.5 Simulačná schéma mild hybridného pohonu v programe AVL Cruise M.

Elektrický pohon turbodúchadla – prináša niekoľko benefitov. V prvom rade odstraňuje oneskorenie (turbo lag) spôsobené zotrvačnosťou výfukových plynov využívaných na jeho pohon. Ďalšou výhodou je stabilizácia otáčok turba, ktoré vďaka nezávislému elektrickému pohonu nevykazuje kladnú spätnú väzbu bežnú u klasických preplňovaných ICE.

III. SIMULÁCIA MILD HYBRIDNÉHO POHONU

Simulácia mild hybridného pohonu bola robená pre vozidlo s ICE výkonu 110 kW a maximálnym momentom 250 Nm pri 1500 až 3500 rpm. BSG v motorickom režime poskytoval maximálny moment 50 Nm a v generátorickom režime zaťažoval ICE maximálnym momentom 30 Nm. Vozidlo bolo vybavené dvojspojkovou 7 stupňovou prevodovkou. Simulácie boli robené v programe AVL Cruise M. Simulačná schéma je na obr.5.

Program AVL Cruise M umožňuje multimodálnu simuláciu mechanických, elektrických a termodynamických systémov. Obsahuje rozsiahlu knižnicu komponentov fyzikálnych modelov ICE, mechanických komponentov vozidla, batérií, palivových článkov, elektromotorov a pod. Program umožňuje vyskladať trakčný reťazec vozidla po jednotlivých komponentoch, doplniť riadiace funkcie a komunikáciu medzi nimi, modely správania sa vodiča a jazdné scenáre [8].

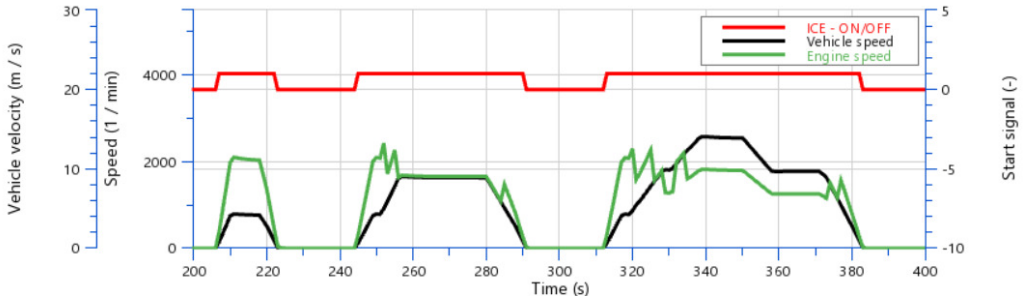
Príspevok obsahuje výsledky simulácií jazdných režimov, o ktoré mild hybridný pohon rozširuje možnosti vozidla s bežným ICE. Ako jazdný cyklus bol použitý NEDC. Tento cyklus má plynulejšie zmeny rýchlosti, čo umožní lepšie demonštrovať pomery v trakčnom reťazci. Podrobnejšie informácie o funkciách a ich realizácii v programe AVL Cruise M sú uvedené v [9].

A. Pokročilý štart/stop systém

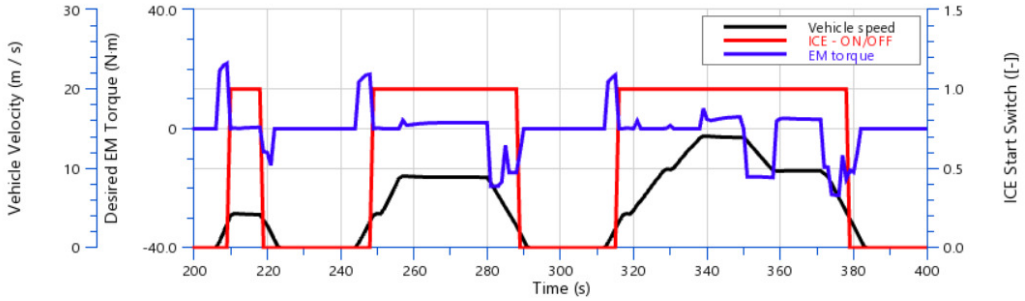
Štart/stop systém sa používa na vypínanie motora v čase, keď vozidlo stojí, napr. na križovatke. Na obr. 6 sú uvedené priebehy veličín vozidla so zohriatym ICE u klasického štart/stop systému: rýchlosť vozidla, otáčky ICE a stav ICE (ON/OFF). Zákmity v priebehu otáčok ICE sú spôsobené prepínaním prevodových stupňov. ICE sa vypína pri zastavenom vozidle a spúšťa sa pri rozjazde. Pri pokročilom štart/stop systéme sa ICE vypína skôr, než vozidlo zastane (obr.7). Pri rozjazde sa najprv aktivuje BSG, ktorý rozbieha ICE aj vozidlo (EM torque - modrý priebeh). ICE sa tak štartuje pri vyšších otáčkach motora a pohybujúcom sa vozidle. To prináša hladší štart, úsporu paliva a emisií. Podobne pri spomaľovaní sa ICE vypína skôr, než vozidlo zastane. Trakčné odpory počas spomaľovania sa prekonávajú pomocou kinetickej energie vozidla a momentu BSG. Aktívne brzdenie vozidla je zabezpečené hlavne prostredníctvom BSG, ktorý kinetickú energiu premieňa na elektrickú a ukladá ju do MHEV batérie.

B. Prídavný moment

Špičkový výkon elektromotora v simulovanom vozidle je cca 10 kW, čo je dostatočný výkon na krytie trakčných strát vozidla pri pomalom rozjazde na veľkom prevode resp. na udržiavanie rýchlosti pri pomalej jazde. Na obr.8 je situácia pri rozbehu a jazde pomalou rýchlosťou (rýchlosť vozidla cca 30 km/h). EM je aktívny na začiatku rozbehu, potom záťaž preberá ICE, ktorý sa štartuje až pri rozbehnutom vozidle a točiacom sa motore. Po rozbehu preberá moment opäť EM, nakoľko pri pomalej jazde sa prejavuje hlavne valivý odpor kolies, odpor vzduchu je malý. ICE by v takomto režime pracoval pri malom momente a s veľmi nízkou účinnosťou. Pri spomaľovaní sa zase využíva možnosť rekuperácie prostredníctvom EM.



Obr.6 Priebehy veličín pri klasickom štart/stop systéme

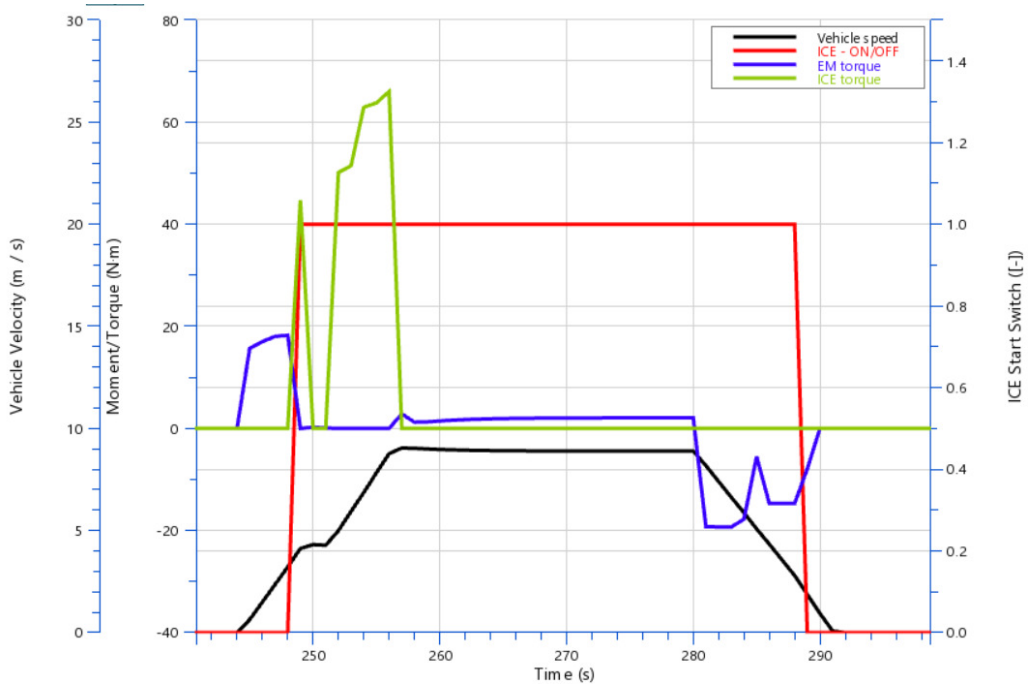


Obr.7 Priebehy veličín pri pokročilom štart/stop systéme (EM torque je vztiahnutý na hriadeľ motora)

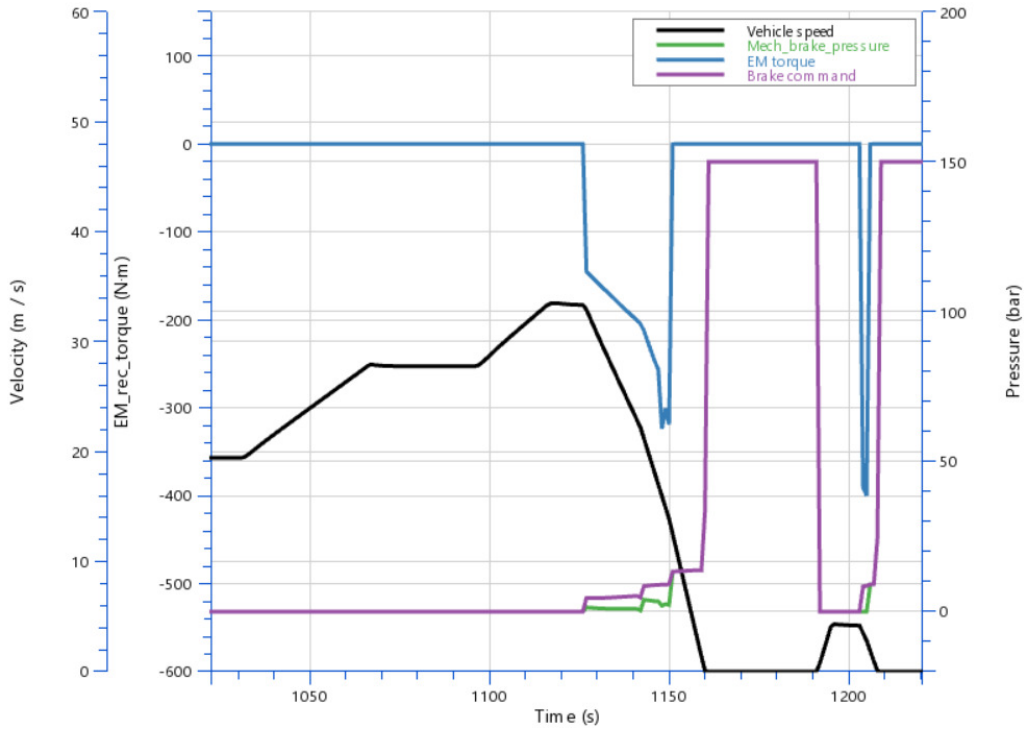
C. Rekuperácia a brzdenie

Pri aktívnom brzdení je snaha o čo najväčšie využitie elektrického brzdenia prostredníctvom BSG, pri ktorom sa energia rekuperuje. Mechanické brzdy sa aktivujú až vtedy, keď sa prekročí brzdný limit EM a pri veľmi malej rýchlosti, resp. stojacom vozidle. Situácia je dokumentovaná na obr. 9 pre NEDC a obr.10 pre WLTC pre tie časti cyklov, kde je intenzívnejšie spomaľovanie.

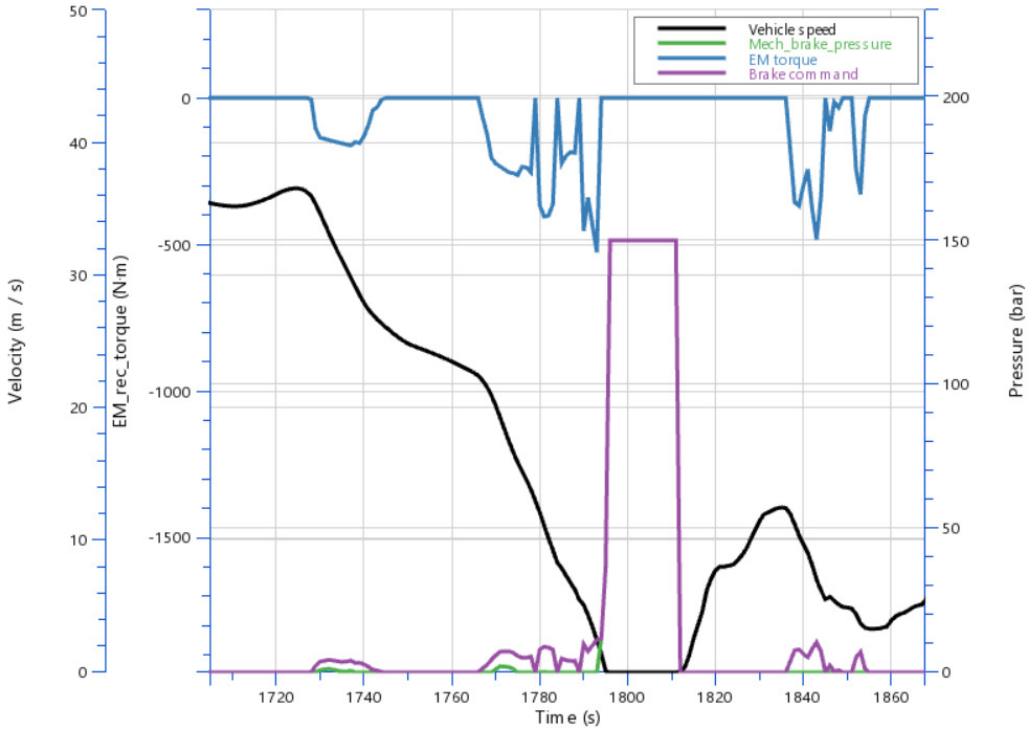
Požiadavka na aktívne brzdenie (Brake command) reprezentovaná brzdným tlakom v baroch (fialový priebeh) je rozdelená medzi EM (modrý priebeh – EM torque) a mechanické brzdy (zelený priebeh – brzdný tlak). Pri stojacom vozidle sú Brake command a žiadosť o mechanický brzdný tlak rovnaké. Nakoľko sa nejedná o veľmi intenzívne brzdenie, hlavný podiel na spomaľovaní má EM a mechanické brzdy sa aktivujú až pri malej rýchlosti vozidla.



Obr.8 Priebehy veličín pri jazde pomalou rýchlosťou (EM torque je vztiahnutý na hriadeľ motora)



Obr.9 Brzdenie a rekuperácia, cyklus NEDC (EM torque je vztiahnutý na hriadeľ kolesa)



Obr.8 Brzdenie a rekuperácia, cyklus WLTC (EM torque je vztiahnutý na hriadeľ kolesa)

IV. ZÁVER

Mild hybridné pohony predstavujú jeden z nových trendov v oblasti znižovania emisií vozidiel so spaľovacím motorom. Pri MH pohone v konfigurácii P0 sa jedná o výmenu pôvodného alternátora za BSG a doplnenie elektrickej sústavy vozidla o druhú elektrickú sieť s napätím 48V. Pri relatívne malých nákladoch a jednoduchej zástavbe do vozidla prináša toto riešenie redukciu emisií CO_2 v rozsahu 7÷15%. Okrem toho rozširuje jestvujúce funkcie vozidla o pokročilý štart/stop systém, prídavný moment, plachtenie a možnosť rekuperácie elektrickej energie pri brzdení. V článku sú jednotlivé funkcie podrobne opísané a simulačne overené pomocou programu AVL Cruise M. Na overenie vlastností MH pohonu v rámci výučby plánujeme zostaviť

stand, ktorý by využíval komponenty MH pohonu z reálneho vozidla. ICE bude nahradený striedavým regulovaným pohonom. Riadenie standu a komunikácia medzi komponentami bude realizovaná prostredníctvom CAN zbernice a programu CANoe [10]. Stand bude možné doplniť o riadenú záťaž, či už na strane 48V alebo 12V siete a o meranie reálnych prúdov a napätí v systéme. Vďaka programu CAPL bude možné jednoducho upraviť riadenie standu a simulovať rôzne jazdné režimy vozidla. Od uvedeného zariadenia očakávame zatraktívnenie výučby v rámci predmetu Elektromobilita.

Poďakovanie

Tento príspevok vznikol s podporou Kultúrnej a edukačnej grantovej agentúry Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky v rámci projektu KEGA 059TUKE-4/2024 “Rozvoj odborných a tímových zručností študentov v študijnom programe priemyselná elektrotechnika” a projektu KEGA 032TUKE-4/2024 „Podpora rozvoja praktických zručností v oblasti merania a spracovania veličín pre autonómne systémy.“

Použitá literatúra

[1] J. Dornoff, J. German, A. Deo, A. Dimaratos, “Mild-hybrid Vehicles: A Near Term Technology Trend for CO_2 Emission Reduction,” International Council on Clean Transportation. (4 July 2022). Access: https://theicct.org/publication/mild-hybrid-emissions-jul22/
[2] S. Lauer, M. Perugini, R. Weldle, F. Graf, “48 Volt High Power – Much More than a Mild Hybrid,” Vitesco Technologies, Regensburg, Germany. 29th Aachen Colloquium Sustainable Mobility 2020, pp.1-17.
[3] 48-Volt Electrical Systems –A Key Technology Paving the Road to Electric Mobility. ZVEI - German Electrical and Electronic Manufacturers’ Association. 2016.
[4] https://efficiencywins.nexperia.com/innovation/the-48-volt-starter-generator
[5] T. Eckenfels, K. Florian, S. Lehmann, W. Neugebauer, C. Manuel, “48V Hybridization: A Smart Upgrade for the Powertrain,”. Schaeffler Symposium, 2018, Documentation. Mobility for tomorrow.
[6] M. Melaika, S. Mamikoglu, P. Dahlander, “48V Mild-Hybrid Architecture Types, Fuels and Power Levels Needed to Achieve 75g CO_2/km,”. SAE Paper 2019-01-0366. https://doi.org/10.4271/2019-01-0366.
[7] M. Werra, A. Ringleb, J. Muller, F. Kucukay, “Topology Comparison of 48 V Hybrid Drivetrains with Manual Transmission,” CTI Symposium 2018. https://doi.org/10.1007/978-3-662-58866-6_29.
[8] https://www.avl.com/en/simulation-solutions/software-offering/simulation-tools-a-z/avl-cruise-m
[9] E. Podracký, “Simulácia mild-hybridného pohonu”. Diplomová práca, FEI TU Košice 2025.
[10] https://www.vector.com/int/en/products/products-a-z/software/canoe/ento

Sieťová steganografia s využitím protokolu ICMP

Jakub ORAVEC, Ľuboš OVSENÍK, Zuzana LIŠČINSKÁ

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

{jakub.oravec, lubos.ovsenik, zuzana.liscinska}@tuke.sk

***Abstrakt* – V tomto príspevku bude predstavená technika sieťovej steganografie využívajúca vlastnosti správ protokolu Internet Control Message Protocol (ICMP). Použitie správ tohto protokolu je zväčša obmedzené na konkrétne úlohy ako overenie konektivity medzi dvoma sieťovými rozhraniami alebo určenie cesty ku cieľovému rozhraniu. Keďže na takéto úlohy nie je potrebné prenášať veľké množstvo dát, tak tieto správy sú umelo predĺžené, aby rámce obsahujúce tieto správy dosiahli minimálnu dĺžku. Na predĺženie sa používa tzv. padding, ktorý je v navrhnutej steganografickej technike pozmenený tajnou správou. Praktickým overením teoretického návrhu sa dosiahne odhalenie niektorých problémov, pre ktoré príspevok bližšie predstavuje jedno z možných riešení.**

***Kľúčové slová* – Internet Control Message Protocol, modifikácia sieťovej premávky, Python, Scapy, steganografia**

I. Úvod

Steganografia je vedná disciplína zaoberajúca sa ukrývaním komunikácie (angl. information hiding). Zaujímavou skutočnosťou pri použití steganografie je to, že nedochádza ku zmene prenášanej informácie tak, aby bola pre iných používateľov nečitateľná (ako pri kryptografii), ale steganografia ukrýva samotnú prítomnosť informácie v dôverne pôsobiacich dátach. Tieto dáta, ktoré sú do istej miery modifikované prenášanou tajnou správou sa vo všeobecnosti označujú ako krycie dáta.

Ako krycie dáta sa dajú použiť rôzne druhy dát. Z minulosti sú známe najmä systémy obrazovej steganografie [1], [2], kde sa tajnou správou modifikovali najmenej významné bity jasových úrovní obrazových prvkov alebo hodnoty transformačných koeficientov diskrétnej kosínusovej transformácie. Okrem systémov obrazovej steganografie existujú aj riešenia využívajú zvukové dáta, textové dáta a v neposlednej rade aj dátové jednotky sieťovej komunikácie.

Sieťová steganografia vďaka vývoju v oblasti hardvéru a softvérových nástrojov zažíva v posledných piatich rokoch rozvoj po určitej dobe stagnácie. Ukazuje na to aj skutočnosť, že za roky 2021 až 2025 bolo v databáze Scopus zaznamenaných viac ako 50 príspevkov s kľúčovým slovom „network steganography“, z čoho väčšina prezentovala novú techniku sieťovej steganografie. Okrem toho je možné nájsť stále viac návodov na sieťovú steganografiu určených hlavne pre laickú verejnosť [3], [4].

Zvyšok príspevku je členený nasledovne: v kapitole II je prezentovaná stručná história sieťovej steganografie spolu s aktuálnym stavom problematiky, možné implementačné problémy a používané softvérové nástroje. Kapitola III bližšie predstavuje steganografickú techniku využívajúcu vlastnosti sieťového protokolu Internet Control Message Protocol (ICMP), experimentálne zistené poznatky použité na úpravu navrhnutého riešenia a obsahuje aj dosiahnuté výsledky. V kapitole IV sú stručne zhrnuté vlastnosti celého riešenia a takisto možné úlohy pre prácu v budúcnosti.

II. Aktuálny stav v oblasti

Prvé techniky z oblasti sieťovej steganografie pochádzajú z obdobia okolo roku 2000 [5], [6]. V tejto dobe sa jednalo o čisto teoretické návrhy, ktoré kvôli vtedajším hardvérovým obmedzeniam (jednoduchosť sieťových kariet) a implementačným ťažkostiam (potreba práce priamo s ovládačom sieťovej karty) nebolo možné prakticky otestovať. Prelom pri testovaní techník sieťovej steganografie nastal okolo roku 2010, kedy bol navrhnutý algoritmus RSTEG [7], ktorý bol otestovaný v simulátore sieťovej premávky ns-2.

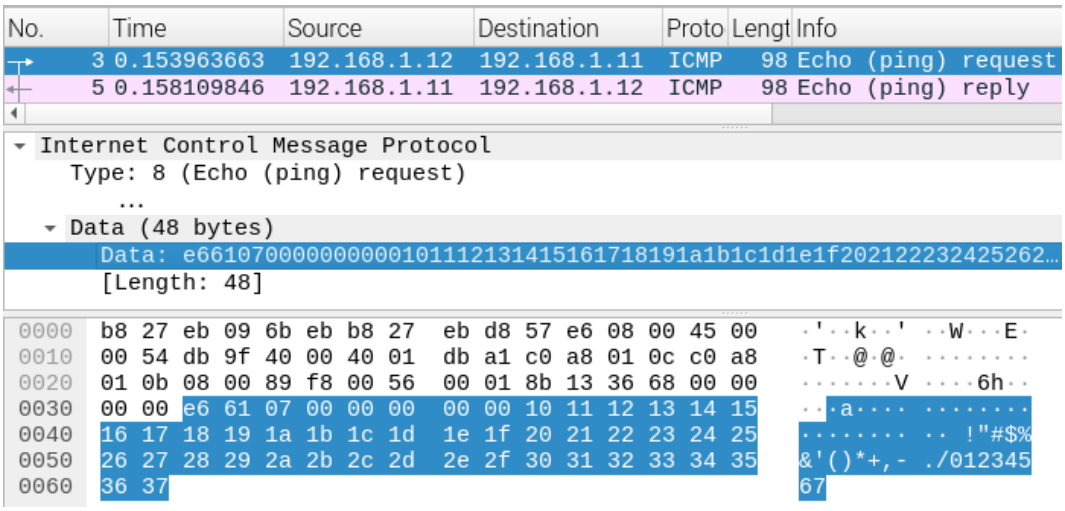
Po prezentovaní popisu algoritmu RSTEG spolu so simulovanými prenosovými rýchlosťami začala vedecká komunita hľadať ďalšie spôsoby, ako by sa dali odmerať vlastnosti techník sieťovej steganografie. V tejto oblasti veľmi pomohol vývoj softvérových nástrojov ako Scapy [8] a NetfilterQueue [9], ktoré sú balíčkami jazyka Python. Kombináciou týchto balíčkov je možné zachytiť sieťovú premávku, vykonať v nej modifikácie (vložiť tajnú správu) a takto pozmenené správy preposlať ďalej do ich cieľa. Po predstavení týchto balíčkov začal rýchlo rásť počet praktických implementácií sieťovej steganografie určených ako pre laických používateľov [10] tak aj pre vedeckú komunitu [11], [12].

Miernym problémom v súčasnosti ostáva tzv. platformová špecifickosť nástrojov na sieťovú steganografiu [12]. Aj keď sú balíčky Scapy a NetfilterQueue a rovnako aj jazyk Python podporované na viacerých platformách (napr. Linux, Windows, ...), tak nie je možné na odlišných zariadeniach dosiahnuť rovnaké výsledky. Táto skutočnosť je spôsobená najmä rozličnými implementáciami sieťových protokolov, ktorých vlastnosti steganografia využíva pre vkladanie tajných správ. Môže sa zdať, že tieto protokoly musia mať jasne určené pravidlá, no väčšina z nich je definovaná pomocou odporúčaní RFC, ktoré vznikali spoluprácou v prostredí odborných komunít (skratka RFC z angl. Request for comments). Pri implementácii sieťových protokolov sa potom odporúčania môžu, ale nemusia zohľadniť.

III. Návrh riešenia

V tomto príspevku bude bližšie predstavené riešenie založené na modifikácii tzv. paddingu (dodatočnej výplne) niektorých správ protokolu Internet Control Message Protocol (ICMP) [13]. Konkrétne sa využijú správy druhu Echo Request a Echo Reply, používané napr. utilitou ping na overenie konektivity medzi dvojicou sieťových rozhraní.

Ukážka disekcie (rozklad na políčka) ICMP správy Echo Request v analyzátore Wireshark je znázornená na Obr. 1. Za povšimnutie stojí skutočnosť, že samotná správa má iba 50 bajtov a pre dosiahnutie minimálnej dĺžky správy je doplnená paddingom, ktorý je v disekcii zvýraznený. Disektor používaný analyzátorom Wireshark zaraďuje padding do políčka ICMP Data.



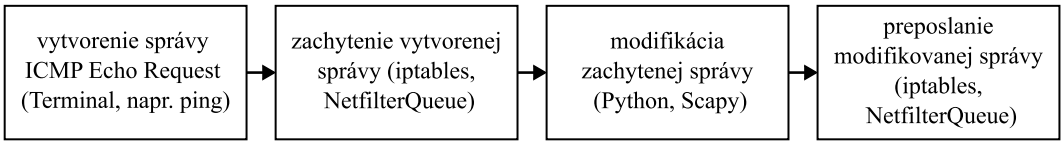
Obr. 1 Disekcia ICMP správy Echo Request v analyzátore Wireshark

Pri použití paddingu na účely prenosu tajných správ je potrebné zohľadniť niektoré z jeho vlastností ako napr. závislosť dĺžky paddingu alebo jeho obsahu na platforme. Ako jednoduchý príklad sa dá uviesť, že niektoré verzie operačného systému Windows do vytvorených správ ICMP Echo Request nevkladajú časové údaje, no na správy Echo Request s takýmito údajmi vedia odpovedať.

Navrhnuté riešenie bude kvôli rozsahu tohto príspevku značne zjednodušené. Ku jeho obmedzeniam bude patriť:

- správna funkcionalita v rámci experimentálneho pracoviska (využíva operačný systém Linux) prezentovaného v tomto zborníku v minulosti [14],
- komunikácia iba v jednom smere (jedno zariadenie tajnú správu odosiela, iné ju prijíma),
- dĺžka tajnej správy max. 40 znakov (veľkosť paddingu v jednej správe ICMP Echo Request).

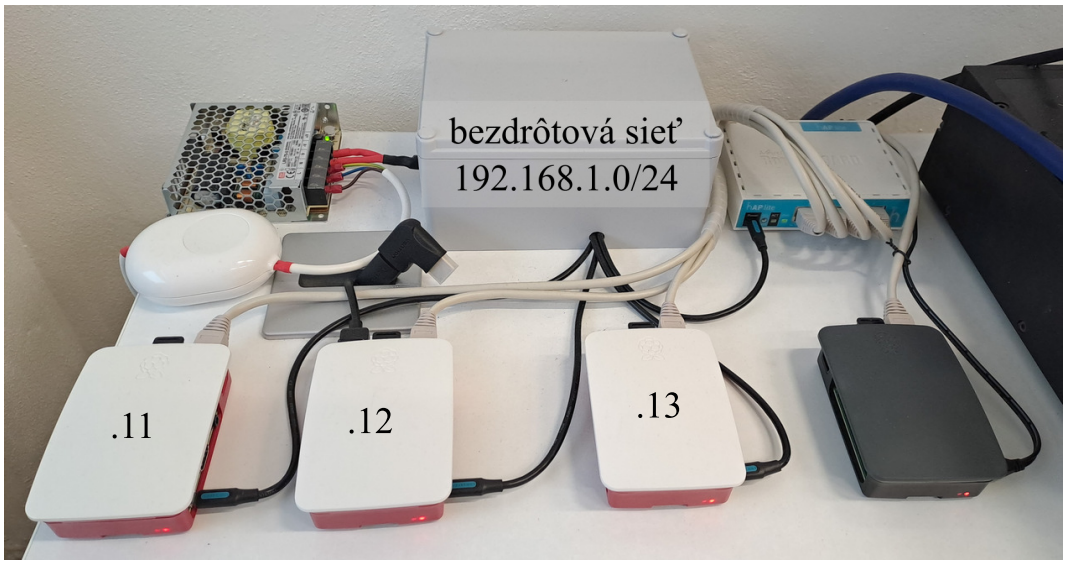
Hlavné kroky vkladania tajnej správy do správ ICMP Echo Request pomocou navrhnutého riešenia sa dajú zhrnúť do blokov schémy demonštrovanej na Obr. 2. V zátvorkách sú uvedené existujúce nástroje použité v jednotlivých blokoch. Proces extrakcie (získania) tajnej správy na strane prijímača je analogický ku procesu vkladania tajnej správy na strane odosielateľa.



Obr. 2 Bloková schéma vkladania tajnej správy pri navrhnutom riešení

A. *Experimentálne poznatky a úprava návrhu*

Navrhnuté riešenie bolo otestované pomocou experimentálneho pracoviska [14] ilustrovaného na Obr. 3 zloženého z mikropočítačov Raspberry Pi, ktoré používajú distribúciu operačného systému Linux. Pri experimentoch prebiehala komunikácia medzi mikropočítačmi .11 (odosielateľ) a .13 (príjemca), pričom mikropočítač .12 túto komunikáciu sledoval pomocou analyzátora Wireshark.



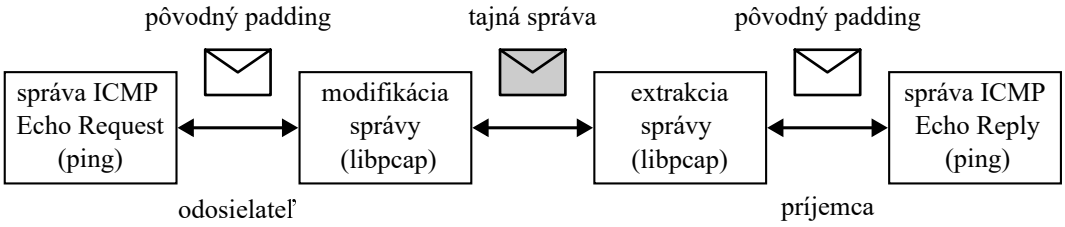
Obr. 3 Ukážka experimentálneho pracoviska – laboratória sieťovej steganografie [14]

Na základe poznatkov získaných počas experimentov bolo potrebné navrhnuté riešenie do určitej miery pozmeniť. Táto skutočnosť bola zapríčinená najmä bezpečnostnými mechanizmami implementovanými v jadre Linuxu, ako napr. kontrolou obsahu paddingu pri prijímaní správ ICMP Echo Reply. V prípade, ak sa v paddingu nachádzal iný obsah ako operačný systém očakával, tak dochádzalo ku zahadzovaniu týchto modifikovaných správ. Príjemca by tým pádom nemohol extrahovať tajnú správu. Okrem toho aj na strane odosielateľa sú zjavné rozdiely (napr. pri použití utility ping nie sú zobrazené časy oneskorenia – nachádzajú sa v správach, ktoré odosielateľovi v takomto prípade vôbec neprídu).

V steganografických systémoch je dôležité, aby stego dáta, ktoré vzniknú vložením tajnej správy do krycích dát nevzbudili žiadne podozrenie. Preto je potrebné upraviť ICMP správy s vloženou tajnou správou tak, aby ich operačný systém zariadení korektne spracoval. Na túto úlohu bola navrhnutá technika maskovania, ktorá spočíva v prepísaní paddingu novým obsahom – na strane príjemcu po extrakcii tajnej správy pôvodným obsahom paddingu a naopak na strane odosielateľa sa v prijatej odpovedi nahradí pôvodný obsah paddingu tajnou správou. Zjednodušená bloková schéma princípu tejto techniky je znázornená na Obr. 4.

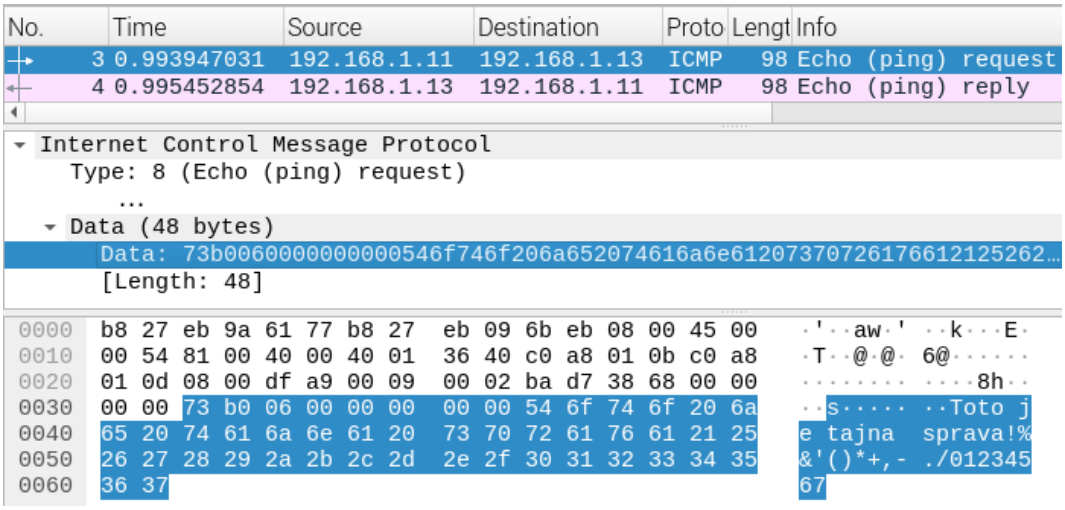
Pri technike maskovania sa využíva skutočnosť, že kombinácia balíčkov použitých na zachytenie premávky (iptables a NetfilterQueue) vie pracovať aj s prichádzajúcou sieťovou premávkou. Pomocou balíčka Scapy sa na strane príjemcu stego dáta modifikujú druhý krát, vďaka čomu má padding správy ICMP Echo Request rovnaký obsah ako pred vložením tajnej správy. Tým pádom je takáto žiadosť spracovaná – pre správu ICMP Echo Request sa vytvorí jej korešpondujúca odpoveď – správa ICMP Echo Reply.

Ďalší problém môže nastať pri odosielaní odpovede naspäť na stranu odosielateľa. Ak by sa odpoveď odoslala s pôvodným obsahom v paddingu, tak všetky nástroje, ktoré používajú rovnakú knižnicu ako balíček Scapy (knižnica libpcap, resp. WinPcap) môžu na strane odosielateľa odhaliť rozdiely medzi odoslanou žiadosťou (s vloženou tajnou správou) a prijatou odpoveďou

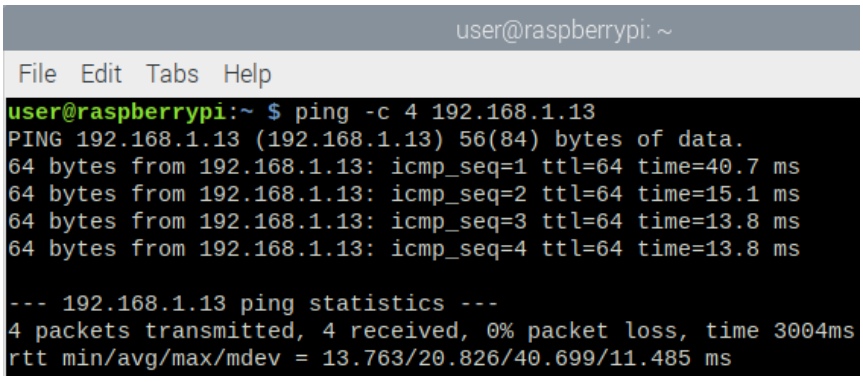


Obr. 4 Princíp maskovania ICMP správ

(s pôvodným paddingom). Takejto situácii sa dá predísť dvoma krokmi – pred odoslaním správy ICMP Echo Reply sa do správy znova vloží tajná správa, aby nenastal rozdiel v paddingu správ na strane odosielateľa a po prijatí a extrakcii správy si odosielateľ nahradí padding správy pôvodným obsahom (znovu sa zamaskuje). Takto sa docieli to, že v analyzátore Wireshark nie je viditeľné žiadne upozornenie a zároveň je správa korektne spracovaná operačným systémom (na výstupe príkazu ping sa zobrazí oneskorenie). Túto situáciu ukazujú Obr. 5 a 6.



Obr. 5 Výstup v analyzátore Wireshark bez upozornení na rozličný padding v dvojici ICMP správ



Obr. 6 Výstup utility ping bez známok manipulácie s ICMP správami

Oko pozorného čitateľa si mohlo na Obr. 5 všimnúť skutočnosť, že výstup analyzátora Wireshark síce neobsahuje žiadne upozornenie, no disekcia správy obsahuje vloženú tajnú správu. V tomto prípade je dôležitejšie nevzbudzovať pozornosť prítomnosťou upozornení. Tajná správa sa dá vkladať do ICMP správ aj v zašifrovanej podobe, vďaka čomu by jej obsah nemusel byť odhalený, no prenos akýchkoľvek doplnkových informácií (napr. šifrovacie kľúče) výrazne znižuje pomerne malú kapacitu steganografickej techniky (pri použití experimentálneho pracoviska [14] je to 40 bajtov na každú správu ICMP Echo Request).

B. Experimentálne výsledky

Keďže technika maskovania robí z pomerne jednoduchého návrhu komplexnejšie riešenie, je vhodné preskúmať aj výpočtovú náročnosť takéhoto riešenia. Z hľadiska koncového používateľa by použitie sieťovej steganografie nemalo viditeľne spomaliť alebo obmedziť poskytované služby. Vďaka tomu, že protokol ICMP sa používa iba pre servisné správy (napr. na overenie konektivity) by prípadné spomalenie takýchto správ nemalo byť pre používateľa obmedzujúce.

Zvýšená výpočtová náročnosť sa môže prejaviť nárastom oneskorenia pri prenose ICMP správ. Oneskorenie sa dá experimentálne určiť viacerými spôsobmi, z ktorých je najjednoduchším meranie tzv. Round Trip Time (RTT), čo je čas potrebný na úspešné prijatie odpovede po odoslaní žiadosti [15]. Hodnoty RTT závisia od zložitosti topológie siete medzi komunikujúcimi zariadeniami – vo všeobecnosti pri väčšom počte medziľahlých zariadení hodnoty RTT rastú.

V prípade použitia experimentálneho pracoviska [14], kde sa topológia siete medzi odosielateľom a príjemcom počas prenosu správ nemení sa RTT dá použiť ako ukazovateľ nárastu oneskorenia. RTT sa v prostredí operačného systému Linux dá zistiť z výpisu príkazu ping, kde sú zobrazené nasledujúce hodnoty:

- hodnota RTT pre aktuálne meranie,
- minimálna hodnota RTT_{min} zo súboru meraní,
- priemerná hodnota RTT_{avg} zo súboru meraní,
- maximálna hodnota RTT_{max} zo súboru meraní,
- stredná kvadratická odchýlka RTT_{mdev} súboru meraní.

Hodnoty RTT_{avg} a RTT_{mdev} sa počítajú pomocou vzorcov (1), resp. (2):

$$RTT_{avg} = \frac{1}{n} \cdot \sum_{i=1}^{n} RTT_i, \tag{1}$$

$$RTT_{mdev} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (RTT_i - RTT_{avg})^2}, \tag{2}$$

kde n je počet meraní RTT.

Pre potlačenie prípadných negatívnych vplyvov operačného systému (napr. prerušenia procesu) sa odporúča vykonať viacero meraní. Tento príspevok obsahuje v Tabuľke 1 kvôli obmedzenému rozsahu príspevku numerické výsledky len pre 5 meraní bez vkladania tajnej správy a pre 5 meraní s vkladaním tajnej správy.

Hodnoty RTT môžu byť negatívne ovplyvnené, ak niektoré z komunikujúcich zariadení nepozná fyzickú (MAC) adresu zariadenia na opačnej strane. V takomto prípade zariadenie čaká kým sa chýbajúca adresa získa pomocou správ protokolu Address Resolution Protocol (ARP) [16]. V niektorých prípadoch môže dôjsť aj ku vypršaniu časovej lehoty na prijatie správy protokolu ICMP skôr ako protokol ARP doručí chýbajúcu adresu, čo pre steganografický systém znamená, že nemá k dispozícii stego dáta.

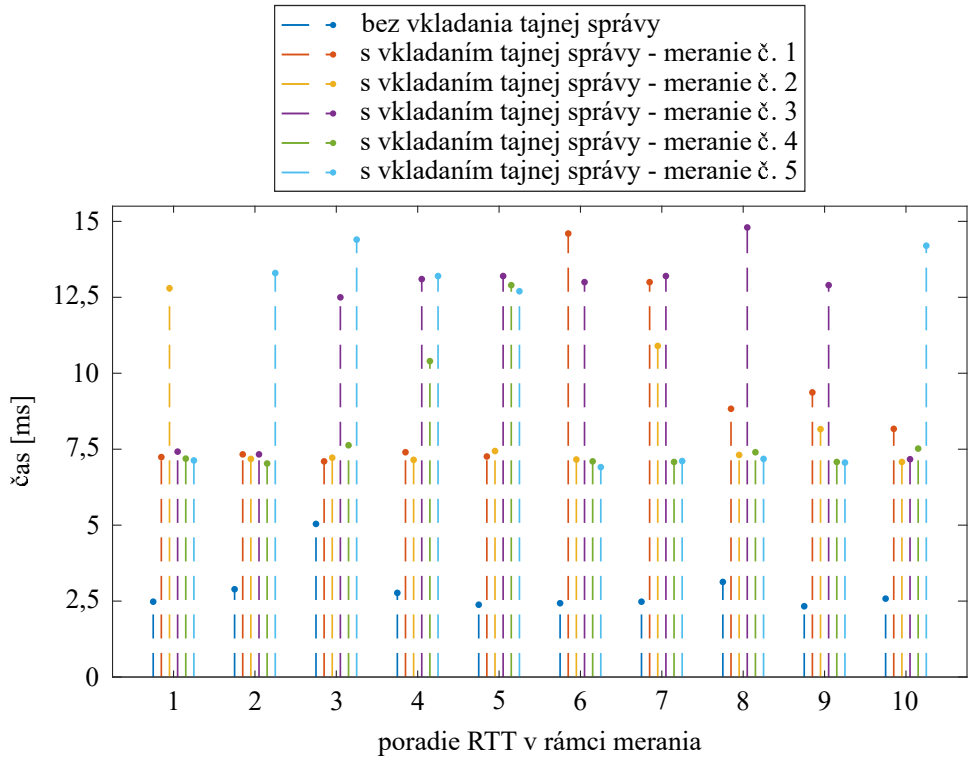
Z dôvodu možného ovplyvnenia prvej získanej hodnoty RTT protokolom ARP sa táto hodnota pri experimentoch nezapisovala. Po nej sa zaznamenalo 10 nasledujúcich hodnôt RTT, z ktorých sa vypočítali hodnoty RTT_{min}, RTT_{avg}, RTT_{max} a RTT_{mdev} pre jednotlivé merania. Správy ICMP Echo Request sa posielali z mikropočítača .11 na mikropočítač .13. Na generovanie týchto správ sa používal príkaz `ping -c 11 192.168.1.13`. V prípade vkladania tajnej správy sa použil textový reťazec „*Toto je tajná správa!*“ s dĺžkou 21 znakov.

Tabuľka 1
Výsledky experimentálneho merania hodnôt RTT s využitím pracoviska [14]

	bez vkladania tajnej správy					**s vkladaním tajnej správy**				
číslo merania	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
RTT_{min} [ms]	2,31	2,26	2,26	2,33	2,37	7,10	7,08	7,17	7,03	6,91
RTT_{avg} [ms]	2,49	2,37	2,83	2,85	2,76	9,03	8,24	11,46	8,13	10,32
RTT_{max} [ms]	2,92	2,66	4,26	5,04	3,66	14,60	12,80	14,80	12,90	14,40
RTT_{mdev} [ms]	0,21	0,11	0,78	0,81	0,39	2,65	1,98	2,93	1,96	3,45

Konkrétne hodnoty RTT z jednotlivých meraní sú kvôli množstvu dát znázornené pomocou grafu na Obr. 7. Kvôli lepšej prehľadnosti boli merania bez vkladania tajnej správy reprezentované len štvrtým meraním, v ktorom bola zistená najvyššia hodnota strednej kvadratickej odchýlky.

Z Tabuľky 1 a Obr. 7 je zjavné, že využitie balíčkov NetfilterQueue a Scapy pre vloženie tajnej správy spôsobilo nárast oneskorenia – hodnoty RTT_{avg} sa zväčšili z približne 3 ms na hodnoty nad 8 ms. V prípade tretieho merania bola hodnota RTT_{avg} až na úrovni 11,46 ms. Na fialovom



Obr. 7 Porovnanie jednotlivých RTT v rámci rôznych meraní

priebehu v Obr. 7, ktorý reprezentuje tretie meranie s vložením tajnej správy je možné vidieť aj výrazné kolísanie jednotlivých RTT – niektoré mali hodnoty približne 7,5 ms, iné aj viac ako 12 ms. Túto skutočnosť ukazuje aj vyššia hodnota RTT_{mdev} pre toto meranie.

Spôsobenie oneskorenia pri vkladaní tajnej správy je samozrejme nežiadúce. Pri servisných protokoloch, ku ktorým patrí aj ICMP to nemusí spôsobovať zreteľné problémy, no môže to vzbudiť pozornosť vnímavých používateľov. Na minimalizovanie takéhoto oneskorenia je možné vynechať niektoré kroky pri spracovaní správ (napr. uvedenú techniku maskovania), no v takom prípade by použitie steganografie mohlo byť odhalené pomerne jednoducho vo výpisoch Terminalu alebo analyzátora Wireshark.

Ďalším problémom je charakter vzniknutého oneskorenia – vlastnosti postupnosti viacerých RTT (napr. kolísanie hodnôt alebo ich stredná kvadratická odchýlka) je možné použiť pri analýze napr. v systémoch detekcie prieniku (angl. Intrusion Detecion System, IDS) alebo moderných firewalloch. Výsledkom môže byť upozornenie na podozrivú sieťovú premávku alebo v horšom prípade aj blokovanie tejto premávky.

IV. Záver

Tento príspevok predstavuje riešenie pre sieťovú steganografiu, ktoré modifikuje padding v správach protokolu ICMP. Testovanie teoretického návrhu v prostredí experimentálneho pracoviska poukázalo na potrebu rozšírenia návrhu o techniku maskovania, vďaka ktorej vkladanie tajnej správy nevyvoláva upozornenia v bežne používaných nástrojoch. Napriek tomu môže aj bežný používateľ s určitými skúsenosťami prítomnosť tajnej správy pomerne jednoducho odhaliť, keďže tajná správa sa prenáša v podobe nezašifrovaného textového reťazca.

Tento nedostatok by bolo možné zmierniť viacerými spôsobmi, ktoré by však znamenali obmedzenie steganografickej techniky – ďalšie spracovanie dát pri vkladaní tajnej správy by spôsobilo väčšie oneskorenie, resp. potreba šírenia ďalších parametrov by znížila kapacitu použitej steganografickej techniky. Takéto, alebo ďalšie vylepšenia sú vhodnou témou na prácu v budúcnosti, no najmä pri iných steganografických technikách.

Okrem toho je možné zamerať sa v budúcnosti na techniky detekcie modifikácie sieťovej premávky. V súčasnosti existuje viacero riešení (spomínané IDS a firewally), ktoré je možné rozšíriť o pravidlá na detekciu špecifických vlastností stego dát, resp. vlastností zapríčinených vkladaním tajnej správy ako je napr. kolísavé oneskorenie.

Poďakovanie

Táto práca bola financovaná pomocou EÚ NextGenerationEU grantu prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00019 – „Kľúčové digitálne technológie v oblasti bezdrôtových širokopásmových odolných prenosových sietí“, EÚ NextGenerationEU grantu prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I03-03-V05-00015 – „Využitie AI v oblasti bezdrôtových širokopásmových odolných prenosových sietí“ a grantu KEGA 041TUKE-4/2025 –„Laboratórium na výskum a vývoj pokročilých IoT senzorových sietí a širokopásmových senzorov“.

Literatúra

[1] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge: Cambridge University Press, 2009, 466 s. ISBN: 978-05-2119-019-0. DOI: 10.1017/CBO9781139192903.
[2] D. Levický, *Multimédiá a ochrana ich obsahu*. Košice: Elfa, 2012, 249 s. ISBN: 978-80-8086-199-5.
[3] SpoofIMEI: Modifying Packets on the Fly (Python) [online]. Citované 26.5.2025. Dostupné na internete: https://medium.com/@SpoofIMEI/modifying-packets-on-the-fly-python-76076c5d6e1e.
[4] crawsecurity: Packet Manipulation With Scapy [online]. Citované 26.5.2025. Dostupné na internete: https://medium.com/@crawsecurity/packet-manipulation-with-scapy-7205dec3dfdc.
[5] C. H. Rowland, Covert Channels in the TCP/IP Protocol Suite, *FirstMonday: Peer-reviewed Journal on the Internet*, 1997, roč. 2, č. 5. ISSN: 1396-0466. DOI: 10.5210/fm.v2i5.528.
[6] G. Fisk, M. Fisk, C. Papadopoulos et al., Eliminating Steganography in Internet Traffic with Active Wardens, v *Proceedings of 5th International Workshop on Information Hiding*. Noordwijkerhout (Netherlands), október 2002, s. 18–35. ISBN: 978-35-4000-421-9. DOI: 10.1007/3-540-36415-3_2.
[7] W. Mazurczyk, M. Smolarczyk, K. Szczypiorski, Retransmission Steganography and its Detection, *Soft Computing*, 2011, roč. 15, č. 3, s. 505–515. ISSN: 1432-7643. DOI: 10.1007/s00500-009-0530-1.
[8] Scapy: the Python-based Interactive Packet Manipulation Program & Library [online]. Citované 26.5.2025. Dostupné na internete: https://scapy.net.
[9] NetfilterQueue [online]. Citované 26.5.2025. Dostupné na internete: https://pypi.org/project/NetfilterQueue.
[10] chrispetrou: NETsteg: Network-layer Steganography [online]. Citované 26.5.2025. Dostupné na internete: https://github.com/chrispetrou/NETsteg.
[11] F. Iglesias, F. Meghdouri, R. Annessi et al., CCgen: Injecting Covert Channels into Network Traffic, *Security and Communication Networks*, 2022, roč. 2022, s. 1–11. ISSN: 1939-0122. DOI: 10.1155/2022/2254959.
[12] S. Bistarelli, A. Imparato, F. Santini, A TCP-based Covert Channel with Integrity Check and Retransmission, *International Journal of Information Security*, 2024, roč. 23, s. 3481–3512. ISSN: 1615-5262. DOI: 10.1007/s10207-024-00879-z.
[13] J. Postel: RFC 792: Internet Control Message Protocol [online]. Citované 27.5.2025. Dostupné na internete: https://www.rfc-editor.org/rfc/rfc792.
[14] J. Oravec, Ľ. Ovseník, S. Andrejčík, Hardvérový návrh a realizácia laboratória sieťovej steganografie, v *Proceedings of Electrical Engineering and Informatics 2023*. Košice (Slovensko), júl 2023, s. 153-157. ISBN: 978-80-5534-407-2.
[15] Wikipedia: Round-trip delay [online]. Citované 3.6.2025. Dostupné na internete: https://en.wikipedia.org/wiki/Round-trip_delay.
[16] Wikipedia: Address resolution protocol [online]. Citované 3.6.2025. Dostupné na internete: https://en.wikipedia.org/wiki/Address_Resolution_Protocol.

Využitie laboratória sieťovej steganografie pre účely výuky

Jakub ORAVEC, Ľuboš OVSENÍK, Zuzana LIŠČINSKÁ

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

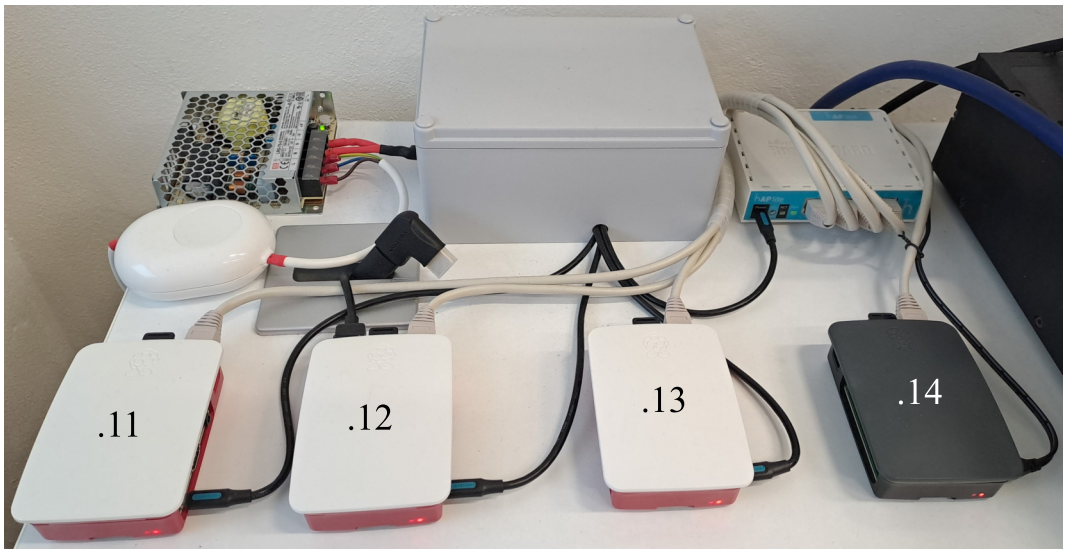
{jakub.oravec, lubos.ovsenik, zuzana.liscinska}@tuke.sk

***Abstrakt* – V tomto príspevku bude demonštrované ako je možné použiť existujúci nástroj v podobe laboratória sieťovej steganografie na praktické ukážky počas výuky sieťovo orientovaných predmetov na Katedre elektroniky a multimediálnych telekomunikácií. Laboratórium na testovanie nástrojov sieťovej steganografie používa veľké množstvo softvérových prostriedkov, ktorých výstup môže byť použitý aj na potvrdenie teoretických poznatkov z prednášok, resp. z materiálov na cvičenia. Príspevok stručne prezentuje prípady, v ktorých sa laboratórium už použilo počas výuky sieťovo orientovaných predmetov, výhody a nevýhody použitia laboratória a možnosti zlepšenia v budúcnosti.**

***Kľúčové slová* – experimentálne pracovisko, siete, výuka, vzdialený prístup**

I. Úvod

Funkčnosť steganografických techník sa dá overiť viacerými spôsobmi. Okrem simulačných prostredí ako napr. MATLAB, ktoré sú populárne najmä pri obrazovej steganografii sa na overenie funkčnosti techník môžu vytvoriť aj samostatné programy, resp. webové aplikácie. V prípade použitia určitých typov dát, ako je napr. sieťová premávka je však vhodné použité zariadenia do istej miery prispôsobiť. Týmto spôsobom vzniklo aj experimentálne pracovisko – laboratórium sieťovej steganografie prezentované v tomto zborníku v minulosti [1]. Toto pracovisko je znázornené na Obr. 1. Rozdelenie zariadení do sietí a úlohy jednotlivých sietí ilustruje Tabuľka 1.



Obr. 1 Experimentálne pracovisko [1]

Experimentálne pracovisko obsahuje viacero mikropočítačov Raspberry Pi, pričom jeden z nich (označený .14) sa používa len na správu IoT (z angl. Internet of Things) zariadení, ku ktorým patrí napr. relé ovládané mikrokontrolérmi, alebo inteligentná zásuvka. Ostatné mikropočítače (.11, .12 a .13) je možné používať na experimenty so sieťovou steganografiou. Niektoré z mikropočítačov majú nakonfigurované rozhrania pre komunikáciu s vonkajšou sieťou (infraštruktúra univerzity), kde majú adresy priradené protokolom DHCP (z angl. Dynamic Host Configuration Protocol).

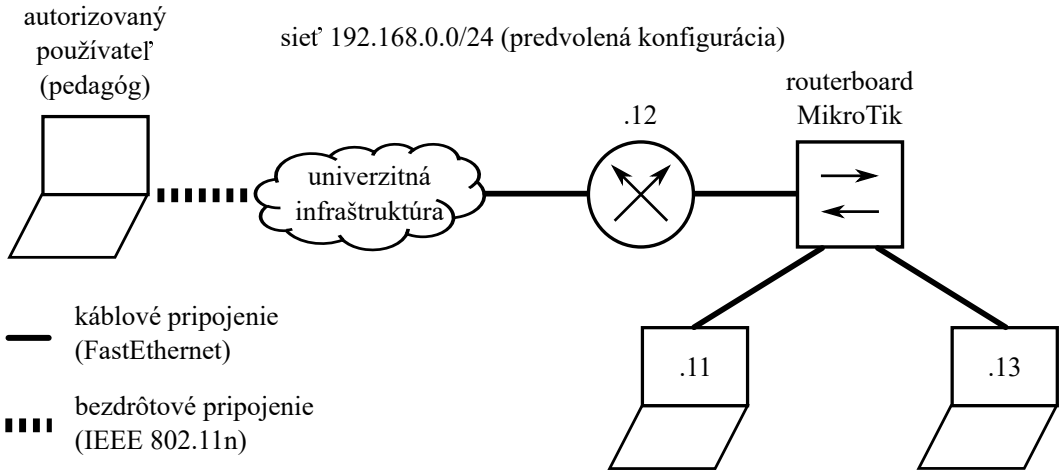
Tabuľka 1
Rozdelenie zariadení do sietí v pracovisku [1]

úloha siete	typ siete	podsieť	zariadenia				
			.11	.12	.13	.14	iné
pripojenie do vonkajšej siete	káblová	DHCP	nie	áno	nie	áno	server pre IoT
riadenie a aktualizácie	káblová	192.168.0.0/24	áno	áno	áno	nie	prepínač
experimenty	bezdrôtová	192.168.1.0/24	áno	áno	áno	nie	–
sieť pre IoT zariadenia	bezdrôtová	192.168.2.0/24	nie	áno	nie	áno	IoT zariadenia

Poznámka: slová áno, resp. nie vyjadrujú, či je zariadenie súčasťou siete.

Topológia pracoviska sa dá pomerne ľahko meniť pomocou vypínania, resp. opätovného spustenia rozhraní alebo aj sieťových služieb na mikropočítačoch. Pri návrhu topológie bola snaha o redundantné (nadbytočné) zapojenie spojení medzi mikropočítačmi tak, aby mikropočítače boli v prípade výpadkov, resp. nevhodnej konfigurácie dostupné po záložných spojeniach. Ďalšou výhodou takéhoto návrhu je aj skutočnosť, že celé pracovisko je možné nakonfigurovať aj na iné účely ako na testovanie techník sieťovej steganografie.

Predvolená topológia siete 192.168.0.0/24 primárne určenej na riadenie a aktualizáciu mikropočítačov .11 až .13 je spolu s pripojeným používateľom znázornená na Obr. 2. Tento obrázok poukazuje na mierne obmedzenie pracoviska – kvôli nedostatku rozhraní Ethernet a USB (angl. Universal Serial Bus) na použitých mikropočítačoch musí byť trojica týchto zariadení prepojená pomocou routerboardu MikroTik. Podobných obmedzení je viac a boli bližšie popísané v [1].



Obr. 2 Predvolená topológia siete 192.168.0.0/24 spolu s pripojeným používateľom

Topológia na Obr. 2 sa môže výrazne zmeniť aj po vypnutí niekoľkých sieťových služieb (napr. preklad IP adries alebo preposielanie IP paketov jadrom operačného systému). V takom prípade sa z mikropočítača .12, ktorý predvolene plní úlohu smerovača medzi vonkajšou sieťou a sieťami 192.168.0.0/24 a 192.168.1.0/24 stane „iba“ koncové zariadenie v posledných dvoch uvedených sieťach. Na tento mikropočítač sa však stále dá pripojiť z bezdrôtovej siete 192.168.2.0/24, v ktorej je tiež koncovým zariadením.

Aj keď sa z uvedených obrázkov a textu môže zdať, že mikropočítače môžu využívať priamo študenti, nie je to reálne z viacerých dôvodov:

- veľký počet študentov v prvom, resp. druhom ročníku študijného programu Počítačové siete (Bc. štúdium) – hardvér pracoviska nie je dosť výkonný ani na pripojenie skupiny cca 20 študentov počas cvičenia,
- využívanie verejných IP adries v univerzitnej infraštruktúre, ktoré by mohli byť cieľmi (ne)úmyselných útokov zo strany študentov,
- nedostatok počítačov v multimediálnych učebniach, resp. ich zastaranosť a potreba inštalácie nástrojov na pripojenie,
- pri niektorých študentoch absencia zručností práce s operačným systémom Linux, resp. v prostredí textovej konzoly.

Najmä z týchto dôvodov sa na prostriedky pracoviska pripája iba pedagóg a s ich využitím realizuje praktické ukážky pre študentov. V súčasnosti (júl 2025) prebieha práca na vytvorení obrazov virtuálnych počítačov, ktoré by po inštalácii mohli využívať aj študenti.

A. *Možnosti experimentálneho pracoviska*

Prispôsobenie tohto experimentálneho pracoviska spočíva v pridaní balíčkov a knižníc do tzv. sieťového stacku. Keďže mikropočítače používajú operačný systém Linux, tak je pomerne jednoduché pridať ďalšie kroky pri spracovaní sieťovej premávky. Na tieto úlohy slúžia balíčky NetfilterQueue [2] a Scapy [3] spolu s knižnicou libpcap [4]. Využitím skriptov programovacieho jazyka Python je potom možné zadržať sieťovú premávku (aj prichádzajúcu aj odchádzajúcu), modifikovať ju a preposlať ju ďalej. Na kontrolu funkčnosti a vlastností steganografických techník je možné použiť viacero nástrojov, ako napr. analyzátor premávky Wireshark [5], resp. balíček na určenie prenosovej rýchlosti iPerf [6].

Experimenty realizované s využitím pracoviska ukázali, že pracovisko sa okrem zamýšľaných cieľov dá použiť aj na iné úlohy. Okrem prezentovania vedeckej činnosti pre záujemcov o témy záverečných prác sa pracovisko začalo využívať aj pri výuke nasledujúcich sieťovo orientovaných predmetov na Katedre elektroniky a multimediálnych telekomunikácií:

- Základy komunikačných technológií (1. ročník Bc. štúdia, letný semester) – základné sieťové utility (ako napr. ping a traceroute), prístup na zariadenia (napr. protokol SSH), administrácia operačného systému a jeho firewallu,
- Úvod do počítačových sietí (2. ročník Bc. štúdia, zimný semester) – skriptovanie v prostredí sietí (jazyk Python a balíček Scapy), zachytávanie sieťovej premávky (analyzátor Wireshark),
- Počítačové siete (2. ročník Bc. štúdia, letný semester) – ukážky jednoduchých útokov ako napr. ARP spoofing [7], [8].

Stručný obsah niektorých z vyššie uvedených praktických ukážok bude bližšie popísaný v podkapitole II-B. Okrem toho sa pracovisko používa aj pre demonštrácie techník sieťovej steganografie, no tento príspevok sa z dôvodu jeho obmedzenej dĺžky sústredí iba na popis niektorých praktických ukážok vo vyššie uvedených sieťovo orientovaných predmetoch.

B. *Možnosti iných výukových nástrojov v oblasti sietí*

Pre zabezpečenie praktickej výuky v oblastí sietí je vhodné mať k dispozícii nástroje ako na spoločné aktivity (v čase cvičení) tak aj pre jednotlivcov na samostatnú prácu (pre záverečné práce, resp. výskum doktorandov). Miernym problémom môže byť skutočnosť, že všetky nástroje, ktoré sa dajú využiť majú určité obmedzenia:

- pri simulátoroch (napr. Cisco PacketTracer) sú používatelia obmedzení možnosťami týchto nástrojov – niektoré procesy sú značne zjednodušené a neodrážajú realitu,
- pri emulátoroch (napr. GNS3) sú používatelia obmedzení ponukou obrazov operačných systémov, resp. firmwaru a navyše emulátory nemusia umožniť emuláciu určitých funkcií (napr. bezdrôtových sietí),
- pri virtualizácii (napr. s hypervízorom VirtualBox) môže používateľov obmedzovať hardvérová konfigurácia hostujúceho zariadenia, prípadne nemožnosť hardvérovej akcelerácie,
- pri práci s reálnymi zariadeniami používateľov obmedzuje najmä malý počet zariadení.

Ako vhodný kompromis sa pri výuke javí kombinácia vyššie uvedených spôsobov, pričom hlavne zložitejšie princípy ako napr. filtrácia premávky firewallom alebo jednoduché útoky je vhodné prakticky demonštrovať s využitím vopred pripravených softvérových prostriedkov na ustálenej zostave (spomenuté experimentálne pracovisko), ktorú neovplyvňujú zásahy študentov tak ako hardvér v učebni, resp. časté aktualizácie softvérových prostriedkov.

Experimentálne pracovisko sa pre takéto demonštrácie môže javiť ako ideálny nástroj, no treba si uvedomiť, že jeho hlavnou úlohou je overenie funkčnosti steganografických techník. Kvôli tomu sú v rámci pracoviska vykonané určité nastavenia (napr. staticky nastavené IP adresy na niektorých rozhraniach, pridané pravidlá vo firewalle, zdieľané SSH kľúče na pripojenie zariadení medzi sebou), ktoré znemožňujú znázornenie niektorých procesov, resp. techník pri výuke.

II. Práca s experimentálnym pracoviskom

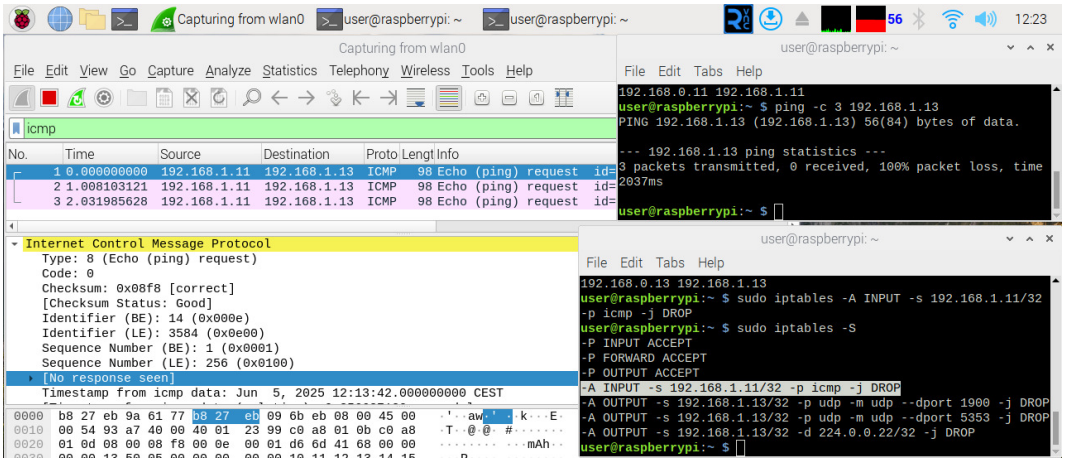
A. *Prístup na pracovisko*

Súčasťou distribúcie operačného systému Linux na mikropočítačoch v experimentálnom pracovisku je balíček OpenSSH umožňujúci pripojenie na zariadenia z lokálnej siete pomocou textovej konzoly. Na pripojenie s využitím grafického užívateľského rozhrania slúži aplikácia RealVNC Server, ktorá umožňuje prístup na tieto zariadenia z lokálnej alebo aj zo vzdialenej siete. Používateľ – pedagóg využíva na prístup z lokálnej siete aplikáciu RealVNC Viewer, do ktorej zadá IP adresu zariadenia na ktoré sa chce prihlásiť. Na prístup zo vzdialenej siete je potrebné prihlásenie sa účtom v tejto aplikácii, pričom sa využíva zašifrovaná komunikácia cez cloud spoločnosti RealVNC. Zriadenie účtu pre skupiny do 3 zariadení vrátane je v čase písania príspevku (jún 2025) bezplatné.

Využitie cloudu externej spoločnosti sa môže javiť nebezpečné, no treba si uvedomiť, že pri prístupe zo zariadení pripojených do univerzitnej infraštruktúry (aj počítače pripojené do pultu v posluchárňach) sa jedná o jednu (veľkú) lokálnu sieť. Tým pádom cloudové riešenie nie je potrebné používať. Iným riešením by bolo otvorenie portov na firewalloch, ktoré však treba zdôvodniť v žiadosti podanej správcovi infraštruktúry.

S využívaním pracoviska súvisí aj iný bezpečnostný problém. Pri IPv4 adresách odlišujeme tzv. verejné adresy, ktoré sú v celosvetovej topológii unikátne a tzv. súkromné adresy, ktoré sa môžu opakovať. Niektoré z adries v rámci univerzitnej infraštruktúry sú verejné. Ak by študenti získali tieto adresy, mohlo by to viesť ku jednoduchým útokom na tieto zariadenia, akým je napr. zahltenie určitým druhom správ. Preto je dôležité vyhnúť sa zobrazovaniu týchto adries. Pri prístupe na pracovisko sa používa istá forma proxy, kedy sa nepristupuje priamo na mikropočítače .11 až .13, ale najprv na mikropočítač .14 a na ostatné počítače sa pripája pomocou lokálnej siete v rámci pracoviska. Takémuto spôsobu prístupu sú prispôsobené aj firewally (na .14 bráni útokom z vonkajších sietí, ostatné mikropočítače ich majú prispôsobené pre steganografické experimenty).

Ukážku pripojenia a práce pomocou aplikácie RealVNC Viewer s využitím grafického užívateľského prostredia znázorňuje Obr. 3. V tomto prípade sa na mikropočítači .13 pridalo pravidlo pre blokovanie správ protokolu ICMP (z angl. Internet Control Message Protocol) prichádzajúcich z mikropočítača .11 čo vidno aj v Terminali (okno dole vpravo) alebo vo výstupe analyzátoru Wireshark (okno vľavo).



Obr. 3 Ukážka nastavenia pravidla firewallu a následného blokovania sieťovej premávky

B. Využitie pracoviska pri výuke

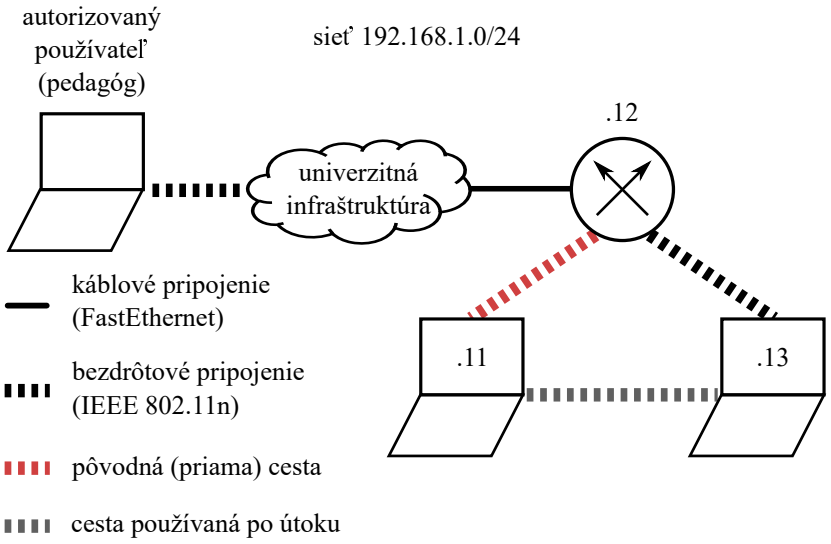
Okrem príkladu zobrazeného na Obr. 3 je možné pomocou experimentálneho pracoviska realizovať viacero ukážok vhodných pre výuku sieťovo orientovaných predmetov.

1) Overenie konektivity: Výhodou experimentálneho pracoviska pri overení konektivity utilitami ako ping, resp. traceroute je skutočnosť, že súčasne je k dispozícii viacero rozhraní (viď Tabuľku 1). Tým pádom je možné použiť jedno z nich na prístup pomocou protokolu SSH a ďalšie vypínať, resp. zapínať počas behu utilít a získavať tak rôzne chybové výpisy – hlásenia ako napr. vypršanie časovej lehoty na príjem správy (angl. Time Exceeded) alebo nedostupnosť cieľa zistená na uzle pred ním (angl. Destination Unreachable). Tieto hlásenia je možné získať aj využitím iných nástrojov (napr. simulátory), no pri pracovisku sa dopredu dajú pripraviť skripty, ktoré riadia vypínanie a zapínanie rozhraní a zobrazujú len podstatné informácie.

2) Zachytávanie premávky analyzátorom Wireshark: Analyzátor Wireshark je síce možné spustiť aj na emulovaných alebo virtuálnych zariadeniach, no treba si uvedomiť že takéto zariadenia nemusia podporovať určité druhy rozhraní (napr. bezdrôtové). Do istej miery sa takéto rozhrania dajú simulovať virtuálnymi rozhraniami, no je náročné dosiahnuť komunikáciu ako v reálnej bezdrôtovej sieti, kde sa rámce môžu zachytiť na všetkých zariadeniach. Pri použití experimentálneho pracoviska je možné sledovať aj ďalšie parametre spojení ako napr. prenosovú rýchlosť alebo počet zahodených paketov, prípadne uložiť zachytenú premávku na neskoršiu analýzu.

3) Manipulácia so sieťovou premávkou: Hlavnou úlohou experimentálneho pracoviska je spúšťanie rôznych skriptov umožňujúcich vkladanie, resp. extrakciu tajných správ zo sieťovej premávky. Jednotlivé kroky tohto procesu je možné aj demonštrovať pomocou Terminalu alebo analyzátoru Wireshark.

Zovšeobecnením tejto úlohy je nastavovanie pravidiel firewallu a následné spracovanie zachytenej premávky. Okrem preposielania správ na iné zariadenie sa dajú realizovať aj jednoduché útoky ako napr. ARP spoofing [7], [8]. V rámci experimentálneho pracoviska je možné pri tomto útoku oklamať zariadenia – poskytnúť im v správach protokolu ARP (Address Resolution Protocol) nesprávne MAC adresy. Týmto spôsobom sa dá zablokovať priama komunikácia medzi dvojicou mikropočítačov a presmerovať ju cez tretí, čo je znázornené na Obr. 4.



Obr. 4 Princíp útoku ARP spoofing v prostredí experimentálneho pracoviska

Miernym nedostatkom pri realizácii útokov v experimentálnom pracovisku je skutočnosť, že použitá distribúcia Linuxu je určená pre laických používateľov. Tým pádom obsahuje bezpečnostné mechanizmy na obranu proti takejto škodlivej činnosti. Pri odosielaní falošných ARP správ s nesprávnymi MAC adresami môže dôjsť ku detekcii duplicitnej MAC adresy priradenej ku IP adrese, resp. zablokovaniu rozhrania pri príjme veľkého počtu ARP správ. Takéto správanie je nutné pre správnu funkciu útoku potlačiť.

4) Práca so sieťovými operačnými systémami: V experimentálnom pracovisku prebieha administrácia mikropočítačov pomocou protokolu SSH, teda textovou konzolou, čo môže byť pre niektorých študentov novinka. Pri ovládaní tejto konzoly je možné demonštrovať rôzne linuxové príkazy, umiestnenie konfiguračných súborov v koreňovom systéme, overenie stavu služieb apod. Niektoré zo služieb sú na mikropočítačoch predvolene spustené kvôli testovaniu vlastností steganografických techník, čo šetrí čas pri praktických ukážkach počas výuky. Takisto sú k dispozícii záznamy (logy) z reálnej prevádzky.

III. Záver

Tento príspevok poukázal na možnosti použitia existujúceho nástroja na výuku študentov, ktorí pri praktických ukážkach môžu pochopiť širšie súvislosti v oblasti sietí. Niektoré činnosti je pre študentov stále vhodné robiť jednotlivo v sieťových simulátoroch, alebo emulátoroch, no ukážka na predstavenom experimentálnom pracovisku môže dobre poslúžiť pri nahradení niekoľkých snímok prezentácie s teoretickým popisom problému alebo viacstranového návodu na cvičenie.

V čase písania príspevku (jún 2025) bola najväčším nedostatkom pri využití experimentálneho pracoviska na výuku skutočnosť, že firewall je na niektorých zariadeniach nastavený „staticky“. Znamená to, že jednotlivé pravidlá je nutné meniť ručne. V ideálnom prípade by na zmeny pravidiel mohol slúžiť skript, ktorý by všetky obmedzenia pre podporné služby bežiace v pracovisku (napr. server na zdieľanie súborov) presmeroval na iné rozhranie. Takéto riešenie si však vyžaduje podrobné testovanie, aby sa nestala situácia pri ktorej sa služba nespustí korektne, resp. v najhoršom prípade sa do pracoviska nebude dať pripojiť.

V budúcnosti by bolo užitočné pripraviť všetky spúšťané softvérové prostriedky (vlastnú distribúciu operačného systému Linux spolu s konfiguračnými súbormi a spúšťateľnými skriptami) tak, aby mal k takýmto pracoviskám s využitím virtualizácie prístup každý študent. Táto problematika si však vyžaduje pomerne veľa práce – od optimalizácie distribúcie operačného systému tak, aby nezaberala veľa miesta na úložisku, cez ladenie skriptov tak, aby nevyžadovali veľké množstvo hardvérových prostriedkov hostujúceho počítača až po optimálnu spoluprácu s virtualizačnými nástrojmi na rôznych platformách (ako napr. Windows, macOS).

Poďakovanie

Táto práca bola financovaná pomocou EÚ NextGenerationEU grantu prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00019 – „Kľúčové digitálne technológie v oblasti bezdrôtových širokopásmových odolných prenosových sietí“, EÚ NextGenerationEU grantu prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I03-03-V05-00015 – „Využitie AI v oblasti bezdrôtových širokopásmových odolných prenosových sietí“ a grantu KEGA 041TUKE-4/2025 –„Laboratórium na výskum a vývoj pokročilých IoT senzorových sietí a širokopásmových senzorov“.

Literatúra

[1] J. Oravec, Ľ. Ovseník, S. Andrejčík, Hardvérový návrh a realizácia laboratória sieťovej steganografie, v *Proceedings of Electrical Engineering and Informatics 2023*. Košice (Slovensko), júl 2023, s. 153-157. ISBN: 978-80-5534-407-2.
[2] NetfilterQueue [online]. Citované 5.6.2025. Dostupné na internete: https://pypi.org/project/NetfilterQueue.
[3] Scapy: the Python-based Interactive Packet Manipulation Program & Library [online]. Citované 5.6.2025. Dostupné na internete: https://scapy.net.
[4] libpcap [online]. Citované 5.6.2025. Dostupné na internete: https://github.com/the-tcpdump-group/libpcap.
[5] Wireshark: The world's leading network protocol analyzer [online]. Citované 5.6.2025. Dostupné na internete: https://www.wireshark.org/.
[6] iPerf: The ultimate speed test tool for TCP, UDP and SCTP [online]. Citované 5.6.2025. Dostupné na internete: https://iperf.fr/.
[7] R. Kaur, E. Guriot Singh, S. Khurana, A Security Approach to Prevent ARP Poisoning and Defensive tools, *International Journal of Computer and Communication System Engineering*, 2015, roč. 2, č. 3, s. 431–437. ISSN: 2312-7694.
[8] Wikipedia: ARP spoofing [online]. Citované 3.7.2025. Dostupné na internete: https://en.wikipedia.org/wiki/ARP_spoofing.

Vplyv antény transpondéra na RF parametre bezkontaktného systému

Ľubomír LIVOVSKÝ

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

lubomir.livovsky@tuke.sk

***Abstrakt* — V tomto článku je opísaná metóda výpočtu parametrov transpondérového systému pre obvody NTAG I²C s anténami menších tvarových faktorov. Dôležité aspekty RF správania sa transpondérov SmartCard v bezkontaktnom vzduchovom rozhraní je možné riešiť postupom, ktorý určuje minimálnu prevádzkovú intenzitu magnetického poľa a maximálnu moduláciu záťaže. Tieto aspekty je možné analyzovať v závislosti od parametrov ekvivalentného elektrického obvodu transpondéra v kombinácii s parametrami antény, elektrických vlastnosti čipu a správaním sa transpondéra v bezkontaktnom systéme.**

***Kľúčové slová* — závitová anténa, NFC bezkontaktná technológia, ISO/IEC14443, NTAG I²C**

I. Úvod

V súčasnosti sa v praxi používa množstvo osobných kariet, pričom sa čoraz väčší dôraz kladie na aplikácie s vysokými nárokmi na bezpečnosť. Komunikačná technológia NFC (Near Field Communication) sa používa ako štandardizované používateľské rozhranie podľa normy ISO/IEC14443 Proximity Cards. Vďaka nej sa bezkontaktné riešenia, ako elektronické pasy, kreditné karty, prístupové systémy alebo cestovné lístky stali súčasťou každodenného života. Tieto aplikácie využívajú väčšinou štandardný formát bezkontaktnej karty ID-1, definovaný v norme ISO/IEC7810.

Pri výrobe takýchto kariet existujú technologické obmedzenia napr. závity antény nemožno umiestniť pod reliéfne alebo kontaktné oblasti, čo zmenšuje dostupnú plochu pre návrh a tvar antény. To si vyžaduje návrh kompaktnejších antén, najmä pre bezkontaktné kreditné karty, či malé zariadenia ako hodinky.

Výzvou je prepojiť tieto kompaktné antény s transpondérovými čipmi s vysokou bezpečnosťou a výkonom pre osobné aplikácie. Norma ISO/IEC 14443 definuje bezkontaktné rozhranie a existujúce čítačky sú s ňou plne kompatibilné, je potrebné zabezpečiť, aby aj miniaturizované riešenia dosahovali parametre porovnateľné s bežne používanými veľkosťami kariet (Class 1).

Kľúčovým krokom pri vývoji a integrácii takýchto systémov je meranie impedancie čipu (admitanciu, paralelnú kapacitu) a odpor na vývodoch antény. Tieto údaje sú nevyhnutné nielen pre overenie návrhu čipu, ale aj pre správny návrh antény a optimalizáciu celého systému podľa aplikačných požiadaviek. Aj keď je technológia 13,56 MHz rozšírená a dostupné LCR prístroje umožňujú zdanlivo jednoduché merania, existuje potreba presnejšieho, spoľahlivejšieho a relevantnejšieho vyhodnotenia čipov v ich reálnom prevádzkovom prostredí. Štandardné LCR prístroje často neberú do úvahy dynamické podmienky, ako je interakcia medzi anténou a čipom v bezkontaktnom režime, citlivosť na nízke budenie, alebo správanie pri vyšších úrovniach výkonu [1].

II. NTAG I²C transpondér

Na transformáciu akéhokoľvek systému na systém komunikácie NFC je potrebný integrovaný obvod, niekoľko pasívnych komponentov a anténa. Vzhľadom na rozsiahle pole použitia NFC existuje množstvo obmedzení, ktoré je potrebné riešiť, rešpektovať a splniť, aby bolo konečné riešenie spoľahlivé. Závitová vysokofrekvenčná anténa je distribuovaný komponent s indukčnosťou L ako hlavným prvkom, kapacitou C a odporom R ako parazitnými prvkami.

K dispozícii sú rôzne publikácie, ktoré opisujú aspekty návrhu závitovej antény, počnúc Maxwellovými analytickými rovnicami pre indukčnosť niektorých špeciálnych prípadov, až po mnoho užitočných aproximácií pre praktické prípady v RFID aplikáciách.

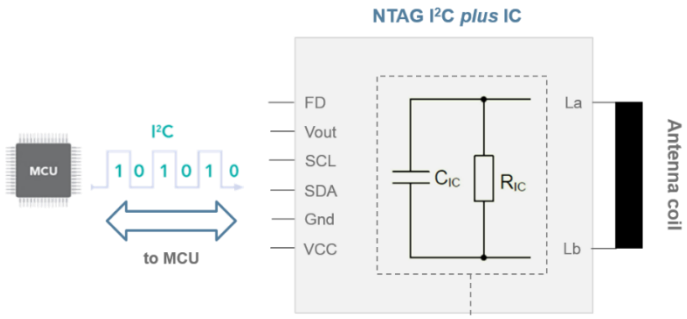
Návrh závitovej antény sa zvyčajne začína definovaním jej geometrie podľa požiadaviek,

pričom cieľom je dosiahnuť požadované hodnoty pre ekvivalentný elektrický obvod. Kľúčovým parametrom je indukčnosť, zatiaľ čo parazitná kapacita a odpor (predstavujúce straty) majú menší význam. V súčasnosti sa na odhad týchto parametrov často používajú analytické vzorce, ktoré poskytujú dostatočnú presnosť.

Ak je potrebná vyššia presnosť, zvyčajne sa vytvorí matica rôznych geometrických konfigurácií, ktoré sa následne zmerajú. Na základe týchto údajov sa model upraví podľa konkrétnych vlastností použitých materiálov alebo výrobných procesov napr. dielektrické straty polymérov môžu byť len približne známe [1].

A. Transpondérový obvod NTAG I²C

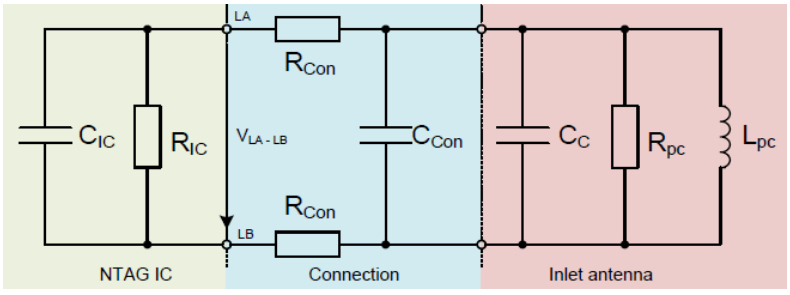
Integrované obvody NTAG od spoločnosti NXP Semiconductors poskytujú platformu pre návrh tagov (transpondérov) kompatibilných s NFC Forum Type 2 na báze štandardu ISO/IEC 14443A. RF rozhranie je v súlade s časťami 2 a 3 štandardu ISO/IEC 14443A. Obvody NTAG I²C vytvárajú mostík medzi bezdrôtovou vysokofrekvenčnou (RF) komunikáciou NFC a komunikáciou po I²C zbernici s mikrokontrolérom (MCU). Vysokofrekvenčné NFC rozhranie pracuje s prenosovou rýchlosťou 106 kb.s^{-1}. Obsahuje NTAG I²C pamäť SRAM (64B) a pamäť EEPROM (1kB alebo 2 kB), ktorá slúži na ukladanie dát vo formáte NDEF (NFC Data Exchange Format). Puzdro XQFN8 s rozmermi 1,6 × 1,6 × 0,5 mm, v ktorom sa NTAG I²C vyrába, uľahčuje integrovanie tohto obvodu do zariadení s podporou NFC, ako napr. rôzna spotrebná elektronika, elektronické regálové etikety alebo zariadenia pre inteligentné domy [2].



Obr.1: Principiálne zobrazenie použitia NTAG I²C obvodu [3]

Principiálnu schému transpondéra možno znázorniť zjednodušeným ekvivalentným obvodom uvedeným na Obr. 2. Vlastnosti čipu sú reprezentované ekvivalentnou vstupnou kapacitou C_{IC} (pre NT3H2111 je 50pF) a odporom R_{IC}, ktorý predstavuje stratový výkon čipu.

Tvarový faktor a parametre antény sú ovplyvnené vstupnou kapacitou pričom vstupná kapacita závisí od typu použitého čipu a veľkosti napätia na čipe. Montáž čipu k anténe môže pridať ďalšie straty na odpore prívodov R_{Con} a dodatočnú malú kapacitu C_{Con}. Anténa je reprezentovaná paralelným rezonančným obvodom, pozostávajúcim z indukčnosti L_{PC}, kapacity C_C a odporu R_{PC}.

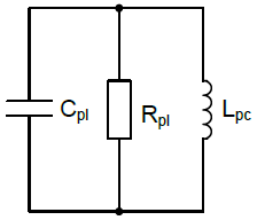


Obr.2: Ekvivalentný obvod tagu [4]

Kapacita C_{IC} spolu s kapacitou antény C_C a parazitnou kapacitou C_{Con} spojenia tvoria rezonančný obvod s indukčnosťou antény. Vstupný odpor R_{IC} spolu so stratovým odporom antény a odporom spojenia R_{Con} určujú činiteľ kvality Q tagu. Tento činiteľ kvality má vplyv na minimálnu (prahovú) veľkosť intenzity magnetického poľa H generovaného čítačkou, aby obvod NTAG IC mohol pracovať.

Ak je hodnota $R_{Con} << 1\ \Omega$, je možné obvod na Obr. 2 nahradiť paralelným obvodom (Obr. 3) s kapacitou C_{PL} a rezistorom R_{PL} a hodnoty kapacity a rezistoru určiť podľa vzťahov [4]:

$$C_{pl} = C_{IC} + C_{Con} + C_c \qquad R_{pl} = \frac{R_{IC} \cdot R_{pc}}{R_{IC} + R_{pc}} \tag{1}$$



Obr.3: Zjednodušený paralelný obvod tagu [4]

Ak budeme uvažovať minimálne prahové napätie potrebné pre činnosť NTAG IC, tak prahovú rezonančnú frekvenciou f_{RT} a prahový činiteľ kapacity Q_T môžeme určiť podľa vzťahov [4]:

$$f_{RT} = \frac{1}{2\pi\sqrt{L_{pc} \cdot C_{plT}}} \qquad Q_T = \frac{R_{plT}}{2\pi f_{RT} L_{pc}} \tag{2}$$

kde: C_{plT} je ekvivalentná kapacita tagu a R_{plT} je ekvivalentný odpor tagu pre prahovú rezonančnú frekvenciu f_{RT} [4].

B. Prahová intenzita magnetického poľa H_T

Veľkosť indukovaného napätia na NTAG IC generované poľom čítačky s prúdom I možno určiť podľa vzťahu (3), pričom podstatný vplyv na veľkosť tohto napätia má vzájomná indukčnosť medzi anténou transpondéra (tagu) a anténou čítačky.

$$V_{LA-LB} = \frac{2\pi f M}{\sqrt{\left(1-\left(\frac{f}{f_R}\right)^2\right)^2 + \left(\frac{2\pi f L_{pc}}{R_{pl}}\right)^2}} . I \tag{3}$$

kde: M - magnetický moment, f_R – vlastná rezonančná frekvencia antény, L_{pc} – indukčnosť antény transpondéra, R_{pl} – paralelný odpor antény transpondéra [4].

$$M = N_C . I . A_C \tag{4}$$

Ak budeme uvažovať strednú plochu závitu antény A_C a N_C počet závitov antény, potom prahovú intenzita magnetického poľa H_T potrebnú pre prevádzku obvodu NTAG IC môžeme vypočítať pomocou [4]:

$$H_T = \frac{\sqrt{\left(1-\left(\frac{f}{f_{RT}}\right)^2\right)^2 + \left(\frac{2\pi f L_{pc}}{R_{plT}}\right)^2}}{2\pi f \mu_0 N_C A_C} . V_{(LA-LB)min} \tag{5}$$

kde, $V_{(LA-LB)min}$ – minimálne napätie potrebné na prácu NTAG IC. Krivka prahovej intenzity magnetického poľa dosahuje minimum pri $f = f_{RT}$ a môže sa určiť z nasledujúceho vzťahu [4].

$$H_{Tmin} = \frac{L_{pc}}{\mu_0 N_C A_C R_{plT}} . V_{(LA-LB)min} \tag{6}$$

Pre konkrétny obvod NT3H2111 je možné použitím vzťahu (5) určiť priebehy magnetickej indukcie H_T v závislosti od zmeny kľúčových parametrov (napätie, indukčnosť antény, ekvivalentného rezistora transpondéra R_{plT}). Hodnoty parametrov transpondérov s veľkosťami antén od Class 1 až po Class 6, na doske plošného spoja (DPS) a flex fólii sú v Tabuľke 1.

Základným predpokladom pre činnosť transpondéra z energetického hľadiska je minimálna hodnota magnetického poľa H_{Tmin} a pre komunikáciu s čítačkou je demodulácia príkazov čítačky a amplitúda modulačného signálu. Vzťah (5) možno použiť na vyhodnotenie energetických požiadaviek na základe charakteristík konkrétneho čipu a antény. Tento vzťah je užitočný pre

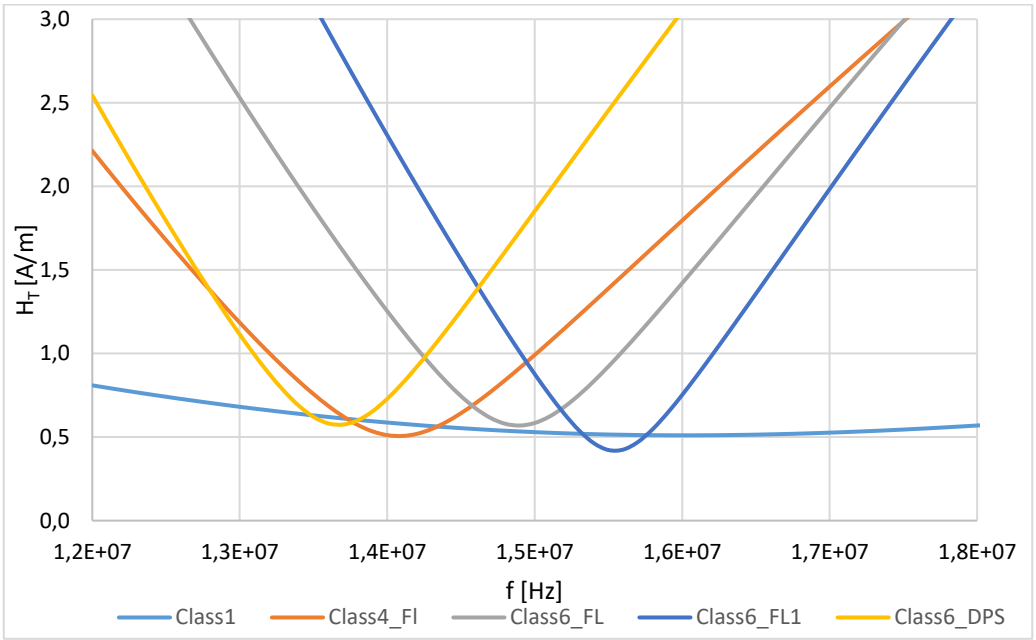
praktické úvahy návrhu transpondérov s NTAG I^2C, ktoré môžu určiť prípustné limity rezonančnej frekvencie za účelom zabezpečenia požadovanej intenzity magnetického poľa [1].

Modulácia záťaže pre čítačku znamená zmenu komplexného napätia antény, iniciovanú zmenou magnetického toku transpondéra, ktorý prechádza cez závitovú anténu čítačky. Môže sa považovať za externú moduláciu, pre ktorú nemožno definovať modulačný index. V typickom prípade väčšej antény čítačky a menšej antény transpondéra je magnetický moment transpondéra M dobrým parametrom nezávislým od tvaru antény, pretože charakterizuje schopnosť transpondéra generovať magnetické pole nezávisle od konkrétnej geometrie antény, a tak umožňuje jednoduchšie porovnanie účinnosti rôznych návrhov. Na hodnotu magnetického momentu má vplyv počet závitov antény N_C, prúd v jednom závite I a priemerná plochy antény A_C podľa vzťahu (4).

Tabuľka 1
Hodnoty parametrov transpondérov s obvodom NT3H211 pre rôzne veľkosti antén a nosných materiálov

Parameter	Class1	Class4 Flex	Class6 Flex	Class6 Flex1	Class6 DPS
L_{pc}	4,72 µH	0,852 µH	1,68 µH	0,989 µH	1,76 µH
C_A	2,8 pF	1,1 pF	1,0 pF	1,5 pF	1,4 pF
R_{LS}	3,66 Ω	1,35 Ω	3,3 Ω	1,5 Ω	2,02 Ω
R_A	117111 Ω	4207 Ω	7487 Ω	6220 Ω	11315 Ω
N_C	6	3	7	5	7
A_C	3069 mm^2	1072 mm^2	332 mm^2	361 mm^2	335 mm^2
f_R	44 MHz	163 MHz	123 MHz	133 MHz	99 MHz
Q_A	246	56	48	64	75
U_C	2,5V	2,4 V	2,4 V	2,4 V	2,4 V
C_C	17pF	50 pF	50 pF	50 pF	50 pF
f_{RT}	15,94 MHz	14,08 MHz	14,89 MHz	15,54 MHz	13,67 MHz
C_T	21 pF	150 pF	68 pF	106 pF	77 pF
R_T	1000	1000	2500	2500	2500
H_{Tmin}	0,51 Am^{-1}	0,51 Am^{-1}	1,42 Am^{-1}	1,04 Am^{-1}	1,43 Am^{-1}
Q_T	2,10	13,27	6,36	10,35	6,61

Nasledujúci obrázok ukazuje priebeh intenzity poľa H_T v závislosti od frekvencie magnetického poľa pre tagy s parametrami uvedenými v Tabuľka 1.



Obr.4: Priebeh intenzity magnetického poľa pre tagy s parametrami podľa tabuľky Tabuľka 1

III. Vplyv zmenšovania antény na výkon transpondérov

Pri optimalizácii analógového výkonu transpondérov s menšími anténami je potrebné sa zamerať na energiu a moduláciu zaťaženia. Z hľadiska energie je možné použiť úvahu, ak sa plocha závitu antény zmenší na polovicu, je možné to kompenzovať dvojnásobným počtu závitov, aby sa dosiahlo rovnaké indukované napätie, pretože Ui ~ N. Pre rezonančný systém

platí, že ak je kapacita čipu stanovená, potom musí mať zmenšená anténa rovnakú indukčnosť, aby sa dosiahla rovnaká rezonančná frekvencia. Avšak z $L \sim N^2$ je zrejmé, že obe podmienky nie je možné splniť súčasne [1].

Praktické riešenie je v použitie nízkoenergetického čipu, čo umožňuje použiť rezonančné frekvencie, ktoré sú vyššie ako nosná frekvencia, napr. odporúčaná je hodnota 16 MHz pre antény Class 1. Hoci vyššia rezonančná frekvencia znamená zvýšeniu intenzitu magnetického poľa potrebného pre prevádzku transpondéra, je tu priestor pre rozladenie rezonancie. Pre menšie antény, je potom možné posunúť rezonančnú frekvenciu na nižšiu hodnotu bližšie k nosnej (napr. 14,9 MHz), pretože šanca na rozladenie v dôsledku vysokej väzby je pre menšie antény menšia [1].

Pri zmenšovaní plochy antény čipu (NTAG IC) a súčasnom využívaní väčších antén čítačky je potrebné zvýšiť prúd v anténe transpondéra, aby bola modulácia dostatočne účinná. S tým súvisí kľúčový parameter kvality transpondéra Q_T. Ak má byť udržaný konštantný stratový výkon, nevyhnutne sa musí zvýšiť energia v systéme. To možno dosiahnuť napríklad zvýšením prúdu v anténe pomocou väčšej kapacity C_T. Zvýšenie kapacitu a zároveň zníženie počtu závitov a indukčnosť antény zvýši záťaž pri zachovaní tej istej rezonančnej frekvencie [1].

IV. Záver

Na návrhu antény transpondérov s použitím obvodov NTAG I^2C so zameraním sa na spoľahlivú NFC komunikáciu definovanej parametrami RF systému je možné použiť analytické vzťahy dostupné v mnohých odborných článkoch. Tieto vzťahy pomáhajú pri riešení vplyvu parametrov čipu, ako je napätie, odpor a integrovaná kapacita, na minimálnu intenzitu magnetického poľa a rezonančný frekvenčný rozsah. V praktickom návrhu transpondéra je nevyhnutné počítať s výrobnými toleranciami parametrov, ktoré môžu viesť k rozladeniu systému a zhoršeniu jeho väzbových vlastností. Významný vplyv na výslednú väzbu medzi čítačkou a transpondérom má najmä veľkosť a geometria závitovej antény.

Vlastnosti transpondérov podľa štandardu ISO/IEC14443 je možné optimalizovať vhodným návrhom antény na úkor šírky tolerančného okna, čo by umožnilo transpondérom s malými rozmermi antén dosiahnuť podobné výsledky, ako transpondérom s anténami štandardnej veľkosti.

Riešenie problému malých antén môže byť úprava pomeru medzi indukčnosťou a kapacitou antény so znížením počtu závitov antény a súčasným pridaním dodatočnej kapacity ku kapacite čipu. Túto dodatočnú kapacitu je možné implementovať do rôznych anténnych technológií (leptané, galvanicky vytvárané alebo tlačené antény) cez prekrižovanie vodičov alebo použiť obojstranný dizajn antény.

Poďakovanie

Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02- 00002.

Literatúra

[1] M. Gebhart, R. Szoncso, "Optimizing Design of Smaller Antennas for Proximity Transponders", Second International Workshop on Near Field Comunication, 2010, 77-82.

[2] https://www.nxp.com/docs/en/data-sheet/NT3H2111_2211.pdf

[3] https://community.nxp.com/t5/NXP-Designs-Knowledge-Base/Guide-to-designing-antennas-for-the-NTAG-I2C-plus/ta-p/1104729

[4] NXP Semiconductors, "NTAG Antenna Design Guide". AN11276, NXP Semiconductors, 2018.

[5] Y. Lee, "Antenna Circuit Design for RFID Applications", AN710, Microchip, 2003.

[6] M. Gebhart, "Analytical considerations for an ISO/IEC14443 compliant SmartCard transponder," Proceedings of the 11th International Conference on Telecommunications, Graz, Austria, 2011, pp. 9-16.

[7] D. Paret, „Antennas Designs for NFC Devices“, Willey, 2016.

[8] M. Gebhart, R. Neubauer, M. Stark and D. Warnez, "Design of 13.56 MHz Smartcard Stickers with Ferrite for Payment and Authentication," 2011 Third International Workshop on Near Field Communication, Hagenberg, Austria, 2011, pp. 59-64.

[9] M. Gebhart, J. Bruckbauer and M. Gossar, "Chip impedance characterization for contactless proximity personal cards," 2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010), Newcastle Upon Tyne, UK, 2010, pp. 826-830.

[10] M. Pesic, W. Pachler, S. Rampetzreiter and H. Arthaber, "Modeling and Design of Small, Passive, and Standard-compliant Proximity Coupling Transponders," 2020 IEEE International Conference on RFID (RFID), Orlando, FL, USA, 2020, pp. 1-7.

[11] H. Lehpamer, "RFID Design Principle”, ARTECH HOUSE, INC., 2008.

[12] https://community.nxp.com/t5/NFC/bd-p/nfc

Usability Evaluation and Optimization of University Websites in the Eastern Slovakia Region

Samuel HLIVA, Sergej CHODAREV

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

samuel.hliva@student.tuke.sk, sergej.chodarev@tuke.sk

Abstract **– The paper presents the results of a usability study conducted on the websites of the Slovak Technical University in Košice and the Pavol Jozef Šafárik University in Košice. The study was conducted using a combination of quantitative and qualitative methods, including surveys and user testing. Based on the results, multiple optimizations are suggested, implemented and tested using the same methodology as in the initial iteration. The results show that the websites in their current state have only sufficient usability level and multiple issues. The implementation of the suggested simple optimizations could substantially improve the usability of the websites.**

Keywords **– System Usability Scale (SUS), usability, user testing, university websites**

I. Introduction

Software developers must thoroughly test their software to minimize potential issues that may arise when the product is released and used by a diverse range of users. In web development, beyond ensuring core functionality, performance, and security, it is equally important to evaluate the usability of the product for different user groups.

Usability directly affects user satisfaction, task success, and overall effectiveness, making it a critical factor in modern software design. The term "*usability*" has been defined by J. Nielsen as *"quality attribute that assesses how easy user interfaces are to use"* [1]. Nielsen has defined [1] 5 key quality components of the usability:

1) Learnability: How easy it is for users to accomplish basic tasks the first time they encounter the design?
2) Efficiency: Once users have learned the design, how quickly can they perform tasks?
3) Memorability: When users return to the design after a period of not using it, how easily can they reestablish proficiency?
4) Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
5) Satisfaction: How pleasant is it to use the design?

With nearly 70% of the global population having access to the internet [2], websites have become essential tools for information dissemination across public and private institutions. For users to effectively retrieve and interact with information, it is crucial to design web interfaces that prioritize not only content delivery but also the overall user experience.

University websites have become a key component of interaction for prospective students, current students, academic staff, and administrative employees. They provide access to study programs, research activities, events, and various academic services. Given their importance, it is essential that these platforms offer high usability.

This paper presents a usability study focused on selected university websites in the Slovak Republic. Building upon existing research and established usability principles, a custom evaluation methodology is proposed. The research includes both quantitative and qualitative usability testing. Identified issues are addressed through the design of improved prototypes, which are subsequently evaluated for usability improvements.

II. Related Work

This section of the paper highlights related studies with a similar goal. During the research we focused on usability studies where the subject was an educational institution. Based on the methodologies and results of described studies this paper has conducted it's own methodology and result analysis.

A. European University of Lefke

In 2010, Caglar and Mentes conducted a usability study [3] of the European University of Lefke's website. The authors utilized widely adopted standardized questionnaire – Website Analysis and Measurement Inventory (WAMMI) [4] – aimed for researching website visitors satisfaction. The questionnaire focused on several key dimensions, including respondent's demography, usability expectations, and overall satisfaction with the website. With 293 respondents, the findings indicated a general dissatisfaction among users, with controllability, usefulness and efficiency identified as the most influential factors contributing to positive user experience, while attractiveness and memorability were highlighted as the weakest aspects.

B. Namik Kemal University

In July 2012, Mentes and Turan conducted a similar study [5] focused on Namik Kemal University. The authors used the same methodology as in the study by Caglar and Mentes [3], described in the subsection II-A, using the standardized WAMMI questionnaire. Collected responses from 339 respondents, including university students and academic staff. Based on the collected data, the researchers proposed usability improvements aimed at enhancing the overall user experience of the university's website.

C. Federal University of Technology Minna and Ibrahim Badamasi Babangida University Lapai

A. Ganiyu et al. conducted a usability study [6] on Federal University of Technology Minna (FUT) and Ibrahim Badamasi Babangida University Lapai (IBBUL) websites. The authors have evaluated 5 factors: content, structure and readability; navigation; design; performance and effectivity; overall usability. The methodology was thoroughly described in the study. The authors had decided to utilize a questionnaire as a data collection tool.

After calculating the sample size, authors conducted a pilot study to create a demography profile of respondents. Followed by performing the full research, they did an analysis of collected responses. For each of 5 evaluated factors they converted it to a numerical representation:

$$\chi = \frac{\sum_{i=1}^{n_f} Q_{fi}}{n_f} \tag{1}$$

where χ represents the numerical representation of each evaluated factor, n_f is the overall number of questions per factor, and $Q_f i$ is the merit for each question of the factor f.

When all the numerical representations had been made, the authors calculated the overall usability level:

$$A = \frac{\sum_{i=1}^{m_f} \chi}{n} \tag{2}$$

where A represents the overall usability level, m_f represents the number of researched factors, χ represents the result of each factor from the previous calculations, and n represents the number of questions per factor.

After computing the overall usability scores, the authors defined a classification scale to interpret the results, as presented in Table 1. This scale categorizes websites into four usability levels ranging from "Bad" to "Excellent" based on the average score obtained.

Table 1
Scale of usability ratings [6]

Score	Usability level
$0 \leq \chi \leq 1.5$	Bad
$1.5 < \chi \leq 2.5$	Poor
$2.5 < \chi \leq 3.5$	Good
$3.5 < \chi \leq 4$	Excellent

Their results of show that IBBUL website obtained a good usability level in terms of content, organization and readability, navigation and links and poor usability level in terms of user interface design and performance and effectiveness from the respondents. While FUT Minna website

obtained good usability level in terms of content, organization and readability, user interface design, and obtained excellent in terms of navigation and links, performance and effectiveness from the respondents.

III. Methodology

In the study we have combined quantitative and qualitative methods. We conducted a survey to gather quantitative data and used usability testing for qualitative results. Based on the findings we prepared the prototypes of the improved websites and conducted a new round of usability testing.

The survey was conducted with current university students as the first target group. The second group consisted of high school students, considered as prospective university applicants. Unlike the first group, prospective students were involved in structured user testing sessions.

A. Selection of the Universities

To ensure objective and diverse results, universities with different academic orientations were selected. As identified in related research, demographic variation significantly influences perceived usability. Therefore, the study included distinct user groups, specifically prospective and current students. Two universities based in the Slovak Republic were selected:

- Technical University of Košice (TUKE)[1],
- Pavol Jozef Šafárik University in Košice (UPJŠ)[2]

At the time of this research, no publicly available usability studies had been conducted on the websites of TUKE or UPJŠ. This, combined with their prominence in Eastern Slovakia and their distinct academic focus – TUKE specializing in technical disciplines and UPJŠ in natural sciences and humanities – made them suitable objects for our usability research.

B. Quantitative Usability Evaluation

Inspired by prior research utilizing the standardized WAMMI questionnaire, this study adopted a similar framework. However, due to licensing restrictions associated with WAMMI questionnaire, the System Usability Scale (SUS) [7] was selected as a freely available alternative. The SUS questionnaire was distributed among current students from each selected university. Its availability, simplicity, and ease of distribution made it a practical choice for gathering a large volume of responses.

The questionnaire was divided into three distinct sections. The first section gathered demographic information to classify respondents into relevant user groups. The second section consisted of the standard SUS questions, designed to measure overall usability perception. The third section included a mix of open-ended and Likert-scale questions (distinct from SUS), allowing respondents to reflect on their experiences and provide subjective feedback on their respective university websites.

C. Sample Size

According to the official website[3] of the Technical University of Košice (TUKE), the current student population is 12,417. In contrast, Pavol Jozef Šafárik University in Košice (UPJŠ) does not disclose this information directly. However, based on statistical data provided by third-party institution[4], the total number of currently enrolled students at UPJŠ is estimated at 7,067.

To determine the minimal sample size for acceptable margin of error, *Slovin's formula* was used, similarly to the approach presented by A. Ganiyu et al. in their usability study [6]. The formula is defined as follows:

$$n = \frac{N}{1 + Ne^2} \tag{3}$$

where n is the sample size, N is the total population, and e is the margin of error (in this case, set to 0.05 or 5%).

By substituting the studied universities population into the formula we obtain a minimum of 388 student responses for TUKE and a minimum of 379 responses for UPJŠ.

[1]Official website: www.tuke.sk
[2]Official website: www.upjs.sk
[3]Technical University of Košice, Basic Information: https://tuke.sk/sk/zakladne-informacie (accessed 2025-01-23)
[4]EduRank, Pavol Josef Šafárik University: https://edurank.org/uni/pavol-josef-safarik-university/ (accessed 2025-01-23)

D. Questionnaire Distribution

The questionnaire was distributed using a combination of online and offline approaches to ensure broad reach and diverse respondent participation. Online distribution included sharing through official social media channels of the Department of Computers and Informatics at TUKE, as well as via student-organized Discord[5] servers, including those associated with specific university courses.

In addition to these channels, the questionnaire was also shared directly by the author – either in person or via student-focused online groups on various social networking platforms. This multichannel strategy was employed to maximize engagement and representation across different faculties and academic years.

E. Qualitative Usability Testing

The qualitative component of this study was conducted through user testing sessions involving high school students, selected as representative prospective university applicants. This phase of the study took place in early February 2025 and included 6 participants between the ages of 18 and 19. Nielsen argues that "The best results come from testing no more than 5 users and running as many small tests as you can afford" [8]. The primary objective was to gather usability insights from users with little to no prior experience navigating university websites.

For purpose of this research method, three task-based scenarios were designed to reflect common actions performed by prospective students. These included searching for information about study programs, academic departments, and the admission requirements for applying to the university. Participants were provided with general instructions outlining the steps needed to complete each task.

F. User Testing Procedure

Prior to each session, participants were equipped with the necessary hardware, including a laptop and an external mouse, along with printed task scenarios and the SUS questionnaire to be completed at the end of the session. Participants were briefed on the structure of the session and informed that both their screen activity and audio would be recorded. They were explicitly told that the purpose of the session was not to evaluate their performance, but to assess the usability of the website.

To minimize distractions, each session was conducted in a controlled environment with only the participant and the moderator present. The moderator adhered to best practices in usability testing, refraining from interfering with the participant's actions in order to observe natural behavior, including how users responded to any encountered errors.

Following the completion of each session, a brief post-task interview was conducted to capture the participant's impressions and experiences. The insights gathered from these interviews were documented in a session protocol, which, alongside screen and audio recordings, served as a source of qualitative data for subsequent analysis. After the interview, each participant completed the SUS questionnaire to provide additional quantitative feedback.

In total, 36 unique tests were conducted across the sample of six users. To establish a referential usability benchmark, a SUS score was calculated for each university website based on the collected responses. The SUS score served as a reference point for the study and prior suggestions created and based on the results of the user testing sessions.

G. Implementing Changes to Prototypes

Based on the findings from the initial user testing, a set of targeted improvements was proposed and subsequently implemented into new prototypes for both university websites. The goal was to validate whether these enhancements could significantly improve usability metrics and task success.

Due to technical constraints, two different implementation strategies were used. For TUKE, the prototype was created using the prototyping tool. The prototype included redesigned versions of all key screens relevant to the tested scenarios, such as the homepage, admission pages, and faculty listings. Consistency with the original visual identity was maintained by reusing the existing color palette, visual elements, and content structure. Reusable components were introduced to simplify navigation, including a unified primary and secondary navigation system and a clearly structured header layout.

In contrast, the UPJŠ prototype was implemented by downloading local copies of the existing website and modifying the static HTML/CSS code. This approach was faster and allowed for

[5]Discord is a widely used communication platform among university students.

direct implementation of small usability improvements, such as clearer program labels, corrected links, and reordered content blocks. Improvements included repositioning the main call-to-action button, enhancing search discoverability, and introducing clear return paths to the homepage from deep-link pages.

Both prototypes were subjected to a second round of qualitative testing with a new group of participants under the same controlled conditions as the initial sessions. The aim was to determine whether the design interventions addressed the previously observed usability problems and to measure any improvement using SUS scores and task completion rates.

IV. Results and Discussion

This section presents the findings from both the quantitative questionnaire and the qualitative user testing. The results are analyzed to identify key usability issues and evaluate the effectiveness of the implemented improvements. Insights from both data sources are discussed to provide a comprehensive understanding of the current state and potential optimizations for the university websites.

A. Questionnaire Results

The questionnaire collected ninety-four (94) responses during the period of distribution. Twenty-seven (27) of them were obtained from the current students of Technical University of Košice and sixty-seven (67) from the current students of Pavol Jozef Šafárik University in Košice. Unfortunately, the number of responses did not meet the requirement of relevant sample size, according to *Slovin's formula*.

Although the sample size is too small to draw definitive conclusions, the results provide a directional understanding of user satisfaction across both websites. These findings should be interpreted as indicative rather than conclusive and are supported by complementary qualitative insights obtained during the user testing sessions.

B. Technical University of Košice

The results from the SUS questionnaire for TUKE indicate a "**Sufficient**" usability level[6], with a mean score of 63.7. Responses to open-ended questions revealed several areas for improvement that could enhance overall usability.

The demographic section of the questionnaire showed that 82% of respondents were male, with most aged between 22 and 25. Over 70% were in their final year of study, suggesting greater familiarity with the university website. The majority were students from the Faculty of Electrical Engineering and Informatics, which may reflect a more technically proficient respondent group, as supported by responses to related question.

Qualitative feedback highlighted general satisfaction with the website's design. However, navigation was frequently cited as a significant weakness. Although information was systematically distributed across different sections, users reported difficulties in locating content efficiently.

In summary, the findings suggest that the website would benefit from improvements in navigational structure and the addition of practical features such as dark mode and quick-access links to frequently visited pages.

C. Pavol Jozef Šafárik University

The results from the SUS questionnaire for UPJŠ indicate a "**Sufficient**" usability level, with a mean score of 62.6%. Open-ended questions revealed that participants generally viewed the website's visual design as satisfactory. However, the search functionality was frequently reported as problematic or entirely non-functional. Several users also noted that the website contained outdated information and described the overall structure as inadequate.

The demographic data showed that nearly 80% of respondents were women. Most were aged between 18 and 21, followed closely by the 22-25 age group. The majority were first-year students enrolled in either bachelor's or master's programs. Most participants were affiliated with the Faculty of Public Administration, while other faculties were represented in roughly equal proportions, with the exception of one that was not represented at all.

In summary, the findings suggested that the website would benefit from improved content maintenance, functional search capabilities, and clearer prioritization of key information.

Table 2
Task Success Rates on Original Websites and Improved Prototypes

Website Version	1st Scenario	2nd Scenario	3rd scenario	Average	Improvement
Original Website					
TUKE	83.33%	83.33%	83.33%	**83.33%**	-
UPJŠ	100%	93.33%	88.10%	**93.81%**	-
Improved Prototype					
TUKE	80.00%	95.83%	88.89%	**91.57%**	+8.24 pp.
UPJŠ	97.22%	100%	97.62%	**98.28%**	+4.47 pp.

D. User Testing

Six high-school students participated in task-based usability testing for both websites. Success rates for original website interactions ranged from ≈ 83% to 100%, as presented in the Table 2, with navigation steps often requiring clarification or repetition.

1) Technical University of Košice: During testing, several problematic areas of the website were observed. A major finding was the confusing navigation—information was too densely packed, making it hard for users to find specific items. Some links were repeated across categories, leading to uncertainty. The section with tables on admissions was also problematic—faculty abbreviations confused users and led to misinterpretation.

On the other hand, users generally rated the website as clear and easy to use. Many appreciated the visual design and clarity. One usability quality—memorability—stood out clearly. It was intentionally included in test scenarios by repeating actions. All users agreed the website's functions were well implemented.

2) Pavol Jozef Šafárik University: Testing revealed several usability issues with the website. The most frequent problem occurred at the transition between the first and second scenario. Information about faculties and their study programs is located on a separate website, which lacks a direct link back to the main university page. This missing link caused difficulties for users in nearly every case—they expected, based on experience with other websites, that clicking the logo in the top left corner would return them to the university's homepage.

Another common issue was identifying links that were more subtle than other elements despite being crucial. Users often overlooked the main action of the page while scrolling down and only discovered it after scrolling back up. This shortcoming confirms findings from the quantitative testing of the UPJŠ website.

E. User Testing on Improved Prototypes

For the TUKE website, the following improvements were implemented in the prototype based on the usability testing results:

- Redesigning the faculty labels in the Admissions table to make them clearly identifiable.
- Placing the main navigation on the same side as the secondary menu.
- Simplifying the navigation content.

The UPJŠ website prototype included the following improvements:

- Implementation of a link that returns the user to the homepage when clicked on the Study Programs Register portal.
- Implementation of a Call-To-Action link on the Contacts subpage.
- Adjusting the layout in the Study Programs Register so that for study programs with the same name but different forms of study, the differences are clearly identifiable.

Prototypes incorporating suggested improvements demonstrated a noticeable increase in task success and satisfaction. For example, TUKE's improved prototype raised the SUS score from 85.83 to 91.66, classified as "**Excellent**" usability, together with the average task success rate increase of 8.24 percentage points, and UPJŠ experienced the raise of SUS score from 85.83 to 89.16, and also reached an increase in average task success rate of 4.47 percentage points. All the quantitative details are presented in Table 2.

V. Conclusion

Although the quantitative sample size limits strong conclusions, the consistency between qualitative and SUS data points to key issues in both websites: unclear structure, lack of search effectiveness, and insufficient personalization of high-priority content. Improvements such as restructured menus, prioritization of admissions links, and implementation of search and accessibility features are recommended.

[6]Based on the adopted SUS interpretation scale [9]

The study demonstrates that even simple, targeted improvements can yield measurable usability gains. The prototype testing showed that addressing navigation issues and implementing basic user interface conventions can improve both task success rates and user satisfaction scores. This suggests that university websites, which serve as critical touchpoints for prospective students, could benefit significantly from systematic usability evaluation and iterative design improvements.

Future work should aim to validate these insights using a statistically significant sample and perform A/B testing with real users on production environments.

ACKNOWLEDGEMENT

This research was supported by the Slovak Research and Development Agency under grant number APVV-23-0408 “Evolving Architectural Knowledge in the Edge-to-Cloud Continuum”.

REFERENCES

[1] J. Nielsen, “Usability 101: Introduction to usability,” 2012, accessed: 2025-05-07. [Online]. Available: https://www.nngroup.com/articles/usability-101-introduction-to-usability/
[2] “Individuals using the internet,” 2 2024. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx
[3] E. Caglar and S. A. Mentes, “The usability of university websites–a study on European University of Lefke,” *International Journal of Business Information Systems*, vol. 11, no. 1, pp. 22–40, 2012.
[4] J. Kirakowski and N. Claridge, “WAMMI: Website analysis and measurement inventory questionnaire,” 2016. [Online]. Available: https://www.wammi.com/samples/index.html
[5] S. A. Mentes and A. H. Turan, “Assessing the usability of university websites: An empirical study on Namik Kemal University.” *Turkish Online Journal of Educational Technology-TOJET*, vol. 11, no. 3, pp. 61–69, 2012.
[6] A. A. Ganiyu, A. Mishra, J. Elijah, and U. M. Gana, “The importance of usability of a website,” *IUP Journal of Information Technology*, vol. 13, no. 3, pp. 27–35, Sep. 2017.
[7] J. Brooke *et al.*, “SUS-a quick and dirty usability scale,” *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.
[8] J. Nielsen, “Why you only need to test with 5 users,” https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/, 2025, 2025-03-07.
[9] D. Derisma, “The usability analysis online learning site for supporting computer programming course using system usability scale (SUS) in a university,” 2020.

Capturing Typing Patterns in Web Applications for Author Identification Research

[1]*Eduard MICHALEK,* [2]*Marek HORVÁTH*

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

[1]eduard.michalek@student.tuke.sk, [2]marek.horvath@tuke.sk

Abstract **– This paper describes the design and implementation of a web based system for collecting and analyzing keystroke dynamics data with the goal of future author identification. The developed application records detailed typing characteristics, including keypress duration, key release time, and inter-key intervals, during programming tasks completed by users in a browser environment. All collected metrics are securely stored in a centralized database for subsequent analysis. After each task, the system automatically compares the new keystroke data with the users existing profile to identify deviations in typing behavior. This approach enables continuous and nonintrusive monitoring of user behavior without requiring additional hardware or complex configuration. The collected dataset is foundation for further research in author identification, behavioral biometrics, and the development of advanced methods for detecting identity misuse or academic dishonesty in online programming environments.**

Keywords **– Author identification, behavioral biometrics, keystroke dynamics, typing patterns, web application**

I. Introduction

The identification of individuals based on their typing behavior is a concept with historical roots reaching back to the 19th century, originating in the era of the telegraph. Telegraph operators were known to recognize each other by the unique rhythm and cadence with which they transmitted Morse code, a phenomenon later termed the ”Fist of the Sender” [1]. This early recognition of motor patterns as behavioral signatures highlighted the potential for distinguishing people based on subtle, habitual actions. Over time, this principle evolved from analog telegraphy to the modern analysis of keyboard input, forming the basis of what is now known as keystroke dynamics. In the digital age, where the boundaries between the physical and virtual worlds continue to blur, the study of these patterns offers new opportunities for strengthening security and verifying user identity with minimal disruption.

Keystroke dynamics has become a prominent behavioral biometric method. The key metrics typically analyzed include hold time, which measures how long a key is pressed, and flight time, which refers to the interval between releasing one key and pressing the next [2]. By capturing these and related features, it is possible to construct a behavioral fingerprint that can be used for authentication or even author identification. Compared to traditional knowledge based methods such as passwords, keystroke dynamics provides a distinct layer of security by relying on involuntary habits that are challenging for impostors to mimic. This approach helps mitigate common vulnerabilities, including brute force attacks and credential theft, by making successful impersonation significantly more difficult.

Biometric authentication methods are generally divided into two broad categories physiological and behavioral [3].

- Physiological methods include fingerprints, facial features, and iris scans, all of which provide high accuracy but usually require dedicated sensors or cameras.
- Behavioral biometrics, such as voice recognition, mouse movement analysis, and keystroke dynamics, operate in the background and typically do not require specialized hardware [4], [2].

This makes them particularly suitable for integration into web applications and services, where security is highly valued. Keystroke dynamics stands out for its accessibility, as it can be implemented using standard input devices already present in nearly all computing environments. At the same time, it enables continuous verification throughout a session rather than relying solely on one time authentication events.

Despite advantages, keystroke dynamics presents several challenges. Typing behavior can fluctuate based on factors such as user stress, physical fatigue, or changes in keyboard hardware [5], [6]. Ongoing research suggests that combining typing metrics with adaptive algorithms, multifactor authentication, or even machine learning techniques can help address these vulnerabilities and improve both reliability and security [7], [8]. In this context, the present paper introduces a web application designed to monitor typing behavior during programming tasks. Typing metrics are recorded and compared to each user historical data after each task, with warnings issued when deviations are detected. This approach enables passive behavioral monitoring without introducing friction or additional steps into the workflow, providing a practical model for author identification and continuous authentication in browser based environments.

II. Related Works

Recent studies have increasingly examined the potential of typing behavior analysis for user authentication and author identification within digital environments. Giot et al. [7] investigated the use of digraph timing for user verification and emphasized the importance of anonymization techniques to safeguard privacy. Their results indicate that combining several timing metrics, such as hold time and flight time, can lead to more accurate user identification. Further research has explored the impact of various factors on the stability of typing patterns, including changes in user mood, fatigue, or keyboard hardware, which may influence the consistency of collected biometric data over time. Idrus et al. [8] studied how keystroke dynamics can be used not only for authentication, but also for the estimation of demographic attributes such as gender and age, thereby broadening its relevance for identity management. Other works have focused on the usability of these methods in everyday scenarios, noting that their integration into web applications is possible without specialized hardware and with minimal disruption to the user experience.

Security considerations remain a central theme in the literature. Hazan et al. [5] proposed injecting decoy data into authentication processes as a means to counteract replay attacks, which are a known vulnerability in behavioral biometric systems. Shadman et al. [6] discussed the emerging risk of artificial intelligence models capable of generating synthetic typing data that closely mimics legitimate user patterns. To address these concerns, researchers have suggested various adaptive approaches and the use of anomaly detection algorithms, as well as the combination of keystroke data with other biometric or behavioral information to increase reliability and security. There is also a growing focus on privacy protection, especially in systems that monitor user behavior continuously in the background. The present work draws on these findings by implementing keystroke dynamics monitoring in a web application for programming tasks, placing emphasis on practical deployment, user privacy, and the challenges of maintaining reliable author identification in real world use.

III. Methodology

This study examines the use of keystroke dynamics for author identification in the context of programming tasks. The application was designed to monitor and analyze typing patterns while users solve assignments in a simple online environment. From the beginning, the emphasis was placed on capturing natural user behavior with minimal distraction or technical obstacles. The interface offers a familiar code editor, and users are not required to install any additional software or perform complex setup. As participants complete tasks, their typing is recorded automatically, focusing on a core set of metrics that reflect their personal style. These metrics include the time each key is pressed, the pause between keys, overall typing speed, the total number of keystrokes, and the timing of specific key pair transitions. The data collection process is transparent for the user, and no personal information is stored. Filtering removes irregular keypresses and accidental modifier use, ensuring the reliability of the dataset.

After each assignment, the captured metrics are processed and compared with the users previous records. All information is saved in a secure database, with each session linked to metadata such as task number and timestamps. For each new input, the system automatically checks for changes relative to established typing patterns. If an unusual deviation is detected, the user receives a clear warning. This procedure allows ongoing behavioral monitoring while avoiding any unnecessary disruption to the users workflow. The analysis is focused on highlighting both general and specific aspects of the users typing. While aggregate values like average key press time and typing speed

show global trends, tracking selected key transitions reveals subtle behavioral differences that can be important for author identification. Throughout the process, the user experience remains consistent, with immediate feedback provided after each task submission.

IV. Results

The effectiveness of the application was evaluated by recruiting multiple users to perform a set of programming tasks under similar conditions. All participants used the same computer and keyboard layout to reduce possible sources of variation. Each person was asked to complete several *C* programming assignments, and their typing was measured across repeated sessions. The analysis included the calculation of average key press times, key intervals, typing speeds, and a comparison of frequently used key combinations. The collected results were visualized to make the observed differences more accessible and to support the interpretation of data.

Figure 1 shows a comparison of general typing metrics for two users completing the same assignment. The chart illustrates that User 2 had longer key press times and higher average intervals between keys, typed more slowly, and pressed more keys overall.

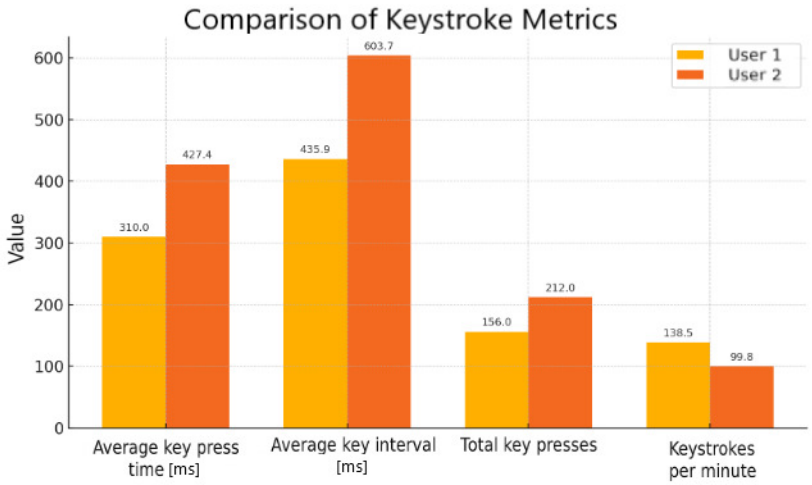


Fig. 1 Comparison of general keystroke metrics

Figure 2 further examines the timing between specific pairs of keys. In this case, differences are visible for several combinations. User 2 exhibited a longer interval between the letters *m* and *a*, as well as *s* and *l*, while both users had similar timings for *c* and *l*. These findings demonstrate that even when working on the same task, users maintain measurable and unique typing rhythms that can be used for identification.

The system was also tested for its ability to detect when an unauthorized user tries to impersonate someone else. In a controlled scenario, one participant logged in under anothers account and attempted to complete a task. The analysis, summarized in Table 1, revealed clear and significant differences in core metrics, such as average key press time, key interval, and typing speed. The system promptly flagged the session as suspicious. The table also compares several key combinations, showing where deviations occurred and confirming that the combination of general and specific metrics improves detection.

The visualizations in Figure 1 and Figure 2 support the finding that typing behavior is both measurable and personal. They show not only overall tendencies, but also highlight the unique signature of each user at the level of detailed key transitions. The success in detecting impersonation further illustrates the applications potential as a behavioral monitoring tool for real world environments, including education or secure workspaces. While the current version was tested on a limited number of participants, the results indicate that author identification using keystroke dynamics is a feasible approach that deserves further investigation and development. Limitations remain, especially in accounting for variability due to stress, device changes, or small datasets, but future work can address these with adaptive analysis methods and larger participant pools.

A. Conclusion

The results of this experiment show that monitoring keystroke dynamics in a simple programming environment enables identification of users based on their natural typing patterns. The

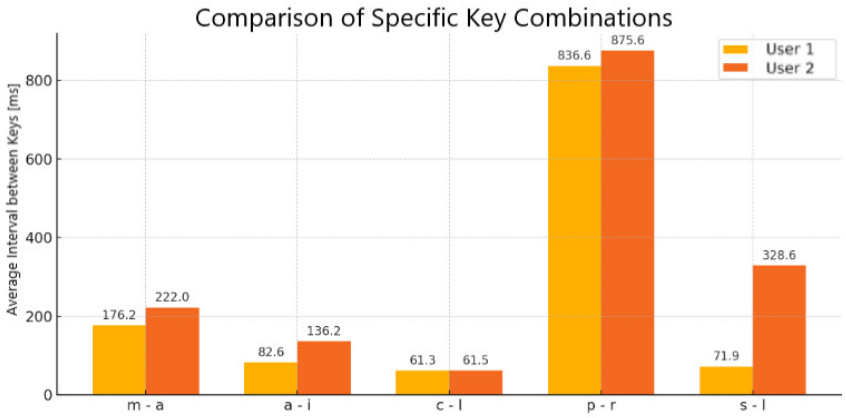


Fig. 2 Comparison of specific key pair timings

Metric	*Legitimate User*	*Impostor*	*Significant Deviation*
Avg. key press [ms]	389.39	233.89	Yes
Inter-key interval [ms]	549.15	322.03	Yes
Total keystrokes	73.00	70.00	No
Keystrokes per minute (KPM)	124.19	189.01	Yes
c → l	111.86	94.00	No
m → a	369.80	210.50	Yes
e → l	341.83	239.90	Yes
p → r	1095.19	1023.69	No
w → o	413.60	216.80	Yes
a → i	152.10	147.50	No

Table 1
Comparison between legitimate user and impostor for selected metrics.

system consistently detected deviations in behavior, flagging possible risks when an impersonator attempted to complete tasks under another users account. At the same time, the analysis confirmed that ordinary fluctuations due to factors such as fatigue or emotional state can influence typing, but careful adjustment of comparison logic and threshold settings helped to limit false alarms. Although the evaluation was carried out on a relatively small sample, the application demonstrated strong potential as a practical biometric tool for continuous user verification, particularly in contexts where seamless operation and minimal user intervention are essential. Future work could focus on extending the method to a larger and more diverse group of users, as well as experimenting with advanced machine learning techniques to further improve accuracy and adaptability. Study highlights that keystroke dynamics can provide an effective layer of behavioral security in online education, secure coding platforms, and remote workspaces without the need for specialized hardware or disruptive authentication steps, making it a promising approach for environments where trust and verification are critical.

Acknowledgment

This work was supported by project KEGA No. 061TUKE-4/2025 Building Bridges between University and High School ICT Education.

References

[1] S. P. Banerjee and D. L. Woodard, "Biometric authentication and identification using keystroke dynamics: A survey," *Journal of Pattern recognition research*, vol. 7, no. 1, pp. 116–139, 2012.
[2] P. S. Teh, A. B. J. Teoh, and S. Yue, "A survey of keystroke dynamics biometrics," *The Scientific World Journal*, vol. 2013, no. 1, p. 408280, 2013.
[3] D. Bhattacharyya, R. Ranjan, F. Alisherov, M. Choi *et al.*, "Biometric authentication: A review," *International Journal of u-and e-Service, Science and Technology*, vol. 2, no. 3, pp. 13–28, 2009.
[4] A. Babich, "Biometric authentication. types of biometric identifiers," 2012.
[5] I. Hazan, O. Margalit, and L. Rokach, "Securing keystroke dynamics from replay attacks," *Applied Soft Computing*, vol. 85, p. 105798, 2019.
[6] R. Shadman, A. A. Wahab, M. Manno, M. Lukaszewski, D. Hou, and F. Hussain, "Keystroke dynamics: Concepts, techniques, and applications," *arXiv preprint arXiv:2303.04605*, 2023.
[7] R. Giot, M. El-Abed, C. Rosenberger *et al.*, "Keystroke dynamics authentication," *Biometrics*, vol. 1, pp. 157–182, 2011.
[8] S. Z. S. Idrus, E. Cherrier, C. Rosenberger, and P. Bours, "Soft biometrics database: A benchmark for keystroke dynamics biometric systems," in *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*. IEEE, 2013, pp. 1–8.

Využitie kamerového systému v paletizačno-montážnych robotických aplikáciách

[1]*Jaroslava ŽILKOVÁ,* [2]*Peter GIROVSKÝ,* [3]*Samuel TOMKO*
[1,2,3] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 1/9, 04200 Košice-Sever, Slovenská republika

[1]jaroslava.zilkova@tuke.sk, [2]peter.girovsky@tuke.sk, [3]samuel.tomko@student.tuke.sk,

***Abstrakt* — Článok prezentuje realizáciu paletizačnej úlohy pomocou kolaboratívneho robota MELFA ASSISTA. Zámerom práce bolo vytvoriť demonštračné pracovisko pre robotický systém vhodný pre presné a spoľahlivé ukladanie objektov na paletu a prípadne aj manipuláciu s paletami. Simulácia pracoviska bola vytvorená pomocou RT ToolBox 3 a následne realizovaná na fyzickom robotickom ramene. Problém s nepresným polohovaním objektov pri realizácii bol vyriešený implementovaním kamerového systému Wenglor. Výsledkom riešenia je flexibilné demonštračné pracovisko vhodné pre automatizáciu paletizačných procesov a manipuláciu s objektmi.**

***Kľúčové slová* — kolaboratívny robot, RT ToolBox 3, kamerový systém Wenglor, paletizácia**

I. ÚVOD

Priemyselná robotika zažíva dynamický rozvoj, ktorý výrazne ovplyvňuje spôsob, akým prebieha výroba v rôznych odvetviach. S rastúcimi požiadavkami na automatizáciu, presnosť, rýchlosť a bezpečnosť práce narastá aj význam nasadzovania robotických systémov do nových výrobných a manipulačných procesov.

Jednou z kľúčových oblastí priemyselnej automatizácie je manipulácia s rôznymi objektmi a ich časťami, ako aj samotná paletizácia, teda proces organizovaného ukladania výrobkov na palety. Jedná sa o monotónnu, fyzicky náročnú prácu s vysokou mierou opakovateľnosti. Vo výrobných pracoviskách vykonávajú túto prácu najmä priemyselné roboty. Automatizácia tejto úlohy prináša podnikom nielen zníženie nákladov na pracovnú silu, ale aj zvýšenie efektivity, spoľahlivosti a bezpečnosti práce.

Na súčasnom trhu existuje pomerne veľké množstvo priemyselných robotov rôznych výrobcov, ktoré sú vhodné pre riešenie úloh paletizácie. Pre tieto priemyselné roboty sú charakteristické určité vlastnosti, ku ktorým patrí:

- vysoká výkonnosť a rýchlosť práce,
- vysoká presnosť a opakovateľnosť,
- ľahká konštrukcia, ktorá umožňuje dosiahnuť vysokú dynamiku, kratšie doby cyklu a vyššiu produkciu,
- všestrannosť, ktorú zabezpečujú rôzne nosnosti, dosahy a univerzálne dimenzované rozhrania a prívody energií,
- kompaktnosť konštrukcie, ktorá umožňuje rýchlu a ľahkú integráciu do už existujúcich zariadení,
- bezúdržbovosť a energetická úspornosť, ktoré sú dosiahnuté vďaka robustnej konštrukcii s dôrazom na malé opotrebenie a použitiu vhodných pohonných systémov, čím umožňujú veľmi dlhé intervaly údržby.

Tak ako výrobcovia priemyselných robotov ponúkajú svoje roboty, rovnako poskytujú aj riadiace systémy, bezpečnostné prvky a podporné služby pre tvorbu robotických pracovísk. Často sa jedná o software umožňujúci jednoduchším užívateľským spôsobom nakonfigurovať robotickú aplikáciu, simulačný software, ktorý umožňuje prejsť od off-line programovania cez simuláciu pracoviska až po virtuálne uvedenie pracoviska do prevádzky a pod.

II. DEMONŠTRAČNÉ PALETIZAČNÉ PRACOVISKO

Vytvorenie demonštračného pracoviska bolo riešené v spolupráci so spoločnosťou Mitsubishi Electric, ktorá ponúka široký rozsah priemyselných robotov, ktoré klasifikuje do niekoľkých skupín:

- vertikálne roboty s kĺbovým ramenom - rady RV-FR a RV-CR, ktoré sú vhodné hlavne pre komplexné montážne práce,
- roboty SCARA - rady RH-FR a RH-CRH , ktoré majú vynikajúce využitie hlavne pri manipulácii s potravinami alebo súčiastkami
- roboty SCARA pre čelnú montáž - rada RH-FRHR , ktoré sú vhodné pre veľmi rýchlu manipuláciu s elektronickými súčiastkami a ich osadzovaním

Pri týchto robotoch je kladený dôraz na vysokú presnosť a rýchlosť pre efektívne použitie vo výrobných bunkách.

Náš prvotný zámer bol spojený s použitím priemyselného robota s kĺbovým ramenom, ale nakoľko našim cieľom bolo vytvoriť demonštračné pracovisko vhodné pre účely výučby a praktické overovanie teoretických znalostí z oblasti robotiky, experimenty a možné ďalšie rozširovanie demonštračného pracoviska, rozhodli sme sa použiť iný - kolaboratívny robot spoločnosti Mitsubishi Electric – MELFA (Mitsubishi Electric Flexible Automation) ASSISTA.

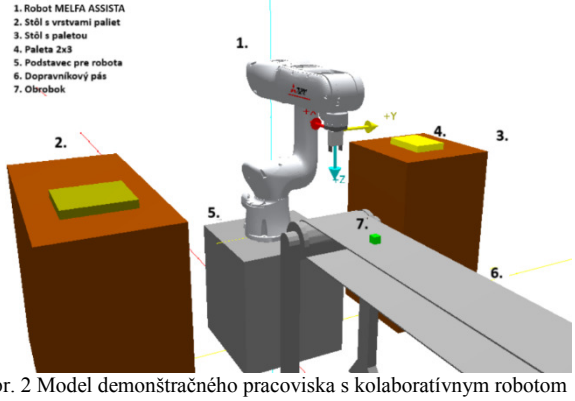
Obr. 1 RV-5AS-D (MELFA ASSISTA)

Robot MELFA ASSISTA (Obr. 1), vďaka pokročilým bezpečnostným technológiám umožňuje zdieľať pracovný priestor s ľuďmi. Medzi tieto technológie patria senzory kolízie s nastaviteľnou mierou citlivosti ako aj kolaboratívny mód, ktorý umožňuje robotu pohybovať sa s maximálnou rýchlosťou 250 mm/s alebo 50 mm/s, čím sú dodržiavané bezpečnostné normy pre prácu v bezprostrednej blízkosti človeka. Okrem toho je MELFA ASSISTA vybavený jednoduchým ovládaním, možnosťou programovania aj bez použitia kódu (napr. pomocou ručného vedenia) a vysokou presnosťou pohybov, vďaka čomu je ideálny pre laboratórne, edukačné alebo rôzne ľahké aplikácie.

A. Štruktúra a model demonštračného pracoviska

Demonštračné pracovisko zahŕňa vertikálny kolaboratívny robot MELFA ASSISTA, ktorý je tvorený sériovým kinematickým reťazcom šiestich rotačných kĺbov, pričom sú použité striedavé servopohony s brzdou pre každú os. Nosnosť ramena je 5kg (vrátane koncového efektora) pričom robot má dosah 910 mm.

Hlavnou úlohou demonštračného pracoviska je manipulácia s objektmi z určeného miesta, ktorým je stôl alebo pásový dopravník, na paletu umiestnenú na stole s paletou, prípadne pri zmene koncového efektora prenesenie palety. Demonštračné pracovisko, ktorého model zobrazený na Obr. 2, bol zostavený pomocou softwaru RT ToolBox 3 spol. Mitsubishi Electric. Pozostáva z viacerých častí, pričom k najdôležitejším patrí: robot MELFA ASSISTA s riadiacim systémom, stôl s vrstvami paliet, stôl s paletou, podstavec robota a pásový dopravník s objektmi, ktorý by mal byť neskôr doplnený aj na reálnom pracovisku.



Obr. 2 Model demonštračného pracoviska s kolaboratívnym robotom

B. Definovanie možných problémov

Základnou požiadavkou zadania je presná manipulácia s objektmi a ich umiestnenie v príslušnej palete. Systém navigácie by mal zabezpečiť dostatočne presné a spoľahlivé polohovanie ramena, uchopenie objektu a jeho presun na presné miesto na palete. Hlavným problémom by mohla byť nepresná poloha manipulovaného objektu. Keď robot vychádza z preddefinovaných súradníc, môže a stať, že objekt nebude na očakávanom mieste, čo by viedlo k chybnému uchopeniu alebo úplnému zlyhaniu úlohy.

Základným, jednoduchším riešením pri manipulácii s objektmi je ich umiestnenie v známej presne určenej polohe v pracovnom priestore robota. Na zabezpečenie takéhoto riešenia sa zvyčajne používajú rôzne mechanické časti zariadení, dorazy, špeciálne prípravky, rôzne vedenia a podobne.

Iný prístup, ktorý nevyžaduje presné polohovanie objektu je lokalizácia objektu pomocou kamerového systému. Výhodou kamerového systému je, že okrem lokalizácie objektu v priestore umožňuje aj ďalšie využitie ako napr. selekciu objektov, kontrolu povrchu, a pod.

Z týchto dôvodov sme rozhodli rozšíriť pôvodné pracovisko o kamerový systém, ktorý umožní rozšíriť možnosti využitia pracoviska o ďalšie funkcie.

III. Demonštračné pracovisko doplnené o kamerový systém

Systém strojového videnia bol implementovaný pomocou kamery Wenglor-typ B50M001 (Obr. 3) a s využitím softvéru Univision 2, ktorý umožňuje spracovanie obrazu a identifikáciu objektov.

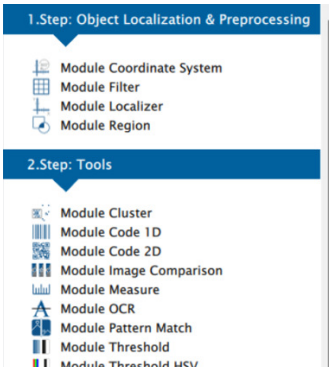
Obr. 3 Kamera Wenglor-typ B50M001.

Táto inteligentná kamera, ktorá je založená na technológii Wenglor MultiCore kombinuje funkcie skenera a zrakových senzorov. Umožňuje preto snímanie všetkých bežných 1D kódov a rôznych typov 2D kódov. Automatické zaostrovanie, výber oblasti záujmu a sledovanie zabezpečujú spoľahlivé a stabilné snímanie obrazu. K dispozícii sú nasledujúce moduly na spracovanie obrazu: kontrola rozmerovej presnosti, triediace procesy, kontrola prítomnosti, počítanie objektov, výstup polohy, počítanie pixelov, optické rozpoznávanie znakov, porovnávanie vzorov, možnosti filtrovania a vyhodnocovanie štatistík.

Kamera bola prepojená s robotom a po rozpoznaní objektu v obraze boli jeho aktuálne súradnice zaslané do riadiaceho systému robota pomocou protokolu TCP/IP. Ten následne vykonal pohyb na získané súradnice a spoľahlivo uchopil objekt aj keď sa nachádzal mimo pôvodne definovanej pozície. V nasledujúcich častiach je opísaný postup pripojenia kamerového systému, podľa ktorého sme postupovali.

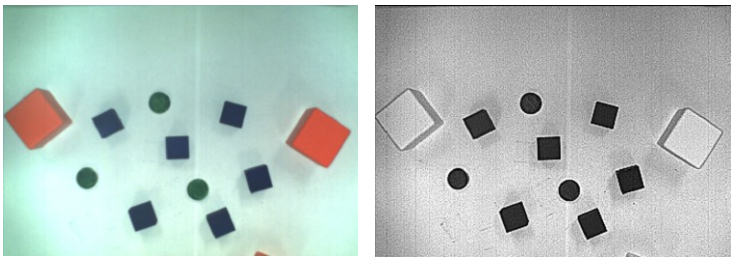
A. Vytvorenie programu pre detekciu objektov pomocou software Univision 2

Univision 2 je software vyvinutý firmou Wenglor, ktorý je kompatibilný s rôznymi kamerovými systémami, ktoré táto firma ponúka. Programovanie sa vykonáva pomocou preddefinovaných modulov, pričom každý z týchto modulov (Obr. 4), poskytuje iné možnosti spracovania snímaného obrazu.



Obr. 4 Dostupné moduly v programe Univision 2

Prvý z modulov *Filter* obsahuje algoritmus na spracovanie vstupného obrazu do formy vhodnejšej pre detekciu tvaru.

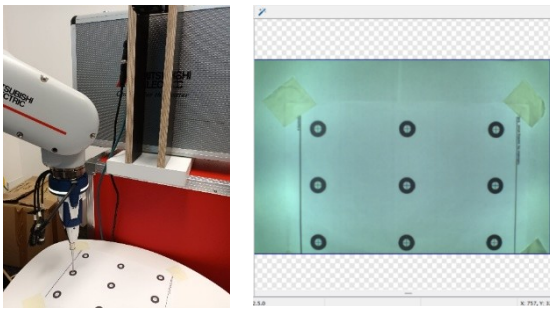


Obr. 5 Pohľad na rôzne typy objektov (vľavo výstupný obraz kamery, vpravo výstup z modulu *Filter*)

Výstupný obraz z kamery je zobrazený na Obr.5 vľavo. Naším cieľom je detegovať správny typ objektu (modré kocky) a zabezpečiť odoslanie súradníc objektu na TCP/IP výstup kamery. Pre overenie rozpoznávacej schopnosti kamery sme medzi modré kocky zamiešali aj iné typy objektov. Prvý modul, ktorý použijeme, je modul *Filter*. Tento modul obsahuje algoritmus na spracovanie vstupného obrazu do formy vhodnejšej pre detekciu tvaru. Výstupný obraz z modulu *Filter* je uvedený na Obr. 5 vpravo. Pomocou modulu *Pattern Match* z výstupu modulu *Filtra* na Obr. 5 kameru naučíme tvar nášho objektu.

B. Kalibrácia medzi robotom a kamerovým systémom

Dôležitým krokom pre zabezpečenie správneho prepojenia kamerového systému s robotickým ramenom je presná kalibrácia. V tomto kroku prebehne stotožnenie súradnicového



Obr. 6 Kalibračný hárok z pohľadu fyzického robota (vľavo) a z pohľadu kamery (vpravo)

systému robota so súradnicovým systémom kamery, za vytvorenia transformačnej matice medzi týmito systémami. Kalibračný hárok pre Cognex vision system s deviatimi terčíkmi bol upevnený k podložke a v strede efektora sme uchytili predmet s ostrým vrcholom pre jednoznačné určenie polôh (Obr. 6 vľavo). Kamera má viditeľných všetkých deväť terčíkov (Obr. 6 vpravo). Robot bol napolohovaný nad stredy jednotlivých terčíkov a pri dosiahnutí stredu bola táto pozícia zaznamenaná. Zároveň boli odčítané aj zodpovedajúce súradnice X a Y v pixeloch (Obr. 7).

Calibration1 *

Teaching Points:

Enable	No.	Robot.X[mm]	Robot.Y[mm]	Camera.X[pixel]	Camera.Y[pixel]
☑	1	218.160	509.600	601.000	172.000
☑	2	222.310	617.580	382.000	170.000
☑	3	226.490	725.050	162.000	170.000
☑	4	158.750	727.630	162.000	310.000
☑	5	154.390	620.020	382.000	310.000

Get the Robot Position | 9 /20

Obr. 7 Stotožnené pozície robota s pozíciami kamery

C. Vytvorenie transformačnej matice

Z deviatich naučených polôh dostávame transformačnú maticu, ktorá reprezentuje našu kalibráciu (Obr. 8).

Result Homography Matrix: (VSCALB1)

M11	M12	M13	M14	=	-0.0215	-0.5154	327.8239	0.0000
M21	M22	M23	M24		-0.5227	-0.0088	844.3850	0.0000
M31	M32	M33	M34		-0.0000	-0.0000	1.0566	-1.6010

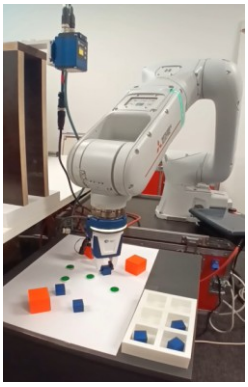
Obr. 8 Transformačná matica zodpovedajúca kalibrácii

IV. REALIZÁCIA DEMONŠTRAČNÉHO PRACOVISKA

Aj keď prvotným cieľom bolo vytvoriť pracovisko s priemyselným paletizačným robotom, po uvážení edukačného zamerania pracoviska bol pri realizácii použitý kolaboratívny robot MELFA ASSISTA. Vytvorené demonštračné pracovisko (Obr. 9) môžeme charakterizovať:

- zvýšenou bezpečnosťou – kolaboratívne roboty sú navrhnuté na bezpečnú spoluprácu s človekom bez nutnosti ochranných bariér, čo je ideálne do laboratórneho a školského prostredia,
- jednoduchšou obsluhou – riešenie ponúka intuitívnejšie programovanie a umožnenie rýchlych zmien a rýchlejšie nasadenie, čo uľahčilo testovanie a ladenie úlohy v reálnych podmienkach,
- vhodnosťou pre demonštračné a edukačné účely – kolaboratívny robot je praktickejší v prostredí, kde sa vyžaduje častá interakcia s operátorom, bezpečné testovanie a pod.

Zároveň bol overený spôsob prenesenia simulovaného riešenia úlohy do samotnej realizácie v prostredí práve s kolaboratívnym robotom, ktorý je určený aj pre bezpečnú prácu mimo plne automatizovaných priemyselných liniek.



Obr. 9 Manipulácia s objektom za použitia kamerového systému

V. ZÁVER

Demonštračné robotizované pracovisko s kolaboratívnym robotom MELFA ASSISTA doplnené o kamerový systém Wenglor je určené hlavne pre demonštračné účely a praktické overenie teoretických znalostí z oblasti robotiky.

Pracovisko je možné zároveň využívať pri riešení rôznych projektov zameraných na riešenie robotických úloh s využitím kamerových systémov. Výsledné riešenie robotického pracoviska zameraného na proces paletizácie a manipuláciu s objektmi spája simuláciu, realizáciu a aplikáciu strojového videnia.

POĎAKOVANIE

Článok bol publikovaný s podporou projektu APVV-19-0210. Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00018.

LITERATÚRA

[1] MITSUBISHI ELECTRIC. Priemyselné roboty [online]. [cit. 2025-06-05]. Dostupné na: https://sk.mitsubishielectric.com/fa/products/rbt/robot
[2] MITSUBISHI ELECTRIC. Kolaboratívny robot ASSISTA [online]. [cit. 2025-06-05]. Dostupné na: https://emea.mitsubishielectric.com/fa/products/rbt/assista/assista
[3] WENGLOR. Popis inteligentnej kamery [online]. [cit. 2025-06-05]. Dostupné na: https://www.wenglor.com/en/Machine-Vision/Smart-Cameras-and-Vision-Sensors/weQube-B50/Smart-Camera-B50/SmartCamera/p/B50M001

Detekcia nenávistných prejavov v slovenskom jazyku s využitím modelov BERT

[1]*Miroslava MATEJOVÁ,* [2]*Ján PARALIČ,* [3]*Anna LEONENKO*

[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]anna.leonenko@student.tuke.sk

Abstrakt **— Toxické komentáre a nenávistné prejavy predstavujú čoraz väčší problém v online prostredí. Hlavným cieľom tohto článku je vytvorenie modelu na klasifikáciu toxických komentárov s rôznymi úrovňami toxicity v slovenskom jazyku. Použité boli dva slovenské datasety a jazykové modely typu transformer BERT a SlovakBERT. Okrem klasifikácie boli implementované aj techniky vysvetliteľnej umelej inteligencie, konkrétne metódy LIME a SHAP, ktoré umožnili zobraziť slová s najväčším vplyvom na rozhodnutie modelu. Práca zároveň upozorňuje na výzvy pri práci so slovenskými dátami, najmä v oblasti anotácie, jazykovej variability a vyváženosti datasetov.**

Kľúčové slová**— BERT, hlboké učenie, nenávistné prejavy, SlovakBERT, toxické komentáre, vysvetliteľná umelá inteligencia**

I. Úvod

V posledných rokoch sa zvyšuje výskyt nenávistných prejavov a toxických komentárov na sociálnych sieťach a v diskusných fórach. Môžu mať rôzne formy, od priamych útokov až po jemnejšie formy diskriminácie. Tieto prejavy môžu mať negatívny vplyv na psychickú pohodu používateľov a na kvalitu verejného diskurzu. V súvislosti s tým sa zvyšuje potreba efektívnych nástrojov, ktoré by umožnili automaticky detegovať a filtrovať nevhodný obsah. Zatiaľ čo v anglicky hovoriacom prostredí existuje množstvo riešení, pre slovenský jazyk sú dostupné nástroje a datasety veľmi obmedzené. Táto skutočnosť zdôrazňuje relevantnosť témy a potrebu vytvorenia riešení špecifických pre slovenský jazyk.

Cieľom tejto práce bolo vytvoriť klasifikačný model, ktorý dokáže spoľahlivo rozlíšiť toxické a netoxické komentáre v slovenčine. Techniky spracovania prirodzeného jazyka (NLP) v spojení s hlbokým učením (DL) zohrávajú v tejto oblasti nezastupiteľnú úlohu. Pomáhajú pri automatickom rozpoznávaní a klasifikácii škodlivého obsahu, ako je nenávistný prejav, kyberšikana, urážky, hrozby a iné formy toxicity.

Okrem základnej klasifikácie sa táto práca ďalej zaoberá vysvetliteľnosťou modelov umelej inteligencie (XAI) prostredníctvom metód LIME (Local Interpretable Model-Agnostic Explanations) a SHAP (SHapley Additive exPlanations). Tieto techniky nám môžu poskytnúť lepšie pochopenie toho, ktoré slová ovplyvňujú výstup modelu a môžu slúžiť ako podpora pre jeho ladenie alebo kontrolu.

II. Analýza stavu problematiky

Detekcii toxických alebo falošných správ v slovenčine sa venovali v posledných rokoch viacerí autori [1], [2]. V článku [1] sa zamerali na klasifikáciu krátkych textov do viacerých stupňov toxicity. Autori vytvorili vlastný dataset, Toxic comments in Slovak social networks, pomocou extrahovania komentárov zo sociálnych sietí. Dataset pozostáva z 3 091 komentárov a má skoro rovnomerné zastúpenie jednotlivých tried: neutrálne (25,2 %), málo toxické (24,49 %), stredne toxické (25,11 %) a veľmi toxické (25,2 %). Komentáre boli anotované na základe obsahu pomocou vytvoreného lexikónu toxických slov. Autori využili modely strojového učenia a učenie súborom metód. Konkrétne to boli Naive Bayes, SVM (Support vector machine), Random forest a Bagging. Najlepšie výsledky dosiahol model SVM, priemernú úspešnosť (accuracy) 89 %. Výsledky úspešnosti ostatných modelov sa pohybovali v rozsahu od 75,9 % po 85,7 %.

V oblasti automatickej detekcie toxických komentárov a nenávistného obsahu zohrávajú kľúčovú úlohu modely DL, predovšetkým architektúry typu transformer. Model BERT

(Bidirectional Encoder Representations from Transformers) priniesol zásadný prielom v spracovaní prirodzeného jazyka vďaka svojej schopnosti učiť sa kontextuálne reprezentácie textu obojsmerne. Zaujímavou prácou je štúdia [3] zameraná na porovnanie rôznych modelov hlbokého učenia na klasifikáciu toxických komentárov. Ide o prácu s anglickými textami, ale autori sa v nej zamerali aj na vplyv rôznych metód predspracovania a reprezentácie textu. Využívali dátový súbor z Kaggle Toxic Comment Classification Competition. Tento dataset obsahuje komentáre z Wikipédie, ktoré môžu byť označené ako toxické, silne toxické, obscénne, urážlivé alebo nenávistné. Komentár môže byť označený viacerými typmi toxicity. V rámci experimentov boli porovnávané rôzne DL modely, ale aj modely typu transformer ako napríklad BERT a jeho modifikácie. Ukázalo sa, že model BERT-base uncased, základná verzia predtrénovaného BERT modelu, dosiahol najlepšie výsledky: úspešnosť 89,71 %. Termín "uncased" znamená, že pri predspracovaní dát sa neberie do úvahy rozdiel medzi veľkými a malými písmenami, čo môže zlepšiť schopnosť modelu zovšeobecňovať pri analýze textu. Tieto zistenia naznačujú, že hlboké učenie, najmä s využitím modelov ako BERT, môže byť veľmi efektívne pri detekcii toxického správania v online diskusiách. Pre slovenský jazyk bol vyvinutý špecifický model SlovakBERT [4], ktorý je trénovaný na veľkom množstve slovenských textov a preto je vhodnejší pre úlohy v slovenčine.

Okrem modelov NLP sa čoraz častejšie uplatňujú metódy XAI. Medzi najpoužívanejšie patria LIME, ktorý lokálne aproximuje správanie modelu pomocou jednoduchších modelov ako napríklad lineárna regresia, a SHAP, ktorý vychádza z teórie kooperatívnych hier a poskytuje kvantifikáciu vplyvu jednotlivých prvkov vstupu na výstup modelu. Tieto prístupy umožňujú lepšie porozumenie rozhodovaciemu procesu zložitých modelov a sú dôležité pri zvyšovaní dôvery v automatizované systémy. Článok [5] sa zaoberá detekciou nenávistných prejavov na sociálnych médiách pomocou vysvetliteľnej umelej inteligencie. Autori ukazujú, ako môžu byť metódy XAI použité na zlepšenie prehľadnosti a spoľahlivosti modelov pri detekcii nenávistných prejavov. Demonštrujú použitie metód XAI, ako je LIME na zlepšenie vysvetliteľnosti modelov, aj SHAP ktorá sa zameriava na poskytnutie globálnych vysvetlení modelov. Tieto metódy umožňujú používateľom lepšie pochopiť, prečo modely dospeli k určitým rozhodnutiam, čo je kľúčové pre transparentnosť automatizovaných riešení.

Výskum v oblasti detekcie toxického obsahu je intenzívny najmä v anglicky hovoriacom prostredí, kde sú dostupné rozsiahle datasety ako Jigsaw [6] alebo HateXplain [7]. V prípade slovenčiny je situácia zložitejšia, a preto je nutné využívať menšie datasety, často vytvárané manuálne alebo upravované zo zahraničných zdrojov. Táto práca preto zároveň poukazuje na potrebu budovania jazykových zdrojov pre menej zastúpené jazyky.

III. NÁVRH RIEŠENIA

Experimentálna časť sa zameriava na praktickú realizáciu klasifikátora toxických komentárov v slovenskom jazyku pomocou hlbokých neurónových sietí a metód vysvetliteľnej umelej inteligencie. Pri implementácii bola použitá metodológia CRISP-DM [8], ktorá viedla postup od pochopenia problému cez spracovanie dát až po finálne vyhodnotenie výsledkov. Táto metodológia bola ešte rozšírená o použitie metód vysvetliteľnosti. **Upozornenie:** Nasledujúce ukážky obsahujú text, ktorý môže byť považovaný za urážlivý alebo nevhodný. Tento materiál je zahrnutý výhradne na výskumné účely v oblasti detekcie toxického obsahu a pochádza z verejne dostupného datasetu.

A. Pochopenie cieľa

Cieľom našej práce je navrhnúť a implementovať efektívny model na detekciu nenávistných komentárov v slovenčine. Takýto model by v budúcnosti mohol prispieť k zlepšeniu online komunikácie, a to zabezpečením bezpečného a pozitívneho prostredia na sociálnych sieťach. Tento model by mal byť schopný nielen identifikovať toxické komentáre, ale aj poskytovať vysvetlenia pre svoje rozhodnutia. Z hľadiska analýzy dát je cieľom vyvinúť klasifikačný model, ktorý dokáže presne rozlíšiť medzi netoxickými a rôznymi stupňami toxických komentárov. Analýza dát nám umožní identifikovať vzory a charakteristické znaky toxického správania, čo vytvorí pevný základ pre návrh presného a spoľahlivého modelu na detekciu nenávistných prejavov v textoch.

B. Pochopenie dát

V experimentoch boli použité dva datasety so slovenskými komentármi. Prvým datasetom je Toxic comments in Slovak social networks [1] z článku popísaného v časti II. Zistili sme, že

komentáre označené ako neutrálne sú v priemere kratšie ako toxické komentáre, pričom najdlhšie komentáre patria do kategórie stredne toxických. V druhej časti analýzy sme sa zamerali na identifikáciu najčastejšie používaných slov v komentároch označených ako toxické. Tento krok nám pomáha pochopiť, aká slovná zásoba je charakteristická pre toxické prejavy v tomto datasete. Analýza najčastejšie používaných slov v toxických komentároch ukázala, že najčastejšie sa používajú tieto slová: ludi – 212, ludia – 181, clovek – 104, hanba – 97, teraz – 92, chore – 80, Matovic – 77, niekto – 72, problem 72, Matelko – 69. Tieto slová naznačujú, že toxické komentáre sa často týkajú aktuálnych udalostí, politiky a verejných osobností.

Druhá dátová množina je dataset toxických komentárov v slovenčine Toxic-sk[1]. Tento dataset sa zameriava na analýzu a klasifikáciu toxických komentárov v slovenčine, pričom obsahuje komentáre z Facebooku. Dataset obsahuje 8 840 manuálne anotovaných komentárov, ktoré sú rozdelené do dvoch hlavných kategórií: toxické (označené hodnotou 1) a netoxické (označené hodnotou 0). Percentuálne zastúpenie tried je rovnomerné: toxické – 50 % a netoxické – 50 %. Každý komentár bol anotovaný skupinou študentov a asistentov z Technickej univerzity v Košiciach, pričom ich rodným jazykom je slovenčina. Anotácia bola vykonaná na základe osobného názoru anotátora, pričom každý komentár bol označený iba raz. Pri analýze datasetu sme porovnali dĺžku komentárov v jednotlivých triedach. Výsledky ukázali, že priemerné dĺžky komentárov boli dlhšie u toxických komentárov. Podobne ako aj pri prvom datasete sme sa zamerali na analýzu najčastejšie používaných slov v toxických komentároch. Výsledky ukázali, že najfrekventovanejšie výrazy v komentároch sú: ludi – 163, ludia – 153, matovic – 126, pan – 124, clovek – 95, premier – 90, k**ot – 85, konecne – 85, igor – 82, uplne – 76. Tieto najčastejšie slová v toxických komentároch naznačujú, že sú zamerané na osobné útoky a politické témy.

C. Príprava dát

Pred aplikovaním modelov na klasifikáciu toxicity v komentároch bolo potrebné vykonať niekoľko krokov predspracovania, ktoré zabezpečia konzistenciu a vhodnosť dát pre trénovanie modelu.

Pre prvý dataset, ktorý obsahuje komentáre, ktoré sú označené podľa úrovne toxicity, sme museli upraviť pôvodné textové označenia tried na číselné hodnoty. Tento krok je dôležitý, pretože väčšina modelov vyžaduje numerické vstupy. Každá trieda bola mapovaná na číslo podľa nasledujúcej schémy:

- neutrálne → 0
- málo toxické → 1
- stredne toxické → 2
- veľmi toxické → 3

Predspracovanie dát zahŕňalo aj čistenie textu. Pre model BERT bola odstránená diakritika. Ďalej boli odstránené špeciálne znaky, čísla a slovenské stop-slová pomocou zbierky stopwords-sk[2]. Bola vykonaná tokenizácia a rozdelenie dát na trénovaciu a testovaciu množinu v pomere 80:20.

D. Modelovanie

Na trénovanie boli použité dve architektúry modelov: BERT a špecializovaný SlovakBERT. Celkovo boli vykonané štyri rôzne experimenty. Prvým experimentom bola kombinácia prvého datasetu a modelu BERT. Model BERT sme inicializovali s použitím predtrénovanej váhy a nastavili počet výstupných tried na 4 zodpovedajúcim rôznym úrovniam toxicity. Na trénovanie bol použitý optimalizátor AdamW s učebným koeficientom (learning rate) 5e-5. Trénovanie bolo nastavené na 10 epoch, pričom sme počas každej epochy vykonávali aktualizáciu váh modelu.

Ďalším experimentom bol model SlovakBERT natrénovaný na prvom datasete. Použili sme váhy, ktoré sú predtrénované na slovenskom texte a opäť nastavili 4 klasifikačné triedy. Pri trénovaní modelu SlovakBERT sme zvolili nižší učebný koeficient 2e-5. Keďže ide o predtrénovaný jazykový model, nižší learning rate umožňuje jemné doladenie bez výrazného narušenia jeho existujúcich váh. Model sa trénoval tiež 10 epoch.

Modelovanie pokračovalo kombináciou druhého datasetu a opäť modelu BERT. Pre model sme použili rovnaké parametre ako v prvom experimente.

Posledný experiment bol spojením druhého datasetu a modelu SlovakBERT. Pre model sme zachovali rovnaké parametre ako v experimente č.2. Počas trénovania sme však zaznamenali známky pretrénovania (overfitting), preto sme počet epoch znížili na 7.

[1] https://huggingface.co/datasets/TUKE-KEMT/toxic-sk
[2] https://github.com/stopwords-iso/stopwords-sk

E. Vyhodnotenie

Po dokončení trénovania modelov sme prešli na fázu vyhodnotenia, aby sme zistili, ako dobre model klasifikuje toxické komentáre. Na tento účel sme použili testovaciu množinu, ktorá nebola zahrnutá v trénovacom procese. Vyhodnotenie modelu je dôležitým krokom, pretože poskytuje informácie o jeho schopnosti generalizovať na neznáme dáta. Porovnávanie výkonu modelov sa realizovalo pomocou metrík ako úspešnosť a miera falošnej negativity (FNR). Výsledky sú zhrnuté v Tabuľke 1, kde vidíme, že model SlovakBERT v oboch prípadoch (experimenty 2 a 4) dosiahol vyššiu úspešnosť a nižšiu mieru falošnej negativity ako model BERT. To potvrdzuje, že model predtrénovaný špecificky na slovenskom jazyku lepšie rozumie nuansám a kontextu slovenských textov, čo vedie aj k presnejšej klasifikácii toxicity.

Tabuľka 1
Výsledky experimentov

Experiment	Dataset	Model	Úspešnosť	FNR
1	Toxic comments in Slovak social networks	BERT-base-uncased	82,85 %	17,15 %
2		SlovakBERT	84,79 %	15,21 %
3	Toxic-sk	BERT-base-uncased	87,90 %	15,30 %
4		SlovakBERT	92,84 %	7,96 %

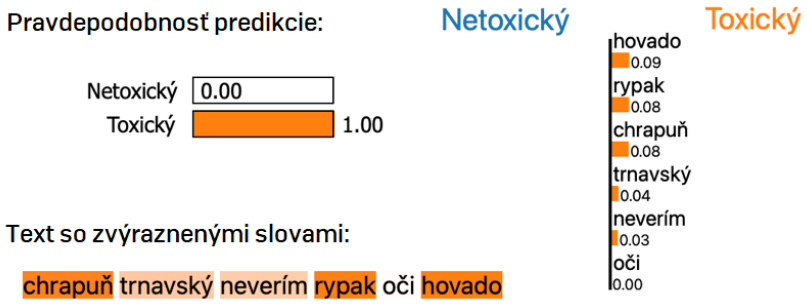
F. Využitie metód vysvetliteľnosti

Na interpretáciu výstupov boli implementované metódy LIME a SHAP, ktoré poskytli vizualizácie vplyvu jednotlivých slov na rozhodnutia modelu. Tieto nástroje umožnili detailnú analýzu klasifikovaných viet a identifikáciu kľúčových výrazov, ktoré model považoval za toxické. Vizualizácie zároveň pomohli pri odhaľovaní možných nedostatkov v predikcii a zvýšili transparentnosť rozhodovacieho procesu.

Na základe vykonaných experimentov a ich porovnania, sme sa rozhodli implementovať metódy LIME a SHAP práve pre experiment č. 4, aby sme mohli podrobne analyzovať, ktoré slová v komentári mali najväčší vplyv na finálnu predikciu modelu SlovakBERT.

1) Metóda LIME

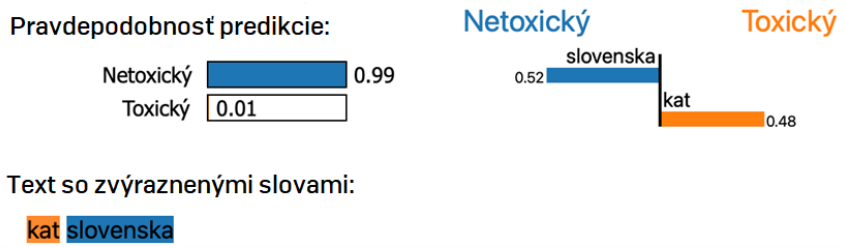
Na ukážku je zobrazené vysvetlenie komentára „chrapuň trnavský neverím rypak oči hovado“ pomocou metódy LIME. Model tento komentár klasifikoval ako toxický s pravdepodobnosťou 100%. LIME vizualizácia (Obr. 1) ukazuje, že najväčší vplyv na klasifikáciu mali výrazy „rypak“ (0,05), „chrapuň“ (0,05), „trnavský“ (0,05) a „hovado“ (0,05), ktoré sú výrazne expresívne, hanlivé a prispeli k zvýšeniu hodnoty pre triedu Toxický. Hoci „trnavský“ samo o sebe nie je hanlivé, v kombinácii s inými slovami môže nadobudnúť urážlivý alebo nepriateľský tón, ktorý model interpretuje ako celok. Slovo „oči“ nemalo vplyv (0,00), čo zodpovedá jeho neutrálnemu významu. Slovo „neverím“ LIME tiež ohodnotil ako 0,00, čo znamená, že v tomto konkrétnom prípade nemalo výrazný vplyv na rozhodnutie modelu. Pravdepodobne preto, že iné výrazy boli pre model oveľa silnejšie signály toxicity. Vysvetlenie potvrdzuje, že rozhodnutie modelu je ovplyvnené najmä silne expresívnymi alebo urážlivými slovami.



Obr. 1 Vysvetlenie pomocou metódy LIME pre toxický komentár

Model nesprávne klasifikoval komentár „kat slovenska“ ako netoxický s vysokou pravdepodobnosťou 99 % (Obr. 2). Vizualizácia LIME ukazuje, že výraz „slovenska“ mal vyšší pozitívny prínos (0,52) smerom k triede netoxický, čo výrazne ovplyvnilo výsledné rozhodnutie. Hoci slovo „kat“ má silne negatívny a potenciálne toxický význam (prínos 0,48 k triede toxický), jeho efekt nebol dostatočne silný na to, aby prevážil vplyv slova „slovenska“. Model tak pravdepodobne nevyhodnotil spojenie týchto dvoch slov ako kontextovo urážlivé alebo hanlivé.

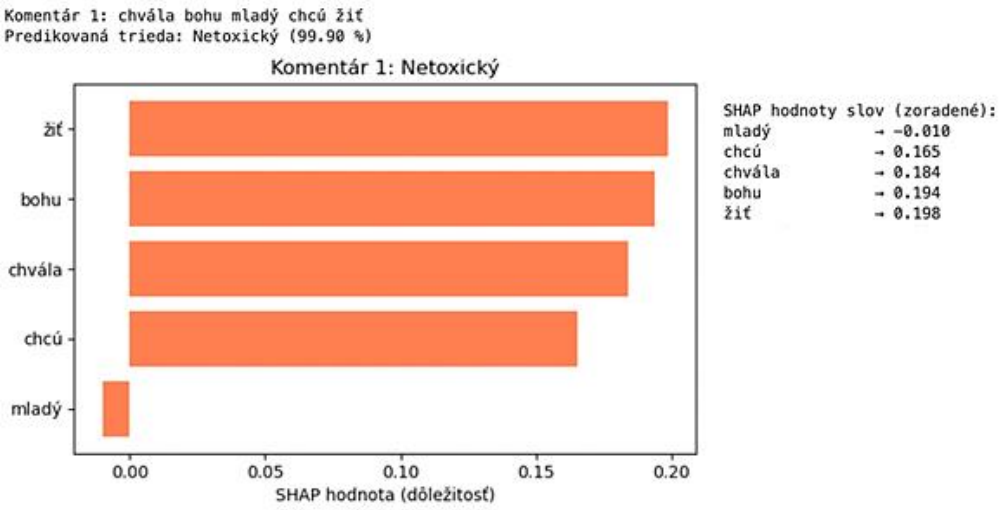
Tento príklad ukazuje slabinu klasifikátora – nedostatočnú citlivosť na implicitnú agresivitu alebo nevhodné spojenia v jazyku. Výsledok tak poukazuje na dôležitosť kontextu, ktorý model pri tejto predikcii nezachytil správne.



Obr. 2 Vysvetlenie pomocou metódy LIME pre nesprávne klasifikovaný komentár

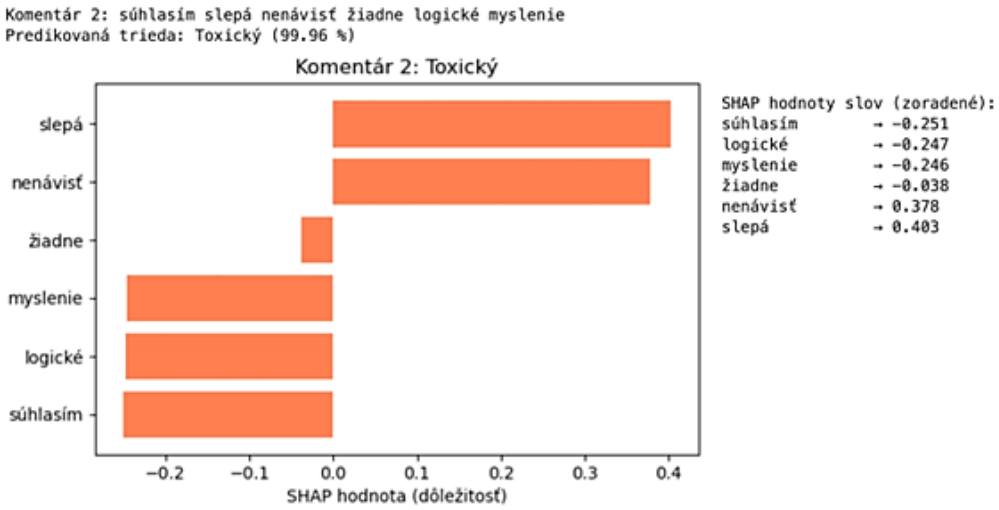
2) Metóda SHAP

Komentár „chvála bohu mladý chcú žiť“ bol modelom klasifikovaný ako netoxický s pravdepodobnosťou 99,90 %. SHAP vizualizácia (Obr. 3) ukazuje, že najvýraznejší pozitívny vplyv na rozhodnutie mali výrazy „žiť“ (0,198), „bohu“ (0,194) a „chvála“ (0,184). Tieto slová sú pozitívne ladené a nesú silný emocionálny význam, ktorý podporuje netoxickú interpretáciu. Slovo „chcú“ (0,165) tiež mierne podporilo túto klasifikáciu. Naopak, výraz „mladý“ (–0,010) mal len minimálny záporný príspevok, ktorý však výsledné rozhodnutie neovplyvnil. Výsledok naznačuje, že model spoľahlivo identifikuje pozitívne ladené komentáre a správne klasifikuje výrazy s podporným, hodnotiacim významom ako netoxické.



Obr. 3 Vysvetlenie pomocou metódy SHAP pre netoxický komentár

Ďalej komentár „súhlasím slepá nenávisť žiadne logické myslenie“ klasifikoval model ako toxický s pravdepodobnosťou 99,96 %. SHAP vizualizácia (Obr.4) ukazuje, že rozhodnutie bolo najviac ovplyvnené výrazmi „slepá“ (0,403) a „nenávisť“ (0,378), ktoré prispeli k vysokému skóre toxickosti. Naopak, výrazy „súhlasím“ (–0,251), „logické“ (–0,247) a „myslenie“ (–0,246) mali negatívne SHAP hodnoty a pôsobili smerom k netoxickej triede. Slovo „žiadne“ (–0,038) malo len slabý vplyv. Výsledok ukazuje, že model síce zachytil toxické prvky, ale nedokázal dostatočne zohľadniť kontext pozitívne ladených slov, a preto komentár nesprávne označil ako toxický, aj keď jeho celkový význam mohol byť obhajobný alebo kritický bez útočného zámeru.



Obr. 4 Vysvetlenie pomocou metódy SHAP pre nesprávne klasifikovaný komentár

ZÁVER

Táto práca preukázala efektivitu kombinácie hlbokého učenia a vysvetliteľnej umelej inteligencie pri detekcii toxických komentárov v slovenčine. Výsledky experimentov ukázali, že model SlovakBERT má výraznú výhodu pri práci s textami v slovenčine. Pri binárnej klasifikácii dosiahol úspešnosť 92,84 %, pričom bol zároveň schopný znížiť mieru falošne negatívnych klasifikácií, čo je kľúčové najmä v kontexte moderovania nenávistného obsahu.

Použitie metód LIME a SHAP poskytlo vizuálne prehľadné výstupy, ktoré umožnili sledovať význam jednotlivých slov a ich prínos k finálnemu rozhodnutiu modelu. Analýzy ukázali, že model spoľahlivo reagoval na silne toxické výrazy, avšak niektoré jemne ironické či sarkastické komentáre boli klasifikované s nižšou presnosťou, čo poukazuje na potrebu ďalšieho zdokonaľovania modelu.

Je dôležité spomenúť, že klasifikácia toxicity môže byť veľmi subjektívna, čo značne obmedzuje riešenie podobných úloh. Zároveň sa počas riešenia objavili výzvy, ktoré poukazujú na potrebu ďalšieho výskumu. Ide najmä o kvalitu slovenských datasetov, problematiku spoľahlivej anotácie a jazykovej variability, ako aj potrebu vytvárať modely, ktoré dokážu pracovať s kontextom a iróniou. Do budúcnosti by bolo vhodné rozšíriť experimenty o väčšie množstvo dát a integrovať systém do reálneho prostredia s cieľom testovania jeho robustnosti v praxi.

POĎAKOVANIE

Táto práca vznikla vďaka podpore Agentúry na podporu výskumu a vývoja na základe Zmluvy č. APVV-22-0414.

ZOZNAM POUŽITEJ LITERATÚRY

[1] K. Machová, M. Mach, and K. Adamišín, “Machine Learning and Lexicon Approach to Texts Processing in the Detection of Degrees of Toxicity in Online Discussions,” Sensors 2022, Vol. 22, Page 6468, vol. 22, no. 17, p. 6468, Aug. 2022, doi: 10.3390/S22176468.

[2] M. Sarnovský, V. Maslej-Krešňáková, and K. Ivancová, “Fake News Detection Related to the COVID-19 in Slovak Language Using Deep Learning Methods,” Acta Polytechnica Hungarica, vol. 19, no. 2, pp. 2022–2065.

[3] V. Maslej-Krešňáková, M. Sarnovský, P. Butka, and K. Machová, “Comparison of Deep Learning Models and Various Text Pre-Processing Techniques for the Toxic Comments Classification,” Applied Sciences 2020, Vol. 10, Page 8631, vol. 10, no. 23, p. 8631, Dec. 2020, doi: 10.3390/APP10238631.

[4] M. Pikuliak et al., “SlovakBERT: Slovak Masked Language Model,” Findings of the Association for Computational Linguistics: EMNLP 2022, pp. 7185–7197, Sep. 2021, doi: 10.18653/v1/2022.findings-emnlp.530.

[5] 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.

[6] C. Adams, J. Sorensen, J. Elliott, L. Dixon, M. McDonald, Nithum, W. Cukierski, “Toxic Comment Classification Challenge,” Kaggle [online]. 2017. [cit. 2024.04.16].

[7] 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, Dec. 2020, doi: 10.1609/aaai.v35i17.17745.

[8] R. Wirth and J. Hipp, “CRISP-DM: Towards a Standard Process Model for Data Mining”.

Vysvetliteľné modely umelej inteligencie pre multimodálnu klasifikáciu toxického obsahu

[1]*Miroslava MATEJOVÁ,* [2]*Ján PARALIČ,* [3]*Karin MAJIRSKÁ*

[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]karin.majirska@student.tuke.sk

***Abstrakt* — Táto štúdia sa zaoberá detekciou toxického obsahu v multimodálnych dátach (obraz a text) s využitím vysvetliteľných modelov umelej inteligencie. Využíva model VisualBERT, ktorý kombinuje textové a vizuálne vstupy z neurónovej siete ResNet-50. Experimenty boli vykonávané na datasete Hateful Memes. Pri analýze výsledkov modelu sme sa zamerali na vizualizáciu pozornosti medzi textovými tokenmi a vizuálnymi patchmi. Cieľom je overiť, do akej miery dokáže VisualBERT správne identifikovať toxicitu obsahu a ako vysvetliteľnosť modelu prostredníctvom vizualizácie jeho rozhodovacieho procesu môže prispieť k lepšiemu pochopeniu chýb modelu.**

***Kľúčové slová* — multimodálne dáta, toxické správanie, umelá inteligencia, vysvetliteľnosť modelov**

I. Úvod

Sociálne siete sú interaktívne technológie, ktoré sa v poslednej dobe stali nevyhnutnou súčasťou našich životov. Približne 45 % svetovej populácie používa sociálne média [1]. Tie sa pre mnohých stali nielen bežnou súčasťou dňa, ale aj formou digitálnej závislosti, čo so sebou prináša viaceré psychologické, spoločenské aj etické otázky. S nárastom objemu používateľov a zdieľaného obsahu zároveň rastie aj riziko šírenia dezinformácií, nenávistných prejavov, toxického správania či iných foriem škodlivého obsahu. Sociálne siete tieto komentáre a diskusie pod príspevkami väčšinou nijako neupravujú, neregulujú ani nemoderujú. Aj keď majú pomerne prísne užívateľské pravidlá a ponúkajú viacero nástrojov na zabezpečenie, kontrolu viditeľnosti či nastavenie súkromia, zasahujú spravidla až po nahlásení porušenia pravidiel, často až po niekoľkonásobne opakovaných podnetoch od viacerých používateľov.

Multimodálne modely umožňujú simultánne spracovanie textových a obrazových vstupov, čím vytvárajú nové možnosti analýzy komplexných dát. Ich využitie je stále viac rozšírené, napríklad v oblasti sociálnych sietí, kde sú schopné analyzovať multimodálny obsah, vrátane meme obrázkov. Model VisualBERT umožňuje spracovanie takýchto dát, no jeho interpretovateľnosť je obmedzená, čo môže viesť k nedôvere. Vysvetliteľnosť modelov umelej inteligencie (XAI) patrí medzi kľúčové témy v oblasti strojového učenia. Výsledky poukazujú na význam vysvetliteľnosti a jej prínos a praktické využitie pri interpretácii rozhodovacieho procesu modelu. V súčasnosti absentujú nástroje, ktoré by dokázali dostatočne vysvetliť rozhodovacie procesy multimodálnych modelov a poskytnúť hlbší pohľad do ich fungovania.

II. Analýza Súčasného Stavu

Detekcia toxicity v multimodálnych dátach je veľmi náročná, pretože toxicita sa môže prejaviť v rôznych formách a kontextoch. Mémy často obsahujú odkazy na internetové fenomény, históriu či populárnu kultúru, ktorým porozumie len recipient s určitými predchádzajúcimi znalosťami. Tieto mémy takisto veľmi často obsahujú iróniu, sarkazmus alebo nejaký skrytý kontext, ktorý ťažko rozpoznať.

Súčasné prístupy k automatickej detekcii toxicity sa delia na metódy strojového učenia a metódy hlbokého učenia. Tradičné metódy strojového učenia využívajú algoritmy ako lineárna regresia [2], [3], metóda podporných vektorov [2], [3] Naive Bayes [2], [4], rozhodovacie stromy [4] alebo náhodné lesy [3], sa zameriavajú hlavne na textové dáta a dosahujú primeranú presnosť,

najmä pri správnom predspracovaní a reprezentácii dát. Tieto prístupy však často nedokážu zachytiť kompletný kontext multimodálneho obsahu. Hlboké učenie sa spolieha na neurónové siete ako napríklad CNN, RNN, Transformery [3], ktoré dokážu automaticky extrahovať významové reprezentácie z dát.

Porovnania medzi klasickými metódami strojového učenia a hlbokým učením ukazujú, že hoci pokročilé modely ako BERTweet dosahujú vyššiu presnosť, klasické algoritmy si udržiavajú pomerne vysokú výkonnosť a sú efektívne aj pri nižších výpočtových nárokoch. V niektorých prípadoch môže rozdiel v presnosti medzi týmito prístupmi byť menší ako 5 %, čo naznačuje, že tradičné metódy môžu byť prakticky využiteľné v reálnych aplikáciách [3]. Zároveň sa ukazuje, že na detekciu sarkazmu a kontextovo závislej toxicity sú modely ako BERT výrazne efektívnejšie, najmä v prípadoch vyžadujúcich pochopenie kontextu.

Multimodálne modely kombinujú text, obraz, zvuk a video s cieľom zlepšiť detekciu nenávistného obsahu a kyberšikany na sociálnych sieťach. Výskumy [5] ukazujú, že využitím viacerých modalít sa zvyšuje presnosť detekcie, pričom modely lepšie identifikujú aj zložité formy nenávisti a implicitný kontext. Multimodálny model VisualBERT, ktorý bol využitý aj v tejto štúdii, kombinuje textové vstupy s vizuálnymi embeddingami extrahovanými z ResNet-50 a vytvára tak kontextuálne reprezentácie, ktoré umožňujú prepojenie významov z textu a obrazu.

Význam vysvetliteľnosti (XAI) v tejto oblasti neustále rastie, pretože rozhodnutia modelov strojového učenia a hlbokého učenia sú často netransparentné. Aj keď dosahujú vysokú presnosť, často nie je jasné, na základe čoho sa rozhodujú. Preto sú potrebné prístupy ako vizualizácia heatmap pozornosti medzi textom a obrazom alebo analýza pozornosti v jednotlivých vrstvách modelu. Tieto techniky umožňujú identifikovať, ktoré prvky majú najväčší vplyv na rozhodnutie modelu a odhaľujú tiež prípadné nedostatky. Autori v štúdii [6] analyzujú využitie XAI pri detekcii nenávistných prejavov a falošných správ na sociálnych sieťach. Autori tiež zdôrazňujú význam XAI, ktorá umožňuje interpretáciu a vysvetlenie rozhodnutí modelov strojového učenia.

Dôležitosť témy detekcie toxického obsahu zdôraznila aj súťaž Hateful Memes Challenge[1], ktorá bola organizovaná spoločnosťou Facebook AI (teraz Meta AI) v roku 2020 a bola zameraná na detekciu nenávistného obsahu v multimodálnych dátach – konkrétne na meme obrázky. Cieľom súťaže bolo navrhnúť modely schopné analyzovať a vyhodnotiť kombináciu textovej a obrazovej modality. Táto výzva reagovala na čoraz častejšie šírenie dezinformácií, falošných správ a nenávistného obsahu pomocou meme obrázkov, ktoré nie je jednoduché odhaliť tradičnými technikami strojového učenia zameranými iba na jednu modalitu. Riešenia, ktoré sa umiestnili na popredných priečkach súťaže využívali najmä architektúry ako VL-BERT, UNITER, LXMERT a VisualBERT [7]. Tieto modely efektívne kombinujú textové a obrazové vstupy a využívajú mechanizmy ako cross-attention, učenie súborom metód a Image-Text Matching (ITM), ktoré zvyšujú presnosť predikcií. Navrhnuté riešenie dosiahlo AUROC skóre [8] 84,5 % a úspešnosť 73,2 %. Hoci sa väčšina riešení sústredila na presnosť a robustnosť modelov, aspektu vysvetliteľnosti sa vo svojich štúdiách primárne nevenovali.

V súčasnej dobe je však vysvetliteľnosť modelov čoraz dôležitejšia, najmä v oblastiach s priamym spoločenským dopadom, ako je moderovanie obsahu alebo detekcia toxického správania. Transparentné modely, ktoré dokážu jasne vysvetliť svoje rozhodnutia, nielen zvyšujú dôveru používateľov, ale tiež pomáhajú odhaliť prípadné nedostatky. Aj keď je vysoká presnosť dôležitá, bez pochopenia toho, ako modely dospeli k svojim záverom, môžu byť ich rozhodnutia pre používateľov nejasné alebo neakceptovateľné – najmä ak sa týkajú citlivých tém.

III. Experimentálna Časť

V praktickej časti práce sa zameriavame na riešenie úlohy detekcie toxických mémov pomocou modelu VisualBERT, ktorý kombinuje textovú a obrazovú modalitu. Prínosom tejto štúdie je následná analýza výstupov modelu pre lepšie pochopenie klasifikácie a chýb modelu.

A. Pochopenie cieľa

Cieľom je klasifikovať mémy ako toxické alebo netoxické a následne analyzovať, ako model funguje. Analýza nám umožní identifikovať faktory ovplyvňujúce jeho rozhodovanie a sprístupniť tieto procesy prostredníctvom vizualizácie mechanizmov pozornosti a techniky Image-Text Matching. Vizualizácia pozornosti umožní sledovať, na ktoré časti textu a obrazu sa model sústreďuje, zatiaľ čo ITM zhodnotí súlad medzi textom a obrazom a odhalí prípady, kde model zlyháva pri interpretácii multimodálneho kontextu.

B. Pochopenie dát

Využili sme vyššie spomínaný dataset Hateful Memes [7], ktorý obsahuje kombináciu

[1] https://ai.meta.com/tools/hatefulmemes/

textových a obrazových dát s popisom, či ide o toxické alebo netoxické mémy. Dataset obsahuje približne 10 000 mémov rozdelených na trénovaciu, validačnú a testovaciu množinu. Každý záznam obsahuje unikátne ID, obrázok, textový popis a štítok či ide o toxický obrázok alebo nie. Súčasťou datasetu sú aj mémy s benígnymi zavádzajúcimi prvkami. Tieto zavádzajúce prvky boli zámerne zahrnuté s cieľom zamedziť tomu, aby modely pri klasifikácii vychádzali len z jednej modality. Príkladom môže byť text, ktorý sám o sebe pôsobí neškodne, no v kombinácii s konkrétnym obrázkom nadobúda toxický význam. V datasete sa nachádzajú aj identické obrázky kombinované s rôznymi textami – zatiaľ čo v jednej kombinácii je výsledný obsah vyhodnotený ako toxický, v inej zase naopak ako netoxický. Táto variabilita zdôrazňuje význam vizuálnej modality a potrebu multimodálneho spracovania, keďže význam obsahu často vyplýva zo súhry textu a vizuálu. Interpretácia multimodálnych dát v tomto datasete je náročná aj z dôvodu prítomnosti rôznych jazykových a vizuálnych prvkov, ktoré vyžadujú širší kontext. Mnohé mémy obsahujú narážky, ktoré predpokladajú predchádzajúce znalosti u pozorovateľa, ako napríklad známe celebrity, historické udalosti alebo kultúrne symboly. Irónia, sarkazmus a kultúrne referencie (napríklad stereotypy spojené s určitými etnickými skupinami, náboženstvami alebo spoločenskými javmi) sú ďalšie prvky, ktoré môžu zmeniť význam textu. Tieto faktory výrazne komplikujú úlohu automatickej detekcie toxicity, keďže model musí byť schopný chápať význam naprieč modalitami aj kontextmi, ktoré nie sú explicitne vyjadrené.

C. Príprava dát

Príprava dát je nevyhnutným krokom pred samotným trénovaním modelov umelej inteligencie, pretože správne pripravené dáta zlepšujú presnosť a robustnosť modelov. Analýza datasetu ukazuje, že dáta sú vo všeobecnosti kompletné a neobsahujú chýbajúce hodnoty.

Textová časť bola tokenizovaná pomocou predtrénovaného BERT tokenizeru, ktorý prevedie vetu na sekvenciu tokenov a následne ich zakóduje do číselnej reprezentácie. Tokeny boli doplnené na maximálnu dĺžku 128 tokenov, čo je štandardná dĺžka pre úlohy klasifikácie.

Obrázky boli spracované pomocou predtrénovaného modelu ResNet-50 z knižnice torchvision.models. Posledná vrstva siete ResNet bola odstránená a model slúžil ako extraktor vizuálnych znakov. Každý vstupný obrázok bol transformovaný na formát 1×2048, čo zodpovedá výstupnej veľkosti ResNet-50 bez klasifikačnej vrstvy. Tieto embeddingy boli následne vložené ako vizuálny vstup do modelu VisualBERT.

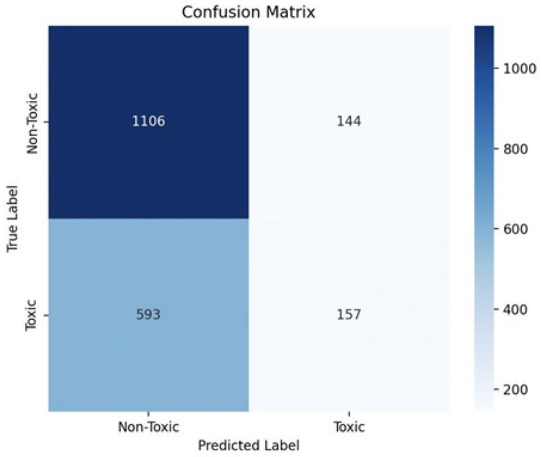
D. Modelovanie

Vzhľadom na povahu datasetu Hateful Memes, ktorý obsahuje kombináciu textu a obrázkov, bol zvolený multimodálny model, ktorý dokáže pracovať s oboma typmi vstupov naraz. Ako základ bola zvolená architektúra VisualBERT, ktorá kombinuje výhody jazykového modelu BERT s možnosťou spracovania vizuálnych znakov z obrázkov. Model bol implementovaný v prostredí PyTorch s využitím knižnice Transformers, pričom architektúra obsahovala predtrénovaný VisualBERT a klasifikačnú hlavičku s lineárnou vrstvou pre binárnu klasifikáciu.

Tréning prebiehal počas 10 epoch s využitím stratovej funkcie CrossEntropyLoss a optimalizátora AdamW. Pre zlepšenie tréningu boli nasadené techniky ako learning rate scheduler a early stopping. Výsledky boli priebežne vyhodnocované pomocou metrík úspešnosť, presnosť, návratnosť a F1-skóre.

E. Vyhodnotenie

Vo fáze vyhodnotenia bol model VisualBERT otestovaný na dátach, ktoré predtým nevidel, pričom preukázal schopnosť identifikovať toxické príspevky, no s obmedzenou citlivosťou, čo naznačuje potrebu ďalšieho zlepšenia. Matica zámen (Obr. 1) ukázala, že správne identifikoval 1 106 netoxických a 157 toxických príspevkov, ale chybne označil 144 netoxických ako toxické a 593 toxických ako netoxické. Citlivosť na toxicitu bola zvýšená znížením rozhodovacieho prahu na 0,35. Model dosiahol úspešnosť 63,15 %, presnosť 52,16 %, návratnosť 20,93 % a F1-skóre 29,88 %.



Obr. 1 Matica zámen pre výsledný model

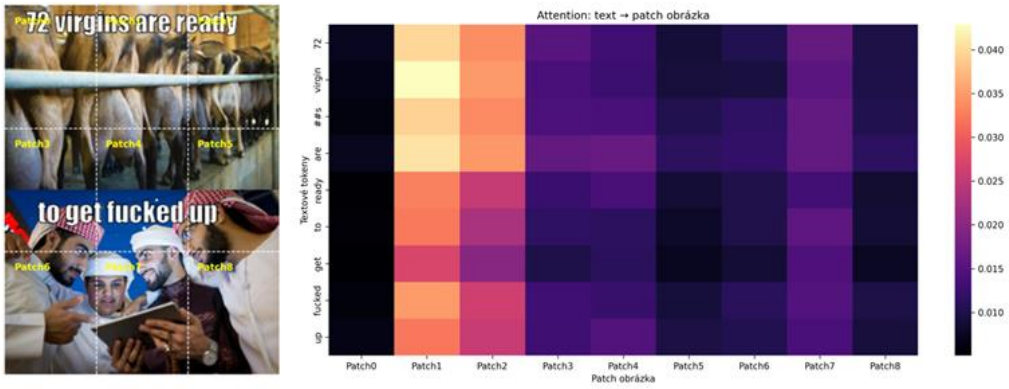
IV. Analýza rozhodnutí modelu

Cieľom je pochopiť, ako model pracuje s kombinovanými textovo-obrazovými vstupmi, ktoré časti vstupu považuje za relevantné pri klasifikácii, a aké vzťahy medzi jednotlivými modalitami si vytvára počas spracovania. Na tento účel sme využili dva prístupy vizualizácie pozornosti. Prvý využíva patch-based reprezentáciu a druhý porovnávanie pozornosti medzi vrstvami. **Upozornenie**: Nasledujúce ukážky obsahujú jazyk/obrázky, ktorý môže byť považovaný za urážlivý alebo nevhodný. Tento materiál je zahrnutý výhradne na výskumné účely v oblasti detekcie toxického obsahu a pochádza z verejne dostupného datasetu.

A. Patch-based reprezentácia

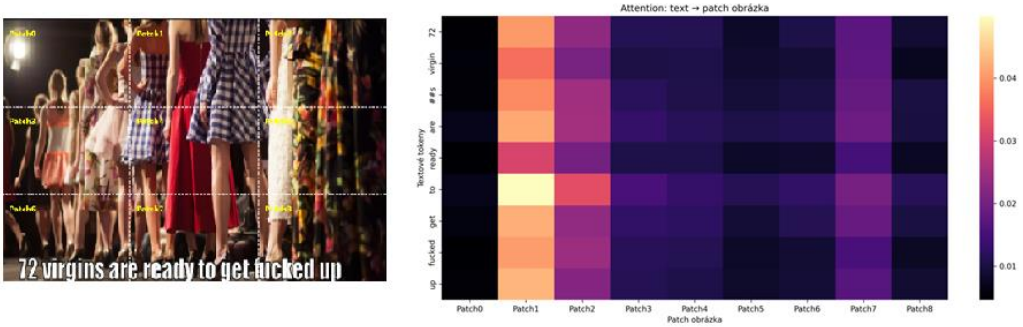
V tomto prístupe bol každý obrázok z validačnej množiny rozdelený na mriežku 3x3 patchov. Každý patch bol samostatne zakódovaný pomocou vizuálneho embeddingu cez ResNet-50, čím sa získala vektorová reprezentácia pre jednotlivé oblasti obrázka. Následne bola z VisualBERT modelu extrahovaná pozornosť z textových tokenov na jednotlivé obrazové patche. Výsledkom je heatmapa, ktorá znázorňuje, na ktoré časti obrázka sa textová informácia najviac zameriava. Tento vizualizačný prístup umožňuje lepšie pochopiť rozhodovanie modelu – napríklad zistiť, či sa zameriava na relevantné časti obrázka, alebo či pozornosť smeruje na nepodstatné časti. V niektorých prípadoch možno pozorovať, že mechanizmus pozornosti modelu VisualBERT sústreďuje takmer všetku svoju pozornosť do jedného dominantného patchu. Tento jav nastáva najmä vtedy, keď je obrázok vizuálne jednoduchý alebo text nie je explicitne prepojený s konkrétnym objektom. Naopak, v prípadoch s viacnásobnými objektmi (napr. osoby, miestnosti, zbrane) sa pozornosť rozprestiera medzi viaceré patche. Takéto „rozptýlené“ heatmapy svedčia o tom, že model využíva viacero vizuálnych vstupov na doplnenie významu textu.

V tejto časti na ukážku analyzujeme dvojicu rôznych obrázkov s rovnakým textom. Prvý obrázok (Obr. 2) je rozdelený na dve polovice – horná časť zobrazuje zvieratá zoradené vedľa seba a dolná časť ukazuje skupinu arabských mužov, ktorý sa zabávajú pri pohľade do tabletu. Vo vrchnej časti obidvoch polovíc obrázka sa nachádza rozdelený text: „72 virgins are ready to get fucked up.“ Význam meme spočíva v narážke na stereotyp o „72 pannách“ a jeho spojení s arabskou kultúrou, čo podporuje šírenie nenávisti a degraduje moslimov. Toxicita obrázka je vysoká, pretože kombinácia vizuálov a textu vytvára obraz zosmiešňujúci konkrétnu skupinu ľudí. Model tento model správne vyhodnotil ako toxický.



Obr. 2 Ukážka 1 obrázka a heatmapy pre patch-based reprezentáciu

Pri druhom obrázku (Obr. 3) vidíme veľmi podobnú heatmapu ako pri tom prvom a je na nej viditeľné, že najviac pozornosti je sústredených na tokeny „72“, „virgins“, „ready“ a „to“, ktoré sú najvýraznejšie na patchoch P1, P2 a P3, teda na časti obrázka, kde sa nachádza zadná časť tiel modeliek. Tieto oblasti model vyhodnocuje ako najrelevantnejšie v kontexte textu. Naopak, patch P7 a P8 (dolná časť obrázka) dostávajú len minimálnu pozornosť. Táto distribúcia poukazuje na fakt, že model sa zameriava najmä na vizuálnu oblasť tela modeliek a text s toxickým významom, čo potvrdzuje, že meme obsahuje sexistický a xenofóbny podtext.



Obr. 3 Ukážka 2 obrázka a heatmapy pre patch-based reprezentáciu

Táto analýza naznačuje, že model primárne kombinuje textové prvky a vizuálne časti, ktoré priamo súvisia s významom textu. V tomto prípade však nedokáže zachytiť hlbší, urážlivý alebo kultúrne problematický význam meme – a preto sa môže rozhodnúť pre nesprávnu klasifikáciu toxicity, ak sa sústredí len na povrchné znaky v obrázku a texte. Tento príklad veľmi dobre ukazuje dôležitosť vysvetliteľnosti modelu, pretože model môže pri samostatnom hodnotení vizuálu a textu vyhodnotiť obsah ako netoxický, avšak kombinácia modalít a interpretácia pomocou attention mechanizmov jednoznačne odhalí diskriminačný podtón.

B. Porovnanie pozornosti medzi vrstvami

V druhom prístupe sa pozornosť analyzovala v rôznych vrstvách modelu. VisualBERT má viacero transformerových vrstiev, pričom každá vrstva môže zachytávať rôzne druhy vzťahov (napr. syntaktické, sémantické, multimodálne). Pre vybrané vstupy boli extrahované pozornostné matice z jednotlivých vrstiev a vizualizované ako heatmapy, ktoré ukazovali, ako sa mení pozornosť medzi slovami v texte naprieč hĺbkou modelu. Tento prístup pomohol odhaliť, v ktorej vrstve sa model najviac „sústredí“ na určité slová alebo vzťahy v texte.

V nižších vrstvách (1-4) bola pozornosť často rozptýlená – viaceré slová sa spájali s rôznymi oblasťami obrázka bez jasného zamerania. V stredných vrstvách (5–8) sa začala pozornosť koncentrovať na konkrétne kombinácie slov a vizuálnych prvkov, čo naznačuje vznik sémantických väzieb. Naopak, vo vyšších vrstvách (9-12) bola pozornosť výrazne selektívnejšia: určité slová smerovali ku konkrétnym vizuálnym prvkom, čo môže naznačovať silnejšiu multimodálnu integráciu alebo rozhodovací mechanizmus modelu.

V. ZÁVER

Táto štúdia analyzovala a interpretovala multimodálne modely na detekciu toxického správania, pričom sa sústredila na vizualizáciu pozornosti v modeli VisualBERT natrénovaného

na dátach Hateful Memes. Výsledky ukázali, že model dokáže zachytiť kľúčové prvky multimodálnych vstupov, no jeho citlivosť na jemné a implicitné formy toxicity (napr. rasizmus, sexizmus) je obmedzená. Kvantitatívna analýza preukázala úspešnosť modelu 63,15 %, pričom vizualizácie odhalili konkrétne slabé miesta, kde model zlyhal v interpretácii skrytého významu, podtónu alebo kontextu.

Práca zdôrazňuje potrebu rozšírenia vysvetliteľnosti modelov pomocou pokročilých metód a rozšírenia analýzy na väčšie množstvo dát a modelov. Je potrebné sa v budúcnosti zamerať na vývoj architektúr schopných lepšie chápať kontext a implicitné vzťahy v multimodálnych vstupoch, čím by sa zvýšila spoľahlivosť a zodpovednosť systémov na detekciu toxického obsahu.

Poďakovanie

Táto práca vznikla vďaka podpore Agentúry na podporu výskumu a vývoja na základe Zmluvy č. APVV-22-0414.

Zoznam použitej literatúry

[1] I. Hameed and B. Z. Irfan, “Social Media Self-Control Failure leading to antisocial aggressive behavior,” Hum Behav Emerg Technol, vol. 3, no. 2, pp. 296–303, Apr. 2021, doi: 10.1002/HBE2.226.

[2] K. K. Kiilu, G. Okeyo, R. Rimiru, and K. Ogada, “Using Naïve Bayes Algorithm in detection of Hate Tweets,” International Journal of Scientific and Research Publications, vol. 8, no. 3, 2018, doi: 10.29322/IJSRP.8.3.2018.p7517.

[3] A. Bonetti, M. Martínez-Sober, J. C. Torres, J. M. Vega, S. Pellerin, a J. Vila-Francés, “Comparison between Machine Learning and Deep Learning Approaches for the Detection of Toxic Comments on Social Networks”, 01. máj 2023, MDPI. doi: 10.3390/app13106038.

[4] E. Omran, E. Al Tararwah, a J. Al Qundus, “A comparative analysis of machine learning algorithms for hate speech detection in social media”, Online J Commun Media Technol, roč. 13, č. 4, s. e202348, okt. 2023, doi: 10.30935/OJCMT/13603.

[5] D. Soni a V. Singh, “See No Evil, Hear No Evil”, Proc ACM Hum Comput Interact, roč. 2, č. CSCW, nov. 2018, doi: 10.1145/3274433.

[6] V. U. Gongane, M. V. Munot, a A. D. Anuse, “A survey of explainable AI techniques for detection of fake news and hate speech on social media platforms”, J Comput Soc Sci, roč. 7, č. 1, s. 587–623, apr. 2024, doi: 10.1007/S42001-024-00248-9.

[7] D. Kiela et al., “The Hateful Memes Challenge: Competition Report,” Proc Mach Learn Res, vol. 133, pp. 344–360, 2021, Accessed: Jun. 06, 2025. [Online]. Available: https://hatefulmemeschallenge.com

[8] M. Tu, “AUROC vs AP: A Deep Dive into Metrics and Their Surprising Differences” Accessed: Jul. 07, 2025. [Online]. Available: https://medium.com/@qrnnis2623891/auroc-vs-ap-a-deep-dive-into-metrics-and-their-surprising-differences-af86cf484a05

Vývoj interaktívneho robota KEMŤÁKa s jazykovým rozhraním na báze LLM

[1]Jozef KIJAC, [2]Zuzana SOKOLOVÁ, [3]Patrik JURÍK, [4]Eva KUPCOVÁ

[1,2,3,4]Katedra elektroniky a multimediálnych telekomunikácií,
Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Letná 9, 04001 Košice, Slovenská republika

[1]jozef.kijac@student.tuke.sk, [2]zuzana.sokolova@tuke.sk, [3]patrik.jurik@tuke.sk,
[4]eva.kupcova@tuke.sk

Abstrakt **– Predmetom tohto článku je návrh a implementácia interaktívneho robota s názvom KEMŤÁK. Projekt je postavený na výpočtovej platforme Raspberry Pi 5, ktorá slúži ako riadiaca jednotka, a zahŕňa všetky nevyhnutné komponenty potrebné na zabezpečenie multimodálnej komunikácie. Softvérová architektúra robota bola vyvinutá v programovacom jazyku Python a je navrhnutá tak, aby sa spúšťala automaticky pri štarte zariadenia. Robot KEMŤÁK disponuje funkciami ako detekcia prítomnosti osoby, rozpoznávanie hlasových príkazov, generovanie odpovedí prostredníctvom aplikačného programového rozhrania a ich následné prehrávanie spolu s vizuálnym výstupom na displeji. Okrem elektronických a softvérových prvkov bola súčasťou projektu aj tvorba fyzickej konštrukcie robota. Integrácia prvkov umelej inteligencie a robotiky predstavuje inovatívny prístup, ktorý má potenciál zvýšiť atraktivitu technického vzdelávania, najmä pre mladšiu generáciu.**

Kľúčové slová **– grafické používateľské rozhranie, komunikácia s API, Raspberry Pi 5, robotika, umelá inteligencia, veľký jazykový model, 3D tlač**

I. Úvod

V posledných rokoch sme svedkami narastajúcej konkurencie medzi technologickými spoločnosťami, ktoré sa predbiehajú vo vývoji najpokročilejších systémov umelej inteligencie. Nástroje ako ChatGPT od OpenAI, Gemini od Googlu či Grok od XAi sa čoraz viac stávajú bežnou súčasťou každodenného života mladých ľudí. Popri tomto vývoji však nezaháľa ani oblasť robotiky. Práve spojenie umelej inteligencie a robotiky otvára nové možnosti pre formovanie budúcnosti celej spoločnosti. Tento článok má za cieľ priblížiť mladým generáciám, ako môže vyzerať súhra týchto dvoch dynamických odborov. Predstavuje interaktívneho robota, ktorý bude pôsobiť ako maskot Katedry KEMT na Technickej univerzite v Košiciach. Jeho úlohou je zaujať budúcich študentov, ukázať im možnosti modernej techniky a motivovať ich k vlastnej tvorivej činnosti. Vďaka tomuto projektu môžu získať cenné poznatky aj praktické skúsenosti už v raných fázach svojho akademického života. Projekt zahŕňal kompletný návrh a realizáciu robota KEMŤÁKa – od výberu komponentov, cez 3D modelovanie jeho fyzickej konštrukcie, až po vývoj softvéru, ktorý zabezpečuje interakciu s používateľom. Výsledkom je funkčný prototyp, ktorý kombinuje kľúčové prvky z oblastí ako počítačové videnie, spracovanie zvuku, práca s aplikačným programovým rozhraním (z angl. Application Programming Interface, API), grafické rozhrania a dokonca aj mechanická výroba. Robot tak predstavuje praktickú ukážku toho, ako môže vyzerať robotika budúcnosti prepojená s inteligentnými technológiami.

II. Rozvoj robotiky a umelá inteligencia

V súčasnosti možno sledovať intenzívny rozmach robotiky, podmienený najmä rýchlym napredovaním technológií umelej inteligencie. Táto súhra je kľúčová pre modernizáciu priemyslu aj vzdelávania. Zatiaľ čo robotika sa sústreďuje na fyzickú interakciu, AI rieši abstraktné úlohy. Výsledkom sú roboty, ktoré optimalizujú výrobu alebo zjednodušujú každodenný život [1]. Táto práca sa venuje návrhu robota KEMŤÁKa – maskota katedry KEMT, ktorý využíva umelú inteligenciu na komunikáciu, popularizáciu technológií a interaktívne vzdelávanie. Podobne ako

ChatGPT, aj KEMŤÁK má potenciál zatraktívniť výučbu a motivovať mladú generáciu [2]. KEMŤÁK využíva platformu Raspberry Pi a jazykový model ChatGPT od OpenAI, čo umožňuje hlasovú interakciu, zobrazovanie informácií a moderné AI funkcie [3].

Zásadný pokrok v spracovaní prirodzeného jazyka nastal zavedením architektúry *Transformer* [4], ktorá využíva mechanizmus *self-attention* na pochopenie kontextu. OpenAI túto architektúru využila pri vývoji modelov GPT-1 až GPT-4, pričom GPT-3.5 a GPT-4 výrazne vylepšili konverzačné schopnosti a podporu rôznych jazykov [5], [6], [7]. GPT modely sú jednoducho integrované cez REST API, ktoré podporuje HTTP požiadavky a je kompatibilné s Raspberry Pi. Pre projekt KEMŤÁK sú vhodné modely GPT-4o alebo GPT-4o-mini, ktoré ponúkajú dobrý pomer výkonu a ceny (pozri Tab. 1) [8].

Model	Cena vstup (za 1M tokenov)	Cena výstup (za 1M tokenov)
OpenAI GPT-4o mini	$0.15	$0.60
OpenAI GPT-4o	$2.50	$10.00
OpenAI GPT-3.5-turbo	$0.50	$1.50
Anthropic Claude 3 Haiku	$0.25	$1.25
Google Gemini 1.5 Flash	$0.35	$1.05

Tab. 1 Porovnanie ChatGPT API modelov a konkurencie v roku 2025 [8]

Ross a kol. [9] vytvorili asistenta, ktorý okrem klasických úloh ako preklad kódu zvláda aj interaktívnu komunikáciu s používateľom. Štúdie sa zúčastnilo 42 programátorov s rôznymi skúsenosťami. Asistent dosiahol mieru prijatia (z angl. rate of acceptance) 66,30%, výrazne viac než GitHub Copilot (27,00%). Štúdia sa zameriavala na empirické ukazovatele ako produktivita a používateľský komfort, nie na technické metriky ako presnosť (z angl. precision), návratnosť (z angl. recall), F1-skóre (z angl. F1-score) a správnosť (z angl. accuracy).

V štúdii Jiang a kol. [10] predstavili nástroj Graphologue, ktorý prevádza odpovede z GPT-4 do grafických diagramov. Pri 50 generovaných témach obsahujúcich 4 360 slov bolo extrahovaných 1 043 entít, z toho 43 nesprávne, čo viedlo k presnosti 96,04%, návratnosti 95,60% a F1-skóre 95,82%. Po automatickej korekcii tieto metriky stúpli na 97,11%, 97,37% a 97,24%. Pri extrakcii relácií dosiahli po korekcii F1-skóre 92,39% (z pôvodných 84,77%). Benary a kol. [11] skúmali použitie LLM pri návrhu onkologickej liečby. ChatGPT-3, Galactica, Perplexity a BioMedLM hodnotili 10 fiktívnych pacientov. Najlepšie skóre malo Galactica (F1-skóre = 19,00%). Kombináciou modelov sa dosiahlo jednotné F1-skóre 29,00%. V dvoch prípadoch navrhli modely unikátne, odborníkmi uznané liečby. Lai a kol. [12] testovali RumorLLM v detekcii dezinformácií. Na BuzzFeed dosiahol správnosť 82,55%, presnosť 70,00%, návratnosť 58,33% a F1-skóre 63,64%. Na PolitiFact boli výsledky lepšie: F1-skóre 86,79% a správnosť 93,91%.

V štúdii Rony a kol. [13] porovnávali ChatGPT-4 a Gemini v detekcii zavádzajúcich titulkov (60 príkladov). ChatGPT-4 dosiahol F1-skóre 90,00% pri nezavádzajúcich a 85,00% pri zavádzajúcich titulkoch. Celková presnosť bola 88,00%. Gemini mal horšie výsledky: F1-skóre 73,00% (nezavádzajúce) a 57,00% (zavádzajúce), celková presnosť 67,00%. Štúdia zdôraznila potrebu etiky pri vývoji modelov. Hung a kol. [14] testovali nasadenie GPT-4 na vyhľadávanie klinických štúdií pre pacientov s nádormi hlavy a krku v aplikácii LookUpTrials. Model vyhodnotil 56 štúdií a vybral 0–4 pre každého pacienta. Dosiahol návratnosť 100%, presnosť 63% a F1-skóre 77%.

Autori Xu a kol. [15] skúmali personalizované odporúčania v e-commerce. Kým klasické algoritmy dosiahli F1-skóre 71%, LLM dosiahol 79%, čím zlepšil používateľský zážitok aj predaje. Babu a Boddu [16] navrhli medicínskeho chatbota založeného na BERT, ktorý prekonal BI-LSTM model vo všetkých metrikách. Nový chatbot dosiahol správnosť 98%, presnosť 97%, návratnosť 96% a F1-skóre 98%, pričom sa zameral na terminologickú presnosť a personalizovanú spätnú väzbu. Autori Salloum a kol. [17] vytvorili univerzitného chatbota, ktorý odpovedá na FAQ. Trénovaný bol na dátach z Kaggle a dosiahol výnimočné výsledky pomocou metódy Viacvrstvový perceptrón (z angl. Multi-Layer Perceptron, MLP): správnosť 99,75%, presnosť 99,76%, návratnosť 99,75% a F1-skóre na úrovni 99,75%.

V štúdii Chan a kol. [18] navrhli dvojstupňového chatbota pre detekciu samovražedných myšlienok. Prvý stupeň využíva BERT na detekciu rizika (presnosť 92,00%, správnosť 94,00%, F1-skóre = 92% a návratnosť 92,00%), druhý stupeň klasifikuje zámer pomocou zero-shot learningu. Korpusom bolo 318 797 príspevkov z Reddit fóra "SuicideWatch". Sato a Mitsukura [19] spojili analýzu výrazu tváre a pulzovej frekvencie pri detekcii klamstva. Pomocou algoritmu Náhodné lesy (z angl. Random Forest, RF) dosiahli priemerné výsledky: správnosť 87,00%, presnosť 84,00%, návratnosť 91,00% a F1-skóre 88,00%.

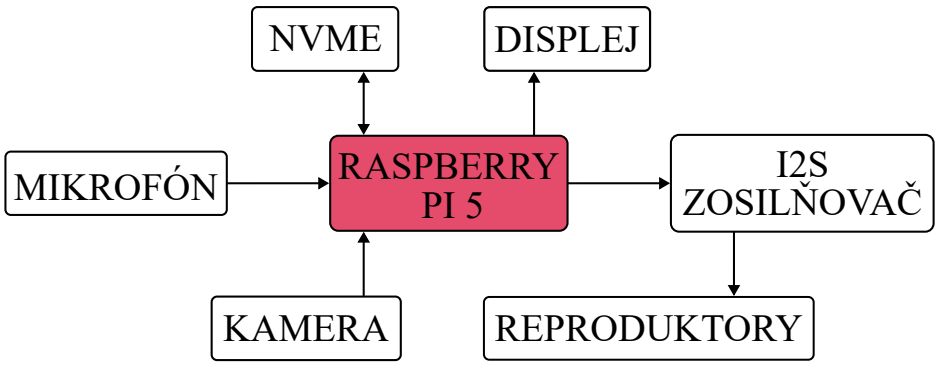
Štúdia	Jazykový model	Presnosť	F1-skóre	Návratnosť	Správnosť	Úloha
[10]	ChatGPT-4	96,04%	95,82%	95,60%	X	transformácia
	ChatGPT-4	97,11%	97,24%	97,37%	X	textu
	ChatGPT-4	91,09%	84,77%	79,27%	X	na
	ChatGPT-4	94,10%	92,39%	90,75%	X	graf
[11]	ChatGPT-3, Galactica, Perplexity, BioMedLM	29,00%	29,00%	29,00%	X	AI diagnóza
[12]	RumorLLM	70,00%	63,64%	58,33%	82,55%	detekcia
	RumorLLM	85,19%	86,79%	88,46%	93,91%	dezinformácií
[13]	ChatGPT-4	85,00%	90,00%	97,00%	0.88	detekcia
	ChatGPT-4	95,00%	85,00%	77,00%	0.88	zavádzajúcich
	Gemini	68,00%	73,00%	79,00%	0.67	názvov
	Gemini	65,00%	57,00%	50,00%	0.67	článkov
[14]	ChatGPT-4	63,00%	77,00%	100,00%	X	onkologická aplikácia
[15]	LLM	75,00%	71,00%	68,00%	X	internetové
	LLM	82,00%	79,00%	77,00%	X	obchody
[16]	BI-LSTM	90,00%	91,00%	92,00%	91,00%	zdravotnícky
	BERT	97,00%	98,00%	96,00%	98,00%	chatbot
[17]	MLP	99,76%	99,75%	99,75%	99.75%	univerzitný chatbot
[18]	BERT	92,00%	92,00%	92,00%	94,00%	identifikácia rizika samovraždy
[19]	RF	84,00%	88,00%	91,00%	87,00%	detekcia klamstva

Tab. 2 Sumarizácia údajov zo súvisiacich štúdií

III. Použitý hardvér

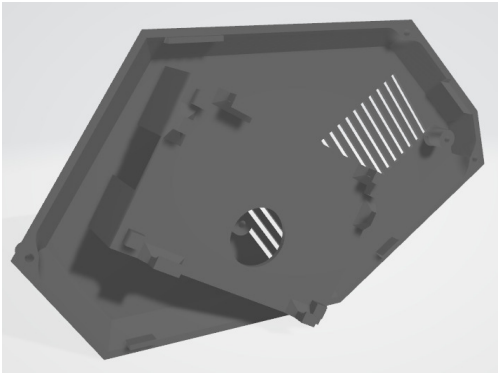
Tento článok prezentuje návrh a vývoj robota KEMŤÁKa, ktorý integruje umelú inteligenciu s fyzickým robotickým systémom. Na základe funkčných požiadaviek bolo zvolené nasledovné technické vybavenie:

- výpočtová jednotka – Raspberry Pi 5,
- úložisko – NVMe rozširujúca doska a SSD disk,
- zvukový výstup – reproduktory a I2S zosilňovač,
- napájanie – USB-C zdroj 27W,
- vstupné periférie – mikrofón a kamera,
- výstupná periféria – dotykový displej,
- fyzická konštrukcia – 3D tlačené časti.

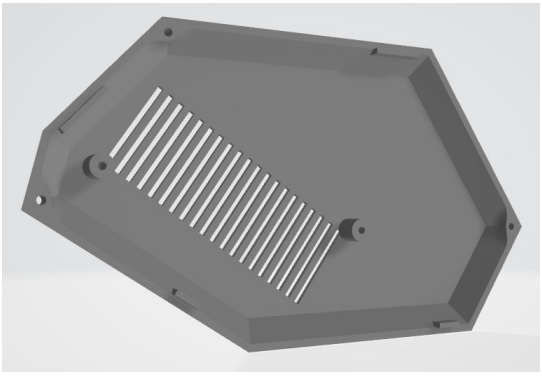


Obr. 1 Bloková schéma zapojenia robota KEMŤÁKa

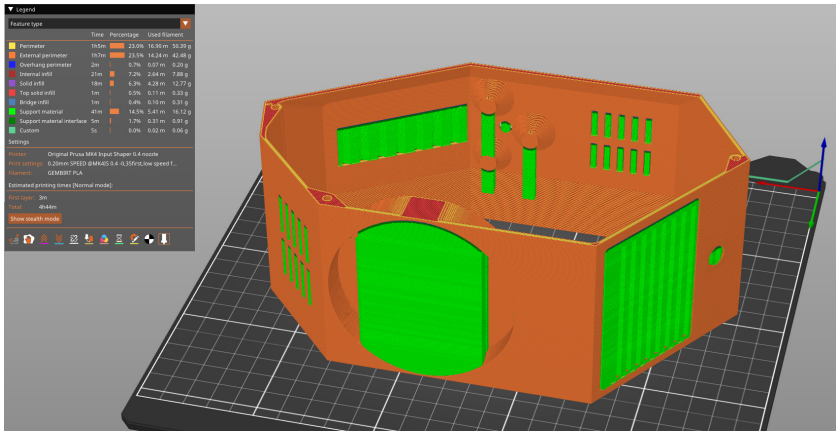
Raspberry Pi 5 tvorí výkonný základ robota, s 64-bitovým ARM Cortex-A76 procesorom (2,4 GHz), 8 GB RAM, Wi-Fi 6, Bluetooth 5.0 a podporou dvoch 4K displejov cez Micro-HDMI. Operačný systém Raspberry Pi OS bol nainštalovaný z bootloadera pomocou Raspberry Pi Imager a beží z NVMe SSD disku pripojeného cez rozširujúcu dosku Pimoroni NVMe Base, čím sa výrazne zvýšila rýchlosť a spoľahlivosť oproti microSD karte. Použitý SSD disk Patriot P310 (240 GB) s rýchlosťou čítania 1700 MB/s a zápisu 1000 MB/s umožňuje rýchle spracovanie dát zo senzora a kamery. Zvukový výstup zabezpečujú dva 8Ω 5W reproduktory Waveshare, pripojené cez Adafruit I2S Speaker Bonnet, pričom digitálne I2S rozhranie minimalizuje šum a overlay `hifiberry-dac` umožňuje plnohodnotné audio výstupy. Na záznam zvuku slúži Respeaker USB Mic Array s potlačením šumu a smerovým snímaním, ktorý je vďaka USB pripojeniu plne funkčný bez nutnosti ovládačov. Vizuálne dáta sníma 12 Mpx kamera tretej generácie s HDR, automatickým zaostrovaním a širokým 120° uhlom záberu, pripojená cez CSI rozhranie pomocou špeciálneho kábla pre Raspberry Pi 5. Na vizuálny výstup slúži 2,8” dotykový IPS displej od Waveshare (480x640 px, 60 Hz), pripojený cez DSI rozhranie. Celé zariadenie je osadené v trojdielnej nosnej konštrukcii (ľavá, pravá, stredná časť), ktorá je navrhnutá pre efektívne rozmiestnenie komponentov, chladenie a káblový manažment, pričom bola vytlačená na 3D tlačiarni Prusa MK4 z PLA materiálu s časom tlače približne 10 hodín (pozri Obr. 2, 3, 4). Komponenty boli zabezpečené katedrou KEMT.



Obr. 2 Ľavá časť fyzickej schránky robota



Obr. 3 Pravá časť fyzickej schránky robota



Obr. 4 Príprava na tlač strednej časti fyzickej konštrukcie robota

IV. Komunikačné rozhrania

Komunikačné rozhranie robota je implementované modulárne v jazyku Python a pozostáva z piatich komponentov riadených hlavným skriptom. Modul spracovania hlasu zabezpečuje záznam a rozpoznanie reči pomocou mikrofónu ReSpeaker USB Mic Array s DSP filtrami. Zvuk sa nahráva, ukladá ako WAV a prevádza na text cez knižnicu SpeechRecognition s modulom Google Speech Recognition, pričom rozpoznaný text sa zobrazuje na displeji a spracúva cez OpenAI API (model gpt-4o-mini). Ak sú rozpoznané ukončovacie frázy ("koniec", "ďakujem", "dovidenia", systém prejde do fázy spätnej väzby. Modul spracovania obrazu (`person_detector.py`) detekuje osobu pomocou Picamera2, OpenCV a predtrénovanej siete MobileNet SSD. Pri detekcii osoby so spoľahlivosťou nad 50% sa spúšťa interakcia. Prevod textu na reč zabezpečuje služba Google Text-to-Speech (gTTS), ktorá generuje MP3 súbor z textu a prehráva ho cez `mpg321`, po čom je súbor zmazaný. Vizuálne rozhranie využíva 2,8" dotykový DSI displej

a knižnicu `pygame`, pričom zobrazuje štyri fázy: šetrič obrazovky s animovanými očami, konverzáciu s vypisovaním vstupov a odpovedí, spätnú väzbu cez tlačidlá like/dislike a zobrazenie QR kódov na web a sociálne siete (pozri Obr. 5, 6, 7, 8. Automatické spustenie robota zabezpečuje skript `wait_for_internet.sh`, ktorý pri zapnutí Raspberry Pi overí pripojenie k internetu a následne spustí `start_chatbot.sh`, pričom inicializácia je pridaná do `.bashrc`. Celý systém tak funguje autonómne bez potreby manuálneho zásahu.

Obr. 5 Šetrič obrazovky robota

Obr. 6 Konverzačná fáza robota

Obr. 7 Spätná väzba

Obr. 8 QR kódy

V. Testovanie robota a dosiahnuté výsledky

Cieľom bolo overiť, či robot dokáže viesť zmysluplnú interakciu pomocou jazykového modelu. Hoci sa testovanie nepodarilo uskutočniť pred verejnosťou, prebehli interné testy jednotlivých funkcií. Kamera spoľahlivo detegovala osobu do 5 metrov pri dobrom osvetlení, no v tme bola neúčinná. Rozpoznávanie reči fungovalo veľmi dobre v tichom aj mierne hlučnom prostredí. Vstup bol spracovávaný na základe parametrov `recognizer.energy_treshold` (280) a `pause_treshold` (1,75 s), ktoré možno doladiť podľa prostredia. Jazykový model OpenAI GPT-4o mini generoval prevažne zrozumiteľné a informatívne odpovede v slovenčine, no nevie udržať kontext ani reagovať na predchádzajúce vstupy. Zvukový výstup bol zrozumiteľný, no čítanie skratiek a dátumov nebolo vždy prirodzené. Displej ovládaný cez `pygame` správne zobrazoval všetky fázy. Počas záťažového testu (1,5 hodiny) nebolo zaznamenané prehrievanie ani žiadna porucha. Overili sme tak spoľahlivosť a funkčnosť jednotlivých komponentov aj ich spoluprácu v automatizovanom cykle interakcie.

Na testovanie bola použitá sada 50 otázok rôznej náročnosti a tém (vtipné, filozofické, politické). Odpovede boli hodnotené na škále 1–5 podľa zrozumiteľnosti, relevancie, porozumenia otázke a celkového dojmu. Väčšina odpovedí bola jasná a správna, s priemerným skóre zrozumiteľnosti 4,91, relevancie 4,66 a rozpoznania otázky 4,59. Problém sa objavil pri skratkách (napr. „KEMT“ zamieňal za inú katedru), čo sa dá riešiť explicitným vyslovením alebo uvedením celého názvu. Model mal obmedzenú kreativitu pri tvorbe básní a nedokázal odpovedať na subjektívne otázky typu „Obľúbené jedlo?“. Celkové hodnotenie dosiahlo 4,3. Tieto výsledky ukazujú, že model vie odpovedať zrozumiteľne a relevantne, pokiaľ mu je otázka jasne formulovaná.

Č.	Vstupná otázka / požiadavka	Zrozumiteľnosť odpovede	Relevancia odpovede	Rozpoznanie a porozumenie otázky	Celkové hodnotenie
1	Ahoj	5	5	5	5
2	Ako sa máš?	5	5	5	5
3	Čo je to umelá inteligencia?	5	5	5	5
4	Povedz mi vtip	4	4	5	4
5	Kto si?	5	5	5	5
6	Kto ťa vytvoril?	5	5	5	5
7	Čo robíš?	5	5	5	5
8	Aké je tvoje meno?	5	5	5	5
9	Vieš rozprávať po slovensky?	5	5	5	5
10	Koľko je 5 + 7?	5	5	5	5
11	Povedz mi niečo o KEMT	4	5	5	5
12	Aké je tvoje obľúbené jedlo?	5	5	5	5
13	Vysvetli mi, čo je algoritmus	5	5	5	5
14	Rozprávaj príbeh o robotovi	5	4	5	4
15	Zaspievaj pesničku	4	4	4	4

Tab. 3 Ukážka testovania robota KEMŤÁKa

Metodika hodnotenia	Priemerné skóre (/5)
Zrozumiteľnosť odpovede	4,91
Relevancia odpovede	4,66
Rozpoznanie otázky	4,59
Celkové hodnotenie (subjektívne)	4,3

Tab. 4 Priemerné hodnotenia interakcie s robotom KEMŤÁKom

Testovaním sme overili technickú aj softvérovú funkčnosť každého komponentu vytvoreného robota KEMŤÁKa. Počas interakcií bude robot zbierať hodnotenia vo forme „páči sa mi to / nepáči sa mi to" (z angl. "like / dislike“), no odporúčame doplniť aj spätnú väzbu od používateľov s možnosťou vyjadriť konkrétne návrhy na zlepšenie. Testovanie preukázalo konzistentnosť výsledkov a ukázalo, v ktorých oblastiach robot KEMŤÁK spĺňa očakávania a kde má rezervy. Overený proces interakcie zahŕňal detekciu človeka kamerou, prevod hlasového vstupu na text cez `speech_recognition`, spracovanie textu jazykovým modelom, syntézu reči cez gTTS a vizualizáciu celej interakcie na displeji. Nasledujúce kroky predstavujú úspešne overený proces interakcie medzi robotom KEMŤÁKom a používateľom:

- detekcia prítomnosti osoby bola úspešne realizovaná pomocou kamery,
- hlasový vstup bol mikrofónom spoľahlivo zachytený a následne konvertovaný na text pomocou knižnice `speech_recognition`,
- získaný text bol odoslaný prostredníctvom API do veľkého jazykového modelu (z angl. Large Language Model, LLM), ktorý vygeneroval odpoveď v slovenčine,
- odpoveď bola prevedená na reč pomocou nástroja gTTS a prehratá cez reproduktor,
- celý priebeh interakcie bol súčasne vizuálne zobrazený na displeji.

Na základe testovania sme identifikovali niekoľko slabín, ktoré by bolo vhodné v budúcnosti zlepšiť. Prvým problémom je knižnica `SpeechRecognition`, ktorá je citlivá na šum, čo výrazne znižuje kvalitu rozpoznávania reči v hlučných prostrediach. Lepšou alternatívou by bol open-source model Whisper od OpenAI, ktorý vďaka kontextovému spracovaniu a normalizácii zvuku lepšie zvláda prácu s nekvalitným audio vstupom, pričom podporuje aj slovenský jazyk. Jeho lokálna implementácia je však výpočtovo náročná a môže zaťažiť či prehrievať zariadenie Raspberry Pi 5, zatiaľ čo používanie prostredníctvom API je spoplatnené, čo predstavuje kompromis medzi výkonom a nákladmi.

Druhým návrhom na zlepšenie je zabezpečenie funkčnosti robota aj pri výpadku internetového pripojenia. To by bolo možné dosiahnuť inštaláciou menšieho lokálneho jazykového modelu, ktorý by zvládol aspoň základné príkazy, alebo implementáciou kontextového okna prostredníctvom uchovávania predchádzajúcich dialógov v pamäti, čo by umožnilo prirodzenejšiu a konzistentnejšiu konverzáciu.

Nakoniec, z hľadiska hardvérových možností by fyzické vylepšenie robota, napríklad zavedením pohybu, vyžadovalo prechod z pevného napájania cez USB-C na prenosnú dobíjateľnú batériu. Tento krok by výrazne rozšíril mobilitu robota a otvoril nové možnosti využitia v oblasti

robotiky a interaktívnych aplikácií. Do budúcna by sa tiež mohlo uvažovať o rozšírení senzorickej výbavy, napríklad o kamerový vstup či detekciu prekážok, čo by výrazne zvýšilo interaktivitu a autonómiu zariadenia.

VI. Záver

Cieľom tohto článku bolo navrhnúť, zostrojiť a naprogramovať komunikačné rozhranie robota využívajúceho umelú inteligenciu. Opísali sme výber a inštaláciu všetkých súčiastok zakúpených katedrou KEMT vrátane konfigurácie Raspberry Pi 5. Vytvorili sme vlastný dizajn robota, ktorý sme zmontovali do funkčnej podoby, a implementovali softvér využívajúci knižnice na detekciu človeka, generovanie odpovedí cez ChatGPT API a syntézu hlasu pomocou Google Text-to-Speech. Súčasťou bola aj tvorba grafického rozhrania a fázovej logiky riadiacej interakciu. Na testovanie sme použili sadu 50 otázok, ktoré potvrdili konzistentnosť, zrozumiteľnosť a užitočnosť robota v edukačnom prostredí. Medzi jeho slabiny patrí absencia kontextu konverzácie a neaktuálne informácie od roku 2021. Do budúcnosti plánujeme rozšíriť model o kontextové udržiavanie dialógu a vylepšiť fyzickú konštrukciu robota na plne funkčné telo.

Poďakovanie

Výskum v tomto príspevku podporilo Ministerstvo školstva, vedy, výskumu a športu SR v rámci výskumného projektu KEGA 049TUKE-4/2024, & KEGA 041TUKE-4/2025, Agentúra na podporu výskumu a vývoja v rámci výskumných projektov APVV-22-0261, APVV-22-0400 & APVV-22-0414. Výskum bol podporný aj v rámci projektu z Plánu obnovy 09I05-03-V02-00019.

Použitá literatúra

[1] M. Ghallab and F. Ingrand, "Robotics and artificial intelligence," *A Guided Tour of Artificial Intelligence Research*, 2020.
[2] T. Adiguzel, M. H. Kaya, and F. K. Cansu, "Revolutionizing education with ai: Exploring the transformative potential of chatgpt," *Contemporary Educational Technology*, 2023.
[3] OpenAI, "Openai api documentation," 2024. [Online]. Available: https://platform.openai.com/docs/overview
[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023. [Online]. Available: https://arxiv.org/abs/1706.03762
[5] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.
[6] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," 2020. [Online]. Available: https://arxiv.org/abs/2005.14165
[7] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, "Training language models to follow instructions with human feedback," 2022. [Online]. Available: https://arxiv.org/abs/2203.02155
[8] OpenAI, "Openai api pricing," 2024. [Online]. Available: https://platform.openai.com/docs/pricing
[9] S. I. Ross, F. Martinez, S. Houde, M. Muller, and J. D. Weisz, "The programmer's assistant: Conversational interaction with a large language model for software development," in *Proceedings of the 28th International Conference on Intelligent User Interfaces*, ser. IUI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 491–514. [Online]. Available: https://doi.org/10.1145/3581641.3584037
[10] P. Jiang, J. Rayan, S. P. Dow, and H. Xia, "Graphologue: Exploring large language model responses with interactive diagrams," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3586183.3606737
[11] M. Benary, X. D. Wang, M. Schmidt, D. Soll, G. Hilfenhaus, M. Nassir, C. Sigler, M. Knödler, U. Keller, D. Beule, U. Keilholz, U. Leser, and D. T. Rieke, "Leveraging Large Language Models for Decision Support in Personalized Oncology," *JAMA Network Open*, vol. 6, no. 11, pp. e2 343 689–e2 343 689, 11 2023. [Online]. Available: https://doi.org/10.1001/jamanetworkopen.2023.43689
[12] J. Lai, X. Yang, W. Luo, L. Zhou, L. Li, Y. Wang, and X. Shi, "Rumorllm: A rumor large language model-based fake-news-detection data-augmentation approach," *Applied Sciences*, vol. 14, no. 8, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/8/3532
[13] M. M. U. Rony, M. M. Haque, M. Ali, A. S. Alam, and N. Hassan, "Exploring the potential of the large language models (llms) in identifying misleading news headlines," *arXiv preprint arXiv:2405.03153*, 2024.
[14] T. Hung, G. Kuperman, E. J. Sherman, A. L. Ho, W. Wong, A. Kriplani, L. Dunn, J. V. Fetten, L. S. Michel, S. S. Baxi, C. Weng, D. G. Pfister, and J. J. Mao, "Performance of a trained large language model to provide clinical trial recommendation in a head and neck cancer population." *Journal of Clinical Oncology*, vol. 42, no. 16_suppl, pp. 11 081–11 081, 2024, pMID:.
[15] W. Xu, J. Xiao, and J. Chen, "Leveraging large language models to enhance personalized recommendations in e-commerce," 2024. [Online]. Available: https://arxiv.org/abs/2410.12829
[16] A. Babu and S. B. Boddu, "Bert-based medical chatbot: Enhancing healthcare communication through natural language understanding," *Exploratory Research in Clinical and Social Pharmacy*, vol. 13, p. 100419, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667276624000143
[17] S. Salloum, K. Shaalan, A. Basiouni, A. Salloum, and R. Alfaisal, *Building and Evaluating a Chatbot Using a University FAQs Dataset*, 07 2024, pp. 207–218.
[18] J. Chan, S.-L. Chua, and L.-K. Foo, *A Two-Stage Classification Chatbot for Suicidal Ideation Detection*, 12 2022, pp. 405–412.
[19] K. Tsuchiya, R. Hatano, and H. Nishiyama, "Detecting deception using machine learning with facial expressions and pulse rate," *Artificial Life and Robotics*, vol. 28, no. 3, pp. 509–519, 2023.

Automatické rozpoznávanie emócií z rečového signálu implementované vo webovej aplikácii

[1]Martin LUKÁČ, [2]Matúš PLEVA, [3]Eva KUPCOVÁ, [4]Kristián SOPKOVIČ

[1,2,3,4]Katedra elektroniky a multimediálnych telekomunikácií,
Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Letná 9, 04001 Košice, Slovenská republika

[1]martin.lukac@student.tuke.sk, [2]matus.pleva@tuke.sk, [3]eva.kupcova@tuke.sk,
[4]kristian.sopkovic@tuke.sk

Abstrakt **– Táto práca sa zaoberá automatickým rozpoznávaním emócií z rečového signálu s využitím neurónových sietí integrovaných do webovej aplikácie. Bol navrhnutý a implementovaný Python modul umožňujúci analýzu rečových nahrávok a ich klasifikáciu podľa emocionálneho obsahu. Modul je nasadený na pozadí webovej aplikácie vytvorenej pomocou frameworku Flask. Používateľ má možnosť nahrať zvukový súbor alebo zaznamenať hlas priamo cez mikrofón, pričom systém následne poskytne predikciu emócie. Navrhnuté riešenie slúži ako praktický demonštrátor využitia umelej inteligencie v oblasti spracovania reči a emocionálnej analýzy v reálnom čase.**

Kľúčové slová **– Docker, Flask, klasifikátory, neurónové siete, Python, rozpoznávanie emócií**

I. Úvod

Reč predstavuje základný prostriedok ľudskej komunikácie a nesie v sebe množstvo informácií, ktoré presahujú samotný verbálny obsah. Emócie zakódované v reči zohrávajú zásadnú úlohu v interpretácii významu hovoreného slova, ako aj v porozumení medzi komunikujúcimi osobami. Oblasť automatického rozpoznávania emócií v reči (z anglického *Speech Emotion Recognition, SER*), sa dynamicky rozvíja s cieľom identifikovať a analyzovať emocionálny obsah reči na základe verbálnych a neverbálnych charakteristík.

Rozpoznávanie emócií však predstavuje výrazne subjektívny problém, nakoľko prejavy emócií sú ovplyvnené viacerými faktormi ako jazyk, pohlavie, vek, zafarbenie a tón hlasu, kultúrne prostredie či osobné skúsenosti. Tieto variabilné aspekty významne sťažujú jednoznačnú interpretáciu emočného stavu rečníka.

Systémy SER nachádzajú široké uplatnenie v rôznych oblastiach. Od zlepšenia interakcie medzi používateľom a inteligentným virtuálnym asistentom až po podporu psychologickej diagnostiky v zdravotníctve. Napríklad pri detekcii úzkosti, depresie alebo stresu môže rozpoznanie emócií prispieť k efektívnejšiemu nastaveniu liečebného postupu. Vzhľadom na komplexnosť úlohy sa väčšina SER systémov [1] zameriava na klasifikáciu obmedzenej množiny najvýraznejších emócií, ako sú: *radosť, hnev, smútok, strach* a *neutrálna reč.*

Na riešenie úloh súvisiacich s rozpoznávaním emócií sa čoraz častejšie využívajú metódy strojového učenia. V poslednom období zohráva kľúčovú úlohu hlboké učenie, ktoré umožňuje modelom automaticky identifikovať komplexné vzory v zvukových dátach. Medzi najúspešnejšie architektúry patria konvolučné neurónové siete (z anglického *Convolutional Neural Networks, CNNs*) a rekurentné neurónové siete (z anglického *Recurrent Neural Networks, RNNs*). CNNs sú obzvlášť efektívne pri extrakcii relevantných príznakov zo spektrogramov (vizuálnych reprezentácií zvukového signálu) [2].

II. Použité metódy

V tejto kapitole predstavujeme použité nástroje, datasetové zdroje a algoritmy, ktoré sme aplikovali pri realizácii rozpoznávania emócií z reči. Zameriavame sa na výber vhodných datasetov, implementáciu testovacích skriptov a využitie moderných metód samostatného učenia na spracovanie zvukových dát.

A. SERAB

Ako základ sme použili nástroj SERAB [3], ktorý obsahuje niekoľko testovacích datasetov a modelov vhodných na tvorbu Python skriptov na rozpoznávanie emócií z reči. Prvým krokom bol výber najúspešnejšieho datasetu na testovanie. Následne sme vytvorili kód umožňujúci testovanie nahrávok priamo z príkazového riadka, čo nám umožnilo vyhodnotiť kvalitu rozpoznávania na dostupných nahrávkach aj na cudzích testovacích datasetoch. Po úspešnom spustení skriptov sme vyvinuli webovú aplikáciu využívajúcu bežné webové technológie ako HTML, JavaScript a Flask. Túto aplikáciu sme nasadili na TUKE cloud a úspešnosť vyhodnocovania sme testovali na vzorke minimálne desiatich respondentov.

B. Databázy emócií

Nástroj SERAB poskytuje deväť rôznych klasifikačných úloh [4] zameraných na rozpoznávanie emócií z reči v šiestich rôznych jazykoch. Použité datasety sú:

- **anglické:** CREMA-D [5], IEMOCAP [6], RAVDESS [7], SAVEE [8],
- **francúzske:** CaFE [9],
- **nemecké:** EmoDB [10],
- **grécke:** AESDD [11],
- **talianske:** EMOVO [12],
- **perzské:** ShEMO [13].

Každý z použitých datasetov obsahuje tri základné atribúty: monofonické zvukové nahrávky vo formáte `.raw`, identifikačné údaje o hovoriacom (pohlavie, vek) a anotáciu prítomnej emócie (napr. radosť, smútok, hnev). Datasety sa líšia veľkosťou, počtom hovoriacich a distribúciou emócií. Zatiaľ čo emócie ako hnev, šťastie a smútok sú prítomné vo všetkých, kategórie ako znechutenie, strach, neutrálna emócia, prekvapenie či nuda sa vyskytujú len v niektorých. Priemerná dĺžka nahrávok sa pohybuje od 2,5 do 4,5 sekundy a ide prevažne o krátke úseky so zameraním na konkrétnu emóciu.

Na účely trénovania modelov, ladenia hyperparametrov a následného nezávislého hodnotenia výkonnosti sú všetky datasety rozdelené do troch častí: trénovacej, validačnej a testovacej. V prípade datasetu CREMA-D je rozdelenie nasledovné: 70 % dát je určených na trénovanie, 10 % na validáciu a 20 % na testovanie. Ostatné datasety sú štandardne rozdelené v pomere 60 % trénovacích, 20 % validačných a 20 % testovacích nahrávok. Takto definované rozdelenie dát zabezpečuje konzistentnosť experimentálneho nastavenia a umožňuje porovnateľné vyhodnotenie výsledkov naprieč rôznymi prístupmi a množstvami trénovacích dát.

C. BYOL-A

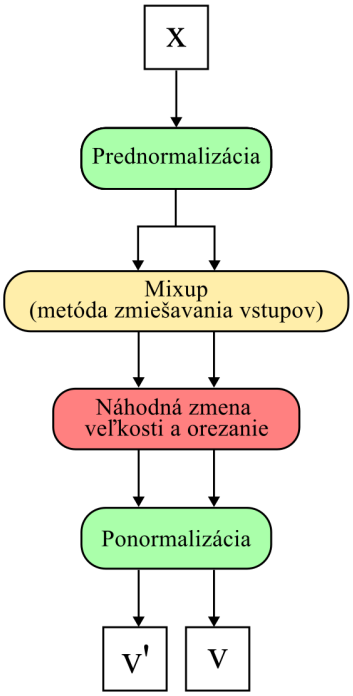
BYOL (Bootstrap Your Own Latent) je algoritmus samostatného učenia reprezentácií obrazu [4], ktorý využíva dve neurónové siete označované ako online a cieľová sieť. Tieto siete sa navzájom ovplyvňujú a učia na základe rôznych rozšírených pohľadov na ten istý obrázok.

BYOL-A je upravená verzia tohto algoritmu pre spracovanie a učenie reprezentácií zvuku. Na vstup prijíma audio dáta predspracované do podoby logaritmického mel-spektrogramu, čo je časovo-frekvenčný znak vhodný pre konvolučné neurónové siete. Modul augmentácie v pôvodnom BYOL je nahradený novým modulom, ktorý umožňuje vytvárať kontrasty v rozšírených zvukových zobrazeniach, čím sa zlepšuje učenie všeobecných reprezentácií zvuku.

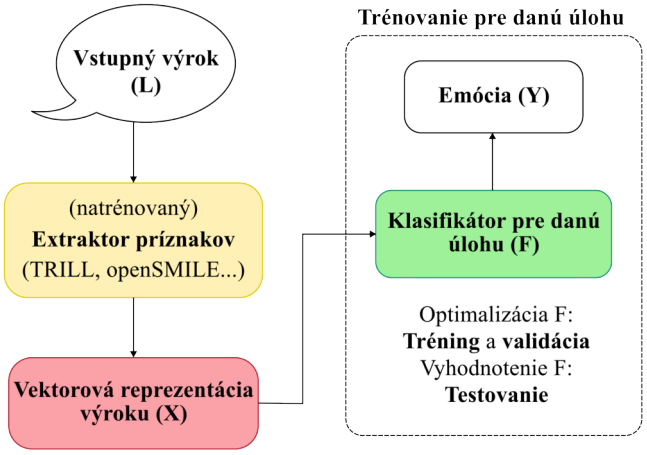
BYOL-A sa skladá zo štyroch blokov, ktoré sú zobrazené na obrázku 1. Najprv blok predbežnej normalizácie normalizuje jeden vstupný zvukový segment, aby sa nasledujúce procesy augmentácie, najmä mixup, stali stabilnými. Normalizovaný vstup sa zdvojí do dvoch kópií a privedie sa do nasledujúceho bloku Mixup. Potom blok Mixup vytvorí dva výstupy, ktoré sú zmesou normalizovaných vstupov a náhodne vybraných minulých normalizovaných vstupov. Nasledujúci blok Random Resize Crop (RRC) náhodne mení veľkosť a orezáva výstupy a nakoniec blok postnormalizácie upravuje štatistické odchýlky spôsobené predchádzajúcimi rozšíreniami. Na učenie všeobecných zvukových reprezentácií sa BYOL-A zameriava na akustické javy v popredí a všetky detaily obsahu nahrávky. Blok Mixup je navrhnutý na vytvorenie kontrastu na učenie reprezentácií akustických javov v popredí a je kombinovaný s blokom prednormalizácie na stabilné zvýšenie výkonu. Blok RRC aproximuje posun výšky tónu a časové rozťahovanie v časovo-frekvenčných funkciách na učenie všeobecných reprezentácií detailov obsahu.

D. Evaluácia výsledkov

Vyhodnocovanie v nástroji SERAB je znázornené na Obrázku 2. Nástroj generuje reprezentácie emócií na základe akustických príznakov extrahovaných z reči pomocou predtrénovaných alebo



Obrázok 1 Bloková schéma BYOL-A



Obrázok 2 Bloková schéma vyhodnocovania emócií v reči

netrénovaných extraktorov. Výsledné vektorové reprezentácie sú následne klasifikované špecializovaným modelom pre danú úlohu.

S cieľom dosiahnutia čo najvyššej presnosti pri vyhodnocovaní emócií boli použité rôzne prístupy ku klasifikácii [14]:

- **logistická regresia** (z anglického *Logistic Regression, LR*): pravdepodobnostný model využívajúci logistickú funkciu na predikciu príslušnosti vstupných dát ku konkrétnej triede,
- **metóda podporných vektorov** (z anglického *Support Vector Machine, SVM*): klasifikátor založený na maximalizácii hranice medzi triedami v transformovanom priestore pomocou tzv. podporných vektorov,
- **kvadratická diskriminačná analýza** (z anglického *Quadratic Discriminant Analysis, QDA*): metóda využívajúca odlišné kovariančné matice pre každú triedu, čím umožňuje modelovanie nelineárnych hraníc medzi triedami,
- **lineárna diskriminačná analýza** (z anglického *Linear Discriminant Analysis, LDA*): štatistický prístup predpokladajúci rovnaké kovariančné matice medzi triedami, ktorý hľadá lineárne hranice rozhodovania,
- **metóda náhodného lesa** (z anglického *Random Forest, RF*): ansámblová metóda založená na kombinovaní viacerých rozhodovacích stromov trénovaných na náhodne vybraných podmnožinách dát a príznakov.

Optimalizácia hyperparametrov jednotlivých modelov prebiehala pomocou metódy mriežkového vyhľadávania (`GridSearchCV(*)`), ktorá systematicky prehľadala rôzne kombinácie nastavení a zvolila tie s najvyššou klasifikačnou presnosťou na validačnej množine. Na porovnanie výkonnosti boli vybrané tri datasety (EmoDB, Emovo, CaFE) a dve metódy normalizácie vstupných vektorov (štandardná normalizácia (*StandardScaler*) a SERAB). Testovanie prebehlo pre štyri klasifikátory: LDA, QDA, LR a RF. Tabuľka 1 sumarizuje výsledky klasifikácie pre všetky kombinácie. Najlepšie hodnoty v jednotlivých stĺpcoch sú zvýraznené šedým pozadím. Najvyššiu presnosť dosiahla logistická regresia, predovšetkým pri datasete CaFE so SERAB normalizáciou (71,4%). Naopak, QDA dosiahla najnižšiu úspešnosť vo všetkých prípadoch. Celkovo sa normalizácia SERAB ukázala ako efektívnejšia než štandardná.

Klasifikátor	**Štandardná normalizácia**			**SERAB normalizácia**		
	EmoDB	Emovo	CaFE	EmoDB	Emovo	CaFE
LDA	83,3 %	**40,8 %**	59,8 %	82,1 %	**41,8 %**	70,5 %
QDA	17,9 %	14,8 %	16,2 %	17,9 %	25,2 %	20,5 %
LR	**85,7 %**	34,2 %	**65,4 %**	**84,5 %**	31,6 %	**71,4 %**
RF	76,2 %	35,2 %	54,7 %	79,8 %	37,2 %	61,1 %

Tabuľka 1 Presnosť klasifikátorov pri rôznych normalizáciách (BYOL-S [1024])

III. Webová aplikácia a jej testovanie

Po implementácii modulu na rozpoznávanie emócií z reči (SER) bola vyhodnotená úspešnosť jednotlivých natrénovaných modelov. Na trénovanie bol použitý dataset RAVDESS, ktorý pri offline experimentoch poskytoval najlepšie výsledky z hľadiska klasifikačnej presnosti.

V ďalšej fáze bolo cieľom overiť generalizačnú schopnosť modelov pri použití nahrávok z iných datasetov, a to najmä z hľadiska jazykovej nezávislosti. Skúmalo sa, či modely natrénované na anglicky hovoriacom datasete RAVDESS dokážu spoľahlivo klasifikovať emócie aj v prípadoch, keď sú vstupné nahrávky v inom jazyku, napríklad v nemčine alebo slovenčine.

Pre nemecký jazyk bol zvolený dataset EmoDB, ktorý obsahuje herecky prednesené emócie v štandardizovanom formáte. Pre slovenský jazyk boli vytvorené nové vlastné nahrávky s cieľom reprezentovať relevantné emócie v kontrolovaných podmienkach.

	Hnev	**Znechutenie**	**Strach**	**Radosť**	**Smútok**	**Prekvapenie**	**Neutrálne**	**Pokoj**
Hnev	**7**	1	0	0	0	0	0	0
Znechutenie	0	**5**	0	0	0	2	0	1
Strach	0	0	**4**	1	3	0	0	0
Radosť	1	0	0	**5**	0	2	0	0
Smútok	0	0	0	2	**3**	0	3	0
Prekvapenie	0	0	2	2	1	**3**	0	0
Neutrálne	0	0	0	0	0	0	**4**	0
Pokoj	0	0	0	1	0	2	0	**5**

Tabuľka 2 Matica úspešnosti klasifikácie emócií pre podmnožinu nahrávok z datasetu RAVDESS

V Tabuľke 2 je znázornená matica úspešnosti zachytávajúca výsledky rozpoznávania emócií pre podmnožinu nahrávok z datasetu RAVDESS. Riadenia reprezentujú skutočné triedy emócií, zatiaľ čo stĺpce zodpovedajú emóciám predikovaným klasifikátorom.

Hodnoty na diagonále (zvýraznené šedou farbou) predstavujú správne klasifikácie, pričom najvyššiu mieru presnosti dosiahli emócie ako *hnev*, *neutrálna emócia* a *pokoj*. Naopak, najčastejšie zámenné prípady sa vyskytli medzi *strachom* a *smútkom* či medzi *radosťou* a *prekvapením*, čo poukazuje na ich akustickú podobnosť v daných vzorkách.

Celková úspešnosť rozpoznávania v tomto teste dosiahla 60 %, čo je približne o 10 % menej ako pri testovaní na plnom datasete RAVDESS. Napriek tomu výsledky naznačujú, že vyvinutý modul na rozpoznávanie emócií vykazuje uspokojivú mieru spoľahlivosti, a to aj v kontexte jeho integrácie do webovej aplikácie určenej na demonštráciu základnej funkcionality systému.

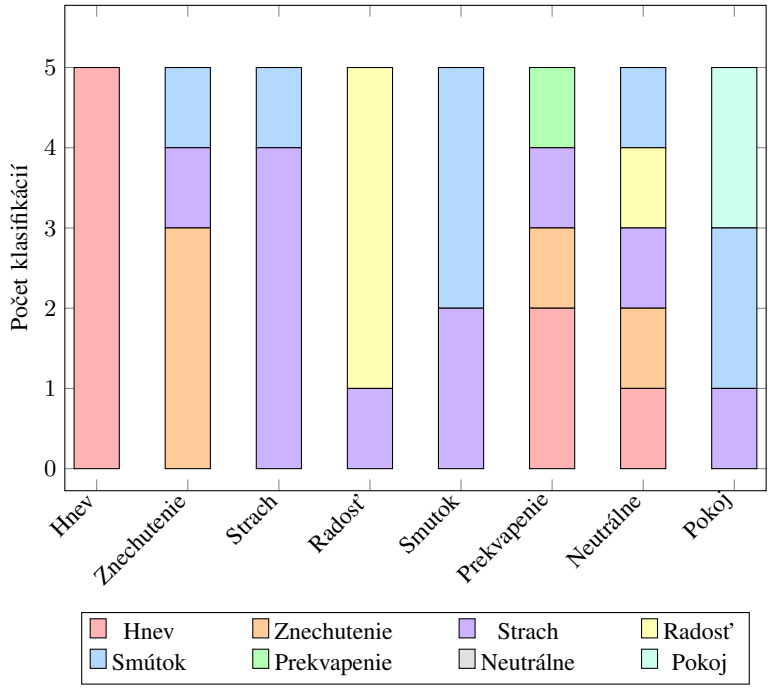
Naopak, v prípade slovenských nahrávok, ktorých výsledky sú znázornené v Tabuľke 3, bola celková úspešnosť rozpoznávania emócií nižšia, pričom dosiahla iba 55 %. Tento výsledok bol však do veľkej miery očakávaný, keďže išlo o tzv. crosslingválne nasadenie – teda aplikovanie modelu trénovaného výlučne na anglickom datasete RAVDESS na nahrávky v slovenskom jazyku. Jazykové a akustické rozdiely medzi angličtinou a slovenčinou môžu významne ovplyvniť presnosť rozpoznávania, čo sa premietlo aj do výsledkov.

	Hnev	Znechutenie	Strach	Radosť	Smútok	Prekvapenie	Neutrálne	Pokoj
Hnev	**5**	0	0	0	0	0	0	0
Znechutenie	0	**3**	1	0	1	2	0	1
Strach	0	0	**4**	0	1	0	0	0
Radosť	0	0	1	**4**	0	0	0	0
Smútok	0	0	2	0	**3**	0	0	0
Prekvapenie	2	1	1	0	0	**1**	0	0
Neutrálne	1	1	1	1	1	0	**0**	0
Pokoj	0	0	1	0	2	0	0	**2**

Tabuľka 3 Matica úspešnosti klasifikácie emócií pre podmnožinu slovenských nahrávok

Z tabuľky je zrejmé, že niektoré emócie boli rozpoznávané výrazne lepšie ako iné. Napríklad emócia hnev bola identifikovaná so 100 % presnosťou, čo naznačuje, že ide o akusticky výrazný emocionálny stav, ktorý si model dokázal spojiť s naučenými vzormi aj napriek jazykovým rozdielom. Na druhej strane však emócie ako prekvapenie, neutrálne a pokoj vykazovali výrazne nižšiu úspešnosť. Obzvlášť pozoruhodný je prípad neutrálnej reči, ktorá nebola správne klasifikovaná ani v jednom prípade, čo naznačuje jej nízku výraznosť a zároveň možnú závislosť na jazykovo špecifických intonačných vlastnostiach.

Tieto výsledky poukazujú na dôležitosť použitia jazykovo relevantných dát pri tréningu modelov určených na rozpoznávanie emócií, ako aj na výzvy spojené s ich generalizáciou naprieč rôznymi jazykmi. Prehľad rozpoznaných emócií podľa úspešnosti je znázornený na Obrázku 3.



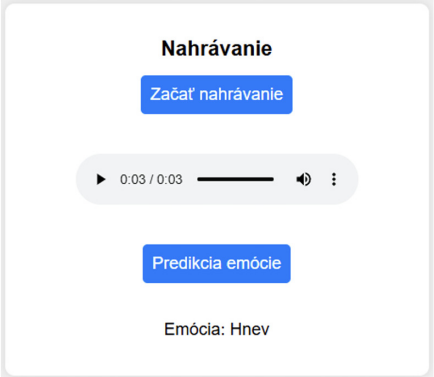
Obrázok 3 Počet klasifikácií jednotlivých emócií pre slovenské nahrávky

Po dokončení testovania rôznych datasetov a vlastných nahrávok pomocou Python skriptov bola vyvinutá webová aplikácia, ktorá umožňuje používateľovi nahrať zvukovú nahrávku prostredníctvom webového prehliadača, a to buď priamo cez mikrofón, alebo importom existujúceho súboru vo formáte `.wav`. Aplikácia následne automaticky vyhodnotí nahrávku a poskytne spätnú väzbu v podobe identifikovanej emócie (Obrázok 4).

IV. Záver

Po úspešnom uvedení aplikácie do prevádzky bol vytvorený krátky dotazník, ktorý bol sprístupnený náhodne vybraným respondentom testujúcim danú aplikáciu. Počas testovania mali možnosť nahrať súbor alebo zaznamenať hlasovú nahrávku.

Dotazník vyplnilo celkovo 11 respondentov, pričom každý z nich testoval osem emócií: *hnev*, *strach*, *prekvapenie*, *radosť*, *neutrálna reč*, *pokoj*, *smútok* a *znechutenie*. Najlepšie rozpoznateľnou emóciou bol jednoznačne *hnev*, ktorý aplikácia správne identifikovala u 82 % respondentov. Naopak, emócie ako *pokoj*, *prekvapenie* a *neutrálna reč* dosahovali najnižšiu úspešnosť rozpoznávania, a to približne 30 %. Emócie *radosť*, *smútok* a *znechutenie* boli rozpoznané so spoľahlivosťou



Obrázok 4 Používateľské rozhranie webovej aplikácie určené na automatické rozpoznávanie emócií z hlasových nahrávok

medzi 50 a 60 %, čo je vzhľadom na malú vzorku respondentov prijateľné. Celková priemerná úspešnosť rozpoznávania emócií pomocou webovej aplikácie dosiahla hodnotu 42,05 %.

Z pohľadu demografie bolo medzi respondentmi 9 mužov a 2 ženy vo vekovom rozpätí od 19 do 34 rokov. Účastníci sa prirodzene rozdelili do dvoch skupín. V prvej skupine boli respondenti, u ktorých aplikácia správne rozpoznala aspoň štyri emócie. V druhej skupine sa nachádzali tí, u ktorých aplikácia vyhodnocovala väčšinu emócií ako jednu konkrétnu, najčastejšie *hnev* alebo *smútok*. Tento jav mohol byť spôsobený špecifickým zafarbením hlasu, slabším prejavom emócií alebo kvalitou záznamového zariadenia. Ďalším možným faktorom je, že použitý model bol trénovaný na nahrávkach v inom jazyku, čo mohlo negatívne ovplyvniť presnosť rozpoznávania v slovenskom jazyku. Do budúcna plánujeme rozšíriť trénovaciu množinu o slovenské zvukové vzorky a optimalizovať model pre lepšiu adaptáciu na slovenský jazyk. Zároveň je cieľom zvýšiť presnosť rozpoznávania menej výrazných emócií.

Poďakovanie

Táto práca bola podporená Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže SR cez projekt KEGA 049TUKE-4/2024, a v rámci projektov podporených Agentúrou na podporu výskumu a vývoja APVV-22-0261 a APVV-22-0414.

Literatúra

[1] K. Mountzouris, I. Perikos, and I. Hatzilygeroudis, "Speech emotion recognition using convolutional neural networks with attention mechanism," *Electronics*, vol. 12, no. 20, p. 4376, 2023.

[2] S. N. Atkar, R. Agrawal, C. Dhule, N. C. Morris, P. Saraf, and K. Kalbande, "Speech emotion recognition using dialogue emotion decoder and cnn classifier," in *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. IEEE, 2023, pp. 94–99.

[3] N. Scheidwasser-Clow, M. Kegler, P. Beckmann, and M. Cernak, "Serab: A multi-lingual benchmark for speech emotion recognition," in *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2022, pp. 7697–7701.

[4] D. Niizumi, D. Takeuchi, Y. Ohishi, N. Harada, and K. Kashino, "Byol for audio: Self-supervised learning for general-purpose audio representation," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[5] H. Cao, D. G. Cooper, M. K. Keutmann, R. C. Gur, A. Nenkova, and R. Verma, "Crema-d: Crowd-sourced emotional multimodal actors dataset," *IEEE transactions on affective computing*, vol. 5, no. 4, pp. 377–390, 2014.

[6] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "Iemocap: Interactive emotional dyadic motion capture database," *Language resources and evaluation*, vol. 42, pp. 335–359, 2008.

[7] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english," *PloS one*, vol. 13, no. 5, p. e0196391, 2018.

[8] P. Jackson and S. Haq, "Surrey audio-visual expressed emotion (savee) database," *University of Surrey: Guildford, UK*, 2014.

[9] V. LoBue and C. Thrasher, "The child affective facial expression (cafe) set. databrary. 2014," *Series The Child Affective Facial Expression (CAFE) set. Databrary,(20142014)*, 2014.

[10] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, B. Weiss *et al.*, "A database of german emotional speech." in *Interspeech*, vol. 5, 2005, pp. 1517–1520.

[11] A. E. Society, "Journal of the audio engineering society," in *Proceedings of the Audio Engineering Society Convention*. Audio Engineering Society, 1957.

[12] G. Costantini, I. Iaderola, A. Paoloni, and M. Todisco, "Emovo corpus: an italian emotional speech database," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*. European Language Resources Association (ELRA), 2014.

[13] M. O. Nezami, P. J. Lou, and M. Karami, "Shemo: a large-scale validated database for persian speech emotion detection," *Language Resources and Evaluation*, 2019.

[14] J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, and A. de Mendonça, "Data mining methods in the prediction of dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests," *BMC research notes*, vol. 4, pp. 1–14, 2011.

Biometrická autentifikácia ako moderný prístup v internete vecí

[1]Eva KUPCOVÁ, [2]Matúš PLEVA, [3]Kristián SOPKOVIČ, [4]Tsimafei LAPANIK

[1,2,3,4]Katedra elektroniky a multimediálnych telekomunikácií,
Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach,
Letná 9, 04001 Košice, Slovenská republika

[1]eva.kupcova@tuke.sk, [2]matus.pleva@tuke.sk, [3]kristian.sopkovic@tuke.sk,
[4]tsimafei.lapanik@student.tuke.sk

Abstrakt **– Tento príspevok sa zaoberá vývojom a analýzou systému biometrickej autentifikácie pre ovládanie zariadení internetu vecí prostredníctvom Android aplikácie využívajúcej komunikáciu Bluetooth Low Energy. Cieľom je spoľahlivo overiť identitu používateľa pred vykonaním príkazov smerom k zariadeniu. Príspevok sa venuje prehľadu biometrických technológií, najmä autentifikácii pomocou odtlačku prsta, a taktiež špecifikám protokolu Bluetooth Low Energy vo vzťahu k ovládaniu jednoduchých zariadení internetu vecí. Výsledkom je funkčná Android aplikácia umožňujúca biometrické overenie používateľa a následné ovládanie zariadenia založeného na mikrokontroléri prostredníctvom Bluetooth. Testovanie ukázalo spoľahlivosť použitej biometrickej metódy a praktickú použiteľnosť riešenia v kontexte jednoduchých scenárov internetu vecí, ako sú inteligentné domáce systémy alebo vzdialené ovládanie zariadení.**

Kľúčové slová **– Android aplikácia, biometrická autentifikácia, Bluetooth Low Energy, internet vecí, ovládanie IoT zariadení**

I. Úvod

Internet vecí (z anglického *Internet of Things, IoT*) sa rýchlo rozširuje a čoraz viac ovplyvňuje priemysel, zdravotníctvo aj bežné domácnosti. S rastúcim počtom miliárd prepojených senzorov a mikrokontrolérov sa však zvyšuje aj riziko, pribúdajú incidenty spojené s únikom údajov alebo neoprávneným ovládnutím zariadení. V takomto prostredí už klasické prihlasovanie cez meno a heslo nestačí. Heslá sú často slabé, opakovane používané alebo pri menších IoT zariadeniach úplne chýbajú. Spoľahlivejšie riešenia preto využívajú viacero spôsobov overenia identity, tzv. viacfaktorovú autentifikáciu (z anglického *Multi-factor authentication, MFA*). Biometria, ako napríklad odtlačok prsta či rozpoznávanie tváre, pritom predstavuje zaujímavý kompromis medzi bezpečnosťou a používateľským komfortom [1].

Energeticky úsporný protokol Bluetooth (z anglického *Bluetooth Low Energy, BLE*) sa stal bežným riešením pre komunikáciu s IoT zariadeniami. Vďaka nízkej spotrebe energie a jednoduchej integrácii s Androidom aj iOS umožňuje dlhú výdrž batérie a pohodlné používanie. Jeho otvorená architektúra a obmedzená priepustnosť však sťažujú nasadenie zložitejších bezpečnostných mechanizmov. Navrhnutý prototyp využíva smartfón ako výpočtové centrum na okraji siete (z anglického *edge computing*), kde v rámci jedného cyklu prebieha biometrická autentifikácia, odvodenie kľúča a jeho následné spracovanie [2]. Periférne zariadenie, ktoré pozostáva zo 4WD podvozku postaveného na platforme Arduino, iba prijíma príkazy a vykonáva ich, čo zodpovedá jeho obmedzeným výpočtovým možnostiam. Tento prístup umožňuje dosiahnuť vysokú funkčnosť bez výrazného nárastu latencie či spotreby energie [3].

II. Metódy autentifikácie

Autentifikácia (z anglického *authentication*) je proces overenia identity používateľa alebo zariadenia pred udelením prístupu k systémom alebo službám. V kontexte moderných technológií zohráva kľúčovú úlohu pri zabezpečení prístupu k citlivým informáciám, najmä v oblasti internetu vecí, kde je potrebné zabezpečiť spoľahlivú komunikáciu medzi fyzickým svetom a digitálnymi platformami [4].

A. Tradičné autentifikačné prístupy

Medzi tradičné autentifikačné metódy patria najmä heslá, PIN kódy alebo odpovede na tajné otázky, teda tzv. vedomostné faktory. Tieto metódy sú jednoduché na implementáciu a dodnes sa široko používajú. Z hľadiska bezpečnosti však trpia viacerými nedostatkami – sú náchylné na phishing, zachytenie alebo uhádnutie, častokrát sa opakovane používajú v rôznych službách a často majú nízku entropiu [5].

Zlepšenie predstavujú metódy založené na niečom, čo používateľ vlastní, napríklad autentifikačné tokeny, čipové karty alebo mobilné zariadenia s generátormi jednorazových kódov. Ich hlavnou výhodou je, že útočník ich nemôže zneužiť bez fyzického prístupu. Nevýhodou je riziko straty alebo odcudzenia takéhoto zariadenia [5].

B. Moderné autentifikačné prístupy

Biometrická autentifikácia predstavuje čoraz bežnejší mechanizmus overovania identity v prostredí inteligentných technológií. Využíva fyziologické a behaviorálne charakteristiky používateľa, ako sú odtlačok prsta, rozpoznávanie tváre, dúhovky či hlasu. Tieto metódy znižujú závislosť od hesiel a fyzických tokenov, čím zvyšujú používateľský komfort. Ich nevýhodou je spracovanie nemenných biometrických údajov, ktoré po úniku nemožno zneplatniť ani nahradiť [6].

Ešte vyššiu úroveň zabezpečenia poskytuje viacfaktorová autentifikácia, ktorá kombinuje viacero nezávislých faktorov (napr. heslo a biometria, alebo token a rozpoznávanie tváre). V prostredí IoT sa MFA často realizuje pomocou mobilného zariadenia, ktoré slúži ako sprostredkovateľ autentifikácie, kombinujúc vlastnícky a biometrický faktor [7].

Významnú úlohu tu zohrávajú mobilné platformy, ktoré už obsahujú vstavané rozhrania na prácu s biometriou (napríklad Fingerprint API, Face ID). Ich kombinácia s bezdrôtovou technológiou otvára cestu k pohodlnej a bezpečnej autentifikácii pre používateľov inteligentných zariadení [8].

III. Význam autentifikácie v prostredí internetu vecí

Rýchly rozvoj internetu vecí mení spôsob, akým interagujeme s technológiami, počnúc inteligentnými domácnosťami, cez priemyselné riadiace systémy až po nositeľné zariadenia v zdravotníctve. Všetky tieto systémy generujú a spracúvajú citlivé údaje, a preto musia byť chránené pred neoprávneným prístupom. Jedným z kľúčových aspektov tejto ochrany je autentifikácia, teda proces overenia identity zariadení a používateľov. V kontexte internetu vecí však autentifikácia naráža na množstvo špecifických výziev, ktoré si vyžadujú netradičné a inovatívne prístupy [9].

A. Špecifiká IoT prostredia

IoT zariadenia často používajú mikrokontroléry ako ESP32, STM32 alebo ATmega328P, ktoré disponujú obmedzeným výpočtovým výkonom (napr. frekvencie pod 200 MHz), malou kapacitou pamäte (desiatky až stovky kB RAM) a sú často napájané batériami. Tieto hardvérové limity komplikujú implementáciu štandardných kryptografických protokolov, ako sú TLS 1.3 či asymetrické šifrovanie s dlhými kľúčmi [9].

Okrem technických obmedzení je IoT prostredie aj mimoriadne heterogénne. Zariadenia môžu komunikovať cez rôzne protokoly (napr. MQTT, CoAP, BLE), byť nasadené v náročných podmienkach (vlhkosť, prach, vibrácie), a často fungujú autonómne bez priameho ľudského zásahu. Všetky tieto faktory ovplyvňujú návrh autentifikačných mechanizmov, ktoré musia byť flexibilné, energeticky úsporné a odolné voči fyzickému kompromitovaniu [10].

B. Bezpečnostné hrozby pri IoT zariadeniach

Vzhľadom na to, že mnohé IoT zariadenia sú permanentne pripojené k internetu a často nedostatočne zabezpečené, stávajú sa atraktívnym cieľom rôznych typov útokov. Slabá alebo chýbajúca autentifikácia predstavuje kritickú zraniteľnosť, ktorá môže viesť k narušeniu dôvernosti, integrity aj dostupnosti služieb. Medzi najčastejšie bezpečnostné hrozby patria [11]:

- **útok hrubou silou** (z anglického *brute-force attack*) – systematické skúšanie kombinácií hesiel za účelom získania prístupu,
- **vydávanie sa za zariadenie** (z anglického *spoofing*) – útočník predstiera identitu legitímneho zariadenia alebo používateľa,
- **znovupoužitie autentifikačných údajov** (z anglického *replay attack*) – opakovanie predtým zachytených dát s cieľom neoprávneného prístupu,
- **útok zo stredu** (z anglického *man-in-the-middle attack*) – odpočúvanie alebo manipulácia komunikácie medzi dvoma stranami bez ich vedomia.

Tieto útoky môžu viesť k strate kontroly nad zariadením, narušeniu integrity dát alebo zneužitiu zariadenia ako vstupného bodu do väčšej siete.

C. Úloha autentifikácie

V prostredí internetu vecí autentifikácia nepredstavuje iba jednoduché zadanie mena a hesla. Zahŕňa širšie spektrum mechanizmov, ktoré zohľadňujú špecifiká tohto prostredia. Už počas počiatočného párovania zariadení dochádza k overovaniu identity, napríklad prostredníctvom výmeny kryptografických kľúčov. Dôležitá je aj vzájomná autentifikácia medzi jednotlivými komponentmi siete, ako je napríklad senzor a centrálna brána. Neoddeliteľnou súčasťou je tiež overenie identity používateľa, ktoré sa v praxi často realizuje prostredníctvom mobilnej aplikácie. Tá môže využívať biometrické údaje alebo jednorazové heslá (z anglického *One Time Password, OTP*), čím sa zabezpečuje spoľahlivý a zároveň používateľsky prijateľný spôsob autentifikácie [12].

S ohľadom na hardvérové obmedzenia IoT zariadení sa do popredia dostávajú tzv. ľahké autentifikačné protokoly, ktoré kladú minimálne nároky na výpočtový výkon, pamäť a spotrebu energie. Medzi takéto riešenia patrí napríklad protokol EAP-NOOB, ktorý je vhodný pre zariadenia bez predinicializovaných údajov, alebo protokoly založené na DTLS s využitím krátkych kľúčov založených na kryptografii eliptických kriviek (z anglického *Elliptic Curve Cryptography, ECC*) [13]. Perspektívne sú aj prístupy, ktoré využívajú fyzikálne vlastnosti samotného hardvéru ako jedinečný identifikátor zariadenia. Ide o tzv. fyzicky neklonovateľné funkcie (z anglického *Physically Unclonable Functions, PUF*), ktoré umožňujú bezpečnú a spoľahlivú identifikáciu bez potreby uloženia citlivých údajov do pamäte zariadenia [14].

D. Smerovanie k používateľsky prijateľným riešeniam

Spojenie IoT zariadení s mobilnými telefónmi otvára priestor pre efektívne a zároveň bezpečné autentifikačné mechanizmy. Využíva sa pritom tzv. *out-of-band* autentifikácia, pri ktorej sa overenie identity neodohráva priamo na IoT zariadení, ale na sprievodnom zariadení používateľa, napríklad na smartfóne. Typickým príkladom je párovanie cez Bluetooth, ktoré sa potvrdzuje biometrickým overením priamo v mobilnej aplikácii. Takýto prístup znižuje výpočtovú a energetickú záťaž na strane samotného IoT zariadenia a zároveň zvyšuje celkovú úroveň zabezpečenia [15].

Tento model autentifikácie nachádza uplatnenie v rôznych oblastiach. V prípade inteligentných zámkov môže používateľ otvoriť dvere jednoducho pomocou odtlačku prsta nasnímaného na svojom smartfóne. V zdravotníctve je zas prístup k citlivým údajom podmienený overením používateľa cez mobilnú aplikáciu, čím sa zabezpečí ochrana súkromia pacienta. V priemyselných systémoch sa autentifikácia kombinuje s viacfaktorovými metódami na strane administrátora, ktorý cez zabezpečené rozhranie riadi a konfiguruje jednotlivé prvky infraštruktúry [15].

Tieto príklady ukazujú, že autentifikácia nemusí byť prekážkou v používateľskej skúsenosti. Naopak, ak je navrhnutá s ohľadom na konkrétne prostredie a potreby, môže komfort a bezpečnosť efektívne skombinovať.

IV. Bluetooth Low Energy v kontexte IoT

Technológia Bluetooth Low Energy predstavuje moderný bezdrôtový komunikačný štandard, ktorý je špeciálne navrhnutý pre aplikácie s nízkou spotrebou energie a krátkym dosahom. V prostredí internetu vecí zohráva BLE kľúčovú úlohu pri zabezpečení efektívnej a energeticky úspornej komunikácie medzi zariadeniami [16].

A. Vývoj a technické špecifiká BLE

Bluetooth je bezdrôtová komunikačná technológia, ktorá využíva rádiové vlny v nelicencovanom ISM pásme (2,4 GHz). Prenos prebieha pomocou technológie FHSS (Frequency Hopping Spread Spectrum), ktorá znižuje pravdepodobnosť rušenia a zvyšuje spoľahlivosť komunikácie. Vďaka svojej univerzálnosti podporuje rôzne komunikačné topológie, ako napríklad bod–bod (z anglického *point-to-point*), vysielanie (z anglického *broadcast*) či sieťovú topológiu typu mriežka (z anglického *mesh*) [17].

Bluetooth štandard sa historicky rozdelil na dva režimy [18]:

- **Bluetooth BR/EDR (Basic Rate / Enhanced Data Rate)** – režim navrhnutý pre vyššie prenosové rýchlosti (do 3 Mb/s), používaný najmä v multimediálnych zariadeniach ako bezdrôtové slúchadlá či reproduktory. Tento režim však vykazuje vyššiu energetickú náročnosť.
- **Bluetooth Low Energy (BLE)** – optimalizovaný pre nízku spotrebu energie, nižšiu prenosovú rýchlosť a občasnú komunikáciu, vďaka čomu je vhodný pre IoT aplikácie ako senzory, nositeľné zariadenia či inteligentné zámky. BLE využíva 40 kanálov so šírkou 2 MHz.

Od svojho zavedenia v roku 1999 sa Bluetooth neustále vyvíja, pričom každá nová verzia priniesla zlepšenia v oblasti dosahu, spotreby energie, priepustnosti či bezpečnosti [19].

BLE používa architektúru klient-server, kde server (napr. IoT zariadenie) poskytuje údaje alebo služby a klient (napr. mobilná aplikácia) k nim pristupuje. Dáta sú organizované v rámci

atribútov a profilov, čo zjednodušuje ich štandardizované spracovanie. Komunikácia prebieha prostredníctvom všeobecného profilu atribútov (z anglického *Generic Attribute Profile, GATT*), ktorý definuje spôsob prístupu ku dátam cez tzv. charakteristiky [20].

Hoci BLE podporuje niekoľko režimov zabezpečenia, vrátane párovania a šifrovania, v praxi sa často stretávame s implementáciami, ktoré tieto možnosti nevyužívajú naplno. Vzhľadom na to, že BLE sa často používa na komunikáciu medzi citlivými komponentmi (napr. zariadenia ovládané cez aplikáciu), je dôležité doplniť komunikáciu o vhodné autentifikačné mechanizmy. V tomto kontexte sa mobilné zariadenia s biometrickým overením javia ako efektívny nástroj na posilnenie bezpečnosti.

V. Návrh a vývoj mobilnej aplikácie

Cieľom navrhnutej implemnetácie je realizácie projektu, ktorý zahŕňa ovládanie autíčka na báze Arduino prostredníctvom aplikácie pre Android. Projekt integruje biometrickú autentifikáciu na strane Android aplikácie, ktorá zabezpečuje, že iba autorizovaný používateľ môže ovládať zariadenie. Použitie BLE modulu HM-10 umožňuje efektívnu a energeticky nenáročnú komunikáciu.

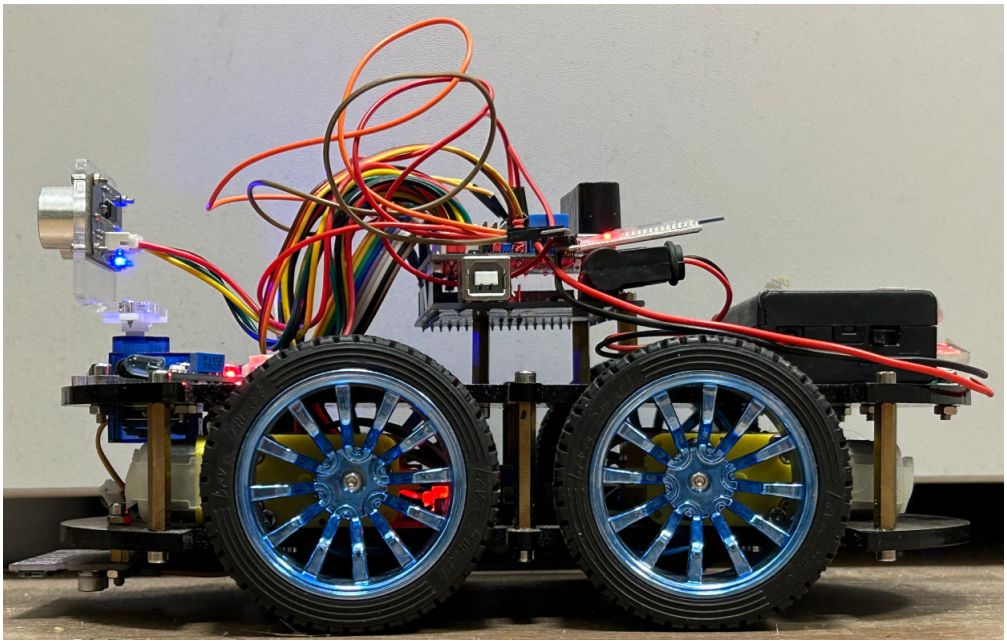
A. Technické komponenty a nástroje

V tejto kapitole podávame ucelený prehľad hardvérových a softvérových prostriedkov, ktoré boli zvolené pri návrhu a implementácii nášho IoT systému.

Hardvérová časť projektu zahŕňa kľúčové komponenty pre implementáciu IoT systému autentifikácie s dôrazom na efektívnosť a kompaktnosť. Použité zariadenia a ich konfigurácia sú nasledovné:

- **Arduino Mega 2560:** hlavná riadiaca jednotka projektu, ktorá poskytuje dostatok vstupno-výstupných pinov a výkon pre spracovanie dát zo senzorov a komunikáciu s mobilnou aplikáciou.
- **Horná prídavná doska:** rozširuje funkcionalitu systému. Do portov 5V a GND je pripojené napájanie, čím zabezpečuje stabilné napätie pre ďalšie moduly.
- **BLE modul HM-10:** pripojený k doske Arduino Mega 2560 cez porty TX14 a RX15 (Serial3). Umožňuje bezdrôtovú komunikáciu s nízkou spotrebou energie, ideálnu pre IoT zariadenia.
- **Modul L298N pre motory:** tento výkonný ovládač motorov riadi pohon kolies modelu vozidla. Podporuje dva motory s napätím 5–35 V a prúdom do 2 A na kanál, čo umožňuje presné ovládanie rýchlosti a smeru. Je pripojený k Arduino Mega 2560 cez piny s pulznou šírkovou moduláciou (z anglického *Pulse Width Modulation, PWM*) na reguláciu otáčok.
- **Batéria:** zabezpečuje napájanie všetkých komponentov, čím zaručuje mobilitu a kompaktnosť systému.

Komponenty sú vzájomne prepojené a optimalizované pre IoT aplikácie s dôrazom na nízku spotrebu a spoľahlivosť. Model autíčka je možné vidieť na Obrázku 1.



Obrázok 1 Model autíčka s hardvérovými komponentmi

Na vývoj a implementáciu projektu bolo použité nasledujúce softvérové vybavenie:

- **Android Studio:** integrované vývojové prostredie (z angl. *Integrated Development Environment, IDE*) pre tvorbu aplikácií pre operačný systém Android. V tomto prostredí bola vyvinutá mobilná aplikácia, ktorá umožňuje ovládanie zariadenia a implementáciu biometrickej autentifikácie.
- **Kotlin:** programovací jazyk používaný na vývoj mobilnej aplikácie. Kotlin ponúka moderný, jednoduchý a efektívny spôsob implementácie funkcionalít.
- **Arduino IDE:** prostredie na písanie a nahrávanie kódu do Arduino Mega 2560. Použité na implementáciu softvéru pre ovládanie motorov a spracovanie príkazov odoslaných z mobilnej aplikácie.
- **Knižnica BiometricPrompt:** knižnica na implementáciu biometrickej autentifikácie.

V rámci implementácie je BLE použité ako komunikačný kanál medzi Android aplikáciou a zariadením ovládaným pomocou mikrokontroléra. Vďaka krátkemu času nadviazania spojenia a nízkej energetickej náročnosti je BLE ideálne pre ovládanie jednoduchých IoT prvkov v reálnom čase. Prenosové možnosti BLE zároveň postačujú na odosielanie príkazov alebo statusových správ medzi aplikáciou a zariadením, čím zabezpečujú stabilnú interakciu v rámci systému.

VI. Realizácia projektu

Táto kapitola popisuje praktickú realizáciu systému na diaľkové ovládanie modelu vozidla (autíčka) pomocou platformy Arduino a mobilnej aplikácie pre operačný systém Android. Návrh a implementácia systému nadväzujú na teoretické poznatky v oblastiach autentifikácie, bezpečnosti IoT zariadení a bezdrôtovej komunikácie.

A. Základná architektúra systému

Navrhnutý systém sa skladá z dvoch hlavných častí – mobilnej aplikácie a riadiacej jednotky postavenej na mikrokontroléri Arduino. Aplikácia zabezpečuje ovládanie autíčka prostredníctvom používateľského rozhrania a zároveň realizuje autentifikáciu používateľa pred samotným spustením komunikácie.

Autentifikácia prebieha prostredníctvom biometrických údajov (napr. odtlačku prsta), využívajúc vstavané mechanizmy platformy Android. Po úspešnej autentifikácii sa používateľovi sprístupní možnosť nadviazať komunikáciu s Arduino zariadením.

Prenos príkazov medzi aplikáciou a Arduinom je zabezpečený prostredníctvom modulu BLE, konkrétne HM-10. Momentálne nie je implementované šifrovanie prenášaných dát, avšak táto funkcionalita je plánovaná ako budúca práca na zvýšenie bezpečnosti komunikácie.

B. Použité technológie a ich výber

Arduino bolo zvolené ako hlavná riadiaca jednotka z dôvodu jeho jednoduchosti, modulárnosti a rozsiahlej podpory zo strany komunity. Umožňuje rýchlu prototypizáciu a prácu s rôznymi senzorickými a komunikačnými modulmi, čo ho robí ideálnym pre IoT aplikácie.

Android predstavuje vhodnú platformu pre vývoj mobilnej aplikácie, keďže poskytuje natívnu podporu biometrickej autentifikácie, správu BLE spojení a vytváranie bezpečného používateľského rozhrania.

BLE bol zvolený ako komunikačné médium vzhľadom na jeho nízku spotrebu energie, podporu v mobilných zariadeniach a spoľahlivosť v krátkom dosahu. Použitý BLE modul HM-10 podporuje sériovú komunikáciu medzi Arduinom a smartfónom, pričom dosahuje dostatočnú rýchlosť a stabilitu prenosu pre potreby riadenia modelu.

VII. Vizualizácia a testovanie aplikácie

Všetky komponenty systému vrátane softvéru boli testované na zariadení Samsung Galaxy A53 5G, ktoré podporuje moderné technológie ako BLE a biometrickú autentifikáciu. Toto zariadenie predstavuje vhodnú platformu na overenie funkčnosti vzhľadom na dostupnosť vstavaných senzorov a schopnosť spracovania BLE komunikácie.

Testovanie spočívalo najmä v overení správneho pripojenia BLE modulu HM-10 k mikrokontroléru Arduino prostredníctvom sériového portu Serial3 (pin TX14 pre vysielanie a RX15 pre príjem). Bola preverovaná stabilita a spoľahlivosť tohto hardvérového prepojenia v reálnych podmienkach.

Funkčné testy prebiehali v interiéri, kde bolo najmä dôležité sledovať priebeh biometrickej autentifikácie používateľa realizovanej na mobilnom zariadení. Ďalej sa overovala schopnosť systému prijímať a spracovať ovládacie príkazy odoslané cez BLE spojenie a následne riadiť pohyb autíčka.

Mobilná aplikácia úspešne identifikovala používateľa prostredníctvom biometrických údajov a následne nadviazala komunikáciu s BLE modulom. Na ovládanie autíčka boli použité jednoduché ovládacie prvky, ktoré umožnili intuitívne riadenie pohybu zariadenia v súlade s požiadavkami používateľa. Na Obrázku 2 je zobrazené používateľské rozhranie aplikácie pre ovládanie autíčka.

Obrázok 2 Používateľské rozhranie mobilnej aplikácie na ovládanie Arduino autíčka

Na účely overenia funkčnosti a praktickej použiteľnosti navrhnutého systému boli vykonané základné merania vybraných parametrov, ktoré sú kľúčové pre bezproblémový chod a efektívnu komunikáciu medzi mobilnou aplikáciou a riadiacou jednotkou Arduino. Merania sa zamerali najmä na aspekty súvisiace s bezdrôtovou komunikáciou prostredníctvom BLE modulu, energetickou náročnosťou systému, odozvou riadenia a časom potrebným na autentifikáciu používateľa.

V interiérových podmienkach bol nameraný dosah BLE spojenia 7 až 10 metrov pri priamom výhľade. Čas odozvy na príkaz je približne 100 ms, čo umožňuje plynulé ovládanie autíčka. Napájanie zabezpečuje 7,4 V lítium-iónová batéria, pričom spotreba BLE modulu v aktívnom režime je približne 8 mA. PWM frekvencia motora je nastavená na 500 Hz, čo zabezpečuje presnú reguláciu rýchlosti. Výdrž batérie pri bežnej záťaži je približne 40 minút. Biometrická autentifikácia trvá 300 až 500 ms.

Podrobný prehľad meraní je uvedený v Tabuľke 1. Výsledky potvrdzujú, že systém dosahuje spoľahlivú komunikáciu a primeranú odozvu, čím spĺňa požiadavky praktického použitia v rámci IoT aplikácií s biometrickou autentifikáciou.

Parameter	**Jednotka**	**Hodnota**	**Poznámka**
Dosah BLE (interiérový)	metrov	7–10	Merané v miestnosti pri priamom výhľade
Latencia príkazu (čas odozvy)	ms	∼100	Zmerané medzi odoslaním príkazu a pohybom
Napájacie napätie (batéria)	V	7,4	2S Li-ion batéria
Prúdový odber BLE modulu v aktívnom režime	mA	∼8	Podľa dokumentácie modulu HM-10
Frekvencia PWM signálu pre motorové ovládanie	Hz	500	Nastavené v Arduino kóde
Prevádzková doba batérie pri strednej záťaži	minút	∼40	Merané pri bežnej prevádzke
Časový interval biometrickej autentifikácie	ms	300–500	Trvanie snímania odtlačku a overenia v aplikácii

Tabuľka 1 Vybrané merateľné parametre systému riadenia Arduino autíčka cez BLE

VIII. Záver

V tomto článku sme predstavili funkčný prototyp mobilnej aplikácie, ktorá umožňuje biometrickú autentifikáciu a ovládanie IoT zariadenia cez BLE. Aplikácia využíva zabudované senzory mobilného telefónu na overenie identity používateľa, čo eliminuje potrebu zadávania hesiel alebo používania ďalších zariadení. Po úspešnej autentifikácii je možné priamo komunikovať s periférnym zariadením a jednoducho ho ovládať.

Testovanie potvrdilo spoľahlivú funkčnosť systému v základných situáciách. Biometria identifikovala používateľa, BLE spojenie bolo stabilné a príkazy sa správne prenášali a vykonávali. Aj keď neboli vykonané detailné merania latencie ani testy výdrže, prototyp je vhodný na použitie v inteligentnej domácnosti alebo vo vzdelávacích aplikáciách.

Do budúcna plánujeme rozšíriť systém o ďalšie biometrické metódy, ako napríklad overovanie hlasu alebo analýzu dotykových pohybov, ktoré môžu zlepšiť bezpečnosť a používateľský komfort. Ďalej chceme implementovať šifrovanie komunikácie medzi aplikáciou a zariadením pomocou AES, aby sme lepšie zabezpečili prenášané dáta. Plánujeme tiež vykonať rozsiahlejšie testy na presnejšie vyhodnotenie odozvy a spoľahlivosti systému v rôznych prevádzkových podmienkach. Výsledky naznačujú, že jednoduchá biometrická autentifikácia môže efektívne zjednodušiť prístup k nízkoenergetickým IoT zariadeniam bez výrazného ovplyvnenia používateľského zážitku.

Poďakovanie

Táto práca bola podporená Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže SR cez projekt KEGA 049TUKE-4/2024, a v rámci projektov podporených Agentúrou na podporu výskumu a vývoja APVV-22-0261 a APVV-22-0414.

Literatúra

[1] C.-W. Lien and S. Vhaduri, "Challenges and opportunities of biometric user authentication in the age of iot: A survey," *ACM Computing Surveys*, vol. 56, no. 1, pp. 1–37, 2023.
[2] J. Fürst, K. Chen, H.-S. Kim, and P. Bonnet, "Evaluating bluetooth low energy for iot," in *2018 IEEE Workshop on Benchmarking Cyber-Physical Networks and Systems (CPSBench)*. IEEE, 2018, pp. 1–6.
[3] W. Yang, S. Wang, N. M. Sahri, N. M. Karie, M. Ahmed, and C. Valli, "Biometrics for internet-of-things security: A review," *Sensors*, vol. 21, no. 18, p. 6163, 2021.
[4] I. Velásquez, A. Caro, and A. Rodríguez, "Authentication schemes and methods: A systematic literature review," *Information and Software Technology*, vol. 94, pp. 30–37, 2018.
[5] M. H. Barkadehi, M. Nilashi, O. Ibrahim, A. Z. Fardi, and S. Samad, "Authentication systems: A literature review and classification," *Telematics and Informatics*, vol. 35, no. 5, pp. 1491–1511, 2018.
[6] S. Alwahaishi and J. Zdrálek, "Biometric authentication security: an overview," in *2020 IEEE international conference on cloud computing in emerging markets (CCEM)*. IEEE, 2020, pp. 87–91.
[7] A. Ometov, S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen, and Y. Koucheryavy, "Multi-factor authentication: A survey," *Cryptography*, vol. 2, no. 1, p. 1, 2018.
[8] A. Rattani and R. Derakhshani, "A survey of mobile face biometrics," *Computers & Electrical Engineering*, vol. 72, pp. 39–52, 2018.
[9] B. B. Gupta and M. Quamara, "An overview of internet of things (iot): Architectural aspects, challenges, and protocols," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 21, p. e4946, 2020.
[10] D. Silva, L. I. Carvalho, J. Soares, and R. C. Sofia, "A performance analysis of internet of things networking protocols: Evaluating mqtt, coap, opc ua," *Applied Sciences*, vol. 11, no. 11, p. 4879, 2021.
[11] G. Rajendran, R. R. Nivash, P. P. Parthy, and S. Balamurugan, "Modern security threats in the internet of things (iot): Attacks and countermeasures," in *2019 International Carnahan Conference on Security Technology (ICCST)*. IEEE, 2019, pp. 1–6.
[12] R. M. J. Shantha, K. Mahender, A. J. M. Jenifer, and A. Prasanth, "Security analysis of hybrid one time password generation algorithm for iot data," in *AIP Conference Proceedings*, vol. 2418, no. 1. AIP Publishing LLC, 2022, p. 030021.
[13] A. Srivastava and A. Kumar, "A review on authentication protocol and ecc in iot," in *2021 International conference on advance computing and innovative technologies in engineering (ICACITE)*. IEEE, 2021, pp. 312–319.
[14] A. Yadav, S. Kumar, and J. Singh, "A review of physical unclonable functions (pufs) and its applications in iot environment," *Ambient Communications and Computer Systems: Proceedings of RACCCS 2021*, pp. 1–13, 2022.
[15] A. Barua, M. A. Al Alamin, M. S. Hossain, and E. Hossain, "Security and privacy threats for bluetooth low energy in iot and wearable devices: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 251–281, 2022.
[16] G. Koulouras, S. Katsoulis, and F. Zantalis, "Evolution of bluetooth technology: Ble in the iot ecosystem," *Sensors (Basel, Switzerland)*, vol. 25, no. 4, p. 996, 2025.
[17] R. Rondón, A. Mahmood, S. Grimaldi, and M. Gidlund, "Understanding the performance of bluetooth mesh: Reliability, delay, and scalability analysis," *IEEE Internet of things journal*, vol. 7, no. 3, pp. 2089–2101, 2019.
[18] R. Katila, T. N. Gia, and T. Westerlund, "Analysis of mobility support approaches for edge-based iot systems using high data rate bluetooth low energy 5," *Computer Networks*, vol. 209, p. 108925, 2022.
[19] H. Bhalla, O. Haggai, H. Bhalla, and O. Haggai, "Bluetooth overview," *Unraveling Bluetooth LE Audio: Stretching the Limits of Interoperable Wireless Audio with Bluetooth Next-Generation Low Energy Audio Standards*, pp. 29–53, 2021.
[20] C. Hirsch, L. Davoli, R. Grosu, and G. Ferrari, "Dyngatt: A dynamic gatt-based data synchronization protocol for ble networks," *Computer Networks*, vol. 222, p. 109560, 2023.

Senzorový uzol s TCP/IP konektivitou a vzdialenou aktualizáciou firmvéru

[1]*Tibor OLEARNÍK,* [2]*Miloš DRUTAROVSKÝ*

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

[1]tibor.olearnik@student.tuke.sk, [2]milos.drutarovsky@tuke.sk

Abstrakt – **Článok predstavuje návrh a implementáciu senzorového uzla s internetovou konektivitou, založeného na mikrokontroléri RP2040 a sieťovom čipe W5100S. Navrhnuté riešenie umožňuje bezkontaktné meranie teploty pomocou infračerveného senzora MLX90632 a poskytuje údaje používateľovi prostredníctvom vstavaného HTTP servera. Webové rozhranie je dynamicky generované vo firmvéri a zobrazuje aktuálne aj historické merania. Kľúčovou vlastnosťou systému je vlastný trojstupňový zavádzač (bootloader) s podporou vzdialenej aktualizácie firmvéru cez TFTP protokol. Softvér využíva dvojjadrovú architektúru RP2040, kde jedno jadro obsluhuje meranie a druhé sieťovú komunikáciu, čím je zabezpečená stabilná a paralelná prevádzka. Experimentálne testovanie preukázalo stabilitu webového servera, merania a zavádzacieho mechanizmu. Navrhnuté riešenie je vhodné pre jednoduché IoT aplikácie s dôrazom na autonómiu, nízku spotrebu a možnosť vzdialenej správy.**

Kľúčové slová – **RP2040, W5100S, MLX90632, HTTP server, TFTP, senzorový uzol, IoT, bootloader, vzdialená aktualizácia, TCP/IP**

I. Úvod

Internet vecí (IoT) čoraz viac vyžaduje zariadenia, ktoré sú schopné spoľahlivo snímať fyzikálne veličiny, komunikovať cez sieťové protokoly a byť jednoducho spravované aj na diaľku. Takéto zariadenia musia byť kompaktné a zároveň energeticky úsporné a schopné autonómnej prevádzky bez nutnosti asistencie používateľa alebo prepojenia s centrálnym serverom. Tento článok sa zameriava na návrh a realizáciu senzorového uzla, ktorý je založený na dvojjadrovom mikrokontroléri RP2040 [1] a sieťovom čipe W5100S [2]. Cieľom bolo vytvoriť zariadenie schopné merať teplotu bezkontaktne pomocou senzora MLX90632 [3], poskytovania údajov prostredníctvom vstavaného HTTP servera a podpory vzdialenej aktualizácie firmvéru pomocou protokolu TFTP. Navrhnutý systém je rozdelený na dve samostatne fungujúce jadrá, jedno vykonáva periodické meranie teploty a generovanie obsahu webovej stránky, druhé obsluhuje sieťové požiadavky. Webové rozhranie zobrazuje aktuálnu teplotu aj časovo označenú históriu meraní. Na zabezpečenie bezúdržbovej prevádzky bol do systému integrovaný trojstupňový zavádzač (bootloader), ktorý umožňuje aktualizáciu firmvéru bez fyzického prístupu k zariadeniu. V článku sú podrobne opísané hardvérové a softvérové komponenty vytvoreného systému, architektúra HTTP servera, návrh zavádzača, ako aj integrácia infračerveného senzora MLX90632. Súčasťou je aj vyhodnotenie funkčnosti prostredníctvom experimentálneho testovania, ktoré potvrdilo spoľahlivosť a vhodnosť systému pre jednoduché IoT aplikácie.

II. Architektúra senzorového uzla

Navrhnutý senzorový uzol predstavuje kompaktné a plne autonómne zariadenie, ktoré prepája bezkontaktné meranie fyzikálnej veličiny, konkrétne teploty, s možnosťou jej vzdialeného monitorovania cez internet. Architektúra systému bola od začiatku navrhnutá s dôrazom na jednoduchosť, spoľahlivosť a rozšíriteľnosť a zároveň podporu diaľkovej aktualizácie firmvéru. Celý systém možno rozdeliť na dve hlavné časti: hardvérovú platformu a softvérovú implementáciu. Ich vzájomná integrácia umožnila vytvoriť riešenie, ktoré je nielen funkčné, ale aj vhodné pre nasadenie do praxe v rámci menších IoT sietí alebo ako samostatné meracie zariadenie.

A. Hardverová platforma

Základom systému je vývojová doska W5100S-EVB-Pico[1], ktorá integruje dva hlavné komponenty:

RP2040[1] je dvojjadrový 32-bitový mikrokontrolér navrhnutý Raspberry Pi Foundation, založený na jadrách ARM Cortex-M0+[2] taktovaných až na 133 MHz. Okrem výkonného procesora ponúka aj 264 kB SRAM a možnosť pripojenia externej SPI Flash pamäte (v prípade dosky W5100S-EVB-Pico má veľkosť 2 MB). Vďaka tomu je schopný spracovávať aj zložitejšie úlohy, ako je napríklad lokálne spracovanie meraní, jednoduché sieťové služby alebo riadenie viacerých periférií súčasne. RP2040 disponuje sadou periférií vrátane dvoch I2C, dvoch SPI, dvoch UART rozhraní, 16 PWM kanálov, ADC a PIO (programmable I/O) blokov, ktoré umožňujú implementáciu vlastných digitálnych protokolov bez zaťaženia hlavného procesora. Vzhľadom na podporu vývoja v C/C++ pomocou oficiálneho Pico SDK, ako aj v MicroPythone, je platforma flexibilná a dobre dokumentovaná, čo výrazne uľahčuje vývoj aplikácií.

W5100S[2] je integrovaný sieťový čip od spoločnosti WIZnet, ktorý poskytuje hardvérovú implementáciu TCP/IP protokolového zásobníka ako napr. TCP spojenia, UDP komunikácie, ARP, ICMP, tie sú vykonávané nezávisle od hlavného procesora, čím sa šetria výpočtové zdroje RP2040. W5100S podporuje štyri simultánne sieťové sockety a obsahuje vlastný vyrovnávací buffer pre príjem a odosielanie dát (až 32 kB), čo umožňuje plynulú komunikáciu bez častého prerušenia CPU. Komunikácia s mikrokontrolérom prebieha cez SPI rozhranie s podporou vysokých prenosových rýchlostí, čip je taktiež vybavený režimom „direct addressing“ pre jednoduchšie mapovanie pamäte.

Komunikácia medzi mikrokontrolérom a sieťovým čipom prebieha prostredníctvom SPI rozhrania, ktoré je plne podporované cez SDK oboch výrobcov. Sieťové rozhranie je ukončené štandardným RJ45 konektorom. Zariadenie je vybavené infračerveným teplotným senzorom MLX90632, ktorý je schopný presne merať teplotu bez fyzického kontaktu s objektom. Tento senzor využíva princíp snímania infračerveného žiarenia, má zabudovanú kalibráciu v EEPROM pamäti a komunikuje cez zbernicu I2C. Vďaka tomu je jeho integrácia jednoduchá a nevyžaduje zložitú konfiguráciu. Pre diagnostiku, ladenie a vývoj je systém doplnený o USB-UART prevodník FTDI TLS-232R-3V3[3], ktorý umožňuje zobrazovať výstupy z mikrokontroléra, prípadne odosielať príkazy počas behu programu. Týmto spôsobom je možné analyzovať a ladiť systém bez zásahu do sieťovej komunikácie.

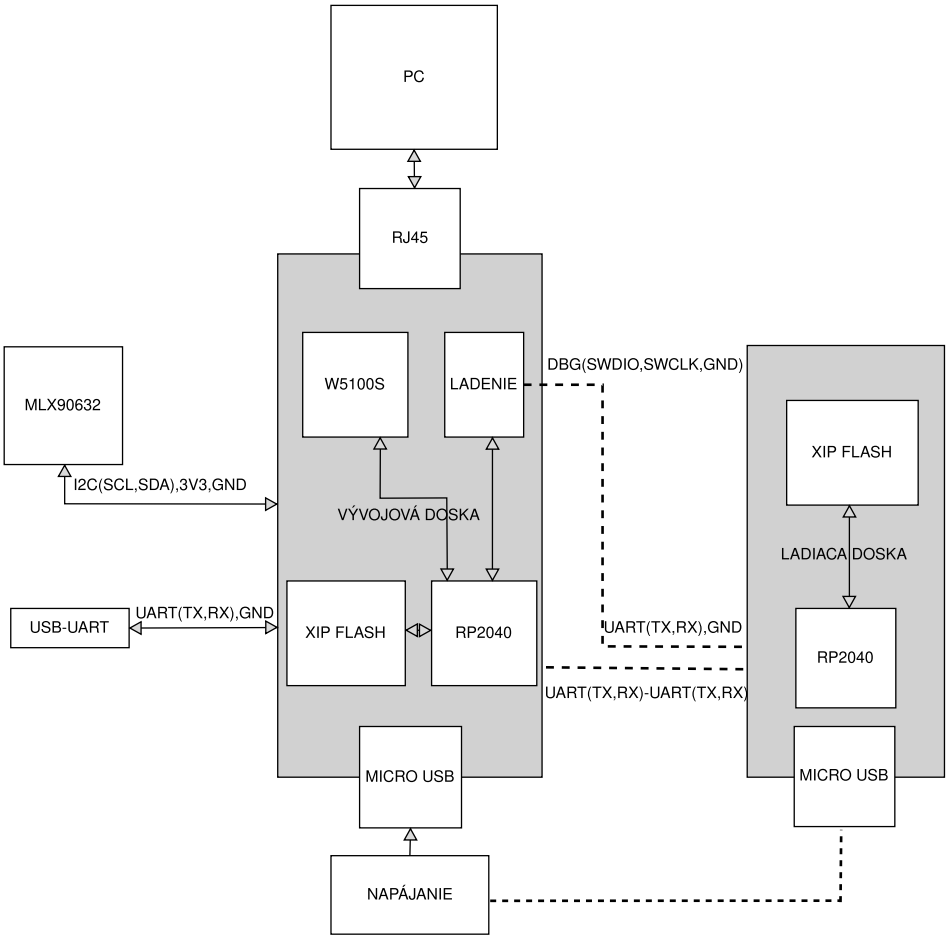
B. Softvérová implementácia

Softvér zariadenia je rozdelený do niekoľkých vrstiev s jasne definovanou hierarchiou a využíva výhody dvojjadrovej architektúry mikrokontroléra RP2040 [1]. Firmvér obsluhuje sieťovú komunikáciu a HTTP server a vykonáva periodické čítanie dát zo senzora a ich spracovanie. Firmvér je napísaný v jazyku C s použitím oficiálneho Pico SDK[4], ktoré poskytuje knižnice pre prácu s GPIO, I2C, SPI, UART, watchdogom, viacjadrovým spracovaním a Flash pamäťou. Firmvér ďalej využíva knižnice Wiznet SDK na prácu s TCP/IP socketmi, implementáciu HTTP servera a klienta protokolu TFTP. Kompletná implementácia senzorového uzla vrátane zdrojových kódov a dokumentácie je dostupná v GitHub repozitári [4]. Súčasťou firmvéru je aj vlastný bootovací mechanizmus, ktorý umožňuje vzdialené nahratie nového firmvéru cez TFTP. Tento zavádzač beží vo vymedzenej časti Flash pamäte, čím je oddelený od aplikačného kódu. Po prijatí firmvéru sa vykoná validácia jeho hlavičky, kontrola vstupného bodu a bezpečné prepnutie do novej verzie softvéru. Dôležitou súčasťou architektúry je aj watchdog, ktorý chráni zariadenie pred zacyklením v prípade chyby. Ak počas definovaného časového intervalu nedôjde k obnoveniu jeho časovača, systém sa automaticky reštartuje, čím sa minimalizuje riziko výpadku. Na účely laditeľnosti a prehľadnosti bol softvér navrhnutý ako modulárny, pričom jednotlivé komponenty (senzor, sieť, TFTP, HTML výstup) sú implementované ako samostatné bloky, ktoré možno ľahko upraviť alebo nahradiť. Implementácia zavádzača so zdrojovými kódmi a dokumentáciou je dostupná v GitHub repozitári [5]. Detailný opis implementácie je k dispozícii v práci [6].

[1] https://docs.wiznet.io/Product/iEthernet/W5100S/w5100s-evb-pico
[2] https://developer.arm.com/Processors/Cortex-M0-Plus
[3] https://ftdichip.com/wp-content/uploads/2023/07/DS_TTL-232R_CABLES.pdf
[4] https://github.com/raspberrypi/pico-sdk

C. Bloková schéma senzorového uzla

Celkovú architektúru vyvinutého systému ilustruje bloková schéma znázornená na obrázku 1. Táto schéma zachytáva prepojenie všetkých hardvérových komponentov, ktoré sú súčasťou funkčného senzorového uzla, ako aj navrhnutý a využitý spôsob jeho ladenia a napájania.



Obr. 1: Bloková schéma vytvoreného senzorového uzla s internetovou konektivitou a možnosťou ladenia

Na hlavnej vývojovej doske sa nachádza mikrokontrolér RP2040, ktorý zabezpečuje všetky výpočtové a riadiace funkcie systému. K nemu je pripojený infračervený senzor MLX90632 [3] prostredníctvom I2C zbernice (vedené ako SDA, SCL, GND a napájanie 3,3 V). Tento senzor je zodpovedný za meranie teploty objektov aj okolitého prostredia. Zároveň je RP2040 [1] prepojený na doske so sieťovým čipom W5100S [2] pomocou SPI rozhrania, ktorý komunikuje cez Ethernet rozhranie (RJ45). Na napájanie zariadenia a jeho základnú komunikáciu sa používa Micro USB konektor, ktorý zabezpečuje aj možnosť aktualizácie firmvéru prostredníctvom BOOTSEL režimu. Pre ladenie a vývoj bol použitý dvojdoskový systém, kde druhá doska RP2040 [1] (označená ako „Ladiaca doska“) vystupuje ako debugger (Picoprobe[5]). Tá je cez rozhranie SWD (Serial Wire Debug) prepojená s cieľovou doskou a zároveň komunikuje s počítačom cez vlastné USB rozhranie. Toto prepojenie umožňuje krokové ladenie, čítanie výstupov pomocou GNU Debuggera[6] alebo iných nástrojov. Diagnostika systému prebieha cez UART, pričom výstupy z vývojovej dosky (TX, RX, GND) sú vedené cez USB-UART prevodník, ktorý zabezpečuje zobrazenie údajov v sériovom termináli na PC. V prípade ladenia sú výstupy zo sériového rozhrania vývojovej dosky prepojené s UART vstupom ladiacej dosky (tzv. spätná väzba UART-TX/RX). Výsledkom takto navrhnutej topológie je možnosť súčasného napájania, ladenia, debugovania aj zobrazovania výstupov bez nutnosti fyzického zásahu do zapojenia. Táto konfigurácia výrazne uľahčuje vývoj a testovanie firmvéru, najmä pri implementácii pokročilých funkcií ako je zavádzač či TFTP klient.

[5]https://www.raspberrypi.com/documentation/microcontrollers/debug-probe.html
[6]https://sourceware.org/gdb/documentation/

III. Dvojjadrová architektúra mikrokontroléra RP2040

Mikrokontrolér RP2040 [1] obsahuje dve nezávislé výpočtové jadrá, ktoré pracujú na takte až 133 MHz. Táto architektúra umožňuje výrazne zlepšiť paralelizmus a výpočtovú efektivitu vstavaných (embedded) aplikácií, najmä pri časovo náročných alebo oddeliteľných úlohách, ako sú spracovanie senzorových údajov a obsluha siete.

Nezávislé jadrá – Každé jadro má vlastný programový čítač (PC), zásobník a registrami riadené kontexty. Zdieľajú však rovnaký pamäťový priestor (SRAM aj Flash) a periférie. Spustenie druhého jadra (core 1) sa vykonáva volaním funkcie `multicore_launch_core1()` zo základného SDK, pričom jadro 0 (core 0) zostáva aktívne od zapnutia systému.

Zdieľaná pamäť – Obe jadrá majú prístup k spoločnej RAM, čo umožňuje výmenu dát, ale zároveň vyžaduje ochranu pred súbežným prístupom. V navrhnutom systéme sa synchronizácia rieši pomocou krátkodobého vypnutia prerušení (`irq_set_enabled()`), ktoré slúži ako jednoduchá mutual-exclusion technika bez použitia mutexov.

Rozdelenie úloh vo firmvéri – V navrhovanom senzorovom uzle bola dvojjadrová architektúra využitá nasledovne:

- Jadro 0 (core 0) – riadi sieťový stack, HTTP server a spracovanie požiadaviek klienta.
- Jadro 1 (core 1) – periodicky číta údaje zo senzora MLX90632, spracúva ich, ukladá do cyklickej vyrovnávacej pamäte a generuje obsah HTML stránky.

Týmto rozdelením sa eliminovalo blokovanie činností a teda čítanie zo senzora a príprava výstupu nezasahujú do obsluhy HTTP požiadaviek. Systém je tak schopný poskytovať webovú stránku okamžite, bez oneskorení spôsobených meraním alebo výpočtami.

Komunikácia medzi jadrami – Pre prenos signálov medzi jadrami poskytujú RP2040 FIFO fronty (`multicore_fifo_push_blocking()`, `multicore_fifo_pop_blocking()`), ktoré sú vhodné na jednoduchú výmenu dát alebo synchronizáciu stavov. V tomto projekte však bola výmena údajov realizovaná cez zdieľané globálne premenné s minimálnou potrebou synchronizácie vďaka jednoduchej architektúre systému.

Výhody a riziká použitia dvoch jadier – Použitie viacjadrovej architektúry umožňuje výrazne zvýšiť stabilitu firmvéru a oddeliť výpočtovo náročné úlohy od úloh očakávajúcich rýchlu odozvu. Na druhej strane vyžaduje dôsledné plánovanie prístupu do spoločnej pamäte a správu synchronizácie, aby sa predišlo nekonzistentným stavom alebo prepisom dát.

IV. Infračervený teplotný senzor MLX90632

Meranie teploty bez fyzického kontaktu je v mnohých aplikáciách nielen praktické, ale často aj nevyhnutné, najmä v prípadoch, keď je potrebná vysoká citlivosť, izolácia meraného objektu, alebo rýchla odozva. V tomto projekte bol použitý digitálny infračervený senzor MLX90632 [3], ktorý poskytuje presné merania v kompaktnom formáte a s veľmi nízkou spotrebou.

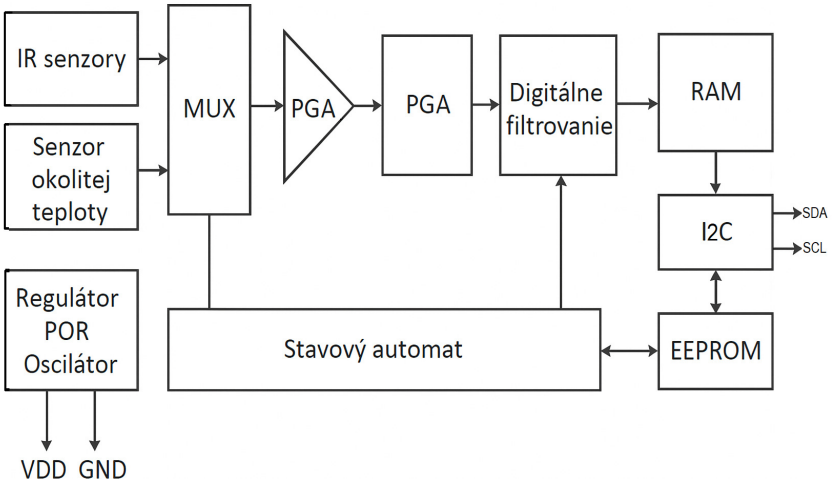
Základné vlastnosti senzora – MLX90632 [3] je mikrosystém od spoločnosti Melexis, určený pre bezdotykové meranie teploty pomocou detekcie infračerveného žiarenia. Ide o Far Infrared (FIR) senzor, ktorý je z výroby kalibrovaný a určený pre aplikácie, kde je potrebná vysoká presnosť pri zachovaní malých rozmerov a minimálnej energetickej náročnosti. Senzor má rozmery 3×3 mm a je dodávaný v puzdre QFN. Vyznačuje sa veľmi nízkym odberom v aktívnom režime (typicky 1 mA) a extrémne nízkym spánkovým prúdom ($< 2{,}5\,\mu$A), čo z neho robí ideálne riešenie pre mobilné alebo batériovo napájané zariadenia.

Merací princíp a interná architektúra – Senzor obsahuje dva detekčné prvky, jeden pre teplotu objektu a druhý pre teplotu vlastného puzdra. Kombináciou týchto údajov a použitím integrovaného matematického modelu je možné kompenzovať vplyv okolia a získať presnú hodnotu povrchovej teploty.

Meranie prebieha v niekoľkých krokoch:

- zachytenie infračerveného žiarenia pomocou mikrotermopílového detektora (thermopile),
- zosilnenie signálu pomocou programovateľného zosilňovača (PGA),
- digitalizácia signálu prostredníctvom interného ADC,
- výstup vo forme digitalizovaných hodnôt prostredníctvom I2C rozhrania.

Súčasťou senzora je aj EEPROM pamäť, v ktorej sú uložené kalibračné konštanty individuálne nastavené pre každý kus počas výroby. Tieto konštanty sú potrebné na presný výpočet výslednej teploty a používajú sa vo výpočtoch firmvéru na strane mikrokontroléra. Vnútorná architektúra senzora je zobrazená na obrázku 2.



Obr. 2: Bloková schéma teplotného senzora MLX90632

Komunikačné rozhranie a integrácia – Senzor komunikuje pomocou štandardného I2C protokolu so 7-bitovou predvolenou adresou 0x3A. Okrem dátových vodičov SDA a SCL je prítomný aj pin ADDR, ktorý umožňuje zmeniť adresu senzora pre prípady, kde sa používa viacero zariadení na jednej zbernici. Pre tento projekt bola použitá oficiálna knižnica Melexis dostupná v jazyku C. Táto knižnica[7] využíva tzv. hardware abstraction layer, v ktorej si vývojár sám implementuje I2C operácie podľa použitej platformy. V prípade RP2040 [1] boli tieto funkcie priamo naviazané na volania `i2c_read_blocking()` a `i2c_write_blocking()` z Pico SDK. Pri čítaní údajov bolo nutné zohľadniť rozdiel v endianite, keďže senzor pracuje s big-endian formátom, zatiaľ čo RP2040 využíva little-endian architektúru.

Presnosť, konfigurácia a použité režimy – MLX90632 [3] podporuje dva hlavné režimy merania. Štandardný režim pre všeobecné použitie, s rozsahom −20 °C až 200 °C a presnosťou ±1 °C. Medicínsky režim určený pre meranie teploty ľudského tela (35 - 42 °C) s presnosťou ±0.2 °C. Senzor umožňuje definovať emisivitu meraného povrchu, čo zvyšuje presnosť merania pri rôznych typoch materiálov. Hodnota emisivity sa nastavuje softvérovo prostredníctvom príslušnej funkcie knižnice `mlx90632_set_emissivity()`. Medzi ďalšie parametre senzora patrí rozlíšenie až 0,01 °C, zorné pole (FOV) približne 50°, optický filter prepúšťajúci iba infračervené spektrum v pásme 2-14 µm, čím sa eliminuje vplyv viditeľného svetla.

Praktické skúsenosti a implementačné poznámky – Počas vývoja sa senzor osvedčil ako stabilný a spoľahlivý. Presnosť merania sa pohybovala v súlade s deklarovanými hodnotami výrobcu, pričom rozdiely oproti referenčnému IR teplomeru boli minimálne. Významnou výhodou je aj nízka energetická náročnosť, ktorá umožňuje jeho využitie aj v zariadeniach s obmedzeným napájaním. Integrácia do systému bola pomerne jednoduchá najmä vďaka dostupnosti knižnice, podpore cez I2C a možnosti ladenia výstupu cez UART. Výsledky meraní boli spracované na druhom jadre mikrokontroléra a prezentované v reálnom čase na dynamicky generovanej webovej stránke.

V. Webové rozhranie a HTTP server

V moderných IoT systémoch sa čoraz častejšie využívajú webové rozhrania ako intuitívny spôsob interakcie s používateľom. V rámci navrhnutého riešenia bol cieľom vytvoriť senzorový uzol, ktorý vystupuje ako samostatný HTTP server, poskytujúci aktuálne aj historické údaje bez potreby cloudovej služby alebo aplikácie.

Implementácia HTTP servera – Na implementáciu webového servera bol použitý hardvérový TCP/IP stack sieťového čipu W5100S, ktorý umožňuje efektívne spracovanie sieťovej komunikácie s minimálnou záťažou hlavného mikrokontroléra. Čip podporuje viacero socketov a základné internetové protokoly (TCP, UDP, ARP, ICMP). Sieťová vrstva je postavená na knižnici WIZnet SDK[8] s pomocou hlavičkového súboru `httpServer.h` z balíka, ktorá poskytuje API na spracovanie HTTP požiadaviek typu `GET`. Odpovede sú generované ako textové HTML dokumenty uložené v RAM a následne odosielané klientovi.

Dynamická generácia HTML obsahu – HTML stránka nie je uložená ako statický súbor, ale je dynamicky generovaná priamo vo firmvéri na základe aktuálnych údajov. Generovanie zabezpečuje druhé jadro mikrokontroléra (core 1), ktoré meria aktuálnu teplotu zo senzora, aktualizuje históriu posledných meraní, vytvára reťazec HTML kódu a registruje stránku

[7] https://melexis.github.io/mlx90632-library/
[8] https://github.com/WIZnet-ioNIC/WIZnet-PICO-C

do HTTP servera. História meraní je implementovaná pomocou cyklickej vyrovnávacej pamäte s maximálnou kapacitou 10 záznamov. Každú minútu sa do nej pridá nový záznam, obsahujúci časovú značku a nameranú teplotu. Obsah stránky je zobrazený na obrázku 3.

Obr. 3: Obsah HTML stránky navrhnutého senzorového uzla

Štruktúra webovej stránky – Webová stránka obsahuje sekciu so zobrazením aktuálnej teploty a tabuľku s poslednými 10 meraniami (čas od spustenia, hodnota teploty). HTML kód je optimalizovaný pre zobrazenie v mobilných aj desktopových prehliadačoch.

Paralelizmus a stabilita systému – Na zabezpečenie spoľahlivosti systému bol využitý dvojjadrový mikrokontrolér RP2040, kde jadro 0 obsluhuje HTTP požiadavky (serverová logika) a jadro 1 vykonáva merania a generuje obsah stránky. Prístup do zdieľaných dátových štruktúr (napr. história meraní) je chránený dočasným zakázaním prerušení počas zápisu, čím sa zabraňuje kolíziám a nekonzistentnému stavu.

Výhody a budúce rozšírenia – Medzi hlavné výhody použitého webového rozhrania patrí nezávislosť od externých aplikácií a cloudových služieb, možnosť okamžitého prístupu k údajom cez bežný prehliadač a jednoduchá rozšíriteľnosť o nové HTML komponenty.

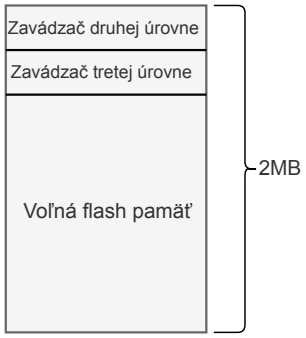
Do budúcnosti je možné rozšíriť funkcionalitu o zabezpečenú komunikáciu (HTTPS), základnú autentizáciu a konfiguráciu siete a senzora cez webové formuláre.

VI. Vzdialená aktualizácia firmvéru pomocou TFTP

Spoľahlivá a jednoduchá aktualizácia firmvéru je kľúčovým prvkom pri nasadení IoT zariadení do reálneho prostredia, najmä ak sa nachádzajú v miestach s obmedzeným fyzickým prístupom. V navrhnutom systéme bol implementovaný vlastný trojstupňový zavádzač (bootloader), ktorý umožňuje vzdialené nahranie nového firmvéru cez sieť pomocou protokolu TFTP. Pri návrhu sa používali časti kódu z SD card bootloadera [7].

Architektúra zavádzača [8] – Zavádzací proces je rozdelený do troch úrovní:

- 1. úroveň – ROM bootloader: pevne zabudovaný v pamäti mikrokontroléra RP2040, aktivuje sa pri každom štarte.
- 2. úroveň – základný SPI zavádzač: inicializuje SPI Flash a RAM, načíta zavádzač tretej úrovne. Schéma Flash pamäte je zobrazená na obrázku 4.
- 3. úroveň – vlastný TFTP bootloader: realizuje kompletnú logiku pre načítanie firmvéru z TFTP servera, overenie platnosti a zápis do Flash pamäte.



Obr. 4: Bloková schéma Flash pamäte vývojovej dosky so znázornením zavádzača druhej a tretej úrovne

Funkcionalita TFTP bootloadera – Tretia úroveň zavádzača je navrhnutá tak, aby minimalizovala zásahy používateľa a maximalizovala bezpečnosť aktualizačného procesu. Po spustení systému bootloader vykoná nasledujúce kroky:

- Inicializuje systémový časovač, UART výstup a LED indikátor.
- Počas 10 sekúnd čaká na prijatie znaku `'p'` cez UART, ak je prijatý, aktivuje TFTP mód.
- V opačnom prípade overí platnosť firmvéru vo Flash pamäti a spustí ho.
- V TFTP móde nadviaže spojenie s pevne nastaveným TFTP serverom a pokúsi sa stiahnuť súbor `firmware.bin`.
- Ak je prenos úspešný, firmvér sa zapíše od offsetu 256 kB (mimo priestoru zavádzača).
- Následne sa vykoná overenie vstupného bodu firmvéru a presmerovanie riadenia.

Bezpečnostné a spoľahlivostné mechanizmy – Zavádzač obsahuje viacero ochranných opatrení:

- Watchdog časovač: aktivovaný na 11 sekúnd, reštartuje systém pri chybe alebo zamrznutí.
- Kontrola linky: bootloader čaká na fyzické pripojenie cez TCP/IP pomocou funkcie `wizphy_getphylink()`.
- Validácia dát: ak je firmvér príliš veľký alebo chýba potvrdenie, zápis sa neuskutoční.
- Spustenie firmvéru z RAM: počas zápisu do Flash sa zavádzač vykonáva z RAM, aby sa predišlo problémom s využitým cache mechanizmom realizovaným hardverovým blokom XIP (eXecute In Flash) integrovaným v RP2040.

Výhody použitého riešenia – Navrhnutý zavádzač umožňuje vzdialenú aktualizáciu bez fyzického zásahu, pričom zachováva bezpečný štart systému. Systém je vďaka nemu vhodný pre reálne nasadenie v autonómnych senzorových sieťach. Výhody zahŕňajú:

- plná kontrola nad procesom aktualizácie,
- jednoduchý protokol (TFTP) bez potreby autentizácie,
- otvorená implementácia prispôsobiteľná iným platformám,
- možnosť rozšírenia o zabezpečený prenos alebo autentizáciu.

VII. Experimentálne testovanie a výsledky

Po implementácii hardvérových aj softvérových komponentov bola vykonaná séria experimentov s cieľom overiť funkčnosť, presnosť a spoľahlivosť navrhnutého senzorového uzla. Testovanie prebiehalo v kontrolovaných aj reálnych podmienkach, pričom boli sledované kľúčové parametre systému: teplotná presnosť, odozva HTTP servera, stabilita webového výstupu, funkčnosť watchdogu a korektnosť TFTP aktualizácie. Detailný opis testovania a vyhodnotenia je k dispozícii v [6].

Metodika testovania – Testovanie pozostávalo z nasledovných úloh:

- Porovnanie výstupov senzora MLX90632 [3] s referenčným infračerveným teplomerom pri rôznych teplotách.
- Meranie odozvy HTTP servera počas dlhodobej prevádzky (12 hodín).
- Simulácia zacyklenia firmvéru a testovanie reakcie watchdog časovača.
- Overenie TFTP aktualizácie firmvéru z lokálneho servera.
- Záťažové testovanie celého systému s nepretržitým výstupom cez UART.

Presnosť teplotného merania – Teploty boli merané v intervale 1 °C medzi 22 - 34 °C. Priemerná absolútna odchýlka a smerodajná odchylka medzi senzorom MLX90632 a referenčným zariadením bola nižšia ako je deklarované výrobcom senzora.

Dostupnosť HTTP rozhrania – Webová stránka bola dostupná počas 12 hodinového testu bez prerušenia. Požiadavky typu `GET` boli obslúžené s typickou odozvou < 120 ms (merané nástrojom `curl -w` cez lokálnu sieť). Počas celej prevádzky boli merania aktualizované každú minútu a správne sa zobrazovali v tabuľke histórie.

Spoľahlivosť watchdogu – Watchdog časovač bol testovaný simulovaným zacyklením hlavného vlákna. Po uplynutí nastaveného intervalu (10 sekúnd) došlo k automatickému reštartu systému. Výpis cez UART potvrdil korektnú činnosť a znovuobnovenie merania aj webového výstupu.

Overenie TFTP aktualizácie – TFTP bootloader úspešne načítal nový firmvér zo statického TFTP servera v lokálnej sieti. Celý prenos súboru veľkosti približne 53 kB trval menej ako 2 sekundy. Po verifikácii vstupného bodu bol firmvér korektne spustený. V prípade výpadku siete (odpojenie Ethernetu) bol prenos detegovaný ako neúspešný a systém zotrval v bezpečnej slučke, ktorú neskôr prerušil watchdog.

VIII. Záver

Článok prezentuje návrh a realizáciu senzorového uzla s podporou Ethernet konektivity, postaveného na dvojjadrovom mikrokontroléri RP2040 a sieťovom čipe W5100S (WIZnet W5100S-EVB-Pico). Navrhnuté riešenie umožňuje presné bezkontaktné meranie teploty pomocou infračerveného senzora MLX90632 [3] a poskytuje jednoduché webové rozhranie bez potreby dodatočného hostiteľského systému. Dôležitou súčasťou implementácie je vlastný trojstupňový bootloader, ktorý zabezpečuje vzdialenú aktualizáciu firmvéru prostredníctvom protokolu TFTP. Vďaka rozdeleniu úloh medzi dve jadrá mikrokontroléra bola dosiahnutá vysoká spoľahlivosť a stabilita systému aj počas dlhodobej prevádzky. Experimentálne testovanie potvrdilo presnosť meraní v rámci špecifikácie senzora, ako aj bezchybnú činnosť watchdogu a webového servera. Navrhnutý systém predstavuje efektívne a energeticky nenáročné riešenie pre jednoduché IoT aplikácie, kde je požadovaná autonómna prevádzka, možnosť vzdialenej správy a dostupnosť údajov cez štandardné webové protokoly. Do budúcnosti je možné riešenie rozšíriť o podporu zabezpečeného HTTP (HTTPS), autentizáciu prístupu, konfiguráciu siete cez webové rozhranie či prechod na dynamickú IP konfiguráciu pomocou DHCP. Rovnako možno uvažovať o rozšírení použitého senzora alebo zapojení viacerých senzorov na jednej zbernici pre komplexnejší zber údajov. Takto upravený systém môže slúžiť ako univerzálna platforma pre vzdialený monitoring v rôznych prostrediach vrátane domácich, priemyselných a výskumných aplikácií.

IX. Poďakovanie

Táto práca bola podporená EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu c. 09I05-03-V02-00019 – „Kľúčové digitálne technológie v oblasti bezdrôtových širokopásmových odolných prenosových sietí“ a Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky v rámci projektu KEGA 041TUKE-4/2025 „Laboratórium na výskum a vývoj pokročilých IoT senzorových sietí a širokopásmových senzorov“.

Literatúra

[1] Raspberry Pi Foundation, *RP2040 Datasheet - A microcontroller by Raspberry Pi*, 2024, [Online]. Dostupné na: https://datasheets.raspberrypi.com/rp2040/rp2040-datasheet.pdf.
[2] WIZnet Co., Ltd., “W5100S,” 2024, [Online]. Dostupné na: https://docs.wiznet.io/Product/iEthernet/W5100S/overview.
[3] Melexis Microelectronic Systems, *MLX90632 FIR sensor*, 2023, [Online]. Dostupné na: https://www.alldatasheet.com/datasheet-pdf/view/1089304/MELEXIS/MLX90632.html.
[4] T. Olearník, *Senzorový uzol s internetovou konektivitou pomocou WIZnet 5100-EVB-Pico a MLX90632*, 2025, [Online]. Dostupné na: https://github.com/TiborOlearnik/senzor.
[5] ——, *Bootloader pre RP2040 pomocou TFTP*, 2025, [Online]. Dostupné na: https://github.com/TiborOlearnik/bootloader.
[6] ——, “Senzorový uzol s internetovou konektivitou na báze raspberry pi pico rp2040,” Bakalárska práca, Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025.
[7] H. H. LAI, *RP2040 Datasheet - A microcontroller by Raspberry Pi*, 2024, [Online]. Dostupné na: https://hsuanhanlai.com/writting-custom-bootloader-for-RPI-Pico/.
[8] B. Starkey, *Pico serial bootloader*, 2021, [Online]. Dostupné na: https://blog.usedbytes.com/2021/12/pico-serial-bootloader/.

Transfer of the Evaluation of the Copymaster assignment to the Arena system

[1]Tomáš BUČEK, [2]Matej ELIAŠ,[3]Ján GENČI

[1]Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic
[1]tomas.bucek@tuke.sk, [2]matej.elias@student.tuke.sk, [3]jan.genci@tuke.sk

Abstract **– This paper presents our implementation of tests for the Copymaster assignment and the use of these tests in the Arena system and other testing contexts. It describes the assignment and the Arena system. It explains the methods and technologies used for implementing the tests, with particular focus on system call tracking with strace and test creation in the Go language. It compares various ways of test execution. The paper presents the effectiveness of the newly implemented tests by showing summarized results of the issues identified in student assignment submissions.**

Keywords **– operating systems, evaluation, testing**

I. Introduction

At the Department of Computers and Informatics of the Technical University of Košice, automatic testing and evaluation of student software projects is being applied in several courses. This evaluation approach is also being used in the Operating Systems course. As part of this course, two coding assignments in C programming need to be completed by the students. One of them is the Copymaster that instructs the students to work with Unix-like syscalls for file management and file manipulation. Students now submit code through the university's GitLab, where the assignments are graded automatically. The Copymaster assignment grading system examines hundreds of solutions annually. Nonetheless, even though it was designed to do this, over the years it has evolved into an undependably documented and progressively problematic system to maintain. Even though modified multiple times, its usage became challenging. Based on these factors, the idea surfaced to replace it with a much more flexible alternative that would be easier to keep up and suitable for today's needs. Among the available options, one which comes into the picture is the Arena system that is developed by department and is already used for testing assignments of other courses that are taught in the department.

II. Prerequisites

A. Copymaster assignment

This assignment was created so students would get practical knowledge about file management on Unix-like operating systems. The core part of this assignment revolves around the usage of the most common simple system calls which are utilized by the operating system. The program created by students should perform actions from a predefined list, and those actions must use a specific set of system calls and functions. These system calls should be invoked through the wrapper functions available in the standard C libraries. The program is operated via the command line, and actions are selected using command-line arguments. If any unexpected situation or error occurs, the program must set the program's return code to the number code that is assigned to the action, and an error message explaining the issue must be written to stderr. It's good practice to develop Copymaster on Linux systems, because the whole Operating Systems course explains concepts on Linux systems.

```
./copymaster [OPTION]... SOURCE DEST
```

Listing 1 Usage of Copymaster program.

In listing 1, the general usage of the program is displayed. The program accepts 13 specific arguments and every argument is corresponding to a particular action. For example, the argument `--fast` starts the copying of data from the source file to the destination file. When this argument is used, the program has to perform the copy operation with a more efficient technique that uses buffer and blocks of data are copied. Some program arguments are compatible and can be used simultaneously. For example, we can copy the content of one file using the `--slow` argument and also delete the source file after the copy is finished with the `--delete` argument. When the assignment project is created, the student can use an already created program skeleton that provides command-line parsing and additional functions. It is not necessary to use this skeleton, but it is recommended. The structure of the project is completely in the student's hands, and the only obligatory file is the `Makefile`, containing instructions for program compilation with `gcc`.

B. Arena system

The Arena system is a platform for evaluating and testing code. It is continuously used and developed by personnel from the Department of Computers and Informatics at the Technical University of Košice. It was build primarily using the Python programming language, and one of its main features is the utilization of Docker containerization technology. Students submit their code to Arena indirectly — they upload their written code to the school's GitLab. Arena then periodically, every three hours, pulls the necessary repositories, and the code is automatically evaluated. Arena is extensible and can incorporate various independent components. The main component responsible for managing testing and communicating with other parts is called Gladiator. All communication with Gladiator is handled via a REST API [1][2]. Another important component is the Arena web UI, where both students and teachers can view the results of code evaluations. This environment makes Arena accessible from the internet and provides a simple, easy-to-use interface for browsing testing results.

III. Technical implementation of tests

A. Technologies used

After conducting a series of experiments with various programming languages and other tools, we ultimately decided to implement our tests using the Go programming language and strace tool.

1) Golang: Go or Golang, is a compiled and statically typed programming language developed in association with Google. The reason why the langue was Go, or Golang, is a compiled and statically typed programming language developed in association with Google. The reason why the language was created is to fix issues present in older languages like C or C++. Two main problems which Go tries to fix are memory management and the ease of parallel programming. Go comes with one of the fastest garbage collectors and unique lightweight threads called goroutines. Go tries to be a very simplistic language and provides a strong and reliable standard library. Nowadays, Golang is a very popular choice for creating high-performance applications, often used in the cloud.[3]. In our work, we take advantage of the testing library in the Go standard library for creating efficient tests. The tests are compiled by the Go compiler into a single binary.

The authors in the book [4] share their opinion that Go is built from the ground up, while still being similar to languages from the C-family. They call Golang "C for the 21st century." The authors also claim that Go code is generally only 20% slower than C++ code, which is not a significant speed reduction, and that Go code scales much better in large codebases. According to them, Go code is memory-efficient compared to other garbage-collected languages; in comparison, similar programs written in Go use up to 70% less memory than their Java equivalents.

2) Strace: strace is a debugging and diagnostic tool for Linux that allows us to trace the system calls made by a program and the signals it receives. It accomplishes this by intercepting and recording the low-level system calls between a process and the Linux kernel. strace attaches to a process (or starts one) using the `ptrace` system call and uses the `PTRACE_SYSCALL` request to ask the kernel to stop the process each time it enters or exits a system call [5][6].

In the code listing 2 below, a part of the strace output is displayed, that is written to the output file after the program's execution. The analyzed program is just a simple C program that opens a file and writes data to it. In our case, the program only opens a file with the `open()` call, writes the text with the `write()` call, and after writing, the file is closed using the `close()` call. Most of the output consists of information about the syscalls used. Each syscall is written on a separate line in form of function call with all its parameters and also return code. Additionally, the output includes the return code of the program, and it may also display a captured signal. If the application misbehaves, it is often terminated by the SIGTERM signal, which is logged in the strace output.

```
openat (AT_FDCWD , "output.txt", O_WRONLY |... , 0666) = 3
write (3, "Example text .", 13)= 13
close (3)= 0
exit_group (0)= ?
+++ exited with 0 +++
```

Listing 2 Output of strace when analyzing simple program.

B. General test structure

Created tests evaluate student programs by running them inside strace. Then, the strace output file and other files accessed by Copymaster are analyzed. We created 93 tests, and in each test, Copymaster is executed under specific circumstances. We identified three main variables that make up unique tests:

1) Existence of source or destination file
2) Permissions of files
3) Correctness of used parameters

We placed all tests that evaluate usage of specific arguments used with Copymaster into one source file. Source files containing tests have names ending with the suffix `_test.go`, and the testing library can automatically find all implemented tests. All source files containing tests are in the package `tests`. We separated the code even further and placed all subtests into the package `tests/subtests`. More details about subtests are in subsection III-D. Other helpful functions used by tests and subtests are in the package `utils`.

Diagram 1 displays the general flow of each test. At the beginning of almost every test, random variables are generated and files specific to the test are created. An example of a needed random variable may be text used as file content or random Unix permissions. After test preparation is finished, Copymaster is run by `strace` as process. More about Copymaster execution is in subsection III-C. After the `strace` process finishes, all subtests are run. In the diagram, strict subtests are separated from regular subtests, but in the code, they are simply placed one after the other. We illustrated it this way because strict subtests may terminate the test before other subtests get a chance to be run.

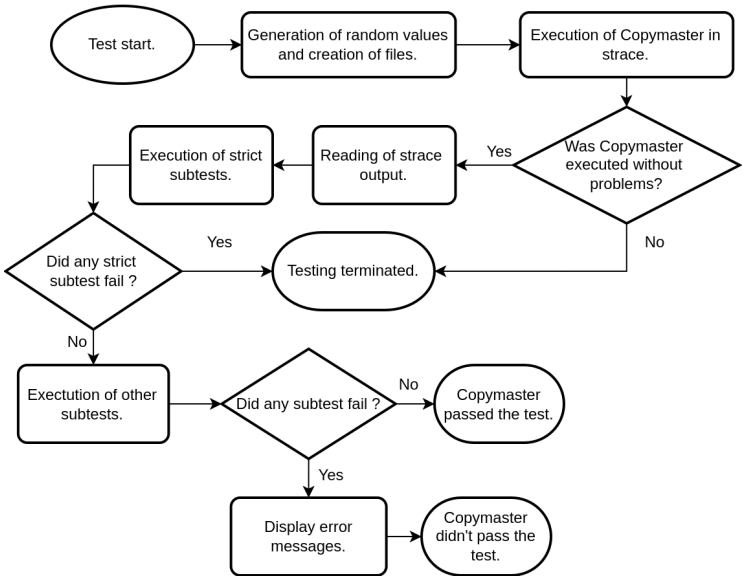


Fig. 1 Diagram describing general test flow.

C. Copymaster Execution and Process Handling

After the successful compilation, the created program must be run and analyzed. Our approach doesn't run the student's program directly, but it is indirectly run by the strace process. During execution, strace captures system calls used by the program and signals sent to the program. All useful data is written to the specified output file after execution. During our code development, we found a sporadic issue. When the problem appeared, the strace process was unresponsive and this process was eventually killed by the OS after a few seconds. The problem was caused by bugs in badly written student code.

In the created Golang code, the `RunCommand` function is responsible for starting the `strace` process and it accepts a variable number of arguments and combines arguments into final the `strace` command. In this function, we also use a strategy that allows us to detect and kill unresponsive processes. The found problem was fixed using features and techniques specific to Go code. The complete process of launching is comprised of these steps:

1) Initialization of the object `context.WithTimeout`, which sets the maximum time for process execution, in our case 0.3 seconds is used.
2) The struct `syscall.SysProcAttr()` for process configuration is used, and the attribute `Setpgid` is set to true. This change assigns a new process group ID to the created process.
3) Starting the strace process and capturing the text that was written to `stderr`.
4) Creating a goroutine that runs independently in background, its purpose is to monitor strace process and kill the process if it is unresponsive.
5) Waiting until the strace process is finished. When the process runs without issues, relevant information is fetched; otherwise, an error explaining the termination is fetched instead.

```
strace -o strace_outputfile
copymaster_executable [copymaster_args]
sourcefile destinationfile
```

Listing 3 Template showing how each strace command is structured.

Code listing 3 shows a template of how each `strace` command is structured. The entire command is assembled from a variable number of `RunCommand` function arguments. This approach allows us to run student program inside strace with any numbers of arguments.

D. Subtests

The main reason why we separated code into subtests is the frequent usage of these code blocks. Each subtest was made to evaluate a specific aspect of an action or work that is performed during the program's run. Some more complicated subtests also evaluate if the correct method was used. In the code, subtests are in the form of helper functions and contain functions that compare data and directly resolve the test result. The Golang testing library doesn't use any functions for assertions like other libraries. That's why, in the code, we have to use control structures such as **if** and **while** in combination with functions available in the testing library. In every test, multiple subtests are called. The final result of the test is positive only if every subtest had a positive result and none of the subtests failed. Each subtest, as a helper function, has a `*testing.T` pointer as its first parameter. This pointer allows the function to determine the test result. In our tests, only these functions available in the `testing` library are used:

1) **t.Fatalf()** – in tests, we use this function in strict subtests. The function prints a formatted error message and also terminates the test. The final result of the tests is negative after the usage of this function.
2) **t.Errorf()** – in subtests, this function is used most commonly. The function also prints a formatted error message, but other subsequent subtests are still run. The final result of the tests is negative after the usage of this function.
3) **t.Logf()** – we use this function only to print formatted messages. The function has no effect on the test result.

In code listing 4, a subtest is shown that is used for comparing the expected return code to the actual return code. The subtest is used in every test and is called like a regular function. This is one of the simpler subtests, in the code, only one comparison is used. More complex subtests may compare more complex data or perform more comparisons. In case when return codes are different, an error message is printed.

```
func CheckExitCode(t *testing.T, expected int, got int) {
  if expected != got {
    t.Errorf("Return code of program is incorrect !!!\n"+
          "expected: %d\n"+
          "     got: %d\n\n", expected, got)
  }
}
```

Listing 4 Subtest whose purpose is to check whether the return code of the program is correct.

E. Function usage tracking

As we mentioned in previous section II-A, the `copymaster` assignment requires the use of certain system calls and functions. From this requirement comes the need to check whether the expected functions were used, and whether any forbidden functions were not used.

- **Search for system calls** — we created functions that read, parse, and analyze the `strace` output. These functions are then used by subtests, which are called when needed. In each test, the reads and writes performed by the program are analyzed with subtests. For example, from the list of `write()` calls made by the program, the exact number of written characters and the total number of calls are calculated. Based on this information, the program judges whether the correct writing method was used. Most commonly, we track system calls related to file opening, reading and writing.
- **Forbidden functions** — the search for these functions is done without strace. Instead, we created a separate test that searches the contents of all source files. Firstly, test reads files and removes comments and string literals from the code text. Then, every line is searched and all function calls present are collected. After the file search, all found functions are compared to a list of forbidden functions. We have divided forbidden functions into two groups:
  - Similar functions that simplify work — a good example is the `fwrite()` or `fseek()` functions. These functions are present in the standard library and make working with files easier. They use system calls under the hood to accomplish the task, but give us a higher abstraction.
  - Functions that allow the creation of processes and execution of other programs.

IV. Usage of tests

Running Go tests may be done in two ways. If the Go compiler is present , we can run all implemented tests with the command `go test`. The second way, and our preferred way, is to compile all available tests into one statically-linked executable . This approach is done with the command `go test -c` and the executable file is named after the package with tests and placed in the package directory. This gives us tremendous flexibility and binary files can be used on other environments that share the same OS and CPU architecture. The Go compiler can even perform cross-compilation and we created executable files for the ARM platform as well. Created executable files don't need any dependencies to function. The created executable file is named `tests.test` and this file is used everywhere where we use implemented tests.Executable allows us to run selected tests via the argument `-test.run`, and all tests that match a regular expression or contain the string will be run.

A. *Putting tests to the Arena*

Putting our tests into Arena wasn't that hard but required some knowledge about the system. We had to create three important configuration files. The most important file was `problemset.yml`. In this configuration file, we defined the order in which the tasks are executed. The first few tasks prepare the submission for testing, and files are extracted from archive and compiled. After all initial tasks, we defined all the tasks for test running. All tests are grouped, except the first one that searches source files for forbidden functions, into suites based on what argument is used when Copymaster is run. Part of one suite with tests evaluating running with argument `--link` is displayed in code listing 5. Each test is of type `test:exec`, meaning that this test is in the form of a command.Each test in the suite has its ID, title, and the importance of the test is expressed by its score.

```
- id: tests_link
  title: "Tests for the -K / --link flag."
  type: suite
  tasks:
  - id: TestLink_A
        title: "Link must be created."
        type: test:exec
        cmd: ./tests.test -test.run TestLink_A -test.timeout 1s
        score: 3
        ...
```

Listing 5 Part of created test suite in configuration file problemset.yml .

We also had to create two important configuration files. The first one is the `Dockerfile`. It holds all information on how to and what base to build the Docker image used by Arena. The usage of statically-linked testing files significantly reduced the packages that we needed to install with package managers. In the file, we added options for both Ubuntu and Alpine bases. For fully functioning testing, we needed to install only the `gcc` compiler, `make` tool, and libraries for the C language. The last configuration file added was `.gladiator.yml`. Inside, we placed general settings for Gladiator. We decided there is no need to write this information in detail.

B. Bash scripts

We also created some bash scripts that utilize the executable with tests. Our intention was to more effectively test our implementation and also showcase the great variability of our testing method. We created scripts that allow us to perform the following tests:

1) Batch testing – The created script is able to test multiple student submissions placed in a specified directory. All directories are sequentially traversed, and programs are compiled, tested, and the results are written into HTML files and a CSV file containing global summary results.
2) Testing on local machine – The script can be used on the student's OS during the coding process. For evaluation, the student just needs to download the repository, which contains only the test executable and the script. After the script is run, all results and outputs of tests are written into a regular text file and an HTML file.

V. Evaluation of tests

A. Speed of tests

In table 1, various types of test execution are presented. Measured results in the table show a significant performance difference between strategies. The fastest strategy ran tests in parallel using 8 concurrent processes and utilized the `xargs` tool. All implemented tests finished in just 0.160 seconds, and the average time for a single test to finish was only 0.0017 seconds. The `xargs` tool is explained after the table.

The fastest way of test launching, when we do not consider parallel execution, was recorded when all tests were run and managed by a binary testing file. All tests finished in 0.375 seconds, and it was only almost two times slower than parallel test execution. This method also preserves the alphabetical order of tests and is generally the most resource-effective. Second worst execution time was recorded when tests were run in a Bash loop. The total execution time was over a second, and the execution time per test was almost four times slower compared to tests managed by the testing executable. The overhead was caused by repeated creation of new processes and repeated context switching.

The worst time was measured when testing was run in Arena on the remote school server. In Docker containers, the overall execution time was 7.907 seconds, and the average time for one test to finish was 0.086 seconds. Go tests also show execution times when the verbose option is enabled or when a test fails. This information was also displayed in Arena in the text output of every failed test. When we calculated the average time for failed tests to finish, it was 0.010 seconds. This calculated number is comparable to the time it took to finish an individual test in the Bash loop. From these numbers, it is obvious that the longer execution time is caused by additional tasks performed by Arena.

Testing type	Time (s)	Average time per test (s)
Sequential execution of tests controlled by binary testing file.	0.375	0.004
Sequential execution of one test at a time in a bash loop.	1.022	0.011
Parallel execution of 8 processes using xargs.	0.160	0.0017
Execution of tests in Arena.	7.907	0.086

Table 1
Comparison of test execution times under different conditions.

Now, the introduction of xargs is needed. The `xargs` application is a command-line tool installed commonly used in most Unix-like operating systems that reads items from standard input, for example through pipes, and allows us to invoke other commands with the arguments received from standard input. Additionally, it offers various options to control the number of arguments passed at once, how inputs are separated, and whether commands are executed in parallel or sequentially [7]. A good example of using `xargs` in conjunction with the `find` command can be found in the article [8].

```
$ ./tests.test -test.list Test | xargs -n1 -P8 \
./tests.test -test.v -test.run
```

Listing 6 Command in shell that allows parallel execution.

In code listing 6, a command is shown that we used for facilitating parallel execution of tests in the shell. For parallel execution, only the testing executable and the xargs command are needed. Firstly, we get the names of all implemented tests from the executable with the `-test.list` argument. All test names are sent via a pipe to the xargs command, and it creates and manages

processes for us automatically. We told xargs with `-n1` that every process gets only the name of one test, and the maximal number of concurrent processes is 8.

B. Effectiveness of tests

To find out how effective the created tests were in detecting errors in student programs, we tested 100 submissions that were submitted last year. We created a few Bash scripts that evaluated all 100 student submissions using the created tests. To obtain more meaningful information, the text printed by the tests was analyzed. The scripts summarized the data mostly with pattern matching using `grep`, and the data were stored in CSV files. It would be too long to include all tables with data from the CSV files in this paper, so we included only the tests for these three arguments:

1) **–append** — copied data must be appended to the destination file.
2) **–overwrite** — data is copied only if the destination file already exists.
3) **–link** — the destination file is created as a link to the source file.

In the table, the following are data gathered after testings are present :

1) Test names.
2) Number of correct programs.
3) Number of incorrect return codes.
4) Number of wrong writes to stderr.
5) Number incorrect actions (excluding other issues).
6) Number of incorrect read/write actions.
7) Number of missing `close()`.
8) Number of terminated programs.

From the results in table 2, we can see that some tests were more successful, and more student programs passed these tests. Tests that reach a **60–70%** success rate and have the _A suffix in their name test basic usages of the Copymaster program. We assume that the success rate is higher because students were testing these usages during coding. When we averaged all the test results, tests where expected program usage is evaluated had an average 64% success rate.

On the other hand, tests evaluating student programs in problematic situations have the suffix _B in their name. A good example of such a problematic situation may occur during the evaluation by `TestOverwrite_B`. This problematic situation is specific only to this test because an error must be reported when the destination file does not exist. On average, only **36%** of programs passed these problematic tests, which is a significant drop.

A quite prevalent issue identified by some tests was caused by missing file closures. This problem most often occurs during tests when Copymaster is run under circumstances where the source file does not exist or permissions do not allow access to the file. In extreme cases, more than **40%** of student programs manager files poorly and failed to close opened files. We consider this issue to be a significant problem. A good example where this occurred during testing was `TestAppend_B`. In this case, **42 out of 100** programs correctly identified the non-existent destination file but forgot to close the opened source file before the program finished.

Test name	Passed	Wrong return code	Wrong stderr	Bad action	Close issue	RW issue	Seg. fault
TestAppend_A	71	3	3	15	10	16	1
TestAppend_B	26	23	31	0	42	17	1
TestLink_A	64	6	5	19	12	15	1
TestLink_B	37	16	36	0	31	6	1
TestOverwrite_A	75	1	2	12	10	14	1
TestOverwrite_B	31	16	27	0	43	10	1
*NoInput	40	43	43	0	8	0	2
*InputNoReadPerms	17	41	64	1	7	0	2
*OutputNoWritePerms	14	52	43	1	49	16	2

Table 2
Summarized results of testings for some Copymaster tests.

We also implemented tests for all Copymaster arguments, and these evaluate Copymaster under problematic, repeatable edge cases. By *repeatable*, we mean that these edge cases can occur when Copymaster is run with almost any argument by itself. Edge cases are caused by special file configurations that make Copymaster's actions impossible to perform. Under these circumstances, Copymaster must not perform the actions selected with the arguments but must stop after file checking and report the detected problems by setting the return code and printing a message to

`stderr`. For better readability, we averaged all test cases of the same type and presented the results for each type on a single line. The calculated average results can be seen in Table 2 in the last three rows.

The first repeatable edge case runs the program when the source file with the specified name does not exist. On average, only **40** student programs passed this test. The most common issues in the code were caused by the use of an incorrect return code and incorrect error message printing. On average, **43** programs per test of this type had these flaws.

The second repeatable edge case runs the program under conditions where the source file cannot be read due to insufficient permissions. Far fewer student programs passed this type of test — on average, only **17** programs handled this problem gracefully. Per test, **64** programs wrote an incorrect error message to `stderr`. On average, **45** programs identified the wrong problem, reporting a missing source file instead of a permission problem. We considered this incorrect error handling.

The last repeatable edge case is covered by tests where the current user cannot write to the destination file due to insufficient permissions. Only **14%** of student programs passed these edge case tests. Nearly half of the programs (**49%**) forgot to close the source file, and on average, **16** programs read data from the source file even though the destination file was not open and could not be written to.

VI. Conclusion

In this article, we describe the implementation of a testing method and tests for the Copymaster assignment, as well as the potential usage of the created tests in the Arena platform and other evaluation approaches. The tests are written in the Go language and compiled into a statically linked executable. For better code analysis, programs are run using `strace` tool, which allows tracking of used system calls. Part of the code is organized into subtests that allow us to separate unrelated issues and logically compose tests.

Evaluation of the tests was performed in two ways. The first approach focused on testing speed. The most effective testing method was found to be using a statically linked Go testing binary. This test execution method was able to run 93 tests in 0.375 seconds. Parallel execution with the `xargs` tool was also tried, which was more resource-intensive but faster. In Arena, the entire testing process took much more time. In Arena, 7.9 seconds were needed to run all tests. According to our measurements, the speed of tests is similar in Docker containers to native test execution, so the time is prolonged due to processes specific to Arena. In real-life scenarios, disk IO operations may become a bottleneck, preventing optimal speeds from being reached.

The second approach focused on the ability of the tests to find problems in student implementations. We evaluated 100 previously submitted student programs with Bash scripts. The results show that, on average, 64% of the programs passed easy, straightforward tests. On the other hand, more problematic situations covered by other tests showed only a 34% success rate.

Acknowledgment

I would like to thank Ing. Tomáš Buček for his support, guidance and help with this project.

References

[1] M. Madeja, M. Biňas, and L. Prokein, "Continuous analysis of assignment evaluation results from automated testing platform in iterative-style programming courses," in *2019 17th International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, 2019, pp. 486–492.
[2] F. Gurbáľ, "Integrácia docker compose do systému aréna," Technická univerzita v Košiciach, p. 152, 2020.
[3] A. Rios, *System Programming Essentials with Go: System calls, networking, efficiency, and security practices with practical projects in Golang*. Packt Publishing Ltd, 2024.
[4] I. Balbaert, *The way to Go: A thorough introduction to the Go programming language*. IUniverse, 2012.
[5] J. Keniston, A. Mavinakayanahalli, P. Panchamukhi, and V. Prasad, "Ptrace, utrace, uprobes: Lightweight, dynamic tracing of user apps," vol. 1, pp. 215–224, 2007.
[6] GNU Project, *strace(1) Linux manual page*, man7.org, 2025, retrieved April 30, 2025. [Online]. Available: https://man7.org/linux/man-pages/man1/strace.1.html
[7] ——, *xargs(1) Linux manual page*, man7.org, 2025, retrieved April 30, 2025. [Online]. Available: https://man7.org/linux/man-pages/man1/xargs.1.html
[8] D. Taylor, "Work the shell: the find— xargs sequence," *Linux Journal*, vol. 2015, no. 249, p. 5, 2015.

Creation of Practical Tasks to Practice Inter-process Communication

[1]Tomáš BUČEK, [2]Alina STANKOVYCH,[3]Ján GENČI

[1]Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic
[1]tomas.bucek@tuke.sk, [2]alina.stankovych@student.tuke.sk, [3]jan.genci@tuke.sk

Abstract **– Effective instruction of inter-process communication (IPC) is crucial for computer-science students, because it delves into the internal logic of operating systems. Traditional teaching methods often leave too little room for hands-on practice, immediate feedback, and systematic assessment. This paper focuses on designing and implementing a web platform that lets students tackle IPC-focused tasks right during class. The platform automatically compiles and evaluates submitted solutions in a secure environment using Docker containers and delivers instant feedback. The goal is to streamline teaching, reduce instructors' administrative load, and give students more opportunity to cultivate practical skills.**

Keywords **– operating systems, evaluation, testing**

I. Introduction

Inter-process communication (IPC) plays a pivotal role in the development of software that has become an integral part of everyday life — from operating systems to complex applications. It is therefore essential for computer-science students not only to grasp the theoretical principles behind these techniques, but above all to be able to apply them in practice. A sound theoretical foundation alone is not enough for a thorough understanding of the subject — hands-on experience is the crucial element. IPC reaches into the very core of an operating system: it covers working with processes, their synchronization and communication. By nature these techniques are abstract, and without direct exposure to their practical use they are often hard to grasp. Regular and systematic practice of these mechanisms enables students to understand how they operate in different contexts, learn to handle errors, and gradually develop the ability to solve problems independently. In ordinary teaching practice, however, students face several constraints that prevent them from fully rehearsing more advanced topics. The number of available tasks is limited, and an even bigger obstacle is the lack of time. Instructors often lack the capacity to examine each solution in detail and to provide feedback that would guide the student and support further development. Consequently, lab sessions frequently serve only as an introduction to the topic, leaving insufficient room for deeper exploration or systematic repetition. Students may thus form only a basic idea while missing the opportunity to practise the topic extensively and truly comprehend it. A solution that automates the assessment process while simultaneously offering a larger pool of tasks can create the conditions needed for meaningful practice outside classroom hours. For these reasons it seems beneficial to design a tool that can respond effectively to the needs of both parties — students and instructors. The platform should give students as much practical practice as possible while relieving instructors of repetitive assessment tasks and allowing them more time for expert guidance. Such a platform will greatly streamline day-to-day lab sessions. Thanks to automatic evaluation of solutions and instant feedback, time is saved on both sides: students see the results of their work immediately and gain a better understanding of their mistakes, whereas teachers are freed from manually checking every assignment.

II. Fundamental Mechanisms of Inter-Process Communication

Inter-process communication (IPC) is a collection of techniques that let processes within an operating system communicate, synchronize, and share data—essential for parallel computing, client–server architectures, and OS design. The most common mechanisms are pipes, shared memory, semaphores, signals, and sockets; mastering them demands an understanding of kernel behavior and low-level programming [1]. Theory alone is not enough: IPC operates "behind the scenes," and without hands-on experimentation students find it hard to grasp what is happening. Teaching must therefore include opportunities to work directly with these mechanisms— the central aim of this project.

A. Core Inter-Process Communication Mechanisms

Inter-process communication (IPC) is a collection of techniques that let processes share data, synchronize, and coordinate their activities. Figure 1 illustrates the most common IPC forms; despite their differences, they all pursue the same goal—reliable information exchange between parallel processes. The choice of mechanism depends on security, flexibility, and speed requirements:

Fig. 1 Inter Process Communications

- **Pipes** - The kernel creates a byte stream that one process writes and another reads. Configuration is quick, but the stream is inherently one-way and large data volumes incur the overhead of copying every byte through the kernel.
- **Shared memory** - The fastest route to data: processes map the same RAM pages and work in place, without copying. This shortcut pays off only if disciplined synchronization (e.g., semaphores or mutexes) prevents inconsistent states.
- **Semaphores** - Synchronization primitives that protect critical sections or shared resources. Binary semaphores (mutexes) grant exclusive access; counting semaphores allow multiple readers. Incorrect locking order can produce deadlock.
- **Signals** - Asynchronous notifications with minimal latency; they often act as triggers for more complex communication through another channel.
- **Sockets** - Suitable for local UNIX-socket links and network communication over TCP/UDP alike. Their versatility adds extra duties—port management, latency handling, and frequently encryption (TLS) [2].

A solid grasp of these mechanisms is crucial. Hands-on work and observation of their behavior in varied scenarios help students choose the right tool for a task and understand the trade-offs involved. Such insight underpins their ability to design software, resolve concurrency issues, and optimize performance in real-world projects.

B. Combining IPC Mechanisms in Practice

In real-world applications, designers rarely rely on a single communication method; instead, multiple IPC mechanisms are combined to address diverse requirements for throughput, latency, and reliability. Early development stages typically favor low-overhead solutions—such as shared memory complemented by semaphores or mutexes—to enable rapid data exchange within one system while keeping latency negligible.

When new requirements demand sending data beyond the local environment, the original logic can be wrapped in a thin network layer. A common pattern routes data first through a local UNIX socket into a helper process, which then transmits it over TCP/UDP to a remote server. This preserves the continuity of the original code base while confining changes to the new layer.

Container platforms adopt a similar multilayer model for log processing: the application writes to a pipe, a collector retrieves the bytes via a UNIX socket, and the host system forwards them over an encrypted TCP connection. Each layer uses the mechanism that best fits its own context.

Key factors when selecting and combining IPC techniques include:

- **Data volume** — small payloads flow reliably through pipes, whereas gigabyte-scale streams call for shared memory.
- **Latency requirements** — tasks needing millisecond responses stay as close to RAM as possible, while less time-critical operations can traverse the network.
- **Scalability and maintainability** — a multilayer architecture lets the system grow by adding a new layer without touching existing logic. With the right combination of mechanisms, one can build a communication infrastructure that meets current demands and remains flexible when new servers or network gateways are introduced.

III. Problematics of Learning IPC

Inter-process communication is one of those topics where theory and practice seem to inhabit different worlds. In the lecture hall everything looks tidy—parallelism, data consistency, semaphores—but the moment a student opens a code editor they face a chain of system calls that behave nothing like intuition suggests [3]. Hence the lingering fear: "On the board it's clear, in a live system it's a fog."

IPC is difficult to observe directly. Communication occurs deep within the kernel and manifests only indirectly through unexpected delays, cryptic errors, or unpredictable hangs [4]. Diagnosing the cause—whether it is a resource race, a deadlock, or merely a context switch—demands insight beyond the API signatures and a view beneath the surface.

One or two lab sessions are nowhere near enough. Real understanding arrives after dozens of attempts—building the intuition for which mechanism fits a given workload and where to look when the system misbehaves.

That's why IPC must be taught "long and wide": solid theory, constantly reinforced by practice that lets students make mistakes, sift through logs, and compare solutions. Only then do the abstract diagrams come alive, giving students the confidence they need to handle parallel processes in real projects [5].

A. The Practice Gap and Feedback Delay in IPC Education

Even with a well-structured course on inter-process communication (IPC), hands-on work usually consists of just a few lab assignments—far too little for students to see how IPC behaves in the wild or to apply it confidently in real projects. True mastery demands frequent, varied practice: only constant experimentation builds intuition, highlights common pitfalls, and reveals how processes change under load.

Here an organisational barrier emerges. Dozens of students share one instructor; when several run into questions simultaneously, it is impossible to give each the time their problem requires. Feedback is delayed: a student does not immediately know whether the chosen approach is sound, whether the program actually works, or where the bug is hiding. This pause slows progress and prevents the gradual formation of a coherent mental model.

Thus, the core difficulty in learning IPC comes from two factors: a shortage of practice covering the full spectrum of real-world scenarios, and the limited availability of an instructor who can provide timely, individual support. Until both issues—more practice and faster feedback—are addressed, students risk leaving with only a fragmented understanding of how IPC operates in practice.

B. Current Online Environments and Their Limitations

Modern web-based editors let students write code directly in the browser, run it instantly in the cloud, and compare the result with community solutions. This model works well for algorithmic problems: read input, print a number, and the "compiler + tests" pipeline declares success. Inter-process communication (IPC), however, hits a technical wall. IPC is about dynamics—multiple processes that share memory, coordinate with semaphores, or signal events in real time [6]. Detecting races, deadlocks, or latency requires the program to have full access to the operating-system kernel.

That requirement collides with the security philosophy of typical online sandboxes. From a platform operator's viewpoint, allowing calls such as `fork()`, `shm_open()`, or `semop()` is risky: a process might exhaust resources, lock shared objects, or break container isolation. Most services therefore block these calls, leaving students with a convenient environment for algorithms but an immediate *"unauthorized access"* error as soon as they try to spawn a process or open shared memory.

This gap means that while beginners have dozens of portals for algorithm practice, those studying IPC often fall back on local virtual machines, ad-hoc scripts, or self-managed containers.

Time and energy that could deepen their understanding of synchronization end up spent on technical work-arounds.

A dedicated online platform for IPC would bridge that gap: a secure container environment that still permits full system calls, lets students observe process behavior in real time, and provides immediate feedback—without the overhead of local installation or configuration.

C. Designing an IPC-Training Platform: Goals and Requirements

A look at today's online editors shows a sharp contrast: what works effortlessly for algorithm drills becomes a real obstacle the moment a student tries to start two processes, share memory, or set up a semaphore. Students expect three things from a web environment—no installation, instant feedback, and rock-solid safety. The platform, meanwhile, wants to block any system call that could touch the kernel.

1) A typical lab session: A class usually starts with theory. The instructor sketches diagrams, explains algorithms, and walks through examples. As soon as the words "now code it yourself" are spoken, the room turns into an impromptu help desk. Dozens of hands go up: one program crashes with an error, another student is unsure which call to use, a third needs confirmation that the chosen synchronization approach makes sense. The instructor can respond only in sequence; while one problem is being solved, everyone else waits. Precious practice minutes slip away, and some questions never get answered.

2) Necessary changes: For practical work to succeed, students need two things. First, immediate feedback—run the code and see at once whether the processes behave correctly. Second, zero setup—open a browser, type the program, and test it without spending time on compilers and configs. When both conditions are met, a lesson stops being a queue to the teacher and becomes a continuous, meaningful experiment.

3) Essential capabilities of the platform:

- Launch multiple isolated processes with a single click, letting them communicate through pipes, shared memory, and semaphores.
- Return a clear verdict within seconds: the solution passes or fails.
- Protect the host system without blocking legitimate IPC calls such as `shm_open()`, otherwise the whole exercise loses its point.
- Keep the run queue short so a "one-second" test never turns into a long pause.
- Present a clean interface—editor, terminal, and the assignment itself—without extra buttons and charts that distract from the main goal: quickly finding out whether the program works.

4) Architectural options: Two approaches are possible. A monolithic server is easy to deploy and serves small groups well, but under concurrent load it soon hits CPU and memory limits. The alternative is an elastic pool of container runners: every code submission runs in its own isolated environment, finishes, frees resources, and makes room for the next one. When more students submit, the pool scales horizontally without changing the grading logic.

The right choice depends on course size. A monolith is fine for a small seminar; with dozens or hundreds of simultaneous submissions, planning for a scalable runner pool is wiser—students won't wait, and instructors can rely on a stable system.

5) Summary: Thus, the proposed platform removes the main technical hurdles—environment setup, configuration differences, and waiting for grading—and immediately displays program results. This enables students to continuously experiment with pipes, semaphores, shared memory, and sockets, while instructors can focus on more complex conceptual issues. The next chapter moves from requirements to the system's practical design and demonstrates how to ensure equally fast and secure operation in a university environment.

IV. System Architecture

Earlier-identified instructional requirements shaped an architecture that balances student convenience with efficient instructor oversight. The platform follows a three-layer design.

Client Layer. A single-page web application runs directly in the browser. Its built-in editor (with POSIX call highlighting), terminal, and submission timeline spare students from installing compilers or configuring local environments.

Server Layer. A REST core handles tasks, authentication, and analytics. Role-based access ensures that students see only their own attempts, while instructors have a full view of class progress.

Execution Layer. Each submission is compiled and executed inside an ephemeral Docker container with strict resource limits; the container is destroyed immediately afterward, guaranteeing security and reproducibility. This separation accelerates feedback for students and frees instructors from manual testing, letting them focus on substantive IPC concepts.

A. Architecture Overview

The system comprises the following components.

Frontend. The user interface is built with React and follows a component-based design [7]. Each screen—task selection, code editor, submission history—is an independent component with well-defined state and props. The Monaco Editor provides in-browser syntax highlighting, bracket matching, and basic error checking. After a solution is submitted, the UI shows its status in real time (queue → compilation → execution) and displays a concise test report when finished. The application is fully responsive, making it equally usable on a campus PC, laptop, or tablet.

Backend. The server uses Spring Boot for its mature ecosystem, built-in REST support, and straightforward database integration. The REST API exposes endpoints for listing tasks, submitting solutions, retrieving feedback, and managing users. Security is two-layered: JWT authentication issued at login and role-based authorization (student or teacher).

Docker Runner. Every solution is evaluated inside an isolated container. The runner pulls the source code, compiles it with gcc, and executes it according to a JSON-defined scenario. After completion, the container is destroyed, ensuring each attempt starts in a clean environment.

Database. PostgreSQL stores all data. The schema includes users (identity, role, invitation code), tasks (problem statement, reference tests), submissions (timestamp, execution state), and results (pass/fail). Indexes on frequently queried columns (user ID, task ID, timestamp) keep look-ups fast even under hundreds of near-simultaneous submissions. Transactions maintain consistency during parallel writes.

Communication Flow. When a student submits a solution, the frontend sends POST /submissions. The backend records the entry, returns the submission ID, and places it in a queue. The runner periodically fetches the first pending record, performs compilation and testing, and then stores the outcome via PATCH /submissions/id. A WebSocket on the client side streams status updates to the interface.

The architecture is designed for long-term maintainability and extensibility. Although the current deployment supports only C and IPC-focused tasks, the database and testing layers are generic, allowing future expansion to other languages and assessment methods. All chosen technologies are well-established and community-supported, minimizing technical debt and simplifying upkeep in an academic setting.

B. User Interface

An intuitive, easy-to-navigate interface is a cornerstone of the entire solution. Everything is arranged clearly and logically so users can focus on what matters—solving tasks, tracking results, or managing students—without endless clicks or searching. The visible controls adapt to the user's role, so each person sees only what is relevant. The goal is to make the system feel natural and stress-free.

1) Student Interface: Students work with assignments directly in the web application. After logging in, they choose a task from the list of available exercises. Each task shows the problem statement, input and output requirements, and any specific limits (e.g., a time-out). Code is written in the built-in editor, which offers syntax highlighting, connection-status indicators, live compilation status, and notifications about test results. Once a solution is submitted, the system automatically evaluates it and instantly displays the outcome, including which tests passed and where errors occurred. The interface also lists previous attempts, allowing students to monitor their own progress.

2) Teacher Interface: The instructor dashboard provides a full set of administrative tools for efficient course management. Teachers can create student groups and assign each a unique registration (invite) code. Sharing this code in class automatically places students in the correct group, giving the teacher a clear view of group composition and each learner's individual progress. The system displays not only the results of individual submissions but also an overall score, making it easier to track performance trends throughout the semester. If needed, a group can be deleted with a single click—for example, after the course ends. Teachers also have unrestricted access to every assignment in the system—no time limits or attempt caps—so they can review the tasks students are working on, gauge their difficulty, and offer targeted guidance.

C. Processing Solutions and Backend Logic

Because the platform's assignments focus on inter-process communication (IPC), students submit their code exclusively in C [7], with direct access to POSIX primitives—pipes, semaphores, shared memory, signals, and sockets.

The backend is built with Spring Boot (Java) [8]. Every submission runs inside an isolated Docker container. Test parameters are stored in simple JSON files: the server loads the configuration, compiles the program, and, after execution, records a concise verdict. Students immediately see which tests passed and how many points were awarded. A typical assignment work is documented in Fig.2.

Fig. 2 Schematic flow of processing a solution: from coding through testing to evaluation.

1) Solution-Processing Workflow: A core feature is in-browser compilation. While working on a task, students can press Run as often as they like; this only checks that the code compiles, without executing tests. No local IDE is needed—writing, debugging, and trial builds all happen in the browser.

When satisfied, the student clicks Submit for final evaluation. The server validates the task, the user's permissions, and the code format, stores the latest version, and queues it for execution.

If the deadline expires, the platform auto-submits the current code so every assignment is graded even if the student forgets to click Submit. This combination of a convenient workspace and a reliable grading engine supports efficient, independent learning.

2) Security Measures during Evaluation: Running code in Docker containers is not just convenient—it solves several practical challenges [9].

- **Security.** Programs run as an unprivileged user in a separate namespace with no access to host files or processes. An infinite loop or crash affects only the container.
- **Uniform environment.** Each image contains the exact library, compiler, and header versions tied to the task, eliminating the "works on my machine" problem.
- **Reproducibility.** Any disputed case can be replayed with the same image and code, simplifying forensic analysis and retesting. Archived tasks need only the Dockerfile or image.
- **Easy maintenance.** Updating the environment means building a new image; older tasks keep their original one, while new tasks inherit fresh libraries, preventing global changes from altering past grades. Containerization is therefore the backbone of a reliable, fair, and maintainable assessment system: programs are tested rigorously and safely without risking the infrastructure [10].

D. Database-model outline

The database is the backbone of the platform, capturing user identity, course structure, and the complete grading trail. Its schema is organized around three logical domains:

1) **Identity and organisation** One table stores core user data plus role, another records teaching groups. A foreign key between them lets the system filter results by class or individual while enforcing that each student belongs to exactly one group.
2) **Task content** Assignments and their test scenarios are stored separately from user data. This separation simplifies versioning, re-using test suites, and adding new exercises without touching the rest of the schema.
3) **Solution and evaluation cycle** Every attempt—source code, intermediate compilations, final verdicts—is logged in linked tables with clear relationships. The design supports both full reconstruction of any attempt and fast aggregation for dashboards and statistics.

This structure delivers three practical benefits:

- **Consistency and fairness** – PostgreSQL foreign keys and transactions ensure that no operation (compilation, result write-back) leaves the database in an inconsistent state, even under parallel grading.
- **Extensibility** – Distinct domains (identity, tasks, evaluation) allow new task types or extra metrics to be added without rewriting existing tables.
- **Performance** – Indexes on hot columns (user ID, task ID, timestamp) keep queries fast even when hundreds of submissions arrive simultaneously. PostgreSQL's strong typing, rich indexing, and trustworthy transactions provide a solid foundation that accommodates new features without accruing technical debt.

E. Authentication and User Management

Secure and reliable identity verification is essential for any academic system. The IPC-practice platform distinguishes two primary user roles—students and teachers—each with its own permissions.

1) Registration and Login: Access requires an account created exclusively with an invite code.

Students receive a code from their teacher.

Teacher use a special code defined in the platform configuration and issued by the system administrator.

During sign-up, the user provides a name, university e-mail address, and password; the password is stored only as a cryptographic hash. Login is performed via the web interface with e-mail and password. After successful validation, the server issues an authentication token that accompanies all subsequent requests.

2) JWT Tokens: The platform employs JSON Web Tokens (JWT) for authentication and authorization. After login, a signed token containing the user ID, role, and expiry time is generated, stored client-side, and sent in an HTTP header with every request.

Key advantages of JWT:

- Stateless server—no session storage required.
- Flexibility—custom claims can be added as needed.
- Broad support in modern front-end frameworks.

3) Role-Based Access Control: A user's role defines system privileges: Students can work on assignments, submit solutions, and view their own results. Teacher can monitor their student group's progress and manage groups (create or delete). Access checks are enforced on the server, preventing any action beyond a user's authorization scope—for example, students cannot view others' grades or modify instructor data.

4) Security Measures: The authentication subsystem follows best practices:

- Passwords are never stored in plain text; secure hashing is mandatory.
- Tokens have limited lifetimes; re-authentication is required after expiry.
- Sensitive operations are protected by dual permission checks to prevent unauthorized use.

V. Assignments and Tests

Students work strictly with the tools and concepts already covered in their university courses—standard C and core POSIX calls—so the assignments should not feel daunting. Instead, they serve as a welcome chance to put freshly acquired knowledge into practical use.

1) **Anonymous pipe**
   Create a pipe, `fork()` the process, and with a single `write()` send the string *"Hello from the parent!"* to the child. Mastery of the trio `pipe() / fork() / read(){write()` is the gateway to the next task.
2) **Signals**
   The child waits three seconds and then sends `SIGABRT` to the parent. Beforehand the parent registers a handler with `sigaction()`, catches the signal, and exits cleanly with code 0. This is the student's first look at how an asynchronous event interrupts a passive wait (`pause()`).
3) **Shared memory**
   When a signal is not enough to transfer a larger block, a segment created with `shm_open()` comes into play. The task is to `mmap()` the segment, write a short text, detach, and remove it with `shm_unlink()`, demonstrating why `shmget()` alone will not do and why permissions matter.
4) **Semaphores**
   Next, a triad of counters is added: one semaphore for the producer, one for the consumer, and one to guard the critical section. Using `semget()` and `semop()`, the student must initialise them so that two processes never overlap while writing to a shared buffer.
5) **Sockets**
   The final exercise combines everything into a miniature TCP *echo* server. The parent creates a socket (`socket()` → `bind()` → `listen()`); when a client connects, it `fork()`s a handler that reads a single line and sends it back.

This progression mirrors the lecture flow: from elementary byte transfer through a pipe, to asynchronous events, shared memory, semaphore synchronisation, and finally a network dialogue. Each step relies only on calls already introduced and prepares the ground for the next, more complex mechanism.

A. Automatic Grading Engine

Every assignment is described by a small JSON file containing the title, time-out, input, expected output, and a list of static and dynamic checks. On submission the code is compiled inside its own Docker container, run with reference input, and the engine verifies:

- error-free compilation,
- exit code 0 and no limit violations,
- output identical to the expected string.

The verdict—*passed* or *failed* plus the score—appears in under a minute, sustaining the *edit* → *run* → *verify* → *edit* rhythm.

B. Competencies Acquired

The incremental task order builds three interlocking layers of skill.

a) Technical layer: Repeated practice cements confidence with

- process management (`fork()`, `exec()`, `waitpid()`),
- asynchronous signalling (`sigaction()`, `kill()`),
- data sharing (`shm_open()`, `mmap()`),
- synchronisation (`semget()`, `semop()`),
- basic networking (`socket()`, `bind()`, `listen()`).

Exposure to pipes, memory segments, and sockets develops the ability to choose the right mechanism for each scenario rather than chasing the "newest library."

b) Methodological layer: Containerised auto-runs deliver feedback in under a minute, reinforcing a tight *test–feedback loop* essential to modern server-side development. A uniform environment for everyone eliminates library- and kernel-version drift, focusing discussion on program logic rather than machine setup.

Together these competencies form a solid foundation for any work that involves concurrency, resource protection, and the need to reproduce results reliably.

VI. Conclusion

The project set out to deliver an integrated environment where students can safely and conveniently practise core IPC mechanisms without any local configuration. After several development iterations and a pilot rollout, the solution has proven effective on multiple fronts:

- **Controlled access via invite codes.** Registration is possible only with a one-time code generated by the instructor for each group, ensuring that only authorised students enter the system, reducing misuse and simplifying progress tracking.
- **Ease of access.** All work happens in the browser: students log in with a university account, choose a task, and immediately write code. No compiler installation, library-version mismatches, or OS incompatibilities interfere with lab time.
- **Rapid feedback.** Each submission is compiled and executed in an isolated container and compared with a JSON-defined reference scenario. A pass/fail verdict appears in under a minute, keeping the *edit* → *run* → *verify* loop tight and productive.
- **Security and isolation.** Containerised execution shields the host system from faulty or malicious programs while still exposing the full POSIX API required for IPC exercises.
- **Scalability.** Runner containers scale dynamically: more instances spin up during peak submission periods, while idle capacity is released during quiet times, ensuring consistent performance without wasting resources.
- **Pedagogical impact.** Automated grading shifts responsibility for debugging to the student. Instructors are freed from manual testing and can focus on deeper topics such as synchronisation design, common pitfalls, and optimisation strategies.

The platform meets its primary goal: it provides a functional, safe, and user-friendly space in which students master the practical side of IPC, while instructors gain an effective tool for reducing administrative load and monitoring progress. By turning a demanding operating-system topic into a systematically organised, hands-on module, the project bridges the gap between theoretical study and real-world application.

References

[1] M. J. Donahoo and K. L. Calvert, *TCP/IP Sockets in C (2nd ed.)*. Burlington, USA: Morgan Kaufmann, 2009.
[2] A. S. Tanenbaum and H. Bos, *Modern Operating Systems (4th ed.)*. Boston, USA: Pearson Education, 2015.
[3] C. M. Arora, *Operating Systems: A Design-Oriented Approach*. Boston, USA: Addison-Wesley, 2000.
[4] M. Kerrisk, *The Linux Programming Interface*. San Francisco, USA: No Starch Press, 2010.
[5] J. S. Gray, *Interprocess Communications in Linux*. Upper Saddle River, USA: Prentice Hall, 2003.
[6] A. Banks and E. Porcello, *Learning React (2nd ed.)*. Sebastopol, USA: O'Reilly Media, 2020.

[7] B. W. Kernighan and D. M. Ritchie, *The C Programming Language (2nd ed.)*. Singapore: Pearson Education Asia, 2002.
[8] C. Walls, *Spring in Action (5th ed.)*. Shelter Island, USA: Manning Publications, 2016.
[9] J. Turnbull, *The Docker Book: Containerization is the New Virtualization*. Melbourne, Australia: James Turnbull Pub., 2014.
[10] I. Miell and A. Hobson, *Docker in Practice*. Shelter Island, USA: Manning Publications, 2016.

Vizualizácia MANET-IoT siete

[1] *Mikuláš Kőveši, Ján PAPAJ*

[1]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]mikulas.kovesi@student.tuke.sk, jan.papaj@tuke.sk

Abstrakt **– Článok sa zaoberá vizualizáciou mobilnej ad hoc siete typu MANET v prostredí internetu vecí (IoT) so zameraním na mestskú oblasť Košíc. Cieľom bolo vytvoriť simulačný model v prostredí NS-3, ktorý zohľadňuje mobilitu vozidiel (SUMO) a útlm signálu spôsobený budovami, pričom vizualizácia prebiehala v nástroji NetAnim. Porovnané boli dva smerovacie protokoly – AODV a OLSR. Model obsahuje stratové modely vhodné pre mestské prostredie. Výsledky ukazujú správanie protokolov pri rôznych hustotách siete a poskytujú odporúčania pre smerovanie v kontexte inteligentných miest.**

Kľúčové slová **– MANET-IoT, vizualizácia, smerovanie, simulátor NS-3**

I. Úvod

Mobilné ad hoc siete (MANET) predstavujú typ bezdrôtových sietí, v ktorých jednotlivé uzly (zariadenia) komunikujú medzi sebou bez potreby centrálnej infraštruktúry. Tieto siete sú schopné dynamicky sa konfigurovať podľa aktuálnej topológie, čo ich predurčuje na využitie v situáciách, kde nie je možné alebo efektívne použiť tradičné sieťové technológie – napríklad v krízových scenároch, v armáde alebo v oblasti internetu vecí (IoT) [1], [2], [3].

S narastajúcim počtom inteligentných zariadení nasadzovaných v mestských prostrediach vzniká potreba realistickej simulácie ich správania v komplexných podmienkach. Kľúčovými faktormi sú mobilita uzlov a vplyv prekážok, ako sú budovy, na šírenie bezdrôtového signálu [4]. Simulácie, ktoré tieto aspekty nezohľadňujú, môžu viesť k nereálnym výsledkom a nesprávnym záverom pri návrhu a optimalizácii sieťových riešení.

Táto práca sa zameriava na vytvorenie simulačného modelu siete typu MANET-IoT v mestskom prostredí Košíc, pričom dôraz je kladený na realistické podmienky mobility a stratového prostredia. Model využíva dopravné dáta generované nástrojom SUMO na simuláciu pohybu vozidiel, pričom samotná sieťová simulácia prebieha v prostredí NS-3 [4], [5]. Porovnávané sú dva smerovacie protokoly – AODV (Ad hoc On-Demand Distance Vector) a OLSR (Optimized Link State Routing) – z hľadiska výkonnostných ukazovateľov ako priepustnosť, strata paketov a priemerné oneskorenie.

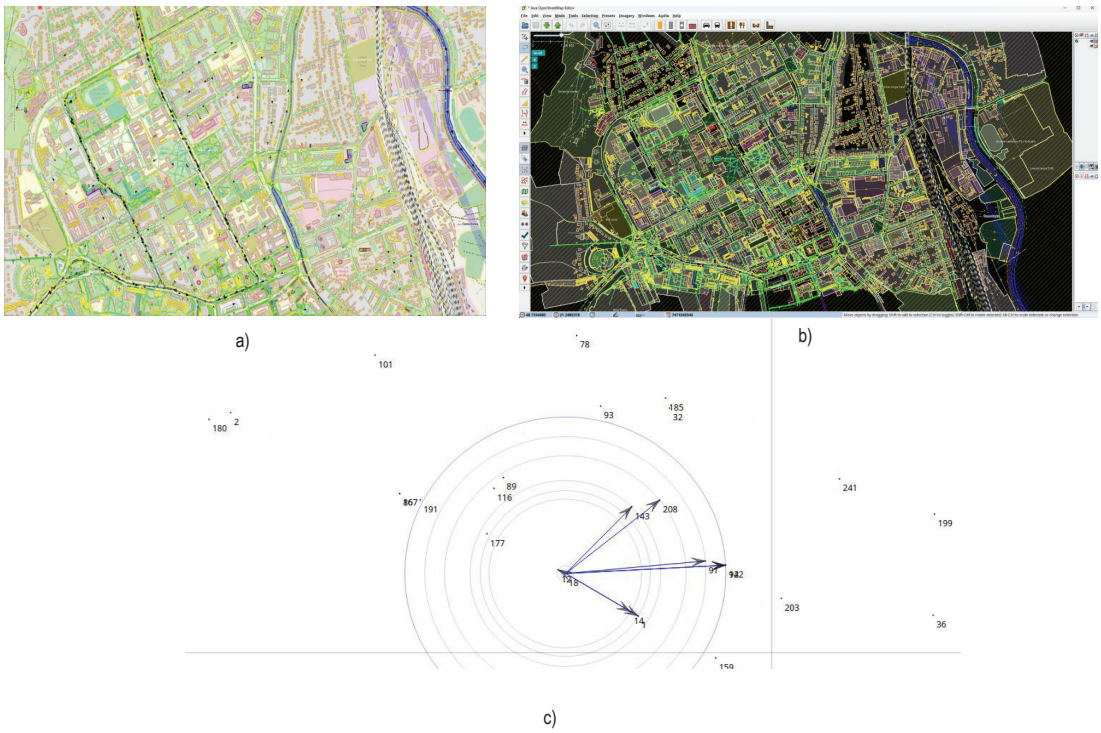
Na zber a vyhodnotenie dát bol použitý nástroj FlowMonitor, vizualizácia prebiehala v programe NetAnime a výsledky boli analyzované pomocou vlastného Python skriptu. Výsledky simulácie prinášajú praktické poznatky o správaní protokolov v rôznych scenároch a môžu slúžiť ako podklad pre výber smerovacieho mechanizmu v kontexte inteligentných miest [4].

A. Použité nástroje a návrh simulačného prostredia

Pri tvorbe simulácie MANET-IoT siete v mestskom prostredí Košíc bolo potrebné integrovať viacero nástrojov na vytvorenie realistického a komplexného modelu (Obr. 1). Simulácia zahŕňala modelovanie pohybu uzlov, vplyvu budov na šírenie bezdrôtového signálu a správania rôznych smerovacích protokolov. Kľúčové nástroje použité v tejto práci boli NS-3, SUMO, NetAnim, OpenStreetMap (OSM) a Java OpenStreetMap Editor (JOSM) [6], [7].

1) Simulátor NS-3: NS-3 je open-source simulačný nástroj určený na výskum a vývoj sieťových protokolov a architektúr. V tejto práci bola použitá verzia 3.39, ktorá poskytuje podporu pre nové Wi-Fi štandardy a realistické modely šírenia signálu [4], [7]. Sieťový model bol vytvorený z uzlov, ktorých mobilita bola importovaná z výstupov nástroja SUMO pomocou modulu Ns2MobilityHelper. Bezdrôtová komunikácia prebiehala na báze štandardu IEEE 802.11n a šírenie signálu bolo modelované pomocou *HybridBuildingsPropagationLossModel*, ktorý umožňuje zohľadniť útlm spôsobený budovami [2].

Na vyhodnocovanie výkonnosti siete bol použitý nástroj FlowMonitor, ktorý generoval údaje o oneskorení, strate paketov a priepustnosti. Výsledky boli exportované do formátov CSV a XML pre ďalšiu analýzu.



Obr. 1 Ukážka z jednolivých simulačných prostredí pre Košice: a) OpenStreetMap,b) Editor JOSM, c) NetAnime

NetAnim je grafický nástroj pre vizualizáciu simulácií vytvorených v NS-3. V tejto práci bola použitá verzia 3.108, kompatibilná s NS-3.39. Vizualizácia prebiehala na základe XML výstupu generovaného triedou *AnimationInterface*. NetAnim umožňuje sledovať pohyb uzlov, vytváranie spojení a výmenu paketov v reálnom čase. Zobrazuje dosah signálu jednotlivých uzlov, umožňuje interaktívne ovládanie prehrávania simulácie a vizuálne overenie funkčnosti siete.

Nástroj bol využitý najmä na kvalitatívne hodnotenie priebehu simulácie, overenie správnosti vstupných údajov a analýzu reakcie protokolov AODV a OLSR v rôznych podmienkach mestského prostredia.

2) Mapové podklady: OSM a editor JOSM: OpenStreetMap (OSM) poskytuje otvorenú databázu mapových údajov, ktorá bola použitá ako základ pre modelovanie infraštruktúry mesta Košice. Na ich úpravu bol využitý editor JOSM, ktorý umožňuje jednoduché čistenie a prispôsobenie mapy – napríklad odstránenie nepodstatných prvkov ako cyklotrasy či chodníky. Upravené dáta boli následne exportované a pripravené na použitie v nástroji SUMO [8].

3) Simulátor dopravnej mobility SUMO: SUMO (Simulation of Urban Mobility) je open-source simulátor určený na modelovanie pohybu účastníkov mestskej dopravy. V tomto projekte bol použitý na generovanie pohybových trajektórií vozidiel na základe reálnej dopravnej siete mesta Košice. Po importe mapových údajov z JOSM a konverzii pomocou netconvert bol pohyb vozidiel náhodne generovaný pomocou skriptu $randomTrips.py$ [6].

Výsledkom bol súbor typu mobility trace, ktorý bol následne importovaný do NS-3. Tento postup umožnil dosiahnuť realistické pohybové scenáre rešpektujúce cestnú sieť, pravidlá dopravy a dynamiku mestského prostredia.

II. Simulovanie MANET siete v prostredí mesta Košice

Simulácia MANET v urbanizovanom prostredí predstavuje komplexný proces, ktorý vyžaduje využitie viacerých softvérových nástrojov, prácu s reálnymi geografickými a dopravnými údajmi, ako aj dôkladné nastavenie simulačných parametrov. Postup navrhnutý v tejto práci nadväzuje na teoretické poznatky uvedené v predchádzajúcej časti, v ktorej boli vysvetlené základné princípy fungovania MANET sietí a predstavené nástroje NS-3, NetAnim, OpenStreetMap a SUMO.

Simulácia bola realizovaná v prostredí operačného systému Ubuntu 24.04, ktorý poskytuje stabilné a flexibilné podmienky na inštaláciu a integráciu spomínaných nástrojov. Práve kompatibilita medzi jednotlivými softvérovými komponentmi je kľúčová pre bezproblémové prepojenie dát, ako aj pre správnu vizualizáciu a vyhodnocovanie výsledkov.

Hlavným cieľom simulačného postupu bolo vytvoriť taký model, ktorý sa čo najviac približuje reálnym podmienkam v mestskom prostredí. To zahŕňa nielen verný geografický podklad mesta

```
UdpEchoServerHelper echoServer(9);
ApplicationContainer serverApps = echoServer.Install(nodes.Get(echoServerNode));
serverApps.Start(Seconds(1.0));

UdpEchoClientHelper echoClient(interfaces.GetAddress(echoServerNode), 9);
echoClient.SetAttribute("MaxPackets", UintegerValue(2000));
echoClient.SetAttribute("Interval", TimeValue(Seconds(0.5)));
echoClient.SetAttribute("PacketSize", UintegerValue(2048));
ApplicationContainer clientApps = echoClient.Install(nodes.Get(echoClientNode));
clientApps.Start(Seconds(2.0));
```

```
OnOffHelper videoStream("ns3::UdpSocketFactory",
    Address(InetSocketAddress(interfaces.GetAddress(videoReceiverNode), 5000)));
videoStream.SetConstantRate(DataRate("500kbps"), 1024);
ApplicationContainer videoApps = videoStream.Install(nodes.Get(videoSourceNode));
videoApps.Start(Seconds(5.0));
```

Obr. 2 Ukážka scriptov pre MANET-IoT v simulátore NS-3

pomocou údajov z OpenStreetMap, ale aj realistické dopravné správanie simulovaných uzlov (vozidiel) na základe SUMO (Simulation of Urban MObility). Tieto údaje následne slúžili ako vstup pre simulačné scenáre v NS-3, v ktorom boli implementované komunikačné protokoly typické pre MANET siete (Obr. 2).

Proces simulácie pozostával z niekoľkých etáp:

- **Získavanie geografických údajov:** Pomocou nástroja OpenStreetMap boli extrahované údaje o cestnej infraštruktúre konkrétneho mestského územia. Tieto údaje boli následne konvertované do formátu podporovaného SUMO pomocou nástroja netconvert.
- **Generovanie dopravného scenára:** V SUMO boli vytvorené realistické dopravné scenáre s využitím nástrojov ako $randomTrips.py$, ktoré umožňujú generovanie náhodných, ale realistických pohybov vozidiel v sieti. Výsledkom bola *.rou.xml* súborová štruktúra, ktorá definovala trasovanie pohybu jednotlivých vozidiel.
- **Prepojenie SUMO s NS-3:** Na základe výstupu zo SUMO boli vygenerované pohybové trajektórie vo formáte kompatibilnom s NS-3. Tie boli spracované prostredníctvom skriptu *sumo2ns3*, čo umožnilo realistickú mobilitu uzlov v rámci simulačného scenára v NS-3.
- **Nastavenie simulačného prostredia v NS-3:** V simulačnom nástroji NS-3 boli implementované potrebné sieťové parametre – typ komunikačného protokolu (napr. AODV, OLSR), a taktiež bol určený počet a rozloženie uzlov v prostredí.
- **Spustenie a vizualizácia simulácie:** Simulácie boli spustené s cieľom analyzovať vybrané metriky výkonnosti siete (napr. priemerné oneskorenie, úspešnosť doručenia paketov, využitie šírky pásma). Na vizualizáciu pohybu a interakcie medzi uzlami bol použitý nástroj NetAnim, ktorý poskytuje vizuálny prehľad o dynamike siete v čase.

Takto definovaný a implementovaný simulačný postup umožnil vytvoriť testovacie prostredie blízke reálnym podmienkam v meste, čo prispelo k vyššej presnosti výsledkov a ich relevantnosti pre budúce aplikácie MANET sietí v reálnom svete – napríklad v systémoch inteligentnej dopravy, záchranných operáciách či pri nasadení autonómnych vozidiel.

Ako základ pre simuláciu bola zvolená reálna mapa mesta Košice, konkrétne oblasť okolo Technickej univerzity a širšie centrum mesta. Táto lokalita bola vybraná pre svoju dopravnú a urbanistickú zložitosť, ktorá zahŕňa hustú zástavbu, viaceré križovatky a členitú cestnú sieť – teda ideálne podmienky pre testovanie mobility uzlov a šírenia signálu v MANET sieti. Mapové dáta boli získané z OpenStreetMap a následne upravené v editore JOSM, kde boli odstránené nerelevantné prvky ako cyklotrasy, parky či vodné plochy. Výsledná dopravná sieť bola prevedená do formátu SUMO pomocou nástroja netconvert a následne overená vizuálne aj pomocou validačných skriptov. Súčasne boli extrahované aj hranice budov, ktoré slúžili ako prekážky ovplyvňujúce šírenie signálu, čím sa zvýšila realistickosť simulácie v prostredí NS-3.

Po úspešnej príprave a validácii cestnej siete v prostredí SUMO nasledovalo generovanie realistického pohybu mobilných uzlov, reprezentujúcich vozidlá alebo zariadenia pohybujúce sa v rámci mestskej infraštruktúry. Namiesto použitia náhodného modelu bola využitá reálna cestná topológia, čím sa dosiahla vyššia presnosť simulácie. Pomocou skriptu $randomTrips.py$ boli vygenerované náhodné, no topologicky korektné trasy, ktoré následne spracoval nástroj duarouter do detailného

```python
import xml.etree.ElementTree as ET

tree = ET.parse("buildings.poly.xml")
root = tree.getroot()

ns3_root = ET.Element("buildings")
BUILDING_HEIGHT = 10.0

for poly in root.findall("poly"):
    shape = poly.get("shape").split(" ")
    x_coords = [float(coord.split(",")[0]) for coord in shape]
    y_coords = [float(coord.split(",")[1]) for coord in shape]
    xmin, xmax = min(x_coords), max(x_coords)
    ymin, ymax = min(y_coords), max(y_coords)

    building = ET.SubElement(ns3_root, "building")
    boundaries = ET.SubElement(building, "boundaries")
    boundaries.text = f"{xmin}, {xmax}, {ymin}, {ymax}, 0, {BUILDING_HEIGHT}"

tree_out = ET.ElementTree(ns3_root)
tree_out.write("buildings.xml")
print("✅ buildings.xml vytvorené úspešne!")
```

Obr. 3 Skript *convert_poly_to_ns3.py* na konverziu polohy budov pre použitie v simuátore NS-3

plánu pohybu. Všetky komponenty boli integrované v konfiguračnom súbore $map2.sumocfg$, ktorý umožnil spustenie simulácie a export podrobných údajov o pohybe (FCD výstup). Tieto dáta boli následne konvertované do formátu kompatibilného s NS-3 pomocou skriptu $sumo2ns3.py$, ktorý vytvoril súbor $sumo_movements.tcl$ pre importovanie cez *Ns2MobilityHelper*. Výsledkom bol presne definovaný a topologicky verný pohyb uzlov v simulácii, ktorý výrazne prevyšuje presnosť tradičných náhodných modelov a umožňuje realistické testovanie sieťovej komunikácie v podmienkach mestského prostredia.

Významným aspektom realistickej simulácie v mestskom prostredí bolo zohľadnenie vplyvu budov na šírenie bezdrôtového signálu. Budovy predstavovali fyzické prekážky, ktoré spôsobovali útlm, odrazy alebo tienenie signálu medzi mobilnými uzlami. Pre ich zapojenie do prostredia NS-3 bolo potrebné extrahovať relevantné mapové dáta z *OpenStreetMap* pomocou nástroja *polyconvert*. Výsledný súbor $map2.poly.xml$ obsahoval polygónové reprezentácie budov, ktoré však neboli priamo kompatibilné s NS-3. Preto bol navrhnutý skript $convert_poly_to_ns3.py$, ktorý tieto polygóny konvertoval na jednoduché obalové obdĺžniky a exportoval ich do formátu $buildings.xml$ (Obr. 3). Po načítaní v NS-3 sa tieto objekty premietli do simulačného prostredia ako samostatné stavebné bloky s definovanými rozmermi a pozíciou, čo umožnilo presnejšie modelovanie šírenia signálu v podmienkach mestskej zástavby.

Po spracovaní dopravnej infraštruktúry, generovaní realistického pohybu z nástroja SUMO a načítaní budov ako prekážok do prostredia NS-3 mohla byť zostavená samotná simulačná topológia. Simulácia bola implementovaná v jazyku C++, ktorý je natívnym programovacím jazykom simulačného nástroja NS-3. Na realizáciu bola použitá verzia NS-3.39, ktorá v čase spracovania poskytovala stabilné prostredie s podporou všetkých potrebných modulov – vrátane mobility modelov, bezdrôtovej komunikácie a smerovacích protokolov. Tento základ umožnil vytvoriť komplexný a realistický model MANET siete v mestskom prostredí.

III. Analyzovanie dosiahnutých výsledkov

Cieľom realizovaných simulácií bolo vytvoriť realistické mestské prostredie pre testovanie MANET-IoT, v ktorom sa overovalo správanie a výkonnosť smerovacích protokolov AODV a OLSR. Simulácie prebiehali v dynamickej sieti bez centrálnej infraštruktúry, kde sa mobilné uzly

Parameter	Hodnota
Počet uzlov	150, 200, 250
Trvanie simulácie	140 s
Fyzikálna vrstva	IEEE 802.11n (2.4 GHz)
Výkon vysielača	30 dBm
Model šírenia vĺn	HybridBuildingsPropagationLossModel + LogDistance (exponent 3.0)
Aplikácie	UDP Echo, videostream (500 kbps), TCP BulkSend (300 000 B)
Smerovací protokol	AODV, OLSR

Tabuľka 1
Simulačné parametre v prostredí simulátora NS-3

Parametre	Protokol	Počet uzlov		
		150	200	250
Odoslané smerovacie pakety [pkt]	OLSR	3061	3142	3340
	AODV	3333	2980	2484
Prijaté smerovacie pakety [pkt]	OLSR	3054	2359	3340
	AODV	3340	2799	2484
Priepustnosť [kbps]	OLSR	139,04	141,81	3340
	AODV	185,93	169,31	2484
Počet stratených paketov [%]	OLSR	24,41	24,14	22,68
	AODV	8,28	12,02	16,08
Oneskorenie doručenia paketov [ms]	OLSR	0.56	6,03	6,48
	AODV	195,2	412,16	432,41

Tabuľka 2
Porovnanie dosiahnutých priemerných hodnôt smerovacích protokolov AODV a OLSR pre MANET-IoT

pohybovali podľa reálne generovaných trás z prostredia SUMO, a ich komunikácia bola ovplyvňovaná prítomnosťou budov ako fyzických prekážok šírenia signálu. Zamerali sme na vytvorenie realistického a opakovateľného testovacieho prostredie pre hodnotenie výkonnosti MANET-IoT v mestských podmienkach. Zámerom bolo porovnať sieťové správanie pri rôznych hustotách uzlov, pričom sa sledovali kľúčové výkonnostné metriky ako priepustnosť, oneskorenie, miera straty paketov a počet úspešne doručených správ. Simulácie vychádzali z reálnej cestnej siete mesta Košice a využívali identické aplikačné zaťaženie, aby sa zabezpečila objektívnosť porovnania. Prostredníctvom tohto prístupu bolo možné analyzovať, ako sa mení efektivita komunikácie v závislosti od počtu aktívnych uzlov v sieti.

Testovacie scenáre vychádzali z reálnych mapových podkladov mesta Košice a boli navrhnuté tak, aby umožnili systematické porovnanie protokolov pri rôznych hustotách siete. Konkrétne boli vytvorené tri varianty s počtom uzlov 150, 200 a 250, pričom všetky ostatné parametre (rozmiestnenie budov, pohybové trajektórie, prenosový výkon a fyzické podmienky) ostali zachované. V každom scenári boli nasadené tri druhy aplikačného zaťaženia: UDP Echo pre meranie oneskorenia, UDP videostream s konštantnou bitovou rýchlosťou (500 kbps) a TCP BulkSend pre simuláciu prenosu objemných dát (viď. Tab. 1).

Výstupom zo simulácie boli podrobné záznamy o sieťovej prevádzke získané prostredníctvom modulu FlowMonitor. Tieto údaje boli následne spracované v jazyku Python, čím sa získali kľúčové výkonnostné metriky ako priemerné oneskorenie, priepustnosť, miera straty paketov a celková efektivita prenosu (priemerný počet odoslaných a prijatých paketov). Navrhnutý simulačný rámec zároveň umožňoval jednoduchú zmenu parametrov bez nutnosti zásahu do jadra topológie, čo zabezpečilo opakovateľnosť experimentov a porovnateľnosť výsledkov medzi jednotlivými protokolmi a scenármi.

Analýza dosiahnutých výsledkov

Výsledky simulácií ukazujú jasný vplyv zvyšujúcej sa hustoty rozloženia mobilných uzlov v sieti na proces vyhľadávania komunikačných ciest a výkonnosť smerovacích protokolov AODV a OLSR (viď. Tab. 2). So zvyšujúcim sa počtom uzlov dochádza k rastu sieťovej záťaže, čo sa prejavuje nárastom oneskorenia a stratovosti paketov, obzvlášť pri reaktívnom protokole AODV. Napriek tomu AODV v náročnejších topológiách vykazuje vyššiu priepustnosť a konzistentnejšie doručovanie dát v porovnaní s OLSR. Tento výsledok naznačuje, že AODV je robustnejší voči dynamickým zmenám v topológii a lepšie využíva dostupnú sieťovú kapacitu v prostrediach s vyššou hustotou.

Naopak, protokol OLSR si v celom spektre testovaných konfigurácií zachoval veľmi nízku latenciu. Táto vlastnosť ho predurčuje na použitie v aplikáciách citlivých na oneskorenie, ako sú hlasové alebo multimediálne služby v reálnom čase. Nižšia spoľahlivosť doručenia dát však predstavuje významnú nevýhodu, najmä v preťažených sieťach, kde OLSR čelí vyššej strate paketov v dôsledku periodického šírenia smerovacích správ a obmedzeného prispôsobenia sa aktuálnym sieťovým podmienkam.

Pri porovnaní výsledkov všetkých troch simulovaných scenárov možno pozorovať, že výkonnosť oboch protokolov je silne závislá od kombinácie sieťovej hustoty a prevádzkového zaťaženia. V prípade strednej hustoty (200 uzlov) sa výkony oboch protokolov zlepšili, pričom AODV profitoval z hustejšieho prepojenia znížením latencie a zachovaním vysokej priepustnosti. OLSR si udržal výhodu nízkej odozvy, no rozdiel v spoľahlivosti doručenia zostal markantný.

Pri extrémnej hustote siete (250 uzlov) sa prejavili limity efektivity oboch protokolov. AODV zaznamenal zvýšené oneskorenie, čo poukazuje na narastajúcu režijnú záťaž pri udržiavaní smerovacích tabuliek a spracovaní požiadaviek na smerovanie. OLSR si síce zachoval nízku latenciu, ale za cenu výrazného poklesu priepustnosti, čo môže byť kritické pri dátovo náročných aplikáciách.

Na základe dosiahnutých výsledkov možno konštatovať, že výber smerovacieho protokolu by mal byť determinovaný konkrétnymi požiadavkami danej aplikácie alebo služby. Protokol AODV je vhodnejší pre scenáre, kde je prioritou spoľahlivé doručenie dát a vysoká priepustnosť, zatiaľ čo OLSR lepšie vyhovuje systémom, kde je nevyhnutná nízka latencia. Optimalizácia výkonu môže byť ďalej podporená adaptívnym prístupom, ktorý zohľadní aktuálne podmienky v sieti a dynamicky volí vhodný smerovací mechanizmus.

IV. Záver

Výsledky simulácií potvrdili, že výkonnosť smerovacích protokolov v MANET je silne ovplyvnená topológiou siete, počtom uzlov a úrovňou prevádzkovej záťaže. Protokol AODV preukázal vyššiu priepustnosť a spoľahlivejšie doručovanie dát v rôznych scenároch, čo ho predurčuje na využitie v prostrediach s vysokými nárokmi na dátovú integritu a stabilitu prenosu. Na druhej strane, OLSR sa osvedčil ako vhodná voľba pre aplikácie s požiadavkou na nízku latenciu, pričom si túto výhodu udržal aj pri výraznom náraste hustoty siete.

Analýza zároveň poukázala na limity oboch protokolov pri extrémnej záťaži, čím sa zvýrazňuje potreba vývoja adaptívnych smerovacích riešení schopných reagovať na dynamické podmienky siete. Vzhľadom na diverzitu požiadaviek moderných IoT a MANET systémov sa javí ako efektívna cesta využitie hybridných alebo inteligentných smerovacích mechanizmov, ktoré dokážu kombinovať výhody proaktívnych aj reaktívnych prístupov.

Budúci výskum by sa mal zamerať na návrh smerovacích algoritmov využívajúcich strojové učenie, ktoré umožnia prediktívne rozhodovanie na základe analýzy prevádzkových vzorov v reálnom čase. Takýto prístup by mohol zásadne zvýšiť efektivitu komunikácie v autonómnych a vysoko dynamických sieťových prostrediach.

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] A. Faiz, "Deep optimization based routing protocol for energy efficient routing in manet-iot," in *2023 International Conference on Data Science and Network Security (ICDSNS)*, 2023, pp. 01–08.

[2] V. Narayandas, R. Dugyala, S. Tiruvayipati, and R. Yellasiri, "Necessity of manet implementation over internet of things: The future of dynamic communication among end devices," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, 2019, pp. 359–362.

[3] V. K. Quy, N. T. Ban, D. Van Anh, N. M. Quy, and D. C. Nguyen, "An adaptive gateway selection mechanism for manet-iot applications in 5g networks," *IEEE Sensors Journal*, vol. 23, no. 19, pp. 23 704–23 712, 2023.

[4] S. Biswas, Y. M. K. P, M. Mrinal, E. Bhardwaj, S. Santhosh, and S. M. I, "The firefly unit: A manet based communication system using zigbee and lorawan protocols for iot applications in remote locations," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2021, pp. 642–646.

[5] B. A. Sadek, E. H. Badreddine, T. E. Houssaine, K. A. Barbosa, and A. J. Rojas, "Consensus congestion control for ad hoc networks: Time-delay and saturation," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 4, pp. 1809–1821, 2023.

[6] R. Monga and D. Mehta, "Sumo (simulation of urban mobility) and osm (open street map) implementation," in *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, 2022, pp. 534–538.

[7] R. I. Al-Essa and G. A. Al-Suhail, "An efficacy of transmission power on dymo routing protocol in vanet," in *2022 International Conference on Engineering & MIS (ICEMIS)*, 2022, pp. 1–6.

[8] T. Mantoro and M. Reza, "Performance analysis of aodv and dsdv using sumo, move and ns2," in *2016 International Conference on Informatics and Computing (ICIC)*, 2016, pp. 372–376.

Digital System for Psychodiagnostics

[1]Lukáš PRONER, [2]Marek HORVÁTH

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

[1]lukas.proner@student.tuke.sk, [2]marek.horvath@tuke.sk

Abstract **– This paper presents the design and implementation of a web platform for supporting the diagnosis and monitoring of mental disorders through digital tools. The solution enables psychologists and mental health professionals to assign, administer, and evaluate psychodiagnostic tests online, either remotely or in clinical settings. In addition to standard questionnaires, the platform introduces a drawing canvas, making it possible to apply selected projective methods in a digital environment. The architecture focus on user data privacy, anonymous client identification, and flexibility. The development was guided by consultations with experts and iterative testing with students and practitioners. The platform represents enhancement in digital psychodiagnostics and offers practical tools for early screening, ongoing assessment, and integration of innovative approaches in mental health practice.**

Keywords **– Digital psychodiagnostics, Mental health assessment, Online diagnosis, Projective methods, Web platform**

I. Introduction

Psychodiagnostics is an essential component in the identification and assessment of mental disorders. Despite the recognized importance of early intervention, current practice often faces obstacles that complicate effective and timely diagnostics. Consultations with professionals indicate that traditional approaches continue to dominate, typically requiring face to face meetings and extensive manual data processing. This reliance on conventional methods can result in increased time demands and administrative burden, which may discourage both patients and practitioners from regular participation in diagnostic processes. Need for more accessible and efficient solutions is evident. Survey data suggest that a large proportion of mental health professionals would prefer to conduct at least part of their diagnostic activities online, which can reduce the time for all involved parties. The shortage of qualified specialists and the growing demand for services mean that not everyone requiring psychological assessment is able to access individual consultations. Delays in diagnosis can have substantial negative consequences, including worsening symptoms, the emergence of disorders, and a higher risk of serious outcomes, as noted by Cheung et al. [1]. Reyes et al. [2] emphasize that constraints related to time, resources, staffing, and finances often lead individuals to seek self testing options before arranging inperson care. The adoption of online diagnostic tools is further motivated by improved reach for geographically dispersed populations, increased efficiency in data handling, and the possibility of automated evaluation, which can streamline the process, reduce costs, and limit human error as described by Joinson et al. [3]. Mirea [4] also highlights that many people experiencing psychiatric symptoms do not seek help at all, which underscores the importance of early detection. For such individuals, access to digital diagnostic resources through a web platform may provide a valuable point of entry into the care process and facilitate earlier intervention.

II. Related Work

The development of digital solutions for psychological assessment has been influenced by several established online platforms, each addressing key challenges in modern psychodiagnostics. One frequently referenced system is *PsyDix.org* [5], which provides access to a broad range of standardized psychological tests covering multiple mental health domains. The platform is recognized for its simple and accessible user interface, which supports efficient navigation and a clear structure of menus, documentation, and help resources. Professionals using *PsyDix.org* do not create custom test items, but rather assign established tests to clients, who then complete them using a unique anonymous code. This system prioritizes privacy and security, ensuring

that personal data is protected throughout the diagnostic process. Analytical features enable automatic scoring and reporting, with professionals able to track client progress over time and export results in PDF format. The use of anonymous identifiers supports both compliance with privacy regulations and efficient handling of large datasets, although the inability to customize test content and the limitation of export options can restrict flexibility for certain clinical or research scenarios.

Another relevant tool is *PsyToolkit.org* [6], a freely available web platform designed for administering psychological tests and cognitive experiments. The system supports both preconfigured and user defined assessments, allowing professionals to adapt or create new tests as required for specific study designs or clinical needs. Users can take advantage of categorized tests or conduct cognitive assessments. *PsyToolkit.org* is especially notable for its research orientation, as it provides a specialized scripting language for creating highly customized test. However, this flexibility comes at the cost of increased technical demands, which may challenge less experienced users. Data export is possible in formats such as *.xlsx*, *.csv*, and *.txt*, but these often contain complex encoded information that may require further processing. The platform does not require registration for general users, which supports accessibility and anonymity, but also places responsibility for data protection on the user if personal identifiers are collected. Educational materials and detailed documentation are readily available, and sharing of tests is streamlined via URLs or QR codes, making the platform valuable for both educational and research contexts.

PsychoPy [7] further expands the possibilities for psychological experiment design, targeting both researchers and educators. The platform allows users to build experiments either visually through a graphical interface or by direct programming in Python. This dual approach offers considerable flexibility, enabling precise control over the experimental design while also lowering the barrier for users with less programming experience. Experiments are saved in XML format for compatibility. While highly effective in controlled research settings, the sophistication of the platform may limit its practical application in everyday clinical environments where speed and ease of use are critical.

Among online tools for survey creation and questionnaire based assessment, *LimeSurvey* [8] is recognized for its extensive functionality and adaptability. The platform enables the creation of diverse question types, such as multiple choice, rating scales, matrix formats, and open ended responses, with support for dynamic adaptation of questions based on previous answers. Users can modify the appearance and structure of their surveys through templates or by editing *HTML* and *CSS*, providing a high degree of visual and functional customization. Although *LimeSurvey* does not provide predefined psychological instruments, it facilitates large scale data collection, export, publication, and analysis.

These platforms demonstrate different strategies for addressing the evolving needs of psychodiagnostics, from prioritizing privacy and user anonymity to supporting flexible test design and large scale data management. These comparisons highlight the ongoing development of web based psychodiagnostic solutions and the importance of balancing research functionality, clinical usability, and data security in contemporary digital practice.

III. Methodology

The development of the web platform for digital psychodiagnostics began with a detailed analysis of current needs among psychologists and psychiatrists. An expert questionnaire confirmed that while there is interest in the possibility of conducting diagnostics online, professionals also expressed considerable reservations. Concerns were primarily focused on the cost and accessibility of digital solutions, the reliability and validity of automated test evaluation, and the protection of sensitive personal data. Security was seen as a fundamental requirement, as any breach could jeopardize the confidentiality of client information. Experts were also aware of the challenges involved in verifying that the person completing an online test is truly the intended client, which is a critical factor in clinical practice.

A distinctive feature of the system is its support for both standard questionnaires and projective methods. A core innovation is the implementation of a drawing environment directly on the platform, enabling projective tasks such as the *tree drawing* or *complete the picture* tests. Figure 1 shows the drawing module as seen by the client, including adjustable pen color and thickness, as well as controls for undo, delete, export, and video or screen recording. This functionality allows for more flexible diagnostic possibilities and provides experts with a broader range of tools for online practice. Literature such as Belák [9] notes that the effectiveness of projective methods often depends on expert supervision, but recent studies [10] suggest that, with proper algorithm design and digital archiving, computer evaluated results can be as informative as traditional expert review. For some projective methods, like the tree drawing test, the final drawing itself may be

Fig. 1 Drawing module interface for projective methods.

sufficient for further analysis. In contrast, techniques such as the *Rorschach test* continue to require the direct presence of an expert, highlighting both the potential and limitations of fully online psychodiagnostics. The recording feature for video and screen adds a further dimension, providing behavioral data that can be reviewed in detail and potentially used for later analysis or training purposes.

Figure 2 illustrates the administrator view for creating and managing test templates, including options for image upload, scaling, and question management, which contribute to the platform overall flexibility. Self assessment tools built into the system can provide the general public with a preliminary understanding of their mental state, while professionals can use the same interface for tailored diagnostic processes. In a controlled setting, such as under professional supervision, the platform has the potential to streamline the assessment workflow and accelerate the availability of results. Looking ahead, integration with advanced technologies, such as artificial intelligence for the automated analysis of drawings or more sophisticated anomaly detection, could further improve the reliability, reach, and diagnostic value of the system.

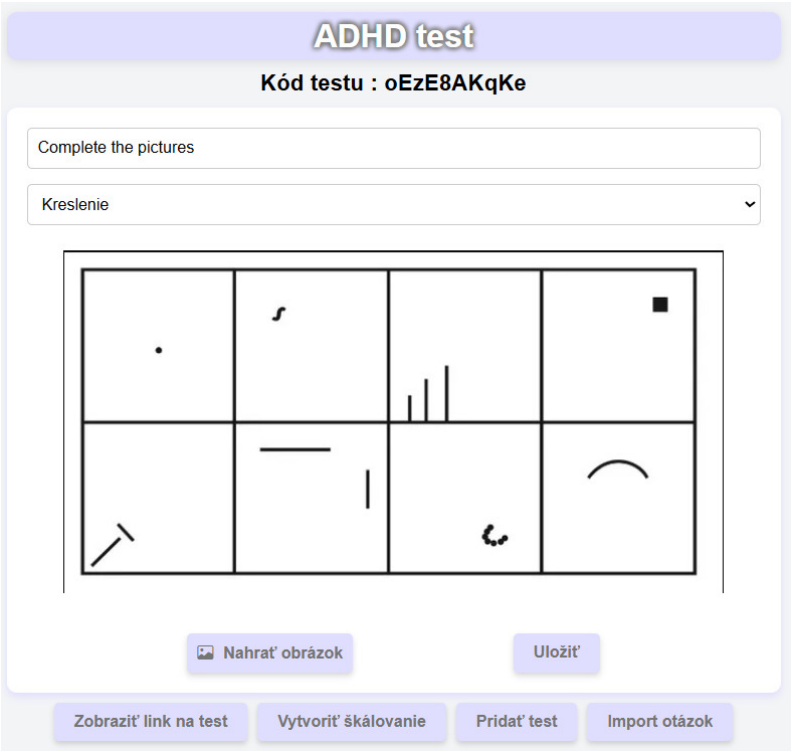


Fig. 2 Administrator interface for test creation.

The platform was also designed to accommodate default self assessment tests for the general public, making it useful for those who may not have access to direct consultations. These tests offer a first insight into mental health for users and can also help experts streamline the diagnostic

process or use provided templates for their own custom tests. Due to the shortage of specialists and growing demand for services, these features help bridge the gap in access and improve efficiency for both users and professionals.

IV. Results and Discussion

The effectiveness and usability of the proposed web platform were evaluated through practical testing with psychology students as well as a practicing psychiatrist, representing both future users and domain experts. Feedback gathered during these sessions highlighted both the strengths of the system and several challenges that remain in digital psychodiagnostics. Students appreciated the simplicity, intuitive navigation, and visual clarity of the platform, especially during the creation and administration of new diagnostic tests.

The interface for drawing projective tasks, as seen in Figure 1, was found to be functional and offered valuable options, such as adjustable pen controls, recording, and export features. However, students also pointed out that using a mouse for drawing projective tests is imprecise and can limit the diagnostic value. They recommended adding support for image upload or for graphic tablets to better replicate natural drawing behavior. Concerns were also raised about screen recording, particularly regarding privacy and the legal implications of recording user activity during assessments. While students acknowledged that online diagnostics will likely play an increasing role in the future, they stressed the importance of clear consent procedures and professional supervision, especially for more sensitive projective methods. For certain types of projective assessments, such as the tree drawing test, students considered remote completion feasible. However, they agreed that tests requiring close observation of client reactions, like the Rorschach test, remain more suited to inperson administration. Overall, the system approach to anonymized client identification, default test templates, and automated data processing was positively received, though students requested further improvements, such as expanded export options, documentation, and additional resources for interpreting results. They suggested that adding detailed test manuals would enhance both the credibility and accessibility of the platform for new users.

The psychiatrist perspective complemented student feedback by emphasizing the importance of maintaining clinical standards and the risks associated with remote testing. He noted that current diagnostic practice still relies heavily on inperson evaluation, given the need to observe subtle nonverbal cues, motor skills, and other contextual factors. According to his experience, these cannot be fully replicated in an online setting. He considered online tools appropriate for initial screenings, especially when working with questionnaires or scales, and noted that a drawing canvas could be valuable for collecting behavioral data if combined with professional supervision. He supported the use of unique identification codes for clients, privacy measures, and the ability to integrate video or screen recording as supplementary data sources. After reviewing the platform, he described its test creation and management functions as intuitive and efficient, echoing student findings. He recognized the system modern design, flexibility for custom test development, and alignment with platforms such as *PsyDix.org*. The SUS questionnaire completed by students resulted in a score of 87.5, confirming the high usability and general acceptability of the system among its intended users.

From a technical standpoint, performance testing identified some scalability concerns. As the number of users increased, so did the latency of requests, indicating a need for further optimization of backend processes and more efficient request handling under load. Additionally, improvements in frontend speed and system response are necessary to support a seamless user experience, particularly for larger or more complex test batteries. Both groups highlighted that the platform assumes basic digital literacy and a stable internet connection, which may present barriers for some populations. Despite these challenges, the results indicate that the platform offers a practical and effective tool for digital psychodiagnostics.

V. Conclusion

Currently, the diagnosis of mental disorders is time consuming and often requires inperson meetings with a specialist. A solution to this problem could be a web platform that enables online evaluation and monitoring of patients mental states. This platform is designed to simplify and accelerate diagnosis through electronic tools and allows patients to fill out tests from the comfort of their homes. The main goal of the platform is to facilitate access to psychodiagnostics and make diagnostic tools available to a broader audience. According to the results, experts would prefer to conduct diagnostics online, as it allows for more efficient data collection and processing while saving time for both them and the patients. The platform also offers self diagnosis options for the general public, which could provide a basic understanding of mental health. The platform

provides various tools for creating and administering psychodiagnostic tests. Professionals can create their own tests, add questions from predefined tests, and assign them to patients. To protect personal data, a system of unique codes for client identification has been implemented. User testing showed that the platform needs improvements in some areas, such as drawing for the diagnosis of projective methods, which was perceived as inaccurate. Another challenge is the legal handling of screen recording while filling out tests and improving data export options. In the future, the use of artificial intelligence for evaluating drawings could be added, which would further increase diagnostic efficiency. The web platform has the potential to improve access to diagnostic services and facilitate the work of professionals.

Acknowledgment

This work was supported by project KEGA No. 061TUKE-4/2025 Building Bridges between University and High School ICT Education.

References

[1] R. Cheung, S. O'Donnell, N. Madi, and E. M. Goldner, "Factors associated with delayed diagnosis of mood and/or anxiety disorders," *Health Promotion and Chronic Disease Prevention in Canada: Research, Policy and Practice*, vol. 37, no. 5, pp. 137–148, 5 2017. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5650019/

[2] A. T. Reyes, R. E. Constantino, R. A. Arenas, J. N. Bombard, and A. R. Acupan, "Exploring challenges in conducting e-mental health research among asian american women," *Asian/Pacific Island Nursing Journal*, vol. 3, no. 4, pp. 139–153, 2018. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC6484146/

[3] A. N. Joinson, K. McKenna, T. Postmes, and U.-D. Reips, *Oxford Handbook of Internet Psychology*. Oxford University Press, 2007.

[4] D.-M. Mirea, N. A. Martin-Key, G. Barton-Owen, T. Olmert, J. D. Cooper, S. Y. S. Han, L. P. Farrag, E. Bell, L. V. Friend, P. Eljasz, D. Cowell, J. Tomasik, and S. Bahn, "Impact of a web-based psychiatric assessment on the mental health and well-being of individuals presenting with depressive symptoms: Longitudinal observational study," *Journal of Medical Internet Research*, vol. 23, no. 2, p. e22738, 2 2021. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7939939/

[5] PsyDix. (2024) Psydix: Psychological tests for automatic scoring and objective results. [Online]. Available: https://psydix.org/en/

[6] G. Stoet. (2024) Faq: Online data. [Online]. Available: https://www.psytoolkit.org/faq.html#onlineData

[7] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv, "Psychopy2: experiments in behavior made easy," *Behavior Research Methods*, vol. 51, no. 1, pp. 195–203, 2019.

[8] LimeSurvey Manual. (2024) Quick start guide - limesurvey 3.0+. [Online]. Available: https://www.limesurvey.org/manual/Quick_start_guide_-_LimeSurvey_3.0%2B/en#Introduction_-_What_is_LimeSurvey

[9] V. Belák, "Projektívne metódy v praxi poradenského psychológa," 2008. [Online]. Available: https://is.muni.cz/th/rpes1/diplomka_FINAL.pdf

[10] T. Ostermann, J. P. Röer, and M. J. Tomasik, "Digitalization in psychology: A bit of challenge and a byte of success," *Patterns*, vol. 2, no. 10, p. 100334, 2021, open access article under CC BY-NC-ND license. [Online]. Available: https://doi.org/10.1016/j.patter.2021.100334

Digitálne dvojča mechatronického systému

[1]*Matej HRIC,* [2]*František ĎUROVSKÝ*

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

[1]matej.hric@tuke.sk, [2]frantisek.durovsky@tuke.sk

***Abstrakt* — Koncept digitálneho dvojčaťa (DD) získal významnú pozornosť v oblasti mechatronických systémov. Ide o virtuálnu repliku fyzického objektu, ktorá zrkadlí správanie a charakteristiky fyzických objektov v reálnom čase. Tento článok prináša komplexný prehľad technológie digitálneho dvojčaťa, porovnanie digitálneho dvojčaťa s inými podobnými technológiami a ukazuje možné riešenia návrhu digitálnych dvojčiat. Prehľad zdôrazňuje výhody v oblasti prediktívnej údržby, optimalizácie výkonu a riadenia životného cyklu. Článok končí návrhom na budúci výskum, pričom zdôrazňuje potenciál digitálnych dvojčiat v mechatronických systémoch.**

***Kľúčové slová* — Digitálne dvojča, mechatronické systémy, data-driven, model-based**

I. Úvod

Simulácia slúži ako modelovací nástroj na vytvorenie infraštruktúry, ktorá monitoruje vlastnosti reálneho systému prenášaním dát z fyzického prostredia do virtuálneho. Umožnením sledovania dát vo výrobných procesoch poskytuje výhody z hľadiska časovej efektívnosti, zníženia nákladov a riadenia rizík. Hlavným cieľom simulácie je predvídať výsledky vo virtuálnom prostredí a uľahčiť potrebné prípravy. Úspešná simulácia sa dosiahne presným replikovaním každého aspektu fyzického systému vo virtuálnom svete [1].

Koncept digitálneho dvojčaťa (DD) sa objavil ako transformačný prístup, ktorý prekonáva priepasť medzi fyzickým a digitálnym svetom v neustále sa vyvíjajúcom technologickom prostredí. „Digitálne dvojča" je virtuálna reprezentácia fyzického systému, procesu alebo objektu, vytvorená zhromažďovaním a integráciou dát v reálnom čase z viacerých zdrojov [2]. Tento digitálny ekvivalent je mocným nástrojom na monitorovanie, analýzu a replikovanie správania jeho reálneho prototypu. Poskytovaním dynamického zrkadlenia fyzických systémov DD zvyšuje porozumenie, umožňuje prediktívne poznatky a podporuje informované rozhodovanie v rôznych priemyselných odvetviach. DD zahŕňa nielen geometrické charakteristiky fyzického objektu, ale aj jeho funkčné atribúty. Používaním reálnych dát zo senzorov, simulácií a historických záznamov poskytujú dynamickú reprezentáciu svojich fyzických náprotivkov. To umožňuje dátovým analytikom a IT špecialistom simulovať a zdokonaľovať návrhy pred výrobou fyzických zariadení. Okrem ich aplikácií vo výrobe hrajú digitálne dvojčatá kľúčovú úlohu pri rozvoji technológií, ako sú Internet vecí (IoT), umelá inteligencia (AI) a dátová analytika [1],[3].

II. Definícia a vývoj

Koncept digitálneho dvojčaťa má svoj pôvod v používaní počítačových modelov NASA na simuláciu a riadenie vesmírnych misií v 60. rokoch 20. storočia [4]. Koncept sa však stal známym až v poslednom období vďaka zlepšeniu snímačových technológií, analýze dát a cloudovými službami. Myšlienka DD vznikla v dôsledku konvergencie rôznych technológií, ktoré umožnili zhromažďovať, integrovať a analyzovať údaje v reálnom čase. DD, ako ho poznáme dnes, predstavil začiatkom 21. storočia Michael Grieves na prednáške o správe životného cyklu produktu [5]. Aj keď bol tento pojem predstavený omnoho skôr, trvalo niekoľko rokov, kým sa objavil v praxi, pričom NASA použila technológiu DD na zrkadlenie podmienok vo vesmíre, aby mohla vykonávať testy na prípravu letov [6].

Počas vývoja modelu DD a jeho vlastností sa objavili rôzne súvisiace koncepty ako napríklad „ultra-high fidelity" [7], „cradle- to- grave" [8], „integrovaný model" [8] a „integrovaný digitálny model (IDMU)" [8]. Tieto pojmy zohrali kľúčovú úlohu pri formovaní vývoja konceptu DD. Avšak vzhľadom na rozšírené prijatie a rastúci význam DD by zavedenie jednotnej a komplexnej definície prinieslo značné výhody. V jadre je DD digitálnym ekvivalentom fyzickej entity. Hoci sa termín „digitálne dvojča" môže javiť ako jednoduchý, jeho definícia je predmetom neustálej

diskusie a zdokonaľovania. Napríklad Abramovici a kol. [9] a Schroeder a kol. [10] považujú DD za konečný produkt, zatiaľ čo Gabor a kol. [12] a Rosen a kol. [13] ho chápu ako koncept pokrývajúci celý životný cyklus produktu. Grieves [14] pôvodne predstavil koncept DD s tromi základnými komponentmi: digitálny (virtuálny) ekvivalent, fyzická entita a spojenie medzi nimi. Tento model sa však odvtedy vyvinul a Tao a kol. [15] ho rozšírili na päť komponentov pridaním dát a služieb ako neoddeliteľných častí DD. Okrem toho Tao a kol. [16] identifikujú overovanie a validáciu ako kľúčové prvky, pričom zdôrazňujú, že dôležitá je integrácia medzi kybernetickým a fyzickým priestorom DD".

Veľmi dôležitou súčasťou popisu technológie DD je porovnanie s inými technológiami, ktoré boli buď jej predchodcami, alebo sú veľmi podobné, pričom toto porovnanie nám umožňuje pochopiť hlavné výhody technológie DD.

Digitálne dvojča stojí na čele reprezentácie reálnych systémov v reálnom čase, poskytujúc dynamický, dátovo riadený digitálny ekvivalent fyzických systémov. Na rozdiel od tradičných simulácií, ktoré sú často statické a špecifické pre konkrétne scenáre, DD neustále aktualizuje dáta zo skutočných senzorov, čo umožňuje prediktívnu údržbu, optimalizáciu výkonu a pokročilé rozhodovanie. Integrácia IoT (Internet of Things) hrá kľúčovú úlohu pri poskytovaní aktuálnych dát do modelov DD, čím zvyšuje ich presnosť a relevantnosť.

Simulácia slúži ako široko využívaný nástroj na analýzu správania systémov za rôznych podmienok bez zásahov do reálneho prostredia [17]. Hoci simulácia aj DD využívajú výpočtové modely, simulácie sa zvyčajne používajú na analýzu diskrétnych udalostí alebo testovanie scenárov bez nutnosti neustáleho prepojenia s fyzickým prototypom. Simulácia je základom inžinierstva, výroby a logistiky, kde je hodnotenie výkonu systému za hypotetických podmienok kľúčové.

Optimalizácia sa zameriava na zlepšovanie parametrov systému s cieľom dosiahnuť maximálnu efektivitu, zníženie nákladov, alebo zlepšenie výkonu. Na rozdiel od DD, ktoré môže zrkadliť systém v reálnom čase, optimalizácia využíva matematické a heuristické algoritmy na určenie najlepšej konfigurácie systému v rámci daných obmedzení. Je široko aplikovaná v alokácii zdrojov a návrhu energeticky efektívnych systémov.

Agent-based modelling (ABM) pre systémy s decentralizovanými interakciami poskytuje jedinečný prístup k simulácii, kde jednotlivé entity (agenti) pracujú autonómne a interagujú v definovanom prostredí. Táto metóda je obzvlášť užitočná na štúdium správania v zložitých systémoch, ako sú dopravné siete, robotika a ekonomické modely [18].

Strojové učenie (Machine Learning- ML) na rozdiel od simulácie a optimalizácie, ktoré sú založené na pravidlách a *umelá inteligencia* (AI) zohrávajú dôležitú úlohu v dátovo riadenom rozhodovaní a automatizácii. ML sa učí zo vzorov vo veľkých súboroch dát a časom sa zlepšuje. ML techniky sa často integrujú do rámcov DD na zlepšenie prediktívnej analýzy, detekcie porúch a autonómneho rozhodovania [19]. Avšak ML je závislé na veľkom množstve tréningových dát a často mu chýba transparentnosť v rozhodovacích procesoch, čo predstavuje výzvy pri kritických aplikáciách.

Digitálny prototyp je ďalším kľúčovým konceptom, ktorý slúži ako virtuálny model produktu alebo systému určený na overenie dizajnu a testovanie pred fyzickou výrobou [20]. Aj keď sa podobá DD v zobrazovaní digitálnej verzie systému, digitálny prototyp je zvyčajne statický a slúži na účely fázy vývoja a testovania. Je široko používaný v produktovom dizajne, automobilovom inžinierstve a leteckom priemysle na urýchlenie inovácií a zníženie nákladov.

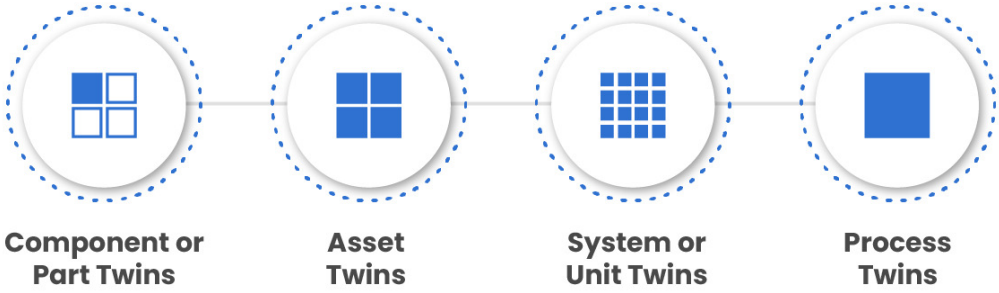
Autonómne systémy- ako sa priemysel posúva smerom k vyššej úrovni autonómie, získavajú na čoraz vyššom význame. Tieto systémy fungujú s minimálnym ľudským zásahom a často sú podporované AI a algoritmami rozhodovania v reálnom čase. Autonómne vozidlá, priemyselné roboty a inteligentná infraštruktúra využívajú kombináciu ML, simulácie a technológie DD na zvýšenie adaptability a bezpečnosti. Avšak zabezpečenie spoľahlivosti v nepredvídateľných reálnych podmienkach zostáva významnou výzvou.

Hoci všetky tieto technológie majú prekrytie v určitých aplikáciách, ich hlavné funkcie sa mierne líšia. DD je nevyhnutné pre integráciu systémov v reálnom čase, zatiaľ čo simulácia, optimalizácia a ABM slúžia ako analytické a prediktívne nástroje. ML zvyšuje dátovo riadenú inteligenciu, digitálny prototyp podporuje overenie dizajnu v raných fázach a autonómne systémy predstavujú budúcnosť nezávislého rozhodovania a prevádzky. Interakcia medzi týmito technológiami poháňa inovácie v oblasti inteligentného priemyslu, digitálnej výroby a návrhu inteligentných systémov, čím zdôrazňuje potrebu interdisciplinárneho výskumu v inžinierstve a informatike. Tab. *1* uvádza prehľad porovnania týchto technológii a ich kľúčových vlastností.

Na lepšie pochopenie technológie DD je dôležité opísať aj rôzne typy digitálnych dvojčiat, ktoré sú znázornené na Obr. *1*. Hlavný rozdiel medzi nimi spočíva v rozsahu pokrytia. Zatiaľ čo niektoré riešenia môžu predstavovať najmenšie jednotky výrobnej linky, iné môžu predstavovať celý fabriku a jej výrobný proces.

Tab. 1Porovnanie technológie digitálneho dvojčaťa s ostatnými technológiami

Technológia	Charakteristické vlastnosti
Digitálne dvojča	Real- time, vzájomne prepojené
Simulácia	Statická, jednosmerná
Optimalizácia	Statická, hľadá najlepšie riešenie
Agent Based Modelling	Statické, bez prepojenia
Strojové učenie	Bez prepojenia
Digitálny prototyp	Statický, najjednoduchšia forma
Autonómne systémy	Bez samo-učenia (len na základe minulých výsledkov)



Obr. 1 Typy digitálnych dvojčiat [21]

Dvojča komponentu (Component twin), nazývané aj dvojča dielov, predstavuje základnú úroveň technológie DD. Zodpovedá najmenším prvkom v rámci systému, ako je senzor, spínač, ventil alebo iný jednotlivý komponent zariadenia alebo produktu. Vytvorením virtuálnej reprezentácie týchto inteligentných častí je možné monitorovať ich výkon a stav a následne simulovať podmienky v reálnom čase, čo umožňuje komplexné testovanie ich odolnosti, stability a efektivity [22]. Hoci tento typ DD replikuje výrobný proces v menšom meradle, významne zvyšuje schopnosť monitorovať komponenty zariadenia a zabezpečiť ich včasnú údržbu. To následne prispieva k celkovej stabilite výrobného procesu a pomáha udržiavať vysokú kvalitu produktov.

Produktové dvojča (Product twin) predstavuje ďalšiu úroveň hierarchie tejto technológie. Zvyčajne sa skladá z viacerých komponentových dvojčiat, alebo využíva dáta na vytvorenie virtuálneho modelu zložitejšieho systému, ako je motor, čerpadlo alebo celá budova. Poskytovaním podrobnej reprezentácie systému umožňuje produktové dvojča analýzu toho, ako jednotlivé komponenty spolupracujú a fungujú ako jednotný systém. Zavedenie produktového DD umožňuje inžinierom získať cenné poznatky o výkonnosti zariadení a identifikovať oblasti na potenciálne zlepšenie. To následne podporuje zvýšenie produktivity, optimalizuje kľúčové prevádzkové ukazovatele ako priemerný čas medzi poruchami (Mean time between failures- MTBF) a priemerný čas opravy (Mean time before repair- MTTR) a prispieva k efektívnejšiemu využívaniu zdrojov [23].

Systémové dvojča (System twin) funguje na úrovni systémov, replikujúc spôsob, akým viaceré systémy spolupracujú a fungujú spolu ako jednotky. Poskytovaním veľkoplošného pohľadu na celú továreň alebo závod umožňujú systémové dvojčatá testovanie rôznych konfigurácií systémov s cieľom maximalizovať efektivitu a objavovať nové obchodné príležitosti vrátane potenciálnych zdrojov príjmov. Systémové dvojča zahŕňa sieť systémov zapojených do konkrétneho procesu, ako je distribúcia energie alebo výroba určitého produktu [23]. Na rozdiel od jednoduchších DD zameraných na jednotlivé komponenty alebo procesy poskytujú systémové dvojčatá monitorovanie v reálnom čase a pokročilé simulácie, ktoré ponúknu viac ako len detekciu porúch. Namiesto toho ponúkajú cenné poznatky pre strategické rozhodovanie, umožňujúce podnikom optimalizovať prevádzku, zlepšiť prehľad o procesoch a podporiť dlhodobé zlepšenie efektivity a produktivity.

Procesné dvojčatá (Process twins)- Na najvyššej úrovni technológie DD spájajú viacero systémových dvojčiat do jednotného modelu, ktorý poskytuje komplexný pohľad na to, ako rôzne systémy spolupracujú a synchronizujú sa. Tento pokročilý prístup ponúka najrozsiahlejší pohľad na pracovné toky a procesy v závode alebo fabrike, čo umožňuje hlbšiu a dynamickejšiu analýzu celkového výkonu a výstupov. Procesné dvojča umožňuje nastavenie kľúčových vstupných premenných, ako sú rýchlosti dávkovania surovín alebo výrobné teploty, pričom zároveň zhromažďuje údaje o výsledných výstupoch. To je možné vykonať bez prerušenia skutočných výrobných operácií alebo zníženia kvality produktu. V dôsledku toho môžu manažéri bezpečne a efektívne testovať rôzne obchodné scenáre, monitorovať kľúčové výkonnostné ukazovatele a robiť informované, dátovo podložené rozhodnutia namiesto spoliehania sa na intuíciu alebo odhady [22].

III. NÁVRH DIGITÁLNEHO DVOJČAŤA

Na to aby bolo možné navrhnúť digitálne dvojča, je dôležité si zvoliť fyzický objekt a zadefinovať ciele ktoré chceme dosiahnuť s využitím tejto technológie. Vybrať možno z dvoch rozdielnych prístupov kde každý ponúkne jedinečné výhody v závislosti od typu aplikácie a dostupných dát.

A. Modelovo orientované digitálne dvojča

Modelovo orientovaný prístup k simulácii (MOPS) je štruktúrovaná metodológia na definovanie požiadaviek, návrh, analýzu a overovanie komplexných systémov [24]. Zameriava sa na používanie modelov na reprezentáciu fyzických systémov, experimentálnych alebo aplikačných, ktoré pozostávajú z prepojených komponentov vykonávajúcich špecifické funkcie. Simulácia takýchto systémov pomocou MOPS vychádza zo základných fyzikálnych zákonov a inžinierskych princípov. Silnou stránkou MOPS je hlboké porozumenie systému alebo procesu, ktoré využíva dobre zavedené vedecké vzťahy.

Modelovo orientované DD rozširujú MOPS o vylepšené snímače a nástroje poháňané umelou inteligenciou. Rôzne štúdie dokumentujú implementáciu modelovo orientovaných DD v rôznych aplikáciách. Korondi a kol. [25] vyvinuli DD mechatronického systému založeného na modeli optimálneho riadenia BLDC motora. Bachelor a kol. [26] aplikovali modelovo orientované DD na systém ochrany lietadla proti námraze. Ward a kol. [27] navrhli DD pre CNC obrábací stroj pomocou MATLAB/Simulink.

Tieto príklady zdôrazňujú, že modelovo orientované DD sú najvhodnejšie pre systémy s dobre definovanými fyzikálnymi vlastnosťami, pričom sa spoliehajú na konvenčné modelovacie a simulačné platformy, techniky umelej inteligencie a integráciu internetu vecí. Ich účinnosť je však obmedzená potrebou zjednodušovania modelov, keďže nie je možné úplne zachytiť vlastnosti systémov. Okrem toho majú problémy so zohľadnením neznámych premenných a dátového šumu, čo predstavuje významné výzvy v aplikáciách s vysokou zložitosťou.

B. Dátovo orientované digitálne dvojča

Implementácia DD umožňuje operátorom monitorovať výrobu, testovať odchýlky v kontrolovanom virtuálnom prostredí a zvyšovať bezpečnosť v procesnom priemysle [28]. Avšak s exponenciálnym rastom objemu procesných dát, tradičné modelovo orientované prístupy nedokážu presne popísať komplexné stavy systémov. Výsledkom je, že modelovanie na základe dát sa objavilo ako schopná alternatíva pre vývoj DD. Tento prístup identifikuje vzťahy medzi vstupnými a výstupnými premennými systému bez potreby presných znalostí fyzikálneho správania systému. V porovnaní s konvenčnými empirickými modelmi ponúkajú prístupy založené na dátach významné pokroky naprieč rôznymi aplikáciami. Modelovanie na základe dát závisí od veľkých, kvalitných dátových súborov, ktoré charakterizujú systém. Tieto dáta umožňujú kľúčové úlohy, ako sú klasifikácia, rozpoznávanie vzorov a prediktívna analýza.

Literatúra dokumentuje rozsiahlu aplikáciu dátovo orientovaných DD, najmä v komplexných systémoch. Wang a kol. [29] vyvinuli DD na základe dát pre trojdoménový systém mobility – zahŕňajúci ľudí, vozidlá a dopravu – využívajúc Amazon Web Services (AWS). Nwoke a kol. [30] navrhli DD založené na FPGA na monitorovanie mechatronických systémov. Táto konfigurácia eliminuje potrebu veľkých dát a cloudového úložiska, čím zabezpečuje ochranu dát a rýchlejšiu, efektívnejšiu implementáciu a aktualizáciu. Článok od Mendiho a kol. [31] popisuje prípadovú štúdiu digitálneho dvojčaťa na automobilovej výrobnej linke, zdôrazňujúc výhody DD pri optimalizácii celého výrobného procesu. Tieto štúdie zdôrazňujú silnú koreláciu medzi dátovo orientovanými DD a vysoko komplexnými systémami s obrovským množstvom dát. Okrem toho poukazujú na to, že AI a platformy Big Data sa bežne používajú pri vývoji dátovo orientovaných DD, čím sa odlišujú od tradičných modelovo orientovaných prístupov. Základné porovnanie dátovo orientovaných a modelovo orientovaných digitálnych dvojčiat je uvedené v Tab. *2*.

Tab. 2 Porovnanie Modelovo orientovaného a dátovo orientovaného DD

Porovnanie	**Modelovo orientované DD**	**Dátovo orientované DD**
Základ	Matematické rovnice	Zozbierane dáta
Cena	Drahšie	Lacnejšie
Doba návrhu	Kratšia	Dlhšia
Aplikácie	Modelovanie systémov (pri návrhu)	Komplexné systémy

Dátovo orientované a modelovo orientované digitálne dvojčatá slúžia na rôzne účely v závislosti od aplikácie. Modelovo orientované DD, ktoré sa spoliehajú na simulácie založené na fyzikálnych princípoch, sú obzvlášť užitočné pri návrhu a optimalizácii nových produktov, pretože umožňujú inžinierom testovať a dolaďovať správanie systému ešte pred výrobou

produktu. Naopak, dátovo orientované DD, ktoré využívajú strojové učenie a analýzu veľkých dát, sú vhodnejšie pre prediktívnu údržbu a detekciu anomálií, pretože sa dokážu neustále učiť zo senzorových dát v reálnom čase a identifikovať vzorce a potenciálne poruchy. Kým modelovo orientované DD poskytujú hlboké porozumenie dynamike systému, dátovo orientované DD vynikajú pri zvládaní komplexných prostredí s vysokou variabilitou, kde je presné fyzikálne modelovanie náročné. V mnohých prípadoch ponúka najkomplexnejšie riešenie hybridný prístup, ktorý kombinuje oba prístupy, avšak jeho návrh je tiež zložitejší.

IV. Záver

Digitálne dvojča zohráva kľúčovú úlohu v mechatronických systémoch. Spája výhody, ktoré môžu pomôcť pri návrhu a monitorovaní takýchto systémov. V súčasnosti sa zameriavame na voľbu mechatronického systému, pravdepodobne letmých nožníc, ktoré sa používané na strihanie materiálu na požadovanú dĺžku v spracovateľských linkách. Tento proces je veľmi krátky a dynamický, a preto je veľmi náročné správne nastaviť riadiaci algoritmus. Vyvinutím digitálneho dvojčaťa letmých nožníc by sme sa chceli zamerať na možnosti prediktívnej údržby, najpravdepodobnejšie na monitorovanie otupenia noža, pretože strihací nôž zohráva dôležitú úlohu v takýchto aplikáciách a môže ovplyvniť kvalitu, efektivitu strihacieho procesu a náklady na finálny produkt. Na implementáciu technológie bude potrebné monitorovať dôležité parametre počas strihacieho procesu – strižnú silu a krútiaci moment motora. Na základe týchto parametrov budeme schopní určiť stav noža včas a zabezpečiť jeho výmenu v prípade potreby.

Poďakovanie

Táto práca bola podporená z projektového grantu pre mladých vedcov na TUKE, číslo projektu 05/TUKE/2025: Návrh digitálneho dvojčaťa letmých bubnových nožníc ako platformy pre vývoj riadiacich algoritmov a potreby prediktívnej údržby s využitím nástrojov umelej inteligencie.

Financované EÚ prostredníctvom iniciatívy NextGenerationEU z Plánu obnovy a odolnosti Slovenskej republiky v rámci projektu č. 09I05-03-V02-00018.

References

[1] O. Korhan, M. Fallaha, Z. M. Çınar, and Q. Zeeshan, The Impact of Industry 4.0 on Ergonomics, London, UK: Intech Open, 2023. doi: 10.5772/intechopen.108864

[2] Crespi N, Drobot AT, Minerva R. The digital twin: What and why? In: N. Crespi, A. T. Drobot, and R. Minerva, "The digital twin: What and why?," in The Digital Twin, Cham: Springer, 2023, pp. 3–20.

[3] M. Soori, B. Arezoo, and R. Dastres, "Digital twin for smart manufacturing, a review," Sustainable Manufacturing and Service Economics, vol. 2, 100017, 2023. doi: 10.1016/j.smse.2023.100017.

[4] E. H. Glaessgen and D. S. Stargel, "The Digital Twin Paradigm for Future NASA and US Air Force Vehicles," in AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Honolulu, HI, USA, 2012, pp. 2012–4343

[5] Grieves, Michael. "Digital twin: manufacturing excellence through virtual factory replication." White paper 1.2014 (2014): 1-7.

[6] Shafto, Mike, et al. "Draft modeling, simulation, information technology & processing roadmap." Technology area 11 (2010): 1-32.

[7] K. Reifsnider and P. Majumdar, "Multiphysics stimulated simulation digital twin methods for fleet management," in 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2013. doi: 10.2514/6.2013-1578

[8] E. Tuegel, "The Airframe Digital Twin: Some Challenges to Realization," in 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 2012. doi: 10.2514/6.2012-1812

[9] J. Ríos, J. C. Hernández, M. Oliva, and F. Mas, "Product avatar as digital counterpart of a physical individual product: Literature review and implications in an aircraft," Advances in Transdisciplinary Engineering, vol. 2, pp. 657–666, 2015. doi: 10.3233/978-1-61499-544-9-657

[10] M. Abramovici, J. C. Göbel, and H. B. Dang, "Semantic data management for the development and continuous reconfiguration of smart products and systems," CIRP Annals–Manufacturing Technology, vol. 65, pp. 185–188, 2016. doi: 10.1016/j.cirp.2016.04.051

[11] G. N. Schroeder, C. Steinmetz, C. E. Pereira, and D. B. Espindola, "Digital Twin Data Modeling with AutomationML and a Communication Methodology for Data Exchange," IFAC-PapersOnLine, vol. 49, pp. 12–17, 2016. doi: 10.1016/j.ifacol.2016.11.115.

[12] T. Gabor, L. Belzner, M. Kiermeier, M. T. Beck, and A. Neitz, "A simulation-based architecture for smart cyber-physical systems," in 2016 IEEE International Conference on Autonomic Computing (ICAC), pp. 374–379.

[13] Rosen, R., Von Wichert, G., Lo, G., Bettenhausen, K.D., 2015. About the importance of autonomy and digital twins for the future of manufacturing, in: IFAC-PapersOnLine, pp. 567–572. doi: https://doi.org/10.1016/j.ifacol.2015.06.141.

[14] M. Grieves, "Digital Twin: Manufacturing Excellence through Virtual Factory Replication," White Paper, 2015. [Online]. Available: http://www.apriso.com/library/Whitepaper_Dr_Grieves_Digital_Twin_ManufacturingExcellence.php.

[15] Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F., 2018. Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology 94, 3563–3576. doi: 10.1007/s00170-017-0233-1.

[16] Tao, F., Zhang, H., Liu, A., Nee, A.Y., 2019. Digital Twin in Industry: State-of-the-Art. IEEE Transactions on Industrial Informatics 15, 2405–2415. doi:10.1109/TII.2018.2873186

[17] Martin, D. (2023, October 17). Simulation 101—What Is It, and How Can It Benefit You? PTC. https://www.ptc.com/en/blogs/cad/simulation-101-what-is-it-and-how-can-it-benefit-you
[18] Salgado, M., & Gilbert, N. (2013). Agent Based Modelling. In T. Teo (Ed.), Handbook of Quantitative Methods for Educational Research (pp. 247–265). Sense Publishers. https://doi.org/10.1007/978-94-6209-404-8_12
[19] Ayodele, T. O. (2010). Machine Learning Overview. In New Advances in Machine Learning. IntechOpen. https://doi.org/10.5772/9374
[20] Autodesk. (2009). Digital Prototyping: Design. Visualize. Simulate. Retrieved from https://images.autodesk.com/adsk/files/dp10_indmach_industry_bro_final.pdf
[21] Vidyatec. (n.d.). The 4 levels of the Digital Twin technology. Retrieved February 27, 2025, from https://vidyatec.com/blog/the-4-levels-of-the-digital-twin-technology/
[22] Program-Ace, "Types of digital twins: Explained in detail," Program-Ace. [Online]. Available: https://program-ace.com/blog/types-of-digital-twins/. [Accessed: 22-Feb-2025].
[23] IBM, "What is a digital twin?" IBM. [Online]. Available: https://www.ibm.com/think/topics/what-is-a-digital-twin. [Accessed: 22-Feb-2025]
[24] S. Kousar, F. Rehman, and M. A. Khan, "Overview of Digital Twin Platforms for EV Applications," ResearchGate, Jan. 2023. [Online]. Available: https://www.researchgate.net/publication/367465526_Overview_of_Digital_Twin_Platforms_for_EV_Applications.
[25] P. Korondi, T. Haidegger, and L. Kovács, "Digital twin of mechatronic drive based on the optimal control model of BLDC motor," ResearchGate, Dec. 2020. [Online]. Available: https://www.researchgate.net/publication/347184908_Digital_twin_of_mechatronic_drive_based_on_the_optimal_control_model_of_BLDC_motor.
[26] G. Bachelor, E. Brusa, D. Ferretto, and A. Mitschke, "Model-Based Design of Complex Aeronautical Systems Through Digital Twin and Thread Concepts," IEEE Systems Journal, vol. 14, no. 2, pp. 1568–1579, Jun. 2020. doi: 10.1109/JSYST.2019.2925627.
[27] Ward, R., Sun, C., Dominguez-Caballero, J. et al. Machining Digital Twin using real-time model-based simulations and lookahead function for closed loop machining control. Int J Adv Manuf Technol 117, 3615–3629 (2021). https://doi.org/10.1007/s00170-021-07867-w
[28] S. Tao, J. Qi, and X. Liu, "Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: A Review," Journal of Manufacturing Systems, vol. 58, pp. 346-361, Jan. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827120314797?via%3Dihub
[29] Ziran Wang. Mobility Digital Twin with Connected Vehicles and Cloud Computing. TechRxiv. February 01, 2022
[30] J. Nwoke, M. Milanesi, J. Viola and Y. Q. Chen, "FPGA-Based Digital Twin Implementation for Mechatronic System Monitoring," 2023 5th International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 2023, pp. 1-6, doi: 10.1109/IAI59504.2023.10327502.
[31] Mendi, A.F. A Digital Twin Case Study on Automotive Production Line. Sensors 2022, 22, 6963

Identifying Industrial Processes from Time Series Data

[1]*Roman HUDÁK,* [2]*Ján MAGYAR*

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

[1]roman.hudak@student.tuke.sk, [2]jan.magyar@tuke.sk

Abstract **– This work presents a general solution for analyzing electrical signals in industrial systems, focusing on anomaly detection and process pattern identification. We compare, test, and evaluate multiple embedding methods for feature extraction and clustering methods for the identification of operational modes. Comparative analysis of these approaches revealed that the combination of statistical embedding with the DBSCAN algorithm provides the best results when applied to real data from a galvanic copper plating machine. The implemented methodology enables reliable anomaly detection and the identification of typical operational modes, while supporting the analysis of transitions between them, providing a comprehensive view of the dynamics of the monitored system.**

Keywords **– anomaly detection, cluster analysis, DBSCAN, embedding**

I. Introduction

Machine learning has found many applications in industrial process supervision, from proactive control [1], through predictive analysis [2], to predictive maintenance [3][4]. These approaches facilitate the detection of changes in how machinery operates, and a timely discovery of potential faults. The use of predictive maintenance minimizes unplanned stoppages, extends a device's longevity and optimizes operational costs.

In this paper, we define a universal solution that can be applied to detect various operational modes of any machine based on energy consumption. During deployment, the method is able to identify normal operational cycles and to detect deviations from these processes. Our approach was tested on data from the observation of galvanic copper plating, but could be applied in any industrial setup where the required input data are available. The main advantage of our proposed method is that it is not necessary to set the number of modes, instead, the algorithm derives it automatically. We compare the combination of multiple embedding and clustering methods, and evaluate their performance for precision in anomaly detection and the detection of industrial processes.

The paper is structured as follows: Section II provides an overview of relevant research in embeddings and clustering. Section III introduces our methodology, the various approaches tested and how they were evaluated on sample data. Final deployment of our solution is described in Section IV, Section V concludes the paper.

II. Literature Review

Energy consumption constitutes time series data, which contain information on the behavior and performance of energetic systems. These signals are characterized by a high rate of variability and non-linearities, complex patterns, daily and seasonal cycles, and sudden amplitudes caused by change in supply or demand. This variance makes it difficult to accurately identify changes caused by an outage or malfunction. When processing such data, it is key to transform raw data into the proper representation and to analyze patterns to determine normal system behavior and divergences from typical patterns.

A. Embedding Methods

Embedding is a technique used to transform complex data structures into a more compact vector space [5]. In energy data, the main goal is to change the representation in a way that makes it possible to describe changes mathematically, and analyze them using advanced analytical methods. On the other hand, the transformation must preserve key time, frequency and statistical characteristics of the signal, while also eliminating redundancies and noise [6]. Embedding methods

can be sorted into a number of categories, such as statistical methods, transformational methods, feature-based methods, and non-linear reduction of dimensionality.

Statistical embedding describes data through their mathematical characteristics and distribution, and stands out with its interpretability and computational efficiency. The main approaches in this group are autoregressive models, which describe data as a linear combination of previous values; ARIMA models, which introduce moving average [7]; exponential smoothing methods, which describe data using recursive actualization and take into consideration seasonality [8].

Transformational embedding maps time series into alternative spaces, where it is more efficient to capture their characteristics. The most frequently used methods in this group are Fourier transform [9], wavelet transform [10], and Empirical Mode Decomposition (EMD) [11]. Fourier transform is a basic approach that cannot detect when a change in frequency occurs, this limitation is addressed both by wavelet transform and EMD.

Feature-based approaches transform raw time series data into vectors of descriptive statistics that capture key signal aspects. These features can be basic statistical metrics such as mean, median, and standard deviation [12], or can be based on time delays such as in Pattern-based Vectorization for Time Series [13]. The method introduced in [14] combines time, frequency, and time-frequency domains to extract the most descriptive features and to capture all relevant characteristics.

Non-linear dimensionality reduction focuses on capturing non-linearities in time series data using methods such as t-Distributed Stochastic Neighbor Embedding (t-SNE) [6], Uniform Manifold Approximation and Projection [15], and autoencoders [16].

B. Clustering Methods

Clustering algorithms are used to identify natural groups in data samples based on the similarity of examples [17]. A data sample is divided into groups so that examples within a group are more similar to each other than to members of other groups. To this end, the similarity or distance between two examples must be clearly defined for the clustering algorithm to be able to quantify their mutual connection. In time series data, it is used to find hidden patterns and anomalies, which is possible only when temporal characteristics such as trends, periodicity, and dynamic features are preserved in the data.

The most basic clustering algorithms, such as k-means [17], k-medoids, and fuzzy c-means [18], use the distance between two arbitrary examples from the dataset to identify clusters. This is done in an iterative way that stops only once the algorithm converges to a certain number of clusters. This number must be set by the user, which limits the methods' applicability.

Clustering can also be done in a hierarchical manner by building tree-shaped clusters. Agglomerative clustering defines the distance between two clusters as the minimal distance between any two members of these clusters [19], complete linkage [20], on the other hand, considers the maximal distance. Another approach is represented by Ward's method, which minimizes within-cluster variance when iteratively merging clusters until conversion [19].

Another approach to clustering is through considering the density of examples in the feature space, meaning that areas with a high density of samples can be considered clusters, which in turn can be separated by areas with low density. One of the basic algorithms in this category is Density-Based Spatial Clustering of Applications with Noise (DBSCAN), that identifies each data sample as a core point, border point, or outlier [21]. Clusters are built by extending them from core points. The Ordering Points To Identify the Clustering Structure (OPTICS) algorithm builds on this by considering the points' reachability [22]. Another algorithm, Hierarchical DBSCAN (HDBSCAN) combines the principles of hierarchical and density-based clustering, while also eliminating some parameters of the base algorithm [23].

III. Experiments

In the first phase of our research, we tested various combinations of embedding and clustering methods on real energy consumption data from an industrial process. To this end, we used the ReSense power device from the company RE-CA, which was configured to monitor all three phases of a galvanic copper plating machine. Measurements were taken with a frequency of 1Hz, the measurement lasted one month, resulting in around 2.6 million samples. In each sample, we observed voltage, current, active, reactive and apparent power, and power factor.

To support embedding, we identified relevant time segments in the data based on significant changes in values. In the context of energy signals, we distinguish four basic types of time intervals bounded by edges: intervals between rising edges (rising–rising), intervals between a rising and a falling edge (rising–falling), intervals between a falling and a rising edge (falling–rising), and intervals between falling edges (falling–falling). Detection was performed on the basis of

Table 1
F1 score for combinations of embedding and clustering methods

		Clustering method			
		k-means	DBSCAN	Ward's	Single linkage
embedding	statistical	0.0	0.588	0.0	0.432
	Fourier	0.0	0.429	0.267	0.364
	wavelet	0.0	0.5	0.0	0.333
	time delay	0.0	0.69	0.0	0.421
	EMD	0.0	0.48	0.0	0.462
	multimodal	0.375	0.545	0.333	0.421

differential changes in active power. The detected time segments were then processed with the goal of feature extraction, these features were then embedded using the following methods:

- **statistical embedding** – features included mean active power and current, their standard deviation, skewness and kurtosis of the distribution of active power values, mean power factor, and the difference between minimum and maximum current.
- **Fourier embedding** – time series data was first mapped into the frequency space, and features were extracted from this representation: average amplitude of component frequencies, their standard deviation, maximum amplitude and the sum of all amplitudes.
- **wavelet embedding** – data was mapped to a time-frequency domain using the Daubechies 4 wavelet transform. For the approximation coefficients, we calculated the mean, standard deviation, and maximum value. For detailed coefficients, we determined the signal's energy as a sum of square roots of coefficients.
- **time delay embedding** – time series was transformed into a multidimensional space, where each point is created from samples from the original signal with a constant time difference between them.
- **EMD embedding** – the input signal was deconstructed into intrinsic mode functions, and features such as the number of local extremes, and relative energetic value, were extracted for each of these functions.
- **multimodal embedding** – features were extracted from the time, frequency, and time-frequency domains independently, and then concatenated. Time domain analysis was based on basic statistical metrics, such as mean, standard deviation, skewness, and kurtosis. Frequency domain analysis was done similarly to Fourier embedding, time-frequency analysis was done by wavelet transform.

For the detection of anomalies and normal processes, we tested the following clustering methods: k-means, DBSCAN, agglomerative clustering with Ward's linkage, and agglomerative clustering with single linkage. We first trained each combination of embedding and clustering on real data with no anomalies, reflecting real-life use case, where in the initial phase the observed machinery functions under standard circumstances. Following this, we introduced synthetic anomalies to real data and carried out clustering. For DBSCAN, samples marked as outliers were considered to be detected anomalies, while for every other method, small clusters were considered to be containing anomalies. Anomaly detection was evaluated using accuracy, precision, recall, and F1 score.

The F1 scores of each combination are shown in Table 1. When comparing the embedding methods based on F1 score, we see that multimodal embedding achieved the highest average F1 score (0.418), followed by time delay embedding (0.277) and Fourier embedding (0.264). While EMD showed a lower average F1 score (0.235), it had the highest accuracy (0.667) (see Fig. 1). The second highest average accuracy was achieved when using wavelet embedding (0.6), and Fourier embedding (0.583). The best single combination proved to be time delay embedding with DBSCAN, with an F1 score of 0.69, and statistical embedding with DBSCAN (0.588 F1 score), the latter also resulting in the highest accuracy. Based on accuracy, hierarchical embedding with Ward's linkage gave the most consistent results.

When analyzing the considered clustering algorithms, the best average F1 score was achieved by DBSCAN (0.538), followed by hierarchical clustering with single linkage (0.405), hierarchical clustering with Ward's linkage (0.1), and k-means (0.062). We also considered time complexity due to our aim to implement a solution that works in real time, the results of our analysis can be seen in Table 2. All combinations were able to meet our expectations, with the combination of statistical embedding and k-means showing a considerable higher computational cost. Hierarchical clustering was shown to be the fastest, with DBSCAN being only slightly slower. By considering accuracy and computational cost, we found that a combination of statistical embedding and DBSCAN was the best option for the data considered in our research, solutions using DBSCAN generally had the best performance/cost ratio.

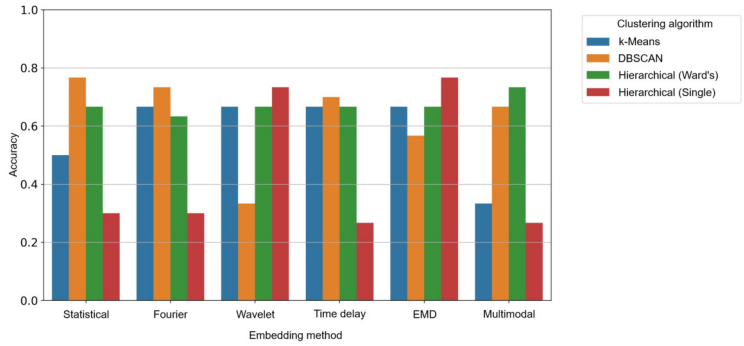


Fig. 1 Accuracy of various combinations of embedding and clustering methods

Table 2
Time complexity of combinations of embedding and clustering methods (in ms)

		Clustering method			
		k-means	DBSCAN	Ward's	Single linkage
embedding	statistical	38	2.6	1.8	1.3
	Fourier	3.7	3.7	1.7	1.3
	wavelet	3.1	2.8	1.7	1.6
	time delay	3.6	2.8	1.6	1.8
	EMD	3.8	2.7	1.5	1.8
	multimodal	3.4	2.9	1.5	1.8

IV. Deployment

A detailed analysis and evaluation of different embedding and clustering methods on synthetic data constituted a crucial step in developing a reliable approach for identifying process cycles in real data. We first focused on anomaly detection, which is a key component of any analytical approach, and enables the method to deal with milder variations in process characteristics.

Based on our initial analysis, we identified DBSCAN as the most suitable clustering algorithm. Its main advantage lies in its ability to determine the number of clusters without prior configuration, and can dynamically adjust its value based on newly processed data. The analysis also showed that DBSCAN in combination with time delay embedding and statistical embedding achieved high precision in detecting anomalies, therefore, we considered these two options.

We tested the two selected approaches in real data acquired in industrial settings as described in Section III. These samples contained data from eight operational modes, with varying levels of representation in the overall dataset. Results of the first set of experiments on real data are shown in Table 3.

Using time delay embedding resulted in faulty clustering and detection, identifying only two operational modes, which is not satisfactory for capturing complex system dynamics. This result was caused by the embedding's high sensitivity to parameter settings and problematic convergence when processing irregular and noisy data. On the other hand, statistical embedding showcased great robustness and precision also on real data. It correctly identified eight process modes, which upon further analysis were confirmed as reflecting real-life scenarios. These modes were represented with varying levels in the data, the number of occurrences is presented in Table 4.

After the detection of process modes, we performed an analysis of system dynamics on data gathered during deployment. In this data, mode 0 was still the most frequent one with 281 occurrences, with modes 1 and 2 represented more evenly at 220 and 221, respectively. Mode 4 was detected 139 times, mode 3 45 times, with modes 5–7 occurring at much lower rates (9, 6, and 9, respectively). We first calculated the entropy of transitions from the most frequent modes, indicating the predictability of the system regularly moving to a specific mode. Mode 0 had the highest entropy with 1.715, meaning that the system could transition into multiple modes from this standard mode. Mode 1 had an entropy of 1.233, Mode 2 had 1.445, and the entropy of Mode 3 was 1.155. Mode 4 had an entropy of 0, as the system always returned to Mode 0 from

Table 3
Comparison of detection results with two selected embeddings on real data

Embedding	**Number of events**	**Number of anomalies**	**Number of modes**
Statistical	968	32	8
Time delay	876	28	2

Table 4
Number of occurrences of identified process modes

Process mode	Number of occurrences
Mode 0	625
Mode 1	123
Mode 2	114
Mode 3	23
Mode 4	70
Mode 5	5
Mode 6	3
Mode 7	5

Table 5
Transition probabilities between different system modes

Mode from/to	0	1	2	3	4	5	6	7
0	0.00	0.59	0.12	0.07	0.19	0.01	0.02	0.00
1	0.07	0.00	0.78	0.03	0.06	0.03	0.00	0.03
2	0.60	0.05	0.00	0.08	0.27	0.00	0.00	0.00
3	0.04	0.74	0.17	0.00	0.00	0.00	0.00	0.04
4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
6	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00

this state.

Transition probabilities are also shown in Table 5. The most common transition was from Mode 1 to Mode 2 (90 times), from Mode 0 to Mode 1 (80 times), from Mode 4 to Mode 0 (70 times), and from Mode 2 to Mode 0 (67 times). The low number of occurrences for modes 5, 6, and 7 indicates that they were typical for the end of process cycles.

Typical process cycles were identified on the basis of an analysis of the ratio of transition probabilities to their expected probability. This analysis showed that some transitions occurred significantly more frequently than they should have with a uniform distribution, hinting at their relevance for normal process cycles. The most frequent cycle with 33 occurrences consisted of modes 0 to 1 to 2 to 0. A slightly less frequent cycle had modes 0 to 4 to 0 (27 times). We also identified one more relevant process cycle with modes 1 to 2 to 0 to 1 (20 occurrences). After comparing these results with the real manufacturing process, we confirmed that the detected cycles reflect typical use of the observed machinery.

V. Conclusion

In this paper we described a general approach to identifying industrial processes and modes of machinery from power consumption data. We tested and evaluated multiple embedding methods for feature extraction, and clustering methods for identifying machine modes. We first evaluated these methods on real data with synthetic anomalies introduced, and compared their performance in anomaly detection. From the tested approaches, we then selected DBSCAN for clustering, and time delay embedding and statistical embedding for feature extraction. The final design was then tested in real deployment with great precision. Our solution detected eight different modes of a galvanic copper plating machine, corresponding to real regimes. We also undertook a study of transitions and identified typical process cycles.

The main advantage of our approach is the small number of parameters and that it does not require the configuration of the number of operational modes, which is instead automatically derived through data analysis. The results of this work have significant potential for practical industrial applications, where they can contribute to increasing energy efficiency, reducing maintenance costs, and extending equipment life. The created solution is scalable and adaptable to various types of industrial systems. In future research, it would be beneficial to extend the methodology to include more advanced deep learning methods that could model complex relationships in energy signals even more accurately.

Acknowledgment

The authors would like to thank RE-CA s.r.o. for their help in the development of the measuring device ReSENSE-POWER used in this study, and Chemosvit, a.s. for providing the data from a real industrial process.

REFERENCES

[1] E. Kuk, S. Bobek, and G. J. Nalepa, "Ml-based proactive control of industrial processes," in *International Conference on Computational Science*. Springer, 2023, pp. 576–589.
[2] W. Strielkowski, A. Vlasov, K. Selivanov, K. Muraviev, and V. Shakhnov, "Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review," *Energies*, vol. 16, no. 10, p. 4025, 2023.
[3] T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, and S. G. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, p. 106024, 2019.
[4] M. H. Abidi, M. K. Mohammed, and H. Alkhalefah, "Predictive maintenance planning for industry 4.0 using machine learning for sustainable manufacturing," *Sustainability*, vol. 14, no. 6, p. 3387, 2022.
[5] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, vol. 26, 2013.
[6] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data mining and knowledge discovery*, vol. 33, no. 4, pp. 917–963, 2019.
[7] R. H. Shumway, D. S. Stoffer, and D. S. Stoffer, *Time series analysis and its applications*. Springer, 2000, vol. 3.
[8] G. E. Batista, X. Wang, and E. J. Keogh, "A complexity-invariant distance measure for time series," in *Proceedings of the 2011 SIAM international conference on data mining*. SIAM, 2011, pp. 699–710.
[9] É. M. Lima, C. M. dos Santos Junqueira, N. S. D. Brito, B. A. de Souza, R. de Almeida Coelho, and H. Gayoso Meira Suassuna de Medeiros, "High impedance fault detection method based on the short-time fourier transform," *IET Generation, Transmission & Distribution*, vol. 12, no. 11, pp. 2577–2584, 2018.
[10] A. H. Khandoker, D. T. Lai, R. K. Begg, and M. Palaniswami, "Wavelet-based feature extraction for support vector machines for screening balance impairments in the elderly," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 4, pp. 587–597, 2007.
[11] M. Barbosh, P. Singh, and A. Sadhu, "Empirical mode decomposition and its variants: A review with applications in structural health monitoring," *Smart Materials and Structures*, vol. 29, no. 9, p. 093001, 2020.
[12] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power signature analysis," *IEEE power and energy magazine*, vol. 1, no. 2, pp. 56–63, 2003.
[13] M. Su, W. Zhao, Y. Zhu, D. Zha, Y. Zhang, and P. Xu, "Anomaly detection of vectorized time series on aircraft battery data," *Expert Systems with Applications*, vol. 227, p. 120219, 2023.
[14] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Computers in Industry*, vol. 105, pp. 182–190, 2019.
[15] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Dbscan revisited, revisited: why and how you should (still) use dbscan," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.
[16] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, "Unsupervised electric motor fault detection by using deep autoencoders," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 2, pp. 441–451, 2019.
[17] T. W. Liao, "Clustering of time series data—a survey," *Pattern recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.
[18] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.
[19] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.
[20] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009.
[21] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.
[22] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "Optics: Ordering points to identify the clustering structure," *ACM Sigmod record*, vol. 28, no. 2, pp. 49–60, 1999.
[23] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013, pp. 160–172.

Vplyv rozhrania Cu/PTF a Ag/PTF na dlhodobú stabilitu vnorených a nevnorených PTF rezistorov

[1]*Igor VEHEC, Lukáš GARNEK*

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

[1]igor.vehec.2@tuke.sk

***Abstrakt* — Tento článok sa zaoberá štúdiom stability polymérnych hrubovrstvových (PTF) rezistorov aplikovaných na substrát FR4 (sklo/epoxid). Cieľom bolo vyhodnotiť vplyv rôznych typov zakončenia PTF rezistorov (priamy kontakt s Cu plôškami bez povrchovej úpravy vs. kontakt s dodatočne nanesenou PTF vrstvou na báze Ag) na ich dlhodobú stabilitu v podmienkach klimatického starnutia a teplotných šokov. Experimentálne vzorky boli vystavené klimatickému testu (85 % relatívna vlhkosť pri 85°C po dobu 1000 hodín) a cyklickým teplotným šokom (–45°C a +125°C počas 1000 cyklov). Výsledky ukazujú, že prítomnosť PTF vrstvy na báze Ag v kombinácii s vnorením PTF rezistorov do FR4 prekrytia vedie k výrazne menším zmenám elektrického odporu. Naopak, pri priamom kontakte PTF rezistora s Cu plôškou bola pozorovaná signifikantná degradácia, najmä v podmienkach so zvýšenou vlhkosťou.**

***Kľúčové slová* — FR4 substrát, polymérne hrubovrstvové rezistory, sieťotlač, skúšky starnutia**

I. Úvod

Polymérna hrubovrstvová technológia (Polymer Thick-Film – PTF) je v oblasti elektroniky dlhodobo etablovaná technológia, používaná pre výrobu pasívnych komponentov, najmä rezistorov a vodivých prepojení, štandardne nanášaných pomocou technológií akými sú napr. sieťotlač alebo tlač cez šablónou. Vďaka svojej jednoduchosti, možnosti spracovania pri nízkych teplotách a kompatibilite so širokou škálou substrátov (vrátane flexibilných, sklenených či polymérnych materiálov) si našla uplatnenie v aplikáciách od jednoduchých elektronických obvodov až po integrované senzory a RFID tagy [1], [2], [3]. Významnou výhodou PTF technológie je možnosť výroby komponentov priamo na konečný substrát bez potreby zložitého montážneho procesu, čo ju robí atraktívnou najmä pri návrhu tenkých, prispôsobiteľných a integrovaných elektronických systémov.

Vzhľadom na rastúci tlak na zmenšovanie rozmerov a zjednodušenie výroby pasívnych komponentov dochádza v posledných rokoch k intenzívnemu výskumu zameranému na spoľahlivosť PTF prvkov pri rôznych typoch záťaže. Zatiaľ čo vplyv zrýchleného starnutia pri rôznych podmienkach na stabilitu PTF rezistorov, vrátane ich porovnania s klasickými SMD súčiastkami, je už do istej miery pokryté v odbornej literatúre [4], [5], [6], [7], stále existuje viacero otvorených otázok, najmä v súvislosti s degradačnými mechanizmami na rozhraní medzi vodivými plochami a PTF vrstvami. Tento aspekt nadobúda zásadný význam najmä pri použití PTF technológie v netradičných aplikáciách alebo pri integrácii do nových typov substrátov či integrácii systémov do puzdier, ktoré nemajú štandardizované ochranné alebo metalizačné vrstvy.

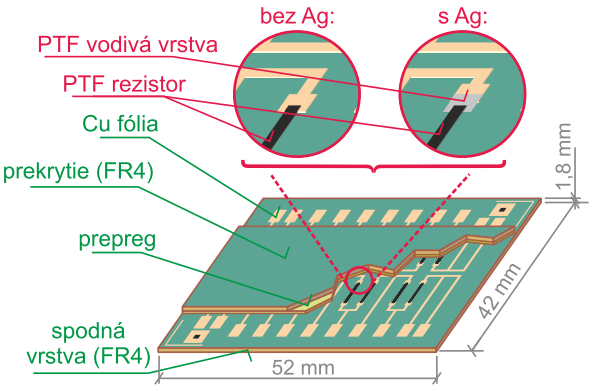
V tomto článku sú porovnávané PTF rezistory tlačené priamo na medené plôšky (bez povrchovej úpravy) s rezistormi, pri ktorých bola medzi medený podklad a PTF odporovú pastu nanesená dodatočná vodivá vrstva z PTF pasty na báze Ag. Pre oba varianty boli realizované konfigurácie s prekrytím (vnorené PTF rezistory) aj bez prekrytia (nevnorené PTF rezistory), čo umožňuje lepšie determinovať vplyv prostredia (vlhkosť, teplotné šoky) na elektrickú stabilitu rezistorov za rôznych podmienok.

II. Príprava experimentálnych vzoriek

Stabilita vnorených a povrchových PTF rezistorov realizovaných na štandardnom FR4 (sklo/epoxid) substráte bola hodnotená pre dva typy zakončení PTF rezistorov:

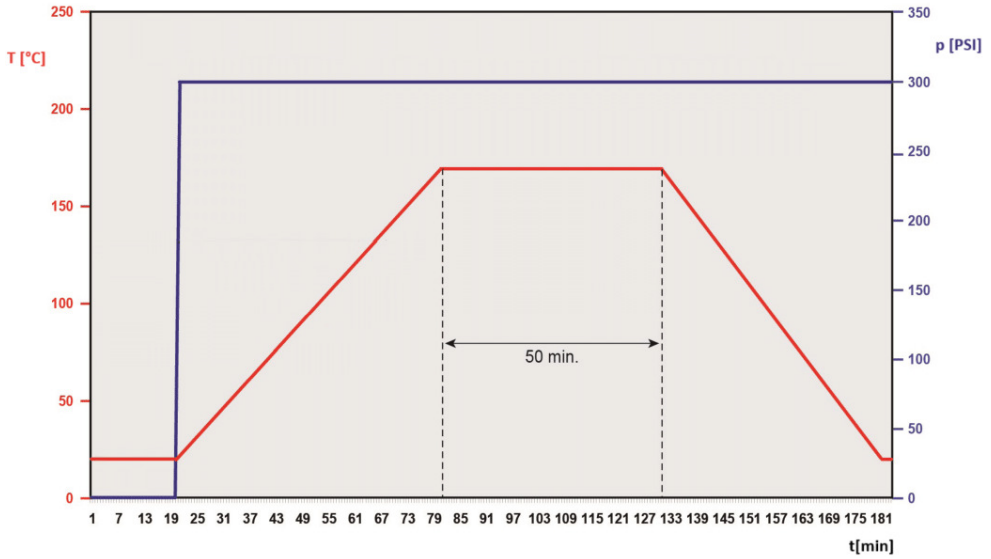
- ***Zakončenie typu Cu/PTF***: PTF rezistor bol priamo tlačený na medenú plôšku bez ďalšej povrchovej úpravy, pričom elektrické spojenie bolo zabezpečené rozhraním Cu/PTF.
- ***Zakončenie typu PTF/Ag***: PTF rezistor bol tlačený na dodatočnú vodivú vrstvu zhotovenú pomocou PTF pasty na báze striebra, čím bolo vytvorené rozhranie PTF/Ag.

Všetky PTF vrstvy boli nanášané sieťotlačou pomocou zariadenia Aurel 900PA. Pre vodivé vrstvy bola použitá komerčne dostupná polymérna pasta ED3000 na báze striebra s plošným odporom 30 mΩ/sq. pri hrúbke 12 μm. Vytvrdzovanie prebiehalo pri teplote 120°C po dobu 30 minút. Odporové vrstvy boli vytvorené z polymérnej pasty ED7500, ktorá pozostáva zo zmesi uhlíkových, strieborných a grafitových práškov. Plošný odpor vrstiev vytvorených pomocou tejto pasty sa pohybuje v rozmedzí od 1 Ω/sq. do 1 MΩ/sq., pričom v našich experimentoch bola použitá pasta s plošným odporom 20 Ω/sq. Tieto vrstvy boli vytvrdzované pri teplote 150°C po dobu 30 minút. Na sieťotlač všetkých typov PTF pást bola použitá šablóna z nehrdzavejúcej ocele s hustotou vlákien 325 mesh. Obr. 1 znázorňuje usporiadanie jednotlivých vrstiev v experimentálnych vzorkách vrátane rozmerov, ako aj umiestnenia vodivých a odporových tlačených vrstiev.



Obr. 1 Usporiadanie vrstiev v experimentálnych vzorkách.

Všetky PTF rezistory mali jednotné rozmery 5,6 × 0,5 mm. Topológia vodivých a odporových vrstiev bola navrhnutá s ohľadom na možnosť presného merania ich elektrického odporu pomocou štvorbodovej (Kelvinovej) metódy. Po vytvrdnutí tlačených vrstiev nasledovalo ich zalaminovanie krycou vrstvou (tiež na báze sklo/epoxidového substrátu FR4). Laminovanie vrchnej a spodnej vrstvy bolo realizované pomocou prepregu DE104, pričom bol dodržaný výrobcom odporúčaný teplotný profil zobrazený na Obr. 2.



Obr. 2 Teplotný profil laminovania pre prepreg DE104.

Vyrobené vzorky boli následne podrobené dlhodobému starnutiu v dvoch rozdielnych testoch za nasledujúcich podmienok:

- ***Klimatický test (vlhké teplo)***: Expozícia vzoriek pri teplote 85°C a 85 % relatívnej vlhkosti

počas 1 000 hodín, uskutočnená v klimatickej komore CTS C-65/100. Merania elektrického odporu boli vykonávané po 20, 50, 100, 200, 500 a 1000 hodinách.

- ***Test teplotnými šokmi***: 1 000 cyklov zahŕňajúcich zotrvanie vzoriek pri teplote –45°C počas 20 minút, nasledované ich presunom do prostredia s teplotou +125°C na 20 minút. Test bol realizovaný v šokovej komore CTS TSS-70/32. Elektrický odpor bol meraný po každých 100 cykloch.

Po ukončení každého testovacieho intervalu bol elektrický odpor PTF rezistorov meraný štvorbodovou metódou pomocou multimetra Agilent 34411A, a to po aklimatizácii vzoriek na laboratórnu teplotu.

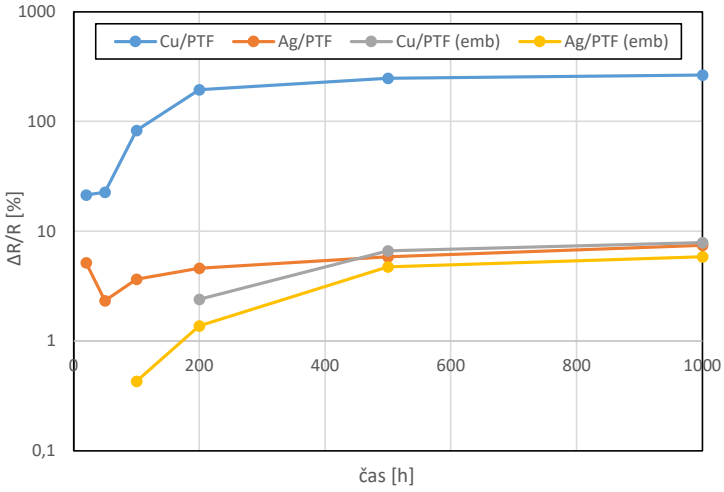
III. ZHODNOTENIE VÝSLEDKOV

Cieľom bolo analyzovať vplyv typu rozhrania medzi polymérnym hrubovrstvovým (PTF) rezistorom a kontaktom na jeho dlhodobú stabilitu. Porovnávané boli dve varianty kontaktov: priamy kontakt PTF rezistora s Cu plôškou a kontakt PTF rezistora s dodatočne vytlačenou PTF vrstvou na báze Ag. Pre každý typ zakončenia boli realizované dve konštrukčné konfigurácie rezistorov: bez prekrytia (nevnorené) a s prekrytím (vnorené). Za predpokladu, že FR4 substrát neprepúšťa vlhkosť (resp. prepúšťa vlhkosť len minimálne), možno rozdiely medzi vnorenými a nevnorenými konfiguráciami interpretovať najmä ako dôsledok pôsobenia vlhkosti na kontakt PTF rezistora s podložkou. Týmto spôsobom je možné oddeliť vplyv vlhkosti od vplyvu samotného rozhrania (Cu/PTF alebo Ag/PTF) a presnejšie vyhodnotiť degradačné mechanizmy v podmienkach klimatického starnutia.

Počas experimentu bolo sledované, do akej miery stabilitu elektrického odporu ovplyvňuje samotné rozhranie (Cu/PTF vs. Ag/PTF), ako aj vnorenie/nevnorenie PTF rezistora. Získané údaje umožňujú posúdiť, či je vhodné a odôvodniteľné použitie dodatočnej Ag vrstvy ako kontaktnej medzivrstvy medzi medeným substrátom a PTF rezistorom. Výsledky sú rozdelené podľa typu testovania – klimatické starnutie a teplotné šoky – a diskutované z hľadiska relatívnej zmeny elektrického odporu v čase, resp. počte cyklov.

A. Starnutie PTF rezistorov v klimatickej komore

Všetky testované vzorky na báze FR4 – vnorené aj nevnorené PTF rezistory – zostali po ukončení klimatického testu plne funkčné, bez výskytu otvorených spojov alebo iných zjavne viditeľných porúch (v prípade nevnorených rezistorov). Výrazné rozdiely však boli pozorované v správaní elektrického odporu v závislosti od typu rozhrania (Cu/PTF alebo Ag/PTF) a konštrukcie PTF rezistora. Na Obr.3 je znázornená časová závislosť relatívnej zmeny odporu ($\Delta R/R_0$) počas starnutia v klimatickej komore pre všetky varianty vzoriek.



Obr. 3 Relatívna zmena elektrického odporu PTF rezistorov počas starnutia v klimatickej komore.

Najväčší nárast elektrického odporu bol zaznamenaný pri nevnorených PTF rezistoroch s priamym kontaktom na Cu plôšku bez povrchovej úpravy (Cu/PTF), kde relatívna zmena dosiahla hodnotu $\Delta R/R_0$ = +264,8 % po 1000 hodinách starnutia. Naopak, vnorenie týchto rezistorov (Cu/PTF(emb)) túto zmenu výrazne znížilo na +7,9 %, čo poukazuje na ochrannú funkciu prekrytia voči okolitému prostrediu.

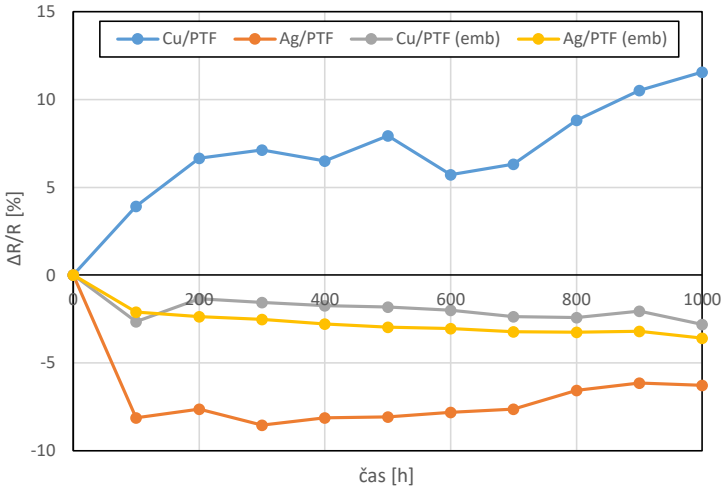
Nakoľko materiál FR4 má veľmi nízku nasiakavosť (~0,01 % hmotnosti) [8], [9], možno

predpokladať, že horná (krycia) vrstva z rovnakého materiálu efektívne bráni prístupu vlhkosti k samotnému rozhraniu medzi PTF pastou a medenou fóliou. To výrazne redukuje riziko korózie Cu [10] podmienenej zvýšenou vlhkosťou (zväčša ide o oxidáciu) a iných degradačných procesov, ktoré môžu nastať pri súčasnom pôsobení tepla a vlhkosti. Tento predpoklad je v súlade s inými publikovanými pozorovaniami [6], kde bol podobný nárast odporu PTF rezistorov na nechránenej Cu vrstve pozorovaný po ich vystavení zvýšenej teplote a vlhkosti.

Z porovnania výsledkov je zrejmé, že vnorenie PTF rezistora aj použitie medzivrstvy z Ag pasty viedli k výrazne nižšej zmene elektrického odporu počas klimatického testu. V prípade nevnorených PTF rezistorov s plôškou na báze Ag (Ag/PTF) bola relatívna zmena po 1000 hodinách +7,5 %, zatiaľ čo pri ich vnorenej verzii (Ag/PTF(emb)) bola relatívna zmena odporu ešte nižšia ($\Delta R/R_0 = +5{,}8$ %). Tieto výsledky naznačujú, že použitie PTF vodivej medzivrstvy na báze Ag môže predstavovať účinné opatrenie na zvýšenie dlhodobej stability rezistora, a to aj v prípade nevnorených PTF rezistorov, pričom vnorenie ich stabilitu ešte mierne zlepšuje.

B. Starnutie PTF rezistorov v šokovej komore

Podobne ako pri klimatickom teste, aj v prípade cyklického teplotného zaťaženia v šokovej komore bola najväčšia relatívna zmena elektrického odporu PTF rezistorov zaznamenaná pri vzorkách s priamym kontaktom PTF odporovej vrstvy na Cu plôšku (Cu/PTF). Po absolvovaní 1 000 teplotných cyklov dosiahla táto zmena hodnotu $\Delta R/R_0 = +11{,}5$ % (Obr. 4), avšak v porovnaní s výsledkami z klimatickej komory bola táto zmena výrazne menšia. Tento rozdiel podporuje predchádzajúcu úvahu, že pri klimatickom teste bola dominujúcim faktorom podmieňujúcim degradáciu prítomnosť vysokej relatívnej vlhkosti (85 % RH), ktorá mohla viesť ku korózii Cu a degradácii rozhrania Cu/PTF. V šokovej komore sa vzorky nachádzali striedavo pri teplotách –45°C a +125°C, pričom relatívna vlhkosť v oboch komorách bola vzhľadom na extrémne podmienky zanedbateľná alebo nulová. Z toho dôvodu tu nebolo možné očakávať vlhkostne indukovanú degradáciu.



Obr. 4 Relatívna zmena elektrického odporu PTF rezistorov počas starnutia v šokovej komore.

Celkovo nebolo počas testovania pozorované zlyhanie žiadneho PTF rezistora. V porovnaní s klimatickým testom bol pri záťaži teplotnými šokmi výraznejší vplyv rozdielnych koeficientov tepelnej rozťažnosti medzi jednotlivými materiálmi, najmä na rozhraní Cu/PTF. Tieto rozdiely môžu viesť ku vzniku mikrotrhlín, ktoré spôsobujú mierne zmeny elektrického odporu. Na druhej strane, vysoké teploty, ktorým boli vzorky opakovane vystavované, môžu spôsobiť dodatočné vytvrdenie polymérnej matice, čo v niektorých prípadoch vedie k stabilizácii elektrických vlastností PTF vrstiev. Z hľadiska kompatibility materiálov je možné predpokladať, že tepelná rozťažnosť PTF odporovej vrstvy je bližšia tepelnej rozťažnosti Ag vodivej vrstvy v porovnaní s Cu fóliou. Tento predpoklad je v súlade s meraniami – menšia relatívna zmena odporu bola pozorovaná pri rozhraní Ag/PTF, v porovnaní s rozhraním Cu/PTF.

Ako je vidieť z grafu na Obr. 4, väčšina vzoriek vykazovala po 1 000 cykloch mierny pokles elektrického odporu, čo kontrastuje s trendom pozorovaným pri klimatickom teste, kde sa elektrický odpor typicky zvyšoval oproti počiatočnej hodnote. Tento jav je u PTF materiálov dobre známy – pri dlhodobom pôsobení zvýšenej teploty (najmä počas fáz s teplotou +125°C) dochádza k dovytvrdeniu polymérnej štruktúry, čo často vedie k miernemu zníženiu elektrického odporu. Jedinou výnimkou z tohto správania tu bola vzorka s nevnorenými PTF rezistormi na Cu

plôškach (Cu/PTF), pri ktorej elektrický odpor napriek možnému dovytvrdeniu vzrástol. To naznačuje, že v tomto prípade degradačný mechanizmus na rozhraní Cu/PTF prevážil nad stabilizačným efektom tepelného dovytvrdenia. Rovnako ako pri klimatických testoch, aj tu vnorené PTF rezistory vykazovali nižší drift elektrického odporu v porovnaní s nevnorenými, čo opätovne potvrdzuje priaznivý vplyv prekrytia na stabilitu PTF rezistorov.

IV. Záver

Dosiahnuté výsledky ukazujú, že stabilita polymérnych hrubovrstvových rezistorov (PTF) závisí výrazne od typu kontaktnej vrstvy a od toho, či sú vnorené alebo nie. Pri aplikácii PTF rezistorov na substrát FR4, s priamym kontaktom na Cu plôšku, došlo k výraznému nárastu elektrického odporu, najmä pri dlhodobom pôsobení zvýšenej teploty a vlhkosti, čo bolo spôsobené degradáciou Cu na rozhraní s PTF pastou. Tento efekt bol pri vzorkách s pridanou PTF vrstvou na báze Ag výrazne znížený, čo naznačuje, že táto vrstva zlepšuje stabilitu PTF rezistorov.

Okrem toho, vnorenie PTF rezistorov pomocou prekrytia na báze FR4 substrátu, ktoré účinne bráni prístupu vlhkosti, malo pozitívny vplyv na stabilitu elektrického odporu počas klimatických testov. V šokových testoch bol pozorovaný menší vplyv tepelnej rozťažnosti medzi materiálmi na zmenu odporu, pričom aj v tomto prípade bola stabilita výrazne lepšia pri Ag/PTF rozhraní v porovnaní s rozhraním Cu/PTF.

Poďakovanie

Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02- 00002.

References

[1] I. Vehec, T. Lenger, S. Kardos, P. Lukacs, a K. Gontkovic, "Polymer Thick-film Piezoresistive Sensor Integrated into the PCB", v *2023 46th International Spring Seminar on Electronics Technology (ISSE)*, Timisoara, Romania: IEEE, máj. 2023, s. 1–8. doi: 10.1109/ISSE57496.2023.10168379.

[2] S. Rajala, S. Tuukkanen, a J. Halttunen, "Characteristics of Piezoelectric Polymer Film Sensors With Solution-Processable Graphene-Based Electrode Materials", *IEEE Sens. J.*, roč. 15, č. 6, s. 3102–3109, jún. 2015, doi: 10.1109/JSEN.2014.2344132.

[3] G. Harsányi, "Polymer films in sensor applications: a review of present uses and future possibilities", *Sens. Rev.*, roč. 20, č. 2, s. 98–105, jún. 2000, doi: 10.1108/02602280010319169.

[4] W. Steplewski, A. Dziedzic, J. Borecki, G. Koziol, a T. Serzysko, "Environmental tests of embedded thin- and thick-film resistors in comparison to chip resistors", *Circuit World*, roč. 40, č. 1, s. 7–12, jan. 2014, doi: 10.1108/CW-10-2013-0039.

[5] A. Dziedzic, A. Klossowicz, P. Winiarski, A. W. Stadler, a W. Steplewski, "Chosen electrical, noise and stability characteristics of passive components embedded in printed circuit boards", v *Proceedings of the 5th Electronics System-integration Technology Conference (ESTC)*, Helsinki, Finland: IEEE, sep. 2014, s. 1–8. doi: 10.1109/ESTC.2014.6962765.

[6] W. Stęplewski, T. Serzysko, G. Kozioł, a A. Dziedzic, "Preliminary assessment of the stability of thin- and polymer thick-film resistors embedded into printed wiring boards", *Microelectron. Reliab.*, roč. 52, č. 8, s. 1719–1725, aug. 2012, doi: 10.1016/j.microrel.2012.03.026.

[7] P. Winiarski a A. Dziedzic, "Stability of thin-film resistors embedded in printed circuit boards", v *2011 International Students and Young Scientists Workshop "Photonics and Microsystems"*, Cottbus, Germany: IEEE, júl. 2011, s. 138–142. doi: 10.1109/STYSW.2011.6155861.

[8] H. Conseil-Gudla, V. C. Gudla, S. Borgaonkar, M. S. Jellesen, a R. Ambat, "Investigation of moisture uptake into printed circuit board laminate and solder mask materials", *J. Mater. Sci. Mater. Electron.*, roč. 28, č. 8, s. 6138–6151, apr. 2017, doi: 10.1007/s10854-016-6292-5.

[9] 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.

[10] P. Gomasang, S. Ogiue, K. Ueno, a S. Yokogawa, "Oxidation Structure Change of Copper Surface Depending on Accelerated Humidity", v *2018 IEEE International Interconnect Technology Conference (IITC)*, Santa Clara, CA: IEEE, jún. 2018, s. 112–114. doi: 10.1109/IITC.2018.8430461.

Metódy odväzbenia MAB meniča

[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, Slovenská Republika

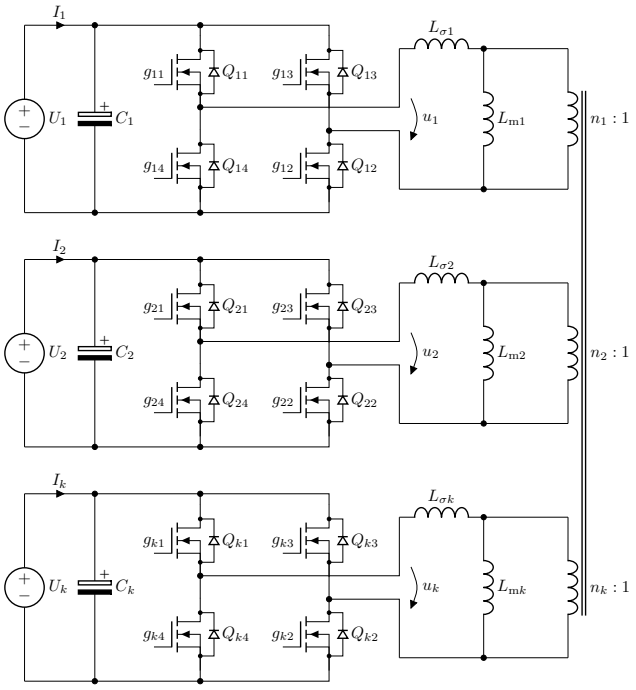
[1]tomas.basarik@tuke.sk, [2]milan.lacko@tuke.sk,
[3]tadeas.kmecik@tuke.sk, [4]daniel.gordan@tuke.sk

Abstrakt **– Tento článok je venovaný rýchlemu prehľadu možných odväzbení pre MAB menič (Multi-Active Bridge). V článku sú uvedené štyri odväzbovacie metódy, ktoré minimalizujú vzájomné pôsobenie medzi jednotlivými portami. Ako prvé je uvedené jedno hardvérové odväzbenie a následne tri softvérové. Tieto štyri metódy sú vhodné pre moderné aplikácie v oblasti hybridných elektrických vozidiel a obnoviteľných zdrojov energie.**

Kľúčové slová **– multi-active bridge, výkonový menič, odväzbenie meniča, riadenie meniča**

I. Úvod

Multiportový DC-DC menič v zahraničnej literatúre známy pod skratkou MAB (Multi-Active Bridge) [1], ktorý je odvodený z topológie DAB (Dual Active Bridge), umožňuje efektívny obojsmerný prenos energie medzi viacerými portami. Každý port pozostáva z jednofázového striedača, jedného vinutia vysokofrekvenčného transformátora a vstupno/výstupného filtra. Výhodou je široký rozsah napätí na portoch, ktorý umožňuje kombinovať vysokonapäťové a nízkonapäťové aplikácie v rámci jedného meniča, ako aj mäkké spínanie, ktoré znižuje straty a zvyšuje účinnosť. Tento menič je tak vhodný pre moderné aplikácie v oblasti hybridných elektrických vozidiel a obnoviteľných zdrojov energie. Nevýhodou týchto meničov je zložité riadenie spôsobené nelineárnymi väzbami a obmedzená možnosť dodatočných úprav, keďže vysokofrekvenčný transformátor neumožňuje jednoduché pridávanie ďalších portov. Topologická štruktúra MAB meniča je schematicky znázornená na Obr. 1.



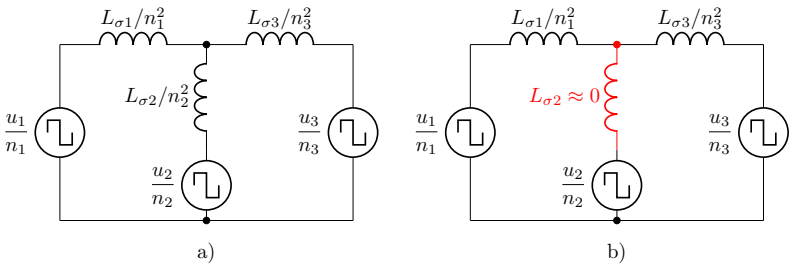
Obr. 1 Topológia Multi-Active Bridge meniča

II. Odväzbovacie metody

Pretože sa v MAB meničoch vyskytujú výrazné nelineárne väzby medzi jednotlivými portami, ich riadenie je zložité a neintuitívne. Na aplikáciu lineárnych regulačných metód je nevyhnutné tieto väzby odstrániť alebo minimalizovať, čo znamená, že systém je potrebné odväzbiť. Odväzbením sa nelineárny MIMO systém rozdelí na viaceré lineárne SISO systémy, čo umožňuje návrh efektívneho lineárneho riadenia pre jednotlivé porty a zjednodušuje dynamickú analýzu meniča. V nasledujúcom texte sú uvedené štyri odväzbovacie metódy pre MAB menič.

A. Hardvérové odväzbenie

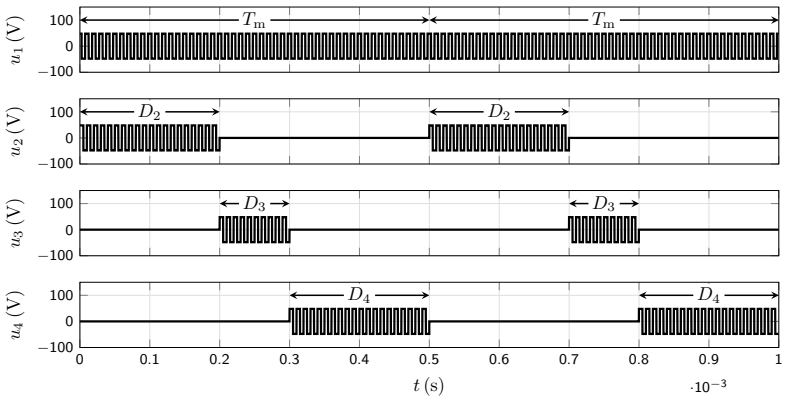
Hardvérové odväzbenie MAB meniča spočíva v úprave jeho konštrukcie tak, aby sa na jednom z portov dosiahla čo najnižšia rozptylová indukčnosť. Ak je táto indukčnosť výrazne menšia než na ostatných portoch, port možno považovať za nadradený a jeho správanie za ekvivalentné pre pevný napäťový zdroj. Ostatné porty sa v tomto prípade správajú ako podradené, pričom výkonové toky medzi nimi a nadradeným portom možno riadiť nezávisle, podobne ako v klasickej DAB topológii. Takéto odväzbenie výrazne zjednodušuje riadenie z pohľadu softvéru, keďže celý systém sa rozpadne na sústavu nezávislých DAB meničov. Obr. 2 zobrazuje náhradnú schému MAB meniča pred a po hardvérovom odväzbení. Na realizáciu tejto koncepcie sa často používa sériový rezonančný obvod tvorený rozptylovou indukčnosťou a prídavným kondenzátorom, pričom pri rezonancii sa impedancia nadradeného portu minimalizuje [2]. Alternatívne možno rozptylovú indukčnosť cielene potlačiť návrhom špeciálneho transformátora, ktorý dosahuje veľmi malé hodnoty rozptylových indukčností, pričom na podradené porty sa pridajú externé rozptylové indukčnosti [3]. Tento princíp sa úspešne uplatnil v aplikáciách ako integrácia batérií, obnoviteľných zdrojov a DC mikrosietí, kde priniesol zlepšenie dynamiky a zjednodušenie riadenia.



Obr. 2 Hardvérové odväzbenie trojportového meniča: a) pred odväzbením, b) po odväzbení

B. Riadenie zdieľaného času

Princíp riadenia zdieľaného času (Time-Sharing control), opísaný v literatúre [4] a [5], je založený na tom, že v každom okamihu sú aktívne len dva porty MAB meniča, zatiaľ čo ostatné zostávajú nečinné a správajú sa ako pasívne diódové usmerňovače. Týmto spôsobom sa eliminujú vzájomné väzby medzi portami, čo výrazne zjednodušuje riadenie. Energia sa v rámci modulačnej periódy T_m prelieva medzi nadradeným portom a aktívnym podradeným portom, pričom pasívne porty vstupujú do riadiacej štruktúry ako porucha. Menič sa tak riadi analogicky ako DAB menič. Napríklad, prvý port môže odoberať alebo dodávať energiu do druhého portu počas doby D_2, do tretieho portu počas doby D_3, a do štvrtého portu počas doby D_4, tak ako to znázorňuje Obr. 3.



Obr. 3 Riadenie zdieľaného času pre štvorportový menič

Podmienkou správneho riadenia je, aby súčet časových úsekov spĺňal $D_2 + D_3 + D4 \leq T_{\mathrm{m}}$, čím sa zabezpečí, že nedôjde k prekrytiu aktívnych intervalov. Výkon a napätie jednotlivých portov sú riadené pomocou aktívnej doby v rámci každej modulačnej periódy, pričom fázové posuny ϕ_1 až ϕ_k zostávajú fixné. Tento princíp sa podobá PWM modulácii. Riadenie možno rozšíriť o dynamickú úpravu fázových posunov počas aktívneho intervalu, čo vedie k lepšiemu využitiu výkonových tokov a k zníženiu zvlnenia napätia. Výpočtovo náročnejšie je len určenie dĺžky aktívnych intervalov a prípadná optimalizácia fázových posunov, pričom samotná koncepcia riadenia ostáva blízka DAB meniču.

C. Riadenie pomocou odväzbovacej matice

V literatúre [6] bol predstavený MAB menič, navrhnutý pre elektromobil. Na dosiahnutie nezávislého riadenia výkonových tokov medzi jednotlivými portami bez vzájomných väzieb bola použitá tzv. odväzbovacia matica, ktorá umožňuje oddelené riadenie každého portu. Tento prístup výrazne zlepšuje dynamickú odozvu systému a je vhodný pre aplikácie s vysokými nárokmi na presnosť. Riadenie vychádza z bezstratového modelu meniča, v ktorom sa uvažujú rozptylové indukčnosti transformátora, zatiaľ čo magnetizačná indukčnosť sa zanedbáva. Keďže model je nelineárny, linearizuje sa okolo pracovného bodu pomocou Taylorovho rozvoja, pričom sa uvažujú len prvé dva členy, čím sa systém zjednoduší na lineárny tvar vhodný na návrh regulátorov.

Pre lepšie pochopenie predpokladajme dvojportový menič s prenosovou maticou $\mathbf{G}$. Vstupom do systému je vektor s fázovými posunmi $\mathbf{\Phi}$ a výstupom systému je vektor s prúdmi jednotlivých portov $\mathbf{I}$. Maticový zápis takéhoto systému je

$$\mathbf{I} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \mathbf{G\Phi}. \tag{1}$$

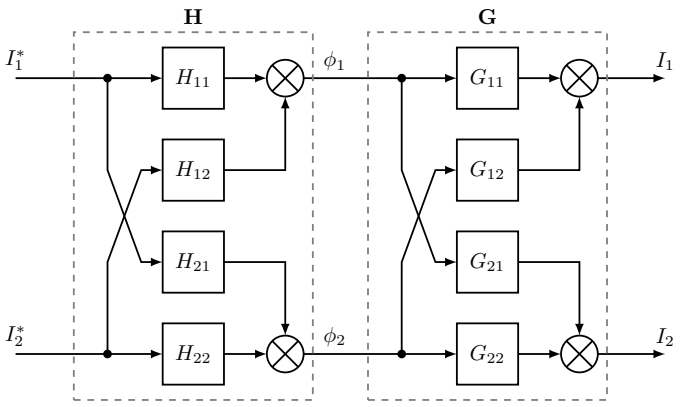
Pre odväzbenie systému je potrebné nájsť takú maticu $\mathbf{H}$, aby po vynásobení s prenosovou maticou $\mathbf{G}$ bola výsledná matica rovná jednotkovej matici $\mathbf{X}$. Odväzbovacia matica $\mathbf{H}$ je tak rovná inverznej prenosovej matici systému

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \mathbf{G}^{-1} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}^{-1}. \tag{2}$$

Po vložení odväzbovacej matice do systému je výsledný prenos rozšíreného systému rovný jednotkovej matici (3), čo znamená, že vstup do systému sa zrkadlí na výstup bez ovplyvňovania ostatných výstupov systému. Grafické znázornenie odväzbenia systému je zobrazené na Obr. 4.

$$\mathbf{X} = \mathbf{GH} = \mathbf{GG}^{-1} \tag{3}$$

Významnou nevýhodou lineárneho modelovania založeného na Taylorovej aproximácii je jeho obmedzená platnosť. Presnosť modelu je zaručená len v blízkom okolí zvoleného pracovného bodu. Pre pokrytie širšieho pracovného rozsahu je preto potrebné vytvoriť viacero linearizácií a ich výsledky uložiť do vyhľadávacích tabuliek. Tento prístup však výrazne zvyšuje nároky na pamäť mikrokontroléra, čo môže byť problémom v aplikáciách s obmedzenými výpočtovými a pamäťovými zdrojmi.



Obr. 4 Odväzbovacia matica dvojportového meniča

D. Newtonova iteračná metóda odväzbenia

Ako alternatíva k Taylorovej aproximácii sa v literatúre [7] uvádza numerický prístup využívajúci Newtonovu iteračnú metódu, ktorá umožňuje výpočet inverzného modelu MAB meniča, a teda určenie požadovaných fázových posunov v celom rozsahu jeho prevádzky bez nutnosti predchádzajúcej linearizácie. Základom metódy je iteratívne riešenie nelineárnej sústavy rovníc definovanej vzťahom medzi požadovanými prúdmi a fázovými posunmi. Sústava je zapísaná v maticovej forme s využitím Jacobiho matice parciálnych derivácií funkcií f_{i}, kde každá funkcia f_{i} predstavuje rovnicu pre prúd I_{i}. Inverzný výpočet potom prebieha podľa vzťahu:

$$\begin{bmatrix} \Delta\phi_1 \\ \Delta\phi_2 \\ \vdots \\ \Delta\phi_k \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial \phi_1} & \frac{\partial f_1}{\partial \phi_2} & \cdots & \frac{\partial f_1}{\partial \phi_k} \\ \frac{\partial f_2}{\partial \phi_1} & \frac{\partial f_2}{\partial \phi_2} & \cdots & \frac{\partial f_2}{\partial \phi_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial \phi_1} & \frac{\partial f_k}{\partial \phi_2} & \cdots & \frac{\partial f_k}{\partial \phi_k} \end{bmatrix}^{-1} \begin{bmatrix} -I_1 \\ -I_2 \\ \vdots \\ -I_k \end{bmatrix}, \tag{4}$$

kde

$$\frac{\partial f_i}{\partial \phi_j} = \begin{cases} U_j K_{ij} \phi_{ij} \left(4 \left|\phi_{ij}\right| - 1\right) & i \neq j \\ \sum\limits_{\substack{j=1 \\ j \neq i}}^{k} U_j K_{ij} \phi_{ij} \left(1 - 4 \left|\phi_{ij}\right|\right) & i = j \end{cases} . \tag{5}$$

Na začiatku iterácie sa zvolia počiatočné hodnoty fázových posunov ϕ_1 až ϕ_{k}. Čím bližšie sa hodnoty nachádzajú k cieľovej hodnote, tým nižší je počet potrebných iterácií. Po určení počiatočných hodnôt sa vypočítajú jednotlivé prírastky pre fázové posuny $\Delta\phi_1$, až $\Delta\phi_k$. Tie sa pripočítajú k pôvodným hodnotám $\phi_{k+1} = \phi_k + \Delta\phi_k$, a následne sa iterácia opakuje. Celý iteračný proces sa opakuje, až kým sa nedosiahne požadovaná presnosť ε pre všetky fázové posuny MAB meniča, tak ako to je uvedené v nasledujúcom vzťahu

$$\Delta\phi_1 < \varepsilon \quad \wedge \quad \Delta\phi_2 < \varepsilon \quad \wedge \quad \ldots \quad \wedge \quad \Delta\phi_k < \varepsilon. \tag{6}$$

III. Záver

V článku boli predstavené štyri odväzbovacie metódy pre MAB menič, ktoré znižujú vzájomné pôsobenie medzi portami. Každá z metód má svoje výhody aj obmedzenia. Hardvérové odväzbenie znižuje nároky na riadiaci softvér, no kladie vyššie výkonové nároky na nadradený port. Prvé softvérové riešenie zas nevyužíva plný výkonový potenciál meniča v dôsledku časového zdieľania riadenia. Druhé softvérové riešenie je efektívne v okolí pracovného bodu, no mimo neho stráca presnosť. Posledná, výpočtovo náročná metóda, nie je vhodná pre aplikácie s rýchlymi regulačnými slučkami. Pre priemyselné využitie je tak dôležité zvoliť metódu podľa konkrétnych požiadaviek aplikácie, pričom výber predstavuje kompromis medzi výpočtovou náročnosťou, flexibilitou a využitím výkonu meniča.

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] Y. Chen, P. Wang, Y. Elasser, and M. Chen, "Lego-mimo architecture: A universal multi-input multi-output (mimo) power converter with linear extendable group operated (lego) power bricks," in *2019 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2019, pp. 5156–5163.

[2] P. Wang, X. Lu, W. Wang, and D. Xu, "Hardware decoupling and autonomous control of series-resonance-based three-port converters in dc microgrids," *IEEE Transactions on Industry Applications*, vol. 55, no. 4, pp. 3901–3914, 2019.

[3] S. Bandyopadhyay, P. Purgat, Z. Qin, and P. Bauer, "A multiactive bridge converter with inherently decoupled power flows," *IEEE Transactions on Power Electronics*, vol. 36, no. 2, pp. 2231–2245, 2021.

[4] Y. Chen, P. Wang, H. Li, and M. Chen, "Power flow control in multi-active-bridge converters: Theories and applications," in *2019 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2019, pp. 1500–1507.

[5] H. Matsuo, W. Lin, F. Kurokawa, T. Shigemizu, and N. Watanabe, "Characteristics of the multiple-input dc-dc converter," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 3, pp. 625–631, 2004.

[6] C. Zhao, S. D. Round, and J. W. Kolar, "An isolated three-port bidirectional dc-dc converter with decoupled power flow management," *IEEE transactions on power electronics*, vol. 23, no. 5, pp. 2443–2453, 2008.

[7] P. Wang and M. Chen, "Towards power fpga: Architecture, modeling and control of multiport power converters," in *2018 IEEE 19th Workshop on Control and Modeling for Power Electronics (COMPEL)*. IEEE, 2018, pp. 1–8.

Procedurálny generátor polygonálnych povrchových modelov skál

[1]*Oleksandr LIASHENKO,* [2]*Branislav MADOŠ*

[1,2] Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 9, 040 01 Košice, Slovenská republika

[1]oleksandr.liashenko@student.tuke.sk, [2]branislav.mados@tuke.sk

***Abstrakt* — Tento článok je venovaný procedurálnemu generovaniu obsahu v oblasti počítačových hier. Cieľom tejto práce bolo vyvinúť nástroj schopný generovať trojrozmerné polygonálne povrchové modely skál, skalných zoskupení a skalných stien. Geometrie vymodelovaných objektov sú exportované do formátu .obj, materiálové vlastnosti sú uložené vo forme Physically Based Rendering (PBR) textúr, vo formáte .png. Procedurálne generovaný obsah je kompatibilný s modernými nástrojmi pre vývoj hier ako je Unity alebo Unreal. Bola vykonaná analýza súčasného stavu a preskúmaná problematika procedurálneho generovania v oblasti počítačovej grafiky a hier. Na realizáciu vlastného riešenia bol zvolený jazyk C++ s využitím knižníc OpenMesh, CGAL a FastNoiseLite. Vyvinuté riešenie podporuje nastavenie parametrov pri vytváraní kamenných geometrií a textúr. Výsledky práce ukazujú, že použitie procedurálneho generovania umožňuje vytvárať rozmanitý a jedinečný herný obsah, čo výrazne optimalizuje pracovnú náročnosť a využitie zdrojov pri vývoji herných projektov.**

***Keywords* — procedurálne generovanie, generovanie obsahu v hrách, 3D modelovanie, PBR textúry, generatívny šum, 3D objekty, generovanie kamenných foriem**

I. Úvod

Procedurálne generovanie obsahu (Procedural Content Generation, PCG) je metóda vytvárania prvkov obsahu pomocou algoritmov a pravidiel namiesto ručnej práce [1]. Takýto prístup umožňuje automatizovať proces vývoja generovaním rôznorodých a jedinečných prvkov obsahu na základe zadaných parametrov, ako aj čiastočne alebo úplne automatizovať proces ich vývoja, čím sa výrazne znižujú náklady na čas, ľudské a finančné zdroje

Procedurálne generovanie obsahu zahŕňa širokú škálu rôznych druhov obsahu, ako sú rôzne druhy povrchov, skál, skalných stien, stromov, kríkov, rastlín, predmetov, ako sú sudy, debničky, rôzne drevené artefakty, stroje, dopravné prostriedky apod. Procedurálne sú často generované aj efekty ako je oheň, dym, hmla, vodná hladina rieky, jazera alebo vodopád. V prípade komplexnejších a pokročilejších procedurálnych generátorov obsahu môžu byť generované aj komplexné modely celých budov, ulíc, štvrtí alebo dokonca celých miest. Vytvárané môžu byť celé otvorené svety (Open World Games), pričom ich rozloha často dosahuje nielen niekoľko km^2 ale dokonca niekoľko desiatok až stoviek km^2. Aj keď je vytvorenie kvalitného procedurálneho generátora náročná úloha, po jej splnení je následne možné generovať príslušný obsah mimoriadne efektívne, šetriac tak finančné, časové a ľudské zdroje. Príklady procedurálne generovaného obsahu – otvorených svetov – je možné vidieť na Obr. 1a až 1d.

Podobne ako vo videohrách, sa procedurálne generovanie často využíva aj v kinematografii na vytváranie vizuálne zaujímavých a detailných priestorov a scén [2]. Procedurálne generovanie virtuálnych scén, napr. komplexných miest s budovami rôznych veľkostí a výšok, je rovnako bežné ako generovanie scén, v ktorých je zachytené veľké množstvo postáv.

Úspešným príkladom procedurálneho generovania obsahu, ktoré pomohlo vytvoriť masívne scény pre filmový priemysel, je využitie softvéru spoločnosti Massive Software, ktorý vznikol pre účely sfilmovania trilógie Pán prsteňov, J.R.R. Tolkiena, režisérom Peterom Jacksonom. Práve vďaka tejto technológii procedurálneho generovania bola dosiahnutá bezprecedentná úroveň masívnosti natočených scén [3], tak ako to ukazuje Obr. 1e.

Obr. 1 Príklady výsledkov použitia procedurálneho generovania obsahu pri vytváraní otvorených svetov počítačových hier a) No Man's Sky (2016), b) Horizon Zero Dawn (2017), c) Starfield (2023) a d) Helldivers 2 (2024), e) až 200 000 bojovníkov v bitkách pri Helovom brale, na Pellenorských poliach a pri hradbách Minas Tirith bolo vytvorených pomocou riešenia spoločnosti Massive Software [4] pre účely filmového priemyslu.

Pre videohry je procedurálne generovaný obsah vytváraný aj s ohľadom na jeho renderovanie v reálnom čase, keď je cieľom okrem vizuálnej príťažlivosti a štylizácie dosahovať aj čo najvyšší počet renderovaných snímok obrazu za sekundu (Frames Per Second - FPS). To si vyžaduje optimalizáciu generovaného obsahu tak z hľadiska počtu polygónov, z ktorých je zložená geometria vytváraných modelov ako aj z hľadiska kvality textúr a nositeľov ďalších materiálových vlastností modelov.

Pri generovaní obsahu pre filmový priemysel nie je dôraz kladený na renderovanie v reálom čase ale na maximalizáciu kvality generovaných modelov a scén, ktoré sú z týchto modelov zostavené. Scény je potom možné generovať v off-line režime a to aj s využitím vysokovýkonných počítačov (High Performance Computing - HPC).

Ďalšou oblasťou, ktorá často využíva procedurálne generovanie obsahu, je vizualizácia architektonických modelov, ktoré sa zasadzujú do ich okolia pre lepšiu názornosť. Generované sú potom aj rozsiahlejšie vizualizácie okolia budov, prípadne v záujme vizualizácie urbanistických štúdií sú generované celé ulice, prípadne štvrte miest. Príkladom je využitie technológie CityEngine spoločnosti ESRI [5]. Skripty CGA umožňujú automatizované generovanie modelov budov (vrátane striech) a umožňujú tak aj zvyšovanie efektivity procesov v rámci urbanistického plánovania.

Tak ako je možné procedurálnym generovaním podporiť vizuálnu stránku obsahu, je možné generovať aj rôzne zvukové efekty, zvuky sprevádzajúce rôzne činnosti, charakterizujúce konkrétne objekty ako aj hudobné nahrávky, ktoré slúžia ako zvuková stopa (soundtrack) videohier alebo filmov [6]. V hrách pomáha procedurálne generovanie obsahu vytvárať zvuk, ktorý sa podrobnejšie prispôsobuje zážitkom hráča, vyhýba sa zbytočnému opakovaniu a poskytuje vyvíjajúci sa, emocionálne inteligentnejší zvukový sprievod hrania.

II. PROCEDURÁLNY GENERÁTOR SKÁL

Navrhnutý procedurálny generátor zahŕňa procedurálne generovanie tak geometrie modelov ako aj ich materiálových vlastností. V súlade s tým je riešenie procedurálneho generátora zostaveného v rámci tejto práce rozdelené na dva samostatné moduly:

Prvý modul je označený ako *RockGenerator,* je určený pre procedurálne generovanie geometrie modelu a je bližšie popísaný v podsekcii článku A. RockGenerator.

Druhý modul je označený ako *TextureGenerator*, je určený pre generovanie textúrovacích máp PBR textúr a je bližšie popísaný v podsekcii článku B. TextureGenerator.

A. RockGenerator

Program *RockGenerator* je určený pre vytvorenie polygonálneho povrchového modelu skaly. Je zodpovedný za tvorbu jedinečných 3D foriem imitujúcich prírodné kamenné štruktúry. Procedurálne generovanie je založené na základnom primitívnom objekte, ktorý sa následne deformuje, čím vzniká variabilita a realistickosť generovanej skaly. Algoritmus pre generovanie polygonálneho povrchového modelu skaly má nasledujúce kroky:

1. **Vytvorenie primitívnej siete** - Základom je objekt ako kocka alebo guľa, ktorý bude slúžiť ako počiatočná geometria. Táto geometria je realizovaná ako 3D sieť trojuholníkov, tvoriacich povrch tohto primitívneho telesa.

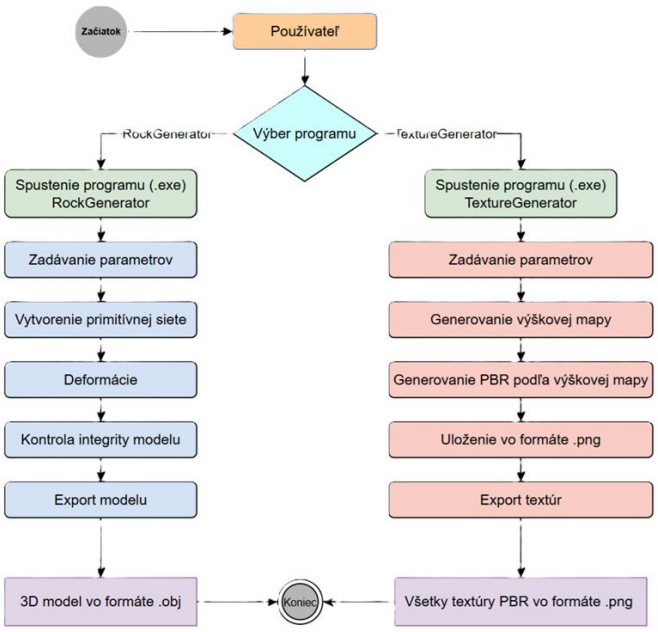
2. **Deformácie primitívnej siete** – Klasické a v oblasti počítačovej grafiky často používané algoritmy pre generovanie šumu, ako je Perlin Noise alebo Worley Noise sa použijú na pseudonáhodné posunutie súradníc vrcholov trojuholníkov tvoriacich sieť primitívneho telesa zostaveného v predchádzajúcom kroku. Tieto algoritmy majú parametre ako je veľkosť bunky, amplitúda, frekvencia, oktáva, útlm, pričom tieto parametre sa nastavujú ručne alebo sú generované ako náhodné čísla. Šum významne ovplyvňuje geometriu povrchového polygonálneho modelu generovanej skaly, dodáva mu náhodnosť a súčasne prírodnosť jeho vzhľadu.
3. **Kontrola integrity modelu** - V prípade, že sa po deformáciách a optimalizáciách objavia chyby, vykonáva sa detekcia týchto chýb a korekcia štruktúry siete a UV mapovania. Na tento účel sa vykonáva kontrola na výskyt geometrických konfliktov, ako sú medzery medzi polygónmi, ich prekrývanie a ďalšie topologické chyby.
4. **Export modelu** - Finálny model sa exportuje vo formáte WaveFront Technologies Object format (.obj), v jeho textovej verzii. OBJ formát bol zvolený ako jednoduchý, populárny a široko podporovaný formát umožňujúci v jednoduchej forme zaznamenať tak geometriu objektov ako aj ich materiálové vlastnosti. Tento formát však nie je plne kompatibilný s modernými Physically Based Renedering (PBR) textúrami, keď nepodporuje všetky mapy, ktoré sú bežnou súčasťou PBR materiálov.

B. TextureGenerator

Program *TextureGenerator* je určený pre vytvorenie PBR (Physically Based Rendering) textúr. V rámci jeho funkcionality je potrebné vytvoriť niekoľko máp, ktoré sa využívajú pri renderovaní. Algoritmus pre generovanie texturovacích máp má nasledujúce kroky:

1. **Výšková mapa (Height Map)** - Syntéza fraktálnych šumov (veľké aj malé detaily), vrátane simulácie žíl.
2. **Farebná textúra (Albedo)** - Farby sa priraďujú na základe výškových hodnôt z palety farieb, s pridaním šumovej variácie. V prípade aktivácie addMoss sa do spodných častí pridáva zelený odtieň pre simuláciu rastu machu.
3. **Drsnosť a metalickosť (Roughness a Metallic)** – Mapa drsnosti je generovaná tak, že sa jej lokálne hodnoty nastavujú podľa lokálnych hodnôt máp výšky a šumu. Metalickosť je nastavená na nulu, keďže generovaný materiál je prírodným materiálom.
4. **Normálová mapa (Normal Map)** - Z výškových rozdielov sa vypočítajú normály a prevedú sa do príslušného formátu využívajúceho farebné kanály používané pri ukladaní obrázkov do obrazového formátu png.
5. **Ambient Occlusion mapa (AO)** - Simuluje tiene v prehĺbeniach podľa zmien výšky.
6. **Export textúr** - Všetky textúry sa ukladajú vo formáte .png.

Takto generovaný PBR balík je kompatibilný s modernými hernými enginmi ako je Unreal Engine 5 alebo Unity a nástrojmi určenými pre modelovanie objektov ako je napr. Blender. Generované PBR textúry poskytujú realistický vzhľad, bez potreby ručnej úpravy. Diagram aktivít oboch modulov generátora možno vidieť na Obr. 2.



Obr. 2 Diagram aktivity zachytávajúci jednotlivé kroky generovania tak geometrie ako aj PBR textúrovacích máp.

III. Testovanie Dosiahnutého Riešenia

Pre otestovanie dosiahnutého riešenia boli osobitne realizované testy generovania geometrie polygonálnych povrchových modelov a testy generovania PBR textúr pre tieto modely.

A. Testovanie generovania geometrie modelov

Cieľom tohto testovania bolo vygenerovanie 100 kameňov s rôznym nastavením detailnosti v troch úrovniach: nízkej, strednej a vysokej, a zistenie vplyvu tohto parametra na čas generovania modelov, zaťaženie procesora počas tohto generovania a spotrebu pamäte. Výsledky testov sú zhrnuté v Tab. 1.

Tabuľka 1
Priemerné ukazovatele efektívnosti a celkový čas vykonávania programu pre generovanie geometrie polygonálnych povrchových modelov skál, podľa úrovne detailnosti pri generovaní 100 rôznych skál.

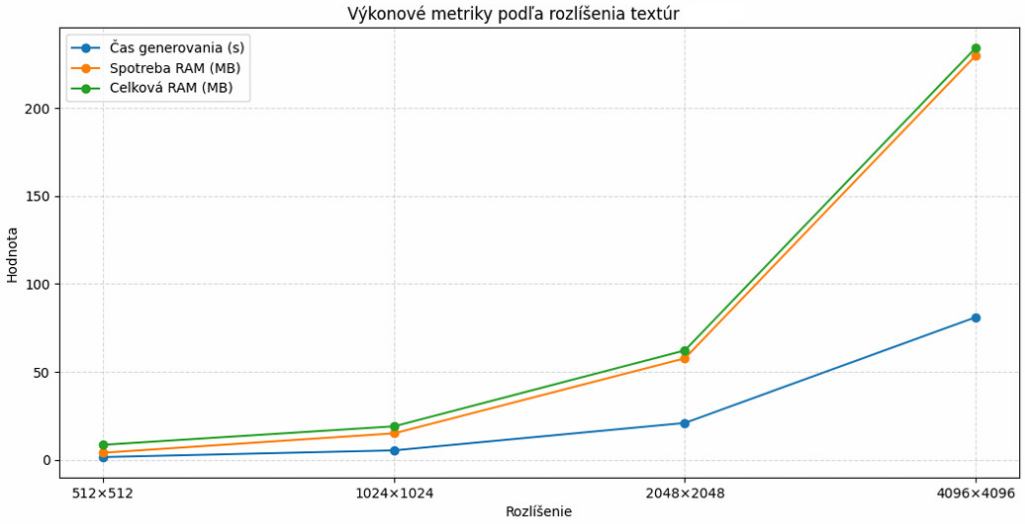
Úroveň detailov	Trojuholníky	Čas generovania [s]		Pamäť [KB]	CPU [%]
		1 skala	100 skál		
Nízka	294	0,20	20	13 663	100
Stredná	2 162	1,24	124	22 899	100
Vysoká	15 302	9,76	966	84 007	100

B. Testovanie generovania PBR textúr modelov

Cieľom tohto testovania bolo overiť, ako stabilne, správne a kvalitne funguje generovanie textúrovacích máp, a zmerať čas generovania, objem operačnej pamäte, ktorá bola použitá v procese generovania textúr, a celkovú spotrebu pamäte aplikáciou (vrátane systémových knižníc, prostredia vykonávania) pri rôznych rozlíšeniach 512×512, 1024×1024, 2048×2048 a 4096×4096 pixelov. Výsledky testov sú zhrnuté v Tab. 2 a Obr. 3.

Tabuľka 2
Parametre generovania textúrovacích máp PBR textúr v závislosti od ich rozlíšenia, nastaveného od rozlíšenia 512×512 až po 4096×4096 (4K).

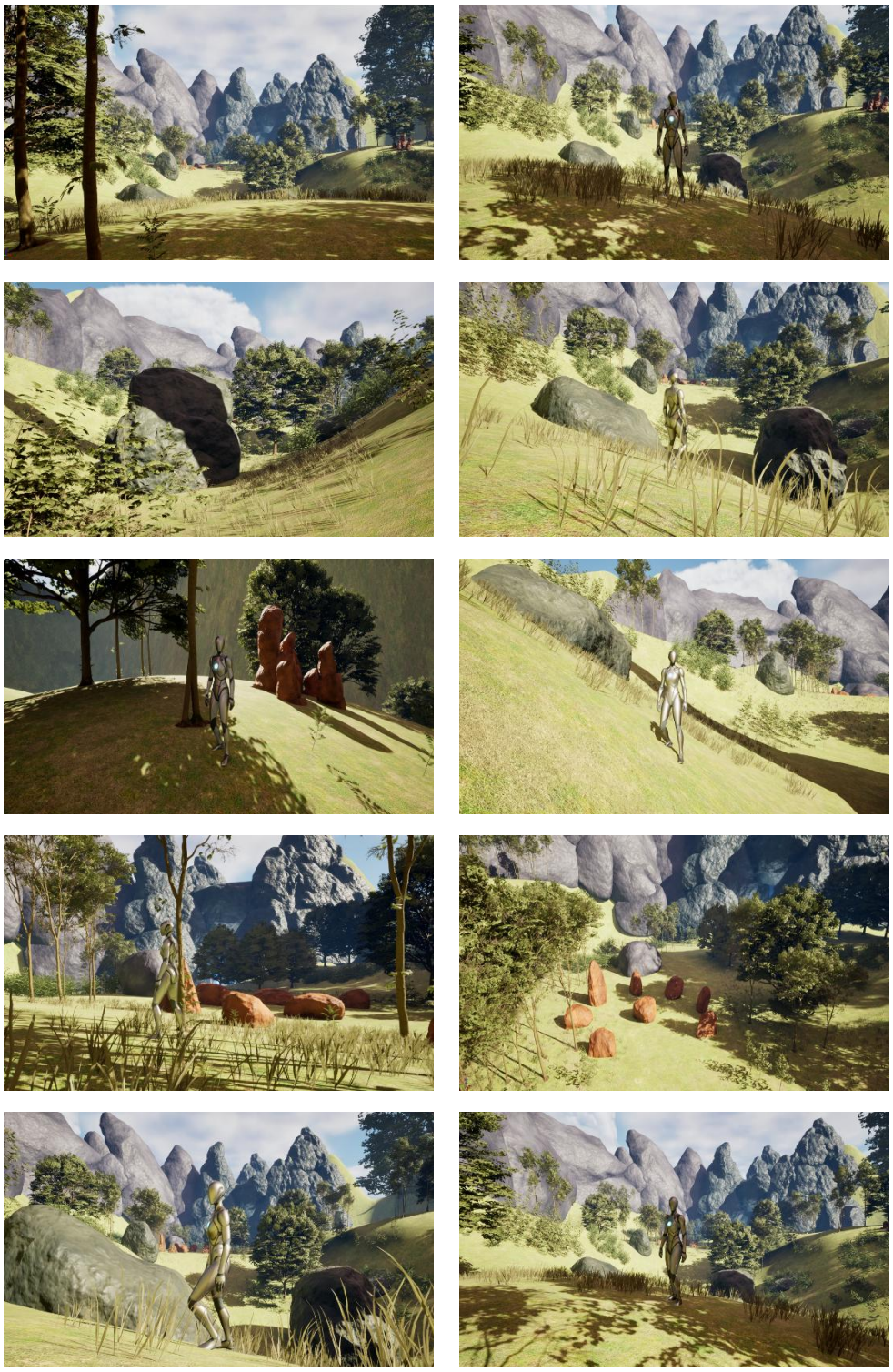
Rozlíšenie	Čas gene-rovania [s]	Spotreba RAM [MB]	
			Celková
512 x 512	1,55	4,0	8,5
1024 x 1024	5,34	15,0	19,0
2048 x 2048	20,92	57,7	62,2
4096 x 4096	81,05	229,9	234,3



Obr. 3 Parametre generovania textúrovacích máp PBR textúr v závislosti od ich rozlíšenia, nastaveného od rozlíšenia 512×512 až po 4096×4096 (4K).

IV. Výsledky Testovania

Na demonštrovanie a hodnotenie vizuálnej kvality generovaných modelov a textúr vo videohre bol zvolený herný engine Unreal Engine 5. Proces integrácie pozostával z niekoľkých fáz: vytvorenie mapy, import generovaných .obj modelov, vytvorenie PBR materiálu s použitím PBR textúr, nastavenie osvetlenia a umiestnenie modelov s textúrami na mapu. Na vytvorenie prirodzenej atmosféry boli použité nielen kamene, ale aj modely stromov, trávy a iných rastlín. Výsledkom týchto krokov bola mapa simulujúca prírodné prostredie. Snímky z vizualizovania scény prostredníctvom Unreal Engine možno vidieť na Obr. 4.



Obr. 4. Vizualizácie zasadenia procedurálne vygenerovaných skál a skalných útvarov do otvoreného sveta vymodelovaného s využitím Unreal 5.

V. Záver

V rámci tejto práce bol vyvinutý systém procedurálneho generovania obsahu v oblasti vývoja počítačových hier. Jedným z hlavných cieľov bolo vytvorenie nástroja na generovanie polygonálnych povrchových modelov trojrozmerných kamenných objektov a príslušných textúr, ktorý môže byť použitý v aktuálnych herných a vizualizačných nástrojoch.

Na základe analýzy bolo navrhnuté vlastné riešenie rozdelené do dvoch pmodulov, ktoré umožňujú ovládať parametre generovania a automaticky vytvárať realistickú 3D geometriu a textúry kameňa. Vykonané testy potvrdili účinnosť navrhovaného riešenia. Systém umožňuje generovať rôzne objekty s rôznym stupňom detailov, prispôsobiteľné požiadavkám herných scén. Demonštrácia výsledkov v Unreal Engine 5 ukázala kompatibilitu výstupných dát, vizuálnu kvalitu textúr a geometrie, ako aj vhodnosť použitia tohto prístupu pri vývoji obsahu pre videohry.

Ďalší rozvoj riešenia môže byť realizovaný pridaním plnohodnotného používateľského rozhrania, podporou rôznych formátov exportu (napr. glTF), ako aj integráciou algoritmov automatického generovania UV rozložení a fyzikálne správnych materiálov.

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] N. Shaker a J. Togelius and M.J. Nelson, "Procedural Content Generation in Games", Springer Cham, 26. Október 2016, ss. 237, ISBN 978-3-319-42714-0, DOI https://doi.org/10.1007/978-3-319-42716-4

[2] M. Scalabrin, L.A. Ripomanti, D. Maggiorini a D. Gadia, "Stereoscopy-based procedural generation of virtual environments", IS&T International Symposium on Electronic Imaging 2016, Stereoscopic Displays and Applications XXVII, 2016, ss. 1–7, DOI: https://doi.org/10.2352/ISSN.2470-1173.2016.5.SDA-042

[3] Massive Software, citované 7.6. 2025, dostupné tiež z: https://www.massivesoftware.com/about.html

[4] K.M. Thompson, "Scale, Spectacle and Movement: Massive Software and Digital Special Effects in The Lord of The Rings", From Hobbits To Hollywood: Essays On Peter Jackson's Lord Of The Rings eds. Ernest Mathijs and Murray Pomerance (Editions Rodopi, The Netherlands), 2006, ss. 283-299, DOI: https://doi.org/10.1163/9789401201513_021

[5] E.M.R.D. Ekanayaka, J. Zejdlik, O. Bittner, S. Popelka, J. Burian, O. Ruzicka, T. Vanicek a O. Jakubec: "Automated roof generation for the city of Olomouc using ArcGIS CityEngine", Abstr. Int. Cartogr. Assoc., 7, 35, DOI: https://doi.org/10.5194/ica-abs-7-35-2024

[6] D. Plans a D. Morelli, "Experience-Driven Procedural Music Generation for Games", IEEE Transactions on Computational Intelligence and AI in Games. September 2012, roč. 4, č. 3, ss. 192–198, DOI: https://doi.org/10.1109/TCIAIG.2012.2212899

Aplikácia pre potreby propagačných akcií na katedre KEMT

[1] *Dušan KANTOR, Ján PAPAJ*

[1]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]dusan.kantor@student.tuke.sk, jan.papaj@tuke.sk

Abstrakt **– Tento článok predstavuje návrh a vývoj multiplatformovej mobilnej aplikácie vytvorenej v technológii React Native, ktorej cieľom je podpora marketingovej komunikácie študijného odboru Počítačové siete na Katedre elektroniky a multimediálnych telekomunikácií. Aplikácia je určená pre stredoškolských študentov ako interaktívny prezentačný nástroj poskytujúci informácie o štúdiu, možnostiach uplatnenia a výhodách odboru, doplnené o multimediálny obsah.**

Súčasťou práce je návrh digitálnej kampane zameranej na zvýšenie záujmu o štúdium a budovanie pozitívneho imidžu katedry. Aplikácia je v testovacej prevádzke, pričom prvotné výsledky potvrdzujú jej prínos pre modernú a efektívnu komunikáciu s generáciou Z. Zároveň bol identifikovaný potenciál ďalšieho rozvoja v oblasti personalizácie, rozšírenia funkcií a analytického vyhodnocovania.

Kľúčové slová **– aplikácia, responzívnosť, react, propagácia**

I. Úvod

Pri návrhu mobilných aplikácií určených pre študentov je kľúčové zohľadniť špecifické princípy UX (user experience) a UI (user interface) dizajnu. Základným predpokladom úspešného návrhu je dôkladné pochopenie cieľovej skupiny, čo predstavuje prvý krok k vytvoreniu atraktívneho a funkčného používateľského prostredia [1].

Súčasné trendy v dizajne mobilných aplikácií sú silno ovplyvnené princípmi minimalizmu, ktoré kladú dôraz na jednoduchosť, prehľadnosť a maximálnu funkčnosť. Minimalistický dizajn pomáha znižovať kognitívnu záťaž používateľa, uľahčuje orientáciu v rozhraní a podporuje intuitívnu interakciu s aplikáciou. Podľa výskumu spoločnosti Google si používatelia vytvárajú prvý vizuálny dojem už do 50 milisekúnd, pričom jednoduché a čisté rozhrania sú vnímané ako esteticky atraktívnejšie [2]. Tento prístup zodpovedá preferenciám mladej generácie, ktorá uprednostňuje rýchly, efektívny a vizuálne jasný prístup k informáciám.

Dizajn zameraný na mladých používateľov by mal vychádzať z využitia dynamických vizuálnych prvkov, intuitívneho ovládania a logickej, plynulej navigácie. Uprednostniť sa odporúča známe používateľské vzory (design patterns), ktoré študenti prirodzene poznajú, čím sa výrazne znižuje krivka učenia pri prvom kontakte s aplikáciou. Nezastupiteľnú úlohu zohráva aj responzívnosť, teda schopnosť aplikácie prispôsobiť sa rôznym veľkostiam obrazoviek a typom zariadení bez straty funkčnosti či estetiky [3], [4].

Prístupnosť predstavuje ďalší zásadný faktor – je potrebné zabezpečiť dobrú čitateľnosť textov, vhodnú veľkosť písma a dostatočný kontrast medzi prvkami používateľského rozhrania [5]. Rovnako kritická je aj optimalizácia výkonnosti aplikácie, keďže mladší používatelia často očakávajú okamžité reakcie a minimálne časy načítania [5], [6] [7], [8].

Nemenej dôležitý je dôraz na personalizáciu obsahu, vizualizácií a samotného používateľského rozhrania, ktoré by mali byť prispôsobené preferenciám a očakávaniam cieľovej skupiny [3]. Takto navrhnutá aplikácia môže výrazne zvýšiť angažovanosť používateľov a efektívnosť komunikácie so študentmi (Viď Obr. 1).

V posledných rokoch narastá popularita vlastných ilustrácií na úkor tradičných fotografií. Ilustrácie podporujú kreativitu, sú ľahko prispôsobiteľné farebným schémam a dizajnovému štýlu aplikácie, čím umožňujú vytvoriť konzistentný a originálny vizuálny zážitok. Ich využitie môže zároveň prispieť k formovaniu výraznej vizuálnej identity, ktorá pomáha odlíšiť aplikáciu – a teda aj reprezentovanú inštitúciu – v konkurenčnom prostredí [4], [7].

Ďalším trendom, ktorý výrazne ovplyvňuje používateľský dojem, je použitie zaoblených tvarov v dizajne prvkov. Zaoblené hrany pôsobia prístupnejšie, vizuálne príjemnejšie a zároveň lepšie ladia s estetickými líniami moderných mobilných zariadení. Tento prvok môže významne zlepšiť celkový používateľský zážitok a zvýšiť atraktivitu aplikácie [1].

II. Návrh aplikácie využiteľnej na propagáciu katedry KEMT

V prostredí súčasnej digitálnej komunikácie zohráva mobilná aplikácia významnú úlohu pri propagácii akademických pracovísk. V prípade Katedry elektroniky a multimediálnych telekomunikácií (KEMT) sa ako efektívna stratégia javí integrácia interaktívnych a vizuálne atraktívnych prvkov, ktoré môžu osloviť a motivovať budúcich študentov k hlbšiemu záujmu o štúdium a výskum na katedre.

Interaktívna mapa patrí medzi kľúčové prvky mobilnej aplikácie, pretože umožňuje používateľom vizuálne objavovať priestory katedry, jednotlivé laboratóriá a ďalšie zaujímavé miesta. Funkcie ako posúvanie, približovanie, vyhľadávanie či zobrazenie doplnkových informácií po kliknutí na konkrétny bod výrazne zvyšujú komfort a atraktivitu používateľského rozhrania [9]. Takýto vizuálny a intuitívny prístup poskytuje potenciálnym študentom možnosť získať prvý dojem z akademického prostredia a môže pozitívne ovplyvniť ich rozhodovanie pri výbere študijného programu [2].

Ďalšou dôležitou funkciou aplikácie je interaktívny kalendár udalostí, v ktorom sú prehľadne zobrazené aktivity ako dni otvorených dverí, workshopy, prednášky, študentské prezentácie či súťaže. Možnosť nastaviť si pripomienky a získať podrobné informácie o jednotlivých podujatiach prispieva k väčšiemu zapojeniu používateľov do života katedry. Tým sa zvyšuje pravdepodobnosť nadviazania prvého kontaktu a vytvárania dlhodobých väzieb s potenciálnymi uchádzačmi. Multimediálny obsah predstavuje kľúčový pilier modernej a efektívnej propagácie. Krátke videá zamerané na prezentáciu výskumných aktivít, rozhovory so študentmi, pohľad do zákulisia výučby či zachytenie každodenného života na katedre ponúkajú dynamický a atraktívny spôsob komunikácie so širokým publikom [2]. Odporúča sa, aby videá nepresiahli dĺžku 5 minút a boli tematicky jasne zamerané, čo prispieva k udržaniu pozornosti a záujmu cieľovej skupiny [1].

Okrem videí je kľúčové aj využitie kvalitných obrázkov, ktoré vizuálne dokumentujú priestory katedry, vybavenie laboratórií, študentské projekty či spoločenské podujatia. V kontexte mobilných zariadení je dôležité zabezpečiť primeranú veľkosť a rozlíšenie obrázkov, ako aj ich tematickú relevanciu [9].

Interaktívne prvky ako tlačidlá, odkazy a výzvy k akcii umožňujú používateľom samostatne objavovať obsah aplikácie a interagovať s jednotlivými sekciami. Intuitívna navigácia a jasne definované prvky interakcie podporujú hlbšie preskúmanie informácií a zvyšujú používateľské zapojenie [1].

Personalizácia mobilnej aplikácie predstavuje významný prvok v oslovovaní mladých ľudí, ktorí sú zvyknutí na individuálne prispôsobené digitálne prostredie. Personalizácia v tomto kontexte zahŕňa prispôsobenie obsahu na základe preferencií, správania a záujmov používateľa [2]. Mobilná aplikácia môže napríklad odporúčať relevantné udalosti, prispôsobovať notifikácie či meniť obsahové rozloženie podľa predchádzajúcej interakcie s aplikáciou.

Takýto prístup nielenže zvyšuje angažovanosť používateľov, ale zároveň vytvára pocit osobného spojenia s katedrou. Z toho vyplýva vyššia miera záujmu o štúdium, aktívnu účasť na podujatiach a celkové pozitívne vnímanie akademického prostredia [2].

Moderný používateľský zážitok výrazne ovplyvňujú jemné animácie a vizuálna spätná väzba. Animácie pri posúvaní, kliknutí či prechodoch medzi obrazovkami dodávajú aplikácii pocit dynamiky a interaktivity, čo pozitívne pôsobí najmä na mladých používateľov zvyknutých na vizuálne bohaté rozhrania [6].

Dizajn pohybovej grafiky zohráva čoraz väčšiu úlohu pri navigácii, udržiavaní pozornosti a budovaní estetického dojmu [1]. Okrem toho je čoraz dôležitejšie prispôsobovať obsah individuálnym potrebám používateľov.

A. Návrh aplikácie využiteľnej na propagáciu katedry KEMT

Súbor $_layout.tsx$ predstavuje jeden z najzásadnejších prvkov celej aplikácie, pretože definuje jej hlavnú architektúru, organizačné usporiadanie a správanie navigácie. Na jeho začiatku sme využili možnosť importovať komponent Tabs priamo z knižnice expo-router, čo nám umožnilo jednoducho integrovať tabovú navigáciu ako nosnú štruktúru aplikácie. Tab navigácia v tomto prípade tvorí rámec, do ktorého sa dynamicky vkladajú jednotlivé obrazovky – ako napríklad Mapa, Kalendár, alebo index, pričom každá z nich je mapovaná pomocou komponentu $< Tabs.Screen/ >$ a identifikovaná cez atribút name. Konfigurácia každého tabu umožňuje individuálne nastaviť titul, ikonu, ako aj jej správanie pri aktívnom stave, čím sa zvyšuje intuitívnosť a použiteľnosť navigácie pre koncového používateľa.

V rámci vizuálneho prispôsobenia sme sa zamerali na responzívny dizajn, ktorý reflektuje preferovanú farebnú schému používateľského systému. Hook $useColorScheme()$ nám umožnil zistiť, či zariadenie používa svetlý alebo tmavý režim, a tomu sme následne prispôsobili vizuálne prvky aplikácie – farby ikon, textov, pozadia tab baru a iné, pričom všetky tieto hodnoty boli

Obr. 1 Ukážka používateľského rozhrania aplikácie pre propagáciu katedry KEMT

centralizovane spravované cez konfiguračný súbor *Colors*. Tento prístup výrazne zjednodušuje údržbu dizajnu a zaručuje konzistentnosť naprieč všetkými obrazovkami aplikácie (Viď Obr. 2).

Z hľadiska správy stavu sme implementovali *React Context API*, konkrétne vytvorením *IntroContext,* ktorý slúži ako zdieľaný komunikačný kanál medzi rôznymi komponentmi aplikácie. V kontexte sme definovali dve kľúčové stavové premenne: *showingIntro* a showingSplash, ktoré kontrolujú, či sa má používateľovi zobraziť úvodná prezentácia, respektíve úvodná obrazovka pri spustení aplikácie. Tieto premenne sme inicializovali s predvolenými hodnotami a spolu so zodpovedajúcimi funkciami *setShowingIntro* a *setShowingSplash* ich sprístupnili cez *IntroContext.Provider*, čo umožňuje flexibilné a synchronizované riadenie úvodných obrazoviek (Viď Obr. 3, Obr. 4).).

Z pohľadu logiky vykresľovania je komponent *AppLayout* postavený na princípe podmieneného

```
import { Tabs } from 'expo-router';
import React, { useState } from 'react';
import { TabBarIcon } from '@/components/navigation/TabBarIcon';
import { Colors } from '@/constants/Colors';
import { useColorScheme } from '@/hooks/useColorScheme';
import SplashScreen from '@/components/SplashScreen';
import { createContext } from 'react';

export const IntroContext = createContext({
  showingIntro: false,
  setShowingIntro: (value: boolean) => {},
  showingSplash: true,
  setShowingSplash: (value: boolean) => {},
});

export default function AppLayout() {
  const colorScheme = useColorScheme();
  const [showingIntro, setShowingIntro] = useState(false);
  const [showingSplash, setShowingSplash] = useState(true);
```

Obr. 2 Ukážka scriptu inicializácie aplikácie.

renderovania. V prípade, že je *showingSplash* nastavené na *true*, aplikácia zobrazí komponent *SplashScreen*, čím sa zabezpečí požadovaný úvodný zážitok pre používateľa. Ak je však táto podmienka neplatná, vykreslí sa samotná navigácia prostredníctvom komponentu *Tabs*. Navyše sme zaviedli dynamické skrytie navigačného menu počas zobrazovania úvodných obrazoviek a prezentácie – konkrétne pomocou štýlu $(displayshowingIntro?'none' :' flex')$, čím sa dosiahne čisté, nerušené prostredie bez navigačných prvkov počas úvodu.

Z pohľadu dizajnovej voľnosti sme tiež vedome deaktivovali globálne zobrazovanie hlavičky jednotlivých obrazoviek $headerShown : false$, čím sme umožnili, aby si každá obrazovka mohla určiť vlastný spôsob zobrazenia hlavičky, alebo ju nezobrazovala vôbec – v závislosti od konkrétneho kontextu a účelu.

Okrem navigácie a úvodných obrazoviek súčasťou tejto architektúry je aj komponent *TabBarIcon*, ktorý sme navrhli tak, aby presne odzrkadľoval požiadavky na vizuálnu konzistenciu a intuitívnosť. Ikony sa prispôsobujú aktívnemu stavu tabu a ich veľkosť a farba sa mení na základe parametrov *focused* a *color*, čo vizuálne zvýrazňuje aktuálne aktívnu položku.

Celkovo možno povedať, že súbor $_layout.tsx$ nie je len technickým vstupným bodom, ale zároveň predstavuje základnú kostru celej aplikácie. Zabezpečuje riadenie navigácie, správu stavu úvodných obrazoviek prostredníctvom kontextu, prispôsobenie vzhľadu aplikácie preferenciám používateľa a integráciu jednotlivých komponentov do jednotného a flexibilného rozhrania. Všetky tieto prvky spolu vytvárajú robustnú a rozšíriteľnú základňu pre ďalší vývoj aplikácie [4].

B. *Interaktívnosť v aplikácii*

Interaktívna mapa hlavného poschodia KEMT bola implementovaná v súbore *Mapa.tsx* ako komponent *MapScreen*. Tento komponent vizualizuje priestory katedry pomocou dotykovej mapy, umožňujúcej používateľovi prezerať si laboratóriá, vedenie katedry a jednotlivé sekcie chodby. Pre túto funkcionalitu boli využité knižnice *react*, *react-native* a *expo-av*.

Z hľadiska architektúry komponent využíva *useState* na riadenie interaktivity (napr. viditeľnosť modálnych okien, index zobrazeného obsahu), *useRef* na prácu s animáciami a videami a Animated pre plynulé prechody pri otváraní alebo zmene multimediálneho obsahu. Typy *LabRoomContent* a *LabRoomKey* zaručujú typovú bezpečnosť dát v TypeScripte [10].

Vizualizácia interiéru bola dosiahnutá kombináciou obrázkov a videí načítavaných pomocou *require()*. Tieto prvky sú prepojené na interaktívne body na mape – laboratóriá a tzv. *bodka chodby*, ktorá umožňuje používateľovi virtuálny pohyb po hlavnej chodbe poschodia. Každý interaktívny prvok má vlastné modálne okno s obsahom a animáciami pre zmenu obrázkov alebo videí (fade-in/fade-out), čo zvyšuje používateľský komfort.

Pre navigáciu medzi sekciami chodby boli implementované animované tlačidlá na pohyb nahor a nadol, ako aj dynamické štýly zobrazujúce pozíciu „bodky“ na chodbe. Interaktívne miestnosti umožňujú zobrazenie podrobného multimediálneho obsahu o laboratóriu (napr. prezentácie vybavenia, študentské projekty).

```
export default function SplashScreen() {
  const { setShowingSplash } = useContext(IntroContext);

  const logoScale = useRef(new Animated.Value(0.5)).current;
  const logoOpacity = useRef(new Animated.Value(0)).current;
  const buttonTranslateY = useRef(new Animated.Value(50)).current;
  const buttonOpacity = useRef(new Animated.Value(0)).current;

  useEffect(() => {
    Animated.parallel([
      Animated.timing(logoScale, {
        toValue: 1,
        duration: 800,
        easing: Easing.elastic(1),
        useNativeDriver: true,
      }),
      Animated.timing(logoOpacity, {
        toValue: 1,
        duration: 600,
        useNativeDriver: true,
      })
    ]).start();
```

Obr. 3 Ukážka scriptu animácie položiek.

```
<Tabs
  screenOptions={{
    tabBarActiveTintColor: Colors[colorScheme ?? 'light'].tint,
    tabBarStyle: {
      backgroundColor: '#F3AD3C',
      display: showingIntro ? 'none' : 'flex'
    },
    headerShown: false,
  }}>
  <Tabs.Screen
    name="index"
    options={{
      title: 'Home',
      tabBarIcon: ({ color, focused }) => (
        <TabBarIcon name={focused ? 'home' : 'home-outline'} color={color} />
      ),
    }}
  />
```

Obr. 4 Ukážka scriptu navigácie v jednotlivých položkách.

Renderovacia logika kombinuje komponenty ako *ScrollView*, *TouchableOpacity*, *Animated.View* a *Modal*, čím vytvára vizuálne pútavé a intuitívne rozhranie. Štýlovanie komponentov je centralizované v súbore $map_style.js$, čo zabezpečuje konzistentný vzhľad celej mapy.

Implementovaná interaktívna mapa predstavuje efektívny nástroj na vizuálnu prezentáciu priestorov katedry prostredníctvom mobilného zariadenia a výrazne prispieva k propagácii katedry medzi potenciálnymi študentmi.

III. Záver

Predstavená multiplatformová mobilná aplikácia, vyvinutá v prostredí React Native a Expo, predstavuje inovatívny nástroj digitálneho marketingu zameraný na propagáciu študijného programu Počítačové siete. S moderným dizajnom, responzívnym používateľským rozhraním a intuitívnou navigáciou cielene oslovuje študentov stredných škôl patriacich do generácie Z, pričom reflektuje ich digitálne očakávania a používateľské návyky.

Prototyp aplikácie je v súčasnosti dostupný v testovacej prevádzke. Hoci je funkčne plne pripravený, jeho ďalší vývoj bude smerovať k validácii prostredníctvom spätnej väzby od používateľov, optimalizácii používateľského zážitku a rozširovaniu funkcionalít. Tento prístup umožní

zabezpečiť dlhodobú udržateľnosť riešenia a jeho relevantnosť v meniacom sa prostredí digitálnej komunikácie.

Funkčný prototyp ponúka interaktívny prehľad o výučbových priestoroch a technickom vybavení katedry a efektívne dopĺňa tradičné formy komunikácie univerzity smerom k uchádzačom. Aplikácia preukázateľne reflektuje aktuálne UX/UI trendy a preukazuje zvýšenú mieru angažovanosti používateľov v porovnaní s pasívnymi informačnými kanálmi [6], [8].

Riešenie zároveň vytvára robustný základ pre ďalší rozvoj, vrátane možností personalizácie obsahu, integrácie interaktívnych prvkov či zavedenia automatizovaných aktualizácií. Tým sa otvára priestor pre jeho dlhodobé uplatnenie v kontexte strategického náboru a budovania značky univerzity.

Záverom možno konštatovať, že prototyp aplikácie naplnil stanovené ciele a priniesol konkrétny, technologicky relevantný výstup. Demonštruje potenciál mobilných technológií v akademickom prostredí a predstavuje významný krok smerom k modernej, digitálne orientovanej komunikácii s budúcimi študentmi.

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] H. Arif, H. S. Muhammad Bilal, F. A. Satti, S. I. Ali, and M. M. Fraz, "Cross-platform integration in legacy mobile apps for code reuse and reduced development efforts," in *2024 International Conference on Frontiers of Information Technology (FIT)*, 2024, pp. 1–6.

[2] V. K. Singh, P. Awasthi, L. Tripathi, S. Thakur, and P. Singh, "Mern (mongodb, express-js, react-js, node-js) stack web-based streaming and broadcasting application," in *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*, vol. 1, 2024, pp. 1–6.

[3] A. S. S. Ansyah, M. Z. Masruri, and S. Rochimah, "Usability testing of user experience and user interface design on mobile map applications: A comparative study of user perception and interaction," in *2023 14th International Conference on Information & Communication Technology and System (ICTS)*, 2023, pp. 7–12.

[4] E. Chang, M. T. Wang, R. Chen, S. P. Tan, and S. Y. Shen, "Design trend of graphical user interface," in *2014 International Symposium on Computer, Consumer and Control*, 2014, pp. 962–965.

[5] M. R. Fachrizal, A. Paramitha Fadillah, and A. Budiarto, "Uiux prototype usability analysis of e-commerce websites," in *2023 International Conference on Informatics Engineering, Science & Technology (INCITEST)*, 2023, pp. 1–6.

[6] U. D. Agency. (2025) Ui/ux design agency. [Online]. Available: https://fuselabcreative.com/mobile-app-design-trends-for-2025

[7] A. D. Gupta, N. K. Gujral, A. M. Yasir, N. Kumar, U. P. Pandey, and S. Kumar, "Ux/ui - companion app using data science," in *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, vol. 10, 2023, pp. 231–234.

[8] A. I. solutions. (20262) Top mobile appdesign in 2025. [Online]. Available: https://aglowiditsolutions.com/blog/app-design-trends/

[9] F. Mushtaq, F. Azam, and M. W. Anwar, "Performance comparison of single code base development tools: Flutter, react native, and xamarin," in *2024 14th International Conference on Software Technology and Engineering (ICSTE)*, 2024, pp. 17–23.

[10] C. Rippon, *Learn React with TypeScript: A beginner's guide to reactive web development with React 18 and TypeScript*, 2023.

Návrh informačného systému pre tieniacu komoru

[1]*Chrispopher NAGY,* [2]*Ján ZBOJOVSKÝ*

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

[1]christopher.nagy@student.tuke.sk, [2]jan.zbojovsky@tuke.sk

***Abstract* — Článok sa zameriava na návrh a realizáciu informačného systému a elektroinštalácie pre tieniacu komoru, ktorá slúži na laboratórne merania čiastkových výbojov. Cieľom práce je zlepšiť bezpečnosť a funkčnosť komory použitím senzorov, softvérového rozhrania a spoľahlivej elektroinštalácie. Systém je postavený na platforme Raspberry Pi a programovací jazyk Python, s použitím knižnice Tkinter, ktorý zobrazuje GUI na displeji. Okrem teoretických východísk sú opísané aj praktické aspekty realizácie a testovania.**

***Keywords* — elektroinštalácia, informačný systém, raspberry Pi, tieniaca komora, Tkinter**

I. ÚVOD

Táto práca sa venuje návrhu a realizácii informačného systému spolu s elektroinštaláciou pre tieniacu komoru, ktorá slúži na laboratórne merania čiastkových výbojov a je umiestnená v laboratóriu čiastkových výbojov na pôde Katedry elektroenergetiky Technickej univerzity v Košiciach.

Cieľom bolo vytvoriť moderné technické riešenie, ktoré zlepší bezpečnosť a prehľadnosť pri laboratórnych meraniach v tieniacej komore, kde je nevyhnutné eliminovať vplyv vonkajšieho elektromagnetického rušenia. Navrhnutý systém využíva jednodoskový počítač Raspberry Pi a bol naprogramovaný v jazyku Python, pričom grafické rozhranie sa vytvára pomocou knižnice Tkinter. Používateľ má k dispozícii dotykový displej, na ktorom môže v reálnom čase sledovať teplotu, vlhkosť, atmosférický tlak, elektrické napätie a stav dverí komory. Súčasťou riešenia je aj vlastnoručne navrhnutý lineárny usmerňovač na meranie napätia a nová elektroinštalácia, ktorá spĺňa požiadavky na bezpečnú prevádzku. Výsledkom je funkčný systém, ktorý je praktický, rozšíriteľný a využiteľný nielen pri výskume, ale aj pri výučbe v laboratóriách.

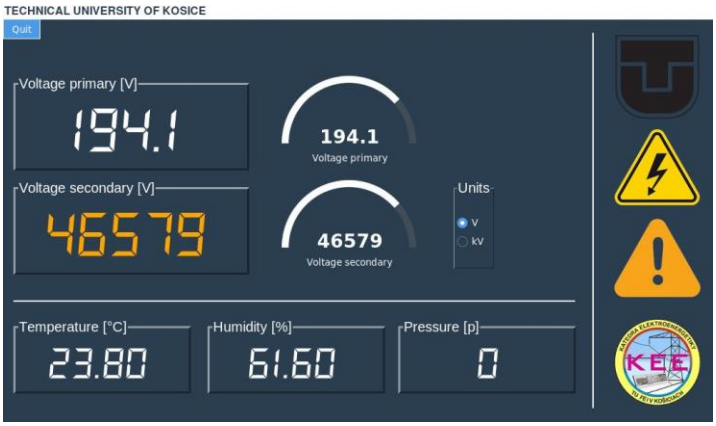
II. NÁVRH A REALIZÁCIA INFORMAČNÉHO SYSTÉMU

Systém beží na jednodoskovom počítači Raspberry Pi 4, ktorý vďaka svojmu výkonu, nízkej spotrebe a množstvu rozhraní predstavuje ideálnu platformu pre vývoj riadiacich a monitorovacích aplikácií. Vstupné údaje sú získavané zo senzorov teploty a vlhkosti DHT22 a tlaku BMP280, ktoré poskytujú digitálne výstupy. Napätie je merané pomocou navrhnutého presného jednocestného usmerňovača s operačným zosilňovačom LM358, ktorého výstup je pripojený na 16-bitový AD prevodník ADS1115. Prepojenie s Raspberry Pi je realizované cez rozhranie I²C, pričom hodnoty sú spracovávané a zobrazované v reálnom čase.

Užívateľské rozhranie, je navrhnuté v jazyku Python s využitím knižnice Tkinter, ktoré je optimalizované pre dotykové ovládanie a umožňuje sledovanie aktuálnych hodnôt veličín (teplota, vlhkosť, tlak, napätie) spolu s vizuálnym znázornením stavu dverí komory. Používateľ má možnosť prepínať medzi jednotkami (V/kV) a v prípade prekročenia prednastavených prahov systém automaticky zobrazí výstražné symboly, čím upozorňuje na rizikový stav. Prehľadný a moderný vzhľad rozhrania bol dosiahnutý vďaka použitiu knižnice ttkbootstrap.

Bezpečnostný prvok vo forme monitorovania otvorenia dverí je realizovaný cez digitálny vstup s galvanickým oddelením, čím sa zabezpečí, že systém nie je v priamom kontakte s napätím 230 V. Prvky informačného systému sú uložené v montážnej krabici s primeraným krytím, s dôrazom

na jednoduchosť údržby. Napájanie systému bolo riešené cez samostatný napájací zdroj s ochranou proti prepätiu.

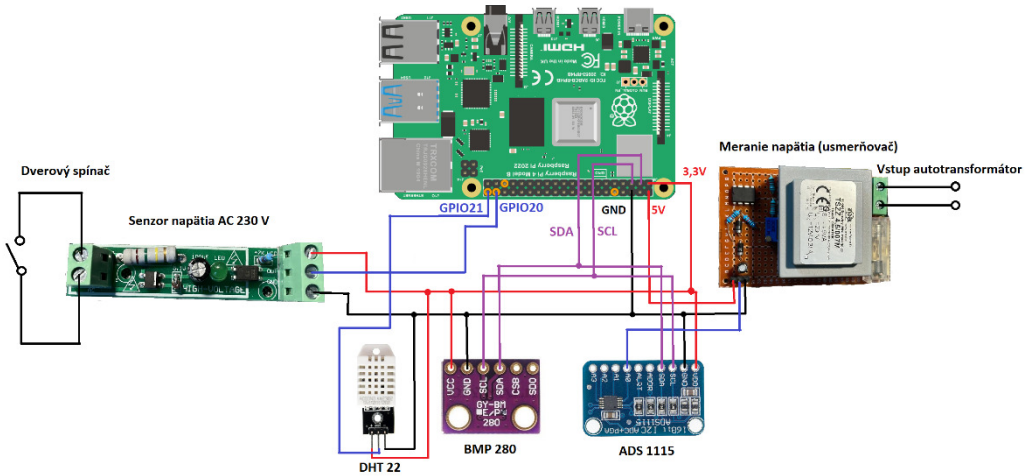


Obr. 1 Vytvorené grafické užívateľské rozhranie

Na meranie vlhkosti senzor DHT22 využíva dve elektródy a substrát, ktorý udržiava vlhkosť medzi nimi. Pri zmene vlhkosti sa mení vodivosť substrátu, a tým aj odpor medzi elektródami. Tieto zmeny sú vyhodnocované integrovaným obvodom, ktorý pripraví výstupný signál na spracovanie mikrokontrolérom. Teplotu senzor zisťuje pomocou termistora typu NTC čo je súčiastka, ktorej odpor klesá so zvyšujúcou sa teplotou. Termistor je vyrobený z polovodičových materiálov, ako je kremík alebo špeciálne polyméry, ktoré umožňujú výraznú zmenu odporu už pri malých zmenách teploty. Meraním tohto odporu sa dá presne určiť aktuálna teplota prostredia.[1]

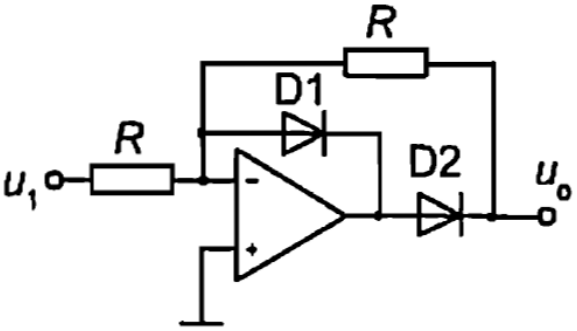
Senzor BMP280 slúži na meranie atmosférického tlaku a teploty. Je už z výroby skalibrovaný, takže je pripravený na použitie hneď po zapnutí. Podporuje komunikačné rozhrania SPI aj I2C, čo umožňuje jednoduchú integráciu do rôznych systémov. Presnosť merania tlaku je približne ±0,12 hPa a senzor je možné využiť aj ako výškomer s presnosťou do ±1 metra.[2]

Na zisťovanie stavu dverí sa používa senzor, ktorý deteguje prítomnosť napätia 230 V na dverovom spínači. Tento senzor je pripojený paralelne k spínaču. Vstupné napätie prechádza ochranným rezistorom s odporom 150 kΩ a následne sa usmerňuje pomocou usmerňovacieho mostíka. Potom je napätie znížené zenerovou diódou na bezpečnú úroveň 5 V. Signál je ďalej galvanicky oddelený optočlenom, ktorý zabezpečuje bezpečné prepojenie so vstupom Raspberry Pi.



Obr. 2 Použité senzory a schéma zapojenia IS

Presný usmerňovač je špeciálny typ usmerňovača, ktorý umožňuje prevod aj veľmi nízkych striedavých napätí na jednosmerné. Na rozdiel od bežných usmerňovačov, ktoré strácajú časť signálu kvôli prahovému napätiu diód (0,7 V pri kremíkových diódach), presný usmerňovač využíva operačný zosilňovač na kompenzáciu tohto úbytku. Vďaka tomu dokáže presne usmerniť aj slabé signály bez výrazného skreslenia. Tento obvod sa preto často používa v meracích zariadeniach a aplikáciách, kde je dôležitá presnosť spracovania nízkoúrovňových signálov.[3]



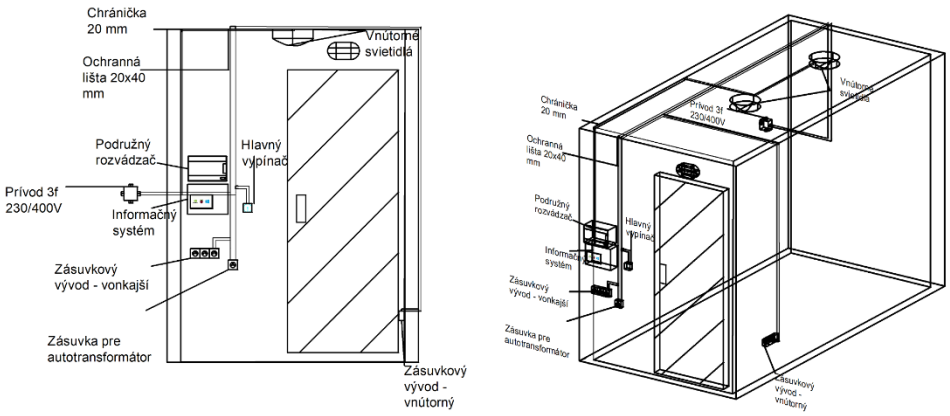
Obr. 3 Jednocestný usmerňovač s OZ a s dvoma diódami[3]

Sieťový transformátor pretransformuje napätie nadol na nižšiu a bezpečnejšiu hodnotu generovaného autotransformátorom. Nižšie napätie z transformátora sa ďalej rozdelí na deličke napätia pozostávajúceho z dvoch rezistorov s hodnotami 12 kΩ a trimrom s hodnotou 10 kΩ na ladenie koncového napätia. Požadovaná hodnota na výstupe trimra je v rozsahu 0 až maximálne 5 V. Zosilnenie invertujúceho zapojenia OZ je dané rezistormi R s hodnotami 10 kΩ.

Na jednodoskovom počítači Raspberry Pi beží webový server vytvorený pomocou frameworku Flask, ktorý je súčasťou jazyka Python. Tento server umožňuje spracovávať požiadavky HTTP a zobrazovať merané údaje v prehľadnej HTML podobe. Raspberry Pi je pripojené do lokálnej IoT siete, čo znamená, že je prístupné cez svoju IP adresu z ktoréhokoľvek zariadenia v tejto sieti. Po zadaní IP adresy Raspberry Pi do internetového prehliadača na akomkoľvek počítači v rovnakej sieti sa používateľovi automaticky zobrazí webová stránka generovaná Flask serverom. Na tejto stránke sú v reálnom čase zobrazované aktuálne údaje zo senzorov, ako napríklad teplota, vlhkosť, tlak vzduchu alebo napätie. Webová aplikácia tak poskytuje pohodlný spôsob vzdialeného monitorovania podmienok v tieniacej komore bez potreby priameho fyzického prístupu k zariadeniu.

III. Elektroinštalácia tieniacej komory

Elektroinštalácia tieniacej komory bola navrhnutá s cieľom zabezpečiť bezpečné, spoľahlivé a prehľadné napájanie jednotlivých častí komory, ako aj prepojenie s navrhnutým informačným systémom. Pôvodná elektroinštalácia už nezodpovedala bezpečnostným normám ani technickým požiadavkám, preto bola kompletne nahradená novým riešením.



Obrázok 4 Pohľady na tieniacu komoru

Systém obsahuje dva zásuvkové, jeden svetelný, jeden obvod pre napájanie autotransformátora a samostatný obvod pre napájanie jednodoskového počítača Raspberry Pi. Všetky obvody sú

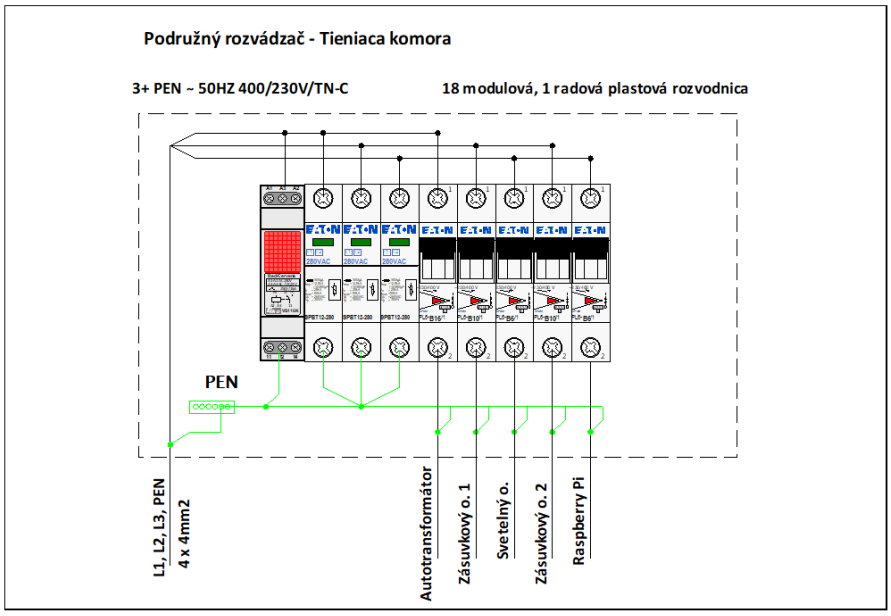
istené podľa príslušného výkonového zaťaženia a sú vedené káblami typu CYKY-J s medeným jadrom. Elektrické vodiče sú vedené povrchovo, v plastových rúrkach s priemerom 20 mm.

Ochrana proti prepätiu je riešená pomocou zvodičov prepätia s rôznymi úrovňami ochrany. Rozvádzač obsahuje prepäťovú ochranu typu C+D, zatiaľ čo obvod napájajúci Raspberry Pi má dodatočnú ochranu typu D, pre zabezpečenie vyššej ochrany počítača, ktorý môže byť zvlášť citlivý voči náhlym výkyvom napätia, ktoré môžu vzniknúť pri spínaní alebo počas meraní.

Súčasťou elektroinštalácie je aj osvetľovací systém pozostávajúci z jedného svetelného obvodu a dvoch svietidiel. Ovládanie osvetlenia je realizované vypínačom vo výške 1,2 m, umiestneným z vnútornej strany komory pri vstupe. Svetelný okruh je istený ističom B6.

Napájanie je zabezpečené z existujúcej trojfázovej zásuvky, pričom podružný rozvádzač je umiestnený na prednej strane tieniacej komory. Všetky komponenty sú zvolené a nainštalované tak, aby spĺňali normy STN a zabezpečili nielen funkčnosť, ale aj ochranu obsluhy a zariadení.

Na zvýšenie bezpečnosti je súčasťou elektroinštalácie aj výrazný trojfázový vačkový vypínač červenej farby, ktorý slúži ako hlavný vypínač a zároveň núdzový prvok pre okamžité odpojenie celej inštalácie.



Obr. 5 Prehľad prvkov podružného rozvádzača

IV. Záver

V tejto práci bol navrhnutý a realizovaný informačný systém a elektroinštalácia pre tieniacu komoru s cieľom zvýšiť bezpečnosť a zefektívniť laboratórne merania. Grafické rozhranie bolo vytvorené v jazyku Python pomocou knižnice Tkinter a beží na Raspberry Pi 4, ktoré sa osvedčilo ako spoľahlivé a výkonné riešenie. Súčasťou systému je aj presný usmerňovač pre meranie napätia a nový rozvádzač s prepäťovou ochranou. Testovanie potvrdilo funkčnosť a stabilitu celého systému. Do budúcnosti sa odporúča zvážiť použitie záložného zdroja pre Raspberry Pi, rozšírenie o ďalšie senzory a doplnenie ochranných prvkov pre vyššiu bezpečnosť pri meraniach.

Referencie

[1] Introduction to DHT22 [online]. [cit. 2025-06-07]. Dostupné na internete: https://www.theengineeringprojects.com/2019/02/introduction-to-dht22.html/?amp=1&fbclid=IwAR0hHflPaNTlVy36xE5mPI2MxSHLW0a4iuh2Cvywd7NaYmyltXwpgfH0GIk
[2] How Does the BMP280 Digital Pressure Sensor Work and how to Interface it with Arduino? [online]. [cit. 2025-06-07]. Dostupné na internete: https://circuitdigest.com/microcontroller-projects/interfacing-bmp280-sensor-with-arduino
[3] DOLEČEK, J.: Moderní učebnice elektroniky. BEN – technická literatura 2007. ISBN: 978-80-7300-187-2

Biometrický prístupový systém riadený a manažovaný cez počítačovú sieť

[1]*Renát HALUŠKA,* [2]*Matúš PLEVA,* [3]*Valér JAKUBČO*

[1,2,3] 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]matus.pleva@tuke.sk, [3]valer.jakubco@student.tuke.sk

***Abstrakt* — Tento článok predstavuje návrh a implementáciu biometrického prístupového systému založeného na rozpoznávaní tváre s využitím jednodoskového počítača Raspberry Pi 5 a kamery Intel RealSense. Navrhnutý systém kombinuje metódy detekcie tváre pomocou algoritmov HOG a CNN s mechanizmom overovania živosti prostredníctvom sledovania žmurknutia (EAR – Eye Aspect Ratio). Dôraz sa kladie aj na bezpečné uchovávanie biometrických údajov pomocou symetrického šifrovania (Fernet – AES v CBC móde s HMAC). Systém bol testovaný v kontrolovaných podmienkach, pričom dosiahol vysokú presnosť rozpoznávania. Výsledky ukazujú potenciál tohto riešenia pre použitie v nízkonákladových bezpečnostných aplikáciách.**

***Kľúčové slová* — tvárová biometria, Raspberry Pi, neurónové siete**

I. Úvod

V súčasnosti sa stretávame s neustále sa zvyšujúcimi požiadavkami na bezpečnosť a spoľahlivosť autentifikačných systémov, ktoré sú súčasťou každodenného života. Autentifikácia založená na klasických metódach, ako sú heslá, PIN kódy alebo RFID karty, už neposkytuje dostatočne spoľahlivú ochranu voči zneužitiu či neoprávnenému prístupu, pretože tieto formy autentifikácie sú náchylné na stratu, krádež alebo jednoduché uhádnutie. V tejto súvislosti sa čoraz častejšie obracia pozornosť na biometrické metódy, ktoré využívajú jedinečné fyzické alebo behaviorálne vlastnosti človeka, a tým zabezpečujú vyššiu mieru spoľahlivosti a bezpečnosti, a používateľského zážitku. Biometrické systémy na rozpoznávanie tváre patria medzi najviac rozvíjajúce sa a perspektívne oblasti biometrie. Výhodou rozpoznávania tváre v porovnaní s ostatnými metódami (ako napríklad odtlačky prstov alebo dúhovka oka) je najmä prirodzenosť a nenáročnosť procesu identifikácie, keďže tvár je možné zachytiť bez fyzického kontaktu a často aj bez aktívnej spolupráce používateľa. Táto jednoduchosť použitia vedie k širokému uplatneniu rozpoznávania tváre v rôznych sférach, od jednoduchého odomykania smartfónov až po komplexné bezpečnostné riešenia vo firmách, štátnych inštitúciách či na verejných miestach, alebo letiskách.

Aj napriek intenzívnemu výskumu pokročilosti technológií však zostáva niekoľko závažných problémov, ktoré bránia širšiemu využitiu biometrických systémov v praxi. Jedným z najväčších problémov v súčasnosti je otázka spoľahlivej detekcie živosti (liveness detection), teda schopnosti systému odlíšiť reálnu tvár živej osoby od falšovanej fotografie, masky, či videa. Vďaka pokročilým technológiám je totiž stále jednoduchšie vytvoriť presvedčivé falzifikáty, ktoré môžu byť úspešne použité na oklamanie autentifikačných systémov.

Ďalším problémom je bezpečné uchovávanie biometrických údajov používateľov. Biometrické údaje sú veľmi citlivé a ich únik môže mať vážne následky, keďže na rozdiel od hesiel alebo PIN kódov ich nie je možné jednoducho zmeniť. Preto je nevyhnutné, aby boli tieto dáta nielen presne zachytené a vyhodnocované, ale aj bezpečne uložené a chránené proti neoprávnenému prístupu. Motiváciou tejto práce je teda vytvorenie takého biometrického systému na rozpoznávanie tváre, ktorý integruje viaceré kľúčové aspekty ako je vysoká presnosť rozpoznávania využitím moderných neurónových sietí, detekciu živosti založenú na behaviorálnych prejavoch (konkrétne žmurkanie používateľa), a súčasne silnú ochranu biometrických údajov pomocou moderných kryptografických metód.

II. 3D Rozpoznávanie Tváre

V oblasti biometrie a rozpoznávania tváre prináša použitie 3D kamier oproti tradičným 2D kamerám niekoľko významných výhod. Jednou z hlavných výhod je schopnosť zachytávať hĺbkové informácie a presnú štruktúru tváre, čo umožňuje vytvárať detailné modely tváre a zlepšuje presnosť rozpoznávania. Táto schopnosť je obzvlášť užitočná pri variáciách v osvetlení a rôznych uhloch pohľadu, kde 2D systémy často zlyhávajú [1].

Ďalšou významnou výhodou 3D kamier je ich odolnosť voči podvodom, ako sú použitie fotografií, videí alebo masiek na oklamanie systému. Keďže 3D systémy zaznamenávajú hĺbkový profil tváre, dokážu efektívne rozlíšiť medzi skutočnou tvárou a plochým obrazom, čo zvyšuje bezpečnosť biometrických systémov [2]. Okrem toho 3D kamery poskytujú vyššiu presnosť a robustnosť rozpoznávania tváre, pretože hĺbkové mapy poskytujú viac informácií o štruktúre tváre, čo znižuje počet falošných pozitív a negatív. Taktiež umožňujú detailnejšiu analýzu biometrických čŕt, ako sú vzdialenosti medzi rôznymi bodmi na tvári, čo je menej citlivé na zmeny vzhľadu, ako sú fúzy, okuliare alebo účes [3].

V neposlednom rade 3D kamery fungujú spoľahlivo aj v nízkych svetelných podmienkach alebo úplnej tme, najmä ak využívajú infračervené senzory. To umožňuje ich použitie v rôznych prostrediach, kde by 2D kamery mohli mať problémy so zachytávaním kvalitných obrazov [6]. Na základe týchto výhod je zrejmé, že 3D kamery prinášajú významné zlepšenia v oblasti rozpoznávania tváre a biometrických aplikácií, čím zvyšujú bezpečnosť a spoľahlivosť týchto systémov [4].

A. Stereoskopické videnie

Stereoskopické videnie je technika, ktorá umožňuje vytvárať ilúziu trojrozmerného obrazu pomocou dvoch zobrazení tej istej scény, každého z iného uhla. Tento princíp je základom mnohých 3D zobrazovacích technológií, vrátane 3D kamier, ktoré sa používajú v rôznych aplikáciách, ako sú robotika, rozšírená realita a počítačové videnie [5]. Podstata stereoskopie spočíva v tom, že každé „oko“ (resp. kamera) získa mierne odlišný pohľad na scénu a rozdiel medzi týmito dvoma pohľadmi (tzv. disparita) je základom na trianguláciu polohy objektov v 3D priestore [6]. V biologickom systéme túto informáciu spracúva mozog a v dôsledku toho vzniká vnem hĺbky. V umelých systémoch je proces hĺbkovej rekonštrukcie zabezpečovaný algoritmami stereo matchingu (párovanie bodov medzi dvoma obrazmi).

Základné kroky pri spracovaní stereoskopického obrazu:

1. Kalibrácia kamier: Každá kamera musí byť kalibrovaná na určenie vnútorných a vonkajších parametrov (ohnisková vzdialenosť, skreslenie atď.).
2. Rectifikácia obrazu: Oba zachytené obrazy (z „ľavej“ a „pravej“ kamery) sa upravujú tak, aby sa vzájomné epipolárne čiary zrovnali - to zjednodušuje hľadanie korešpondencií medzi obrazmi [7].
3. Vyhľadanie korelácií (stereo matching): Algoritmus hľadá zodpovedajúce si pixely alebo oblasti v ľavom a pravom obraze. Rozdiel v horizontálnej polohe (disparite) následne určuje hĺbkovú informáciu.
4. Triangulácia a vytvorenie hĺbkovej mapy: Na základe nájdených korelácií a parametrov kamier sa vypočíta hĺbková mapa (3D interpretácia scény).

V praxi sú pre stereo matching dostupné rôzne algoritmy (lokálne, globálne, založené na hlbokých neurónových sieťach a pod.), ktoré sa líšia presnosťou a výpočtovou náročnosťou [8].

B. Kamery Intel RealSense

Kamery Intel RealSense sú navrhnuté tak, aby poskytovali 3D snímanie pre široké spektrum aplikácií. Pracujú na princípe stereoskopickej rekonštrukcie: majú dvojicu obrazových snímačov (napr. ľavý a pravý infračervený) a doplnkové projektory (napr. infračervené vzory) na zlepšenie kontrastu pri scénach s nízkou textúrou [9].

Hlavné výhody Intel RealSense kamier:

- **Integrované spracovanie:** Mnohé modely obsahujú samostatný hardvérový spracovateľ hĺbky, ktorý vykonáva stereo matching v reálnom čase.
- **Kompaktné rozmery:** Vhodné pre robotické a mobilné zariadenia (napr. drony, autonómne roboty).
- **Rozsiahla softvérová podpora:** Knižnica librealsense (open-source) uľahčuje integráciu do rôznych programovacích prostredí (C++, Python a pod.) [10].

S pribúdajúcimi verziami kamier Intel RealSense sa zvyšuje presnosť a použiteľný hĺbkový rozsah, a zároveň sa zlepšuje adaptácia na náročné scény s horším osvetlením a viacerými objektmi v zábere. Tieto senzory nachádzajú využitie najmä v inteligentnom spracovaní obrazu, kontrole kvality vo výrobných procesoch alebo bezpečnostných systémoch (napr. 3D

rozpoznávanie tvárí).

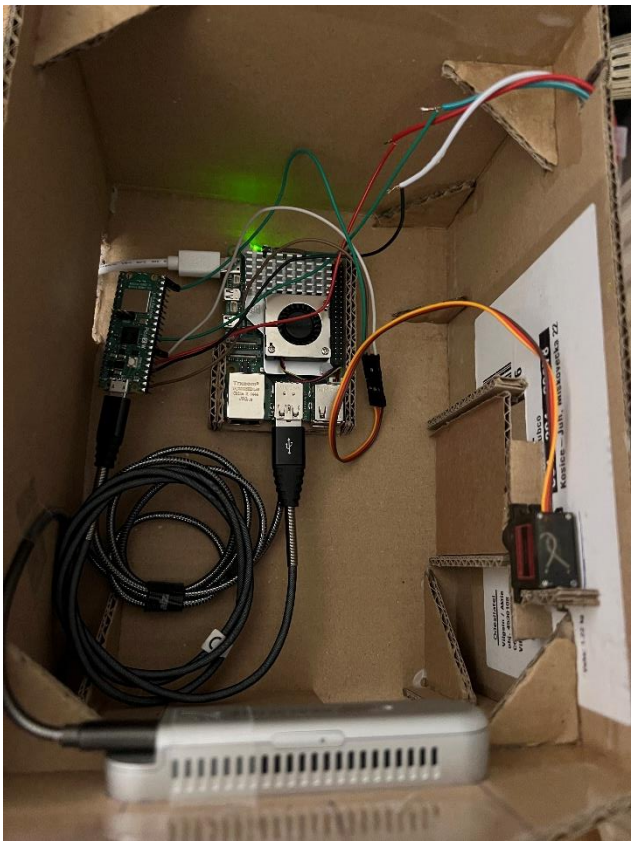
III. Hardvérová a Softvérová Konfigurácia Biometrického Systému

Centrálnym prvkom navrhovaného biometrického systému je jednodoskový počítač Raspberry Pi 5, ktorý disponuje vyšším výkonom oproti predchádzajúcim verziám. Z dôvodu zvýšeného výkonu, a teda aj vyššej produkcie tepla, bolo podľa odporúčania výrobcu potrebné použiť vhodný kryt s aktívnym chladením vo forme integrovaného ventilátora. Tento spôsob chladenia zabezpečuje stabilnú prevádzku zariadenia aj pri zvýšenej záťaži procesora.

Súčasťou systému je aj mikrokontrolér Raspberry Pi Pico, ktorý slúži na generovanie PWM (Pulse Width Modulation) signálu. Tento mikrokontrolér je priamo pripojený k hlavnému počítaču Raspberry Pi 5 prostredníctvom USB rozhrania, ktoré zároveň zabezpečuje aj jeho napájanie. Takto realizované prepojenie minimalizuje potrebu dodatočných zdrojov napájania a výrazne zjednodušuje celkovú hardvérovú architektúru systému. Raspberry Pi Pico je využívaný primárne na riadenie periférnych zariadení, ktoré vyžadujú PWM signály, v našom prípade servo-motor.

Na zariadení Raspberry Pi 5 bol použitý operačný systém Raspberry Pi OS vo verzii založenej na distribúcii Debian Linux (predtým známy ako Raspbian). Výber tohto operačného systému vychádzal z jeho optimalizácie pre jednodeskové počítače Raspberry Pi, dobrej podpory komunity a vysokej stability. Raspberry Pi OS zároveň poskytuje široké možnosti správy balíkov pomocou nástrojov ako apt, čo výrazne zjednodušuje proces inštalácie potrebného softvéru a knižníc.

Celá praktická implementácia systému (Obr. 1) bola realizovaná v programovacom jazyku Python, konkrétne vo verzii 3.10. Táto verzia bola na Raspberry Pi kompilovaná a nainštalovaná z dôvodu lepšej kompatibility s knižnicami tretích strán. Na izoláciu prostredia a správu závislostí bola využitá funkcionalita Python virtual environments, čo umožnilo čisté oddelenie potrebných knižníc a ich jednoduchú správu.



Obr. 1 Ukážka uloženia hardvéru v modeli (Raspberry Pi Pico, Raspberry Pi 5, Intel RealSense kamera)

Pre realizáciu rozpoznávania tváre a spracovania obrazu boli použité tieto hlavné Python knižnice:

- **face_recognition** táto knižnica je postavená na algoritmoch implementovaných v

knižnici dlib a poskytuje jednoduché rozhranie pre detekciu a rozpoznávanie tvárí. Pomocou nej je možné extrahovať embeddings tvárí (vektorové reprezentácie tváre), ktoré sú neskôr využívané na rozpoznávanie. Knižnica umožňuje výber rôznych modelov, napr. CNN alebo HOG, čo umožňuje flexibilné nastavenie presnosti a rýchlosti.

- **dlib** výkonná knižnica napísaná v jazyku C++, poskytujúca nástroje pre strojové učenie, spracovanie obrazu a počítačové videnie. Je známa predovšetkým presnou a rýchlou detekciou landmarkov tváre (charakteristický bod, napr. kútik oka). V tejto práci bola využitá na extrahovanie landmarkov a detekciu živosti prostredníctvom metódy Eye Aspect Ratio (EAR).
- **OpenCV** (Open Source Computer Vision Library) najznámejšia knižnica pre spracovanie obrazu a videa. Použitá bola najmä na načítavanie a spracovanie obrazových dát zo súborov aj z kamery, konverziu farieb medzi rôznymi farebnými priestormi (napríklad BGR na RGB), ako aj na prípadnú vizualizáciu výsledkov spracovania obrazu.
- **cryptography (Fernet)** knižnica cryptography bola použitá na zabezpečenie citlivých biometrických údajov (embeddings tvárí). Šifrovanie je realizované pomocou mechanizmu Fernet, ktorý využíva AES šifrovanie v CBC móde spolu s autentifikáciou HMAC. Použitie Fernetu zabezpečuje dôvernosť a integritu uložených údajov bez nutnosti inštalácie ďalších externých softvérových komponentov.
- **imutils** pomocná knižnica pre jednoduchšiu prácu s obrazovými dátami, ktorá zjednodušuje operácie ako získavanie zoznamu obrázkov z adresárov, úpravy veľkosti obrázkov a podobne.
- **pickle** štandardná knižnica jazyka Python na serializáciu a deserializáciu dátových štruktúr, použitá na uloženie a načítanie embeddings do súboru.

IV. Proces Rozpoznávania Tváre v Reálnom Čase

Úvodným krokom procesu rozpoznávania je načítanie biometrických údajov používateľov, ktoré boli predtým vygenerované počas tréningovej fázy. Proces načítania a dešifrovania prebieha nasledovne:

1. Systém najskôr overí existenciu súboru s embeddings.
2. Ak súbor existuje, načíta jeho obsah a dešifruje údaje pomocou kryptografického kľúča, ktorý je bezpečne uložený v systémových premenných prostredia.
3. Následne sú dešifrované embeddings deserializované pomocou knižnice pickle, čím sa obnovia pôvodné dátové štruktúry.

Proces rozpoznávania tvárí v reálnom čase je realizovaný pomocou video streamu z kamery Intel RealSense D415 pripojenej k zariadeniu Raspberry Pi 5. Na detekciu tváre je použitá rýchla a efektívna metóda HOG (Histogram of Oriented Gradients), ktorá je dostatočne výpočtovo efektívna aj pre menej výkonné zariadenia, ako je Raspberry Pi. Dôležitou súčasťou rozpoznávania je detekcia živosti, ktorá zabezpečuje, že pred kamerou sa nachádza reálna, živá osoba, a nie statická fotografia alebo video. V tomto systéme je implementovaná metóda detekcie živosti založená na sledovaní žmurknutia používateľa pomocou výpočtu EAR (Eye Aspect Ratio). Tento postup využíva tvárové landmarky extrahované knižnicou dlib, pričom žmurknutie je indikátorom reálnosti a aktívnej prítomnosti osoby pred kamerou.

Po úspešnej detekcii žmurknutia sa systém pokúsi o rozpoznanie osoby porovnaním aktuálne extrahovaných embeddings s databázou uložených embeddings. Rozhodovanie o identite osoby sa vykonáva pomocou vypočítania euklidovskej vzdialenosti embeddings a porovnaním výsledku s prahovou hodnotou (threshold). Ak je táto vzdialenosť menšia ako prahová hodnota, osoba je identifikovaná ako známa; v opačnom prípade je označená ako neznáma („Unknown“).

Pri vytváraní a testovaní tvárového rozpoznávacieho systému je nevyhnutné správne definovať postup, ktorým budeme vyhodnocovať funkčnosť a presnosť natrénovaných modelov. Zvolený spôsob testovania významne ovplyvňuje objektivitu získaných výsledkov. V rámci tejto práce bola preto zvolená metóda delenia dátového súboru na trénovaciu a testovaciu množinu. Táto testovacia metóda patrí medzi štandardizované a všeobecne akceptované prístupy v oblasti strojového učenia, predovšetkým v oblasti klasifikácie, ku ktorej tvárové rozpoznávanie prirodzene patrí. Dataset použitý v tejto práci pozostával z fotografií tvárí rôznych osôb, pričom každý subjekt (každá osoba) bol zastúpený rovnakým počtom fotografií, konkrétne 14 snímkami. Dôvodom rovnakého počtu fotografií pre každú osobu je minimalizácia potenciálneho skreslenia výsledkov pri tréningu modelu a jeho následnom vyhodnocovaní.

Zvolený dataset bol rozdelený v pomere približne 70:30, čo predstavuje štandardne používaný pomer v oblasti strojového učenia pri menších datasetoch. Pri väčšom počte dostupných fotografií je možné zvoliť aj vyšší podiel pre tréningové dáta, avšak v našom prípade bol zvolený práve pomer 70:30 z dôvodu dostatočného počtu snímok pre tréning aj testovanie. Konkrétne v prípade 14 fotografií pripadajúcich na jednu osobu bolo rozdelenie nasledovné:

- Trénovacie dáta (70%): 10 fotografií každej osoby. Tieto fotografie slúžili výlučne na učenie modelu, ktorý si vďaka nim vytvoril interné vektorové reprezentácie (tzv. encodings). Model teda využil tieto dáta na to, aby bol schopný neskôr identifikovať konkrétne tváre.
- Testovacie dáta (30%): 4 fotografie každej osoby. Tieto dáta neboli použité počas tréningu, čo zabezpečilo objektivitu vyhodnocovania presnosti modelu. Úlohou testovacích dát bolo nezávisle overiť, či model dokáže správne identifikovať tváre.

Z výsledkov uvedených v Tabuľke 1 vyplývajú jasné rozdiely v čase spracovania pri použití algoritmov HOG a CNN. Algoritmus založený na konvolučných neurónových sieťach (CNN) vykazuje výrazne vyššie výpočtové nároky, čo sa prejavuje vo výrazne dlhších časoch spracovania. Naopak, metóda HOG poskytuje dramaticky rýchlejšie výsledky, s približne 25-krát nižším časom spracovania ako CNN.

Rozdiel medzi small a large modelom sa v prípade CNN aj HOG ukázal byť relatívne malý. Pri použití CNN je rozdiel iba niekoľko sekúnd, podobne aj pri HOG je tento rozdiel zanedbateľný (menej než jedna sekunda). Tento minimálny rozdiel je spôsobený tým, že hlavným zdrojom výpočtovej náročnosti CNN modelu je práve detekcia tvárí, pričom extrakcia bodov samotných (či už 5 alebo 68) prispieva k celkovému času relatívne málo. Pri HOG metóde, ktorá je vo všeobecnosti veľmi rýchla, sa prejavuje ešte menší rozdiel medzi malým a veľkým modelom.

Tabuľka 1
Časy trénovania pre rozdielne modely a prístupy

Model	Trénovací čas
CNN + Large Model	26 m 43.302 s
CNN + Small Model	26 m 36.678 s
HOG + Large Model	1 m 1.316 s
HOG + Small Model	1 m 1.168 s

CNN algoritmus je výrazne pomalší, pretože konvolučné neurónové siete sú výpočtovo veľmi náročné. CNN vyžaduje realizáciu mnohých matematických operácií na rôznych vrstvách neurónovej siete (konvolúcie, pooling, aktivácie). Na druhej strane, algoritmus HOG používa len jednoduchšie výpočty gradientov v pixelových blokoch, ktoré je možné realizovať veľmi efektívne.

A. Malý vs. veľký model

Malý model (small): Vracia iba 5 základných landmarkov, čím výrazne obmedzuje počet bodov potrebných na výpočet a tým aj výpočtovú záťaž pri samotnom určovaní landmarkov. Je vhodný pre menej detailné aplikácie, kde postačuje základná detekcia.

Veľký model (large): Obsahuje detailnú informáciu o 68 landmarkoch, čo umožňuje presnú identifikáciu rôznych častí tváre (napr. pery, obočie, brada). Zvýšený počet bodov znamená síce vyššiu presnosť, no zároveň aj vyššie nároky na výpočtovú kapacitu a mierne zvýšený čas spracovania.

V teste sa však ukázalo, že rozdiel medzi malým a veľkým modelom je relatívne malý v porovnaní s rozdielom medzi CNN a HOG. To naznačuje, že výpočtová záťaž je primárne spôsobená detekciou tváre (CNN vs. HOG) a nie počtom landmarkov (5 vs. 68).

V. ZÁVER

Cieľom tejto práce bolo navrhnúť, implementovať a vyhodnotiť spoľahlivý biometrický systém založený na rozpoznávaní tváre, ktorý kombinuje moderné algoritmy strojového učenia s dôrazom na bezpečné uchovávanie biometrických údajov a detekciu živosti používateľa. V úvode stanovené ciele zahŕňali návrh vhodného hardvérového a softvérového riešenia, realizáciu rozpoznávania tváre v reálnom čase s využitím jednodoskového počítača Raspberry Pi 5 a kamery Intel RealSense D415, ako aj zabezpečenie citlivých údajov pomocou kryptografických metód.

Raspberry Pi 5 sa ukázal ako dostatočne výkonný hardvér, schopný zvládnuť nielen

spracovanie obrazových dát v reálnom čase, ale aj trénovanie neurónových modelov, napriek určitým výpočtovým obmedzeniam. Kombinácia CNN veľkého modelu pri tréningu embeddings a rýchlej metódy HOG pre real-time detekciu tváre umožnila zachovať dostatočne vysokú presnosť aj rýchlosť systému. Jedným z pôvodných cieľov bolo využitie možností 3D kamery Intel RealSense D415 na realizáciu robustnej detekcie živosti (liveness detection) na základe analýzy trojrozmerných obrazových dát. Avšak v priebehu práce sa ukázalo, že implementácia tejto pokročilej metódy je výrazne komplexnejšia, než bolo pôvodne predpokladané, a napriek vynaloženému úsiliu sa nepodarilo dosiahnuť plnohodnotnú a spoľahlivú integráciu 3D detekcie živosti do výsledného systému.

Ako alternatívne riešenie bola preto implementovaná jednoduchšia metóda detekcie živosti založená na sledovaní žmurknutia používateľa pomocou Eye Aspect Ratio (EAR). Táto metóda síce neposkytuje rovnakú mieru robustnosti ako pokročilé 3D techniky, ale dokáže spoľahlivo detegovať základné pokusy o falšovanie, napríklad pomocou statických fotografií alebo obrazoviek. Aj keď sa jedná o menej presnú metódu, ukázalo sa, že je dostatočne efektívna pre praktické použitie v navrhnutom systéme. Ďalším významným prínosom práce je dôkladná implementácia kryptografickej ochrany biometrických údajov používateľov.

Použitá bola symetrická kryptografia AES v CBC móde s autentifikáciou HMAC pomocou knižnice Fernet. Táto implementácia spoľahlivo zabezpečila integritu a dôvernosť uložených biometrických dát, čím sa výrazne posilnila bezpečnosť systému. V rámci testovania systému bola dosiahnutá 100% presnosť rozpoznávania tváre v kontrolovaných podmienkach, čo potvrdzuje spoľahlivosť zvolených algoritmov.

Poďakovanie

Túto prácu čiastočne podporila Vedecká grantová agentúra Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky a Slovenská akadémia vied v rámci projektu VEGA 2/0165/21, Kultúrna a vzdelávacia grantová agentúra Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky prostredníctvom projektu 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] K. W. Bowyer, K. Chang, and P. Flynn, “A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition,” *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, Jan. 2006, doi: 10.1016/j.cviu.2005.05.005.

[2] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," *2011 International Joint Conference on Biometrics (IJCB)*, Washington, DC, USA, 2011, pp. 1-7, doi: 10.1109/IJCB.2011.6117503.

[3] X. Zhang and Y. Gao, “Face recognition across pose: A review,” *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, Nov. 2009, doi: 10.1016/j.patcog.2009.04.017.

[4] A. Zabatani *et al*., "Intel® RealSense™ SR300 Coded Light Depth Camera," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2333-2345, 1 Oct. 2020, doi: 10.1109/TPAMI.2019.2915841

[5] AJ. Zhang, Z. Lu, W. Li and Q. Liao, "A robust and fast 3D face reconstruction method using realsense camera," *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, India, 2017, pp. 2691-2695, doi: 10.1109/WiSPNET.2017.8300251.

[6] M. Rabbani, M. J. Mia, T. Khan and M. I. I. Zarif, "A Survey on RealSense: in context of Research and Application," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225558.

[7] O. Messai and A. Chetouani, "End-to-End Deep Multi-Score Model for No-Reference Stereoscopic Image Quality Assessment," *2022 IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, 2022, pp. 2721-2725, doi: 10.1109/ICIP46576.2022.9897616.

[8] J. Feng, S. Li and Y. Chang, "Binocular Visual Mechanism Guided No-Reference Stereoscopic Image Quality Assessment Considering Spatial Saliency," *2021 International Conference on Visual Communications and Image Processing (VCIP)*, Munich, Germany, 2021, pp. 1-5, doi: 10.1109/VCIP53242.2021.9675338.

[9] A. N. Firdaus, S. Suprijanto and A. G. Risangtuni, "RGB-D Camera-Based Scanner for Hand Measurement," *2023 8th International Conference on Instrumentation, Control, and Automation (ICA)*, Jakarta, Indonesia, 2023, pp. 264-269, doi: 10.1109/ICA58538.2023.10273089.

[10] G. Krastev and I. Ralev, "Gesture Based System for User Interface Control," *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Ankara, Turkey, 2021, pp. 438-441, doi: 10.1109/ISMSIT52890.2021.9604693.

Overenie používateľa na základe dynamiky stlačení klávesov

[1]*Renát HALUŠKA,* [2]*Matúš PLEVA,* [3]*Maroš PAVLUS*

[1,2,3] 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]matus.pleva@tuke.sk, [3]maros.pavlus@student.tuke.sk

***Abstrakt* — V súčasnom digitálnom prostredí rastie potreba spoľahlivých a nenápadných autentifikačných metód. Táto práca sa zameriava na overovanie používateľa prostredníctvom dynamiky stlačení klávesov – behaviorálnej biometrie, ktorá nevyžaduje špecializovaný hardvér a umožňuje kontinuálne overovanie identity na pozadí. Predstavuje návrh a implementáciu modelu strojového učenia založeného na kombinácii konvolučných (CNN) a rekurentných (RNN) neurónových sietí, trénovaného na dátovej množine Buffalo. Model spracováva časové charakteristiky stlačení klávesov a identifikuje používateľa podľa jeho jedinečného štýlu písania.**

***Kľúčové slová* — analýza voľného textu, autentifikácia používateľa, biometrická autentifikácia, dynamika stlačení klávesov, strojové učenie**

I. Úvod

V dnešnej modernej dobe plnej digitálnych technológií sa ochrana osobných údajov stáva čoraz dôležitejšou. Najbežnejšou metódou overovania identity je overovanie heslom alebo PIN kódom. Tieto metódy sú však náchylné na určité problémy, kedy heslá môžu byť uhádnuté, ukradnuté alebo hacknuté. Preto sa na účely overovania identity začali používať aj rôzne biometrické údaje. Medzi takéto metódy patria odtlačky prstov, rozpoznávanie tváre, skenovanie dúhovky alebo rohovky. Tieto metódy overovania však môžu mať aj určité problémy. Jedným z problémov tejto metódy overovania je potreba mať špecifický hardvér. Napríklad snímač odtlačkov prstov alebo snímač rohovky. Keďže nie každý má prístup k takýmto technológiám, lepšou možnosťou by bolo mať systém, ktorý nevyžaduje žiadny ďalší hardvér. Jednou z metód, pri ktorej nepotrebujeme žiadny ďalší hardvér na overenie identity, je analýza dynamiky stlačení klávesov používateľa. To tiež umožňuje nenápadné overenie pri používaní danej aplikácie na detekciu narušení. Môže bežať na pozadí a nijako nebude rušiť používateľa.

Dynamika stlačení klávesov je automatizovaná metóda overovania identity používateľa na základe toho, ako používateľ píše na klávesnici. V tejto overovacej metóde máme dva spôsoby spracovania textu na overenie. Delia sa na metódu s pevným textom, kde používateľ vždy zadáva to isté heslo alebo ten istý text znova a znova. A metódu s voľným textom, ktorá sa zameriava na overenie používateľa pomocou voľného textu. Kde používateľ môže zadať ľubovoľne dlhý text, ktorý sa nemusí opakovať pri každom novom overení.

Spracovanie voľného textu však čelí väčším výzvam ako spracovanie s pevným textom. Keďže pri voľnom texte je variabilita dĺžky textu vždy iná. Zatiaľ čo krátky text nemusí mať dostatok informácií, funkcie sú potrebné na identifikáciu používateľa.

II. Analýza Dynamiky Stlačení Klávesov

Analýza nezohľadňuje to, čo používateľ píše, ale štýl, akým píše. Analýzou vzorov stlačení klávesov je možné používateľa identifikovať na základe jeho charakteristického štýlu písania. Ten je pre každého používateľa jedinečný, a preto sa môže použiť ako metóda identifikácie alebo autentifikácie v bezpečnostných systémoch. Okrem toho táto metóda môže pomôcť aj pri odhaľovaní votrelca, ktorý získal neoprávnený prístup do systému [1]. Mohli by sme ju použiť so súčasnou metódou autentifikácie pomocou hesla alebo PIN kódu. Pri prihlasovaní do systému by používateľ zadal heslo, ale systém by nielen skontroloval, či je heslo napísané správne, ale aj štýl, v ktorom bolo heslo napísané, a či sa zhoduje s osobou, ktorá sa ho snaží overiť [2].

V každom strojovom učení a vývoji modelov zohrávajú vstupné dáta kľúčovú úlohu. Bez nich by model nemal na čom „trénovať“ a jeho presnosť a funkčnosť by neboli správne. Pre potreby

vytvorenia modelu strojového učenia sme sa rozhodli použiť dátový súbor Buffalo [3]. Tento dátový súbor obsahuje dáta o stlačeniach klávesov od 157 testovaných subjektov. Celkovo dátový súbor obsahuje záznamy z troch sedení, z ktorých každá trvala približne 50 minút. Počas jedného sedenia subjekt stlačil priemerne 5700 klávesov. Časový interval medzi sedeniami bol 28 dní. Dátový súbor je rozdelený do troch úloh:

- Prepis prejavu Steva Jobsa na promócii na Stanfordskej univerzite: Táto úloha simuluje typickú činnosť písania textu.
- Odpovedanie na otázky v štýle prieskumu: Táto úloha testuje schopnosť modelu spracovať a odpovedať na otázky.
- Popis obrázka: Táto úloha testuje schopnosť modelu generovať text na základe vizuálnych informácií.

Súbor údajov bol následne rozdelený do dvoch podmnožín, kde v prvej podmnožine sa počas celého zberu údajov používala rovnaká klávesnica, zatiaľ čo v druhej podmnožine sa používali rôzne klávesnice. Formát údajov (Obr. 1) pozostáva z textového dokumentu pre každého používateľa, v ktorom sú uložené tri základné údaje, a to:

- **Kláves:** Prvý údaj predstavuje typ stlačeného klávesu.
- **Akcia:** Ďalej máme akciu, ktorú kláves vykonal, či už bol stlačený pravým tlačidlom myši alebo uvoľnený. Tieto údaje sú uložené ako KeyDown pre stlačenie a KeyUp pre uvoľnenie daného klávesu.
  Časová pečiatka: Posledné údaje sú časovou pečiatkou uvedenou v milisekundách.

```
Space KeyDown 63578429797922
D KeyUp 63578429797953
Space KeyUp 63578429798015
T KeyDown 63578429798109
T KeyUp 63578429798218
O KeyDown 63578429798296
Space KeyDown 63578429798390
```

Obr. 1 Príklad údajov z súboru údajov

III. Popis Modelu Strojového Učenia

V tejto práci používame na vytvorenie modelu kombináciu konvolučných neurónových sietí (CNN) a rekurentných neurónových sietí (RNN). V klasickom modeli by sme vypočítali priemery, rozptyly a ďalšie matematické charakteristiky stlačení klávesov. Pomocou výsledkov týchto operácií by sa systém naučil ako vstupné vzory pomocou algoritmu strojového učenia [4]. Tento prístup je však náročný na overenie komplexného a rozmanitého štýlu písania na klávesnici iba pomocou matematického modelu. Keď systém obsahuje značné množstvo údajov o stlačeniach klávesov, charakterizuje iba časť nášho štýlu písania, čo potom vedie k nízkej presnosti. Škálovateľnosť a spoľahlivosť tohto matematického modelu preto nie sú dostatočné.

Model pozostáva z nasledujúcich vrstiev: Vstupná vrstva: Vstupná vrstva Conv1D s 32 filtrami a veľkosťou jadra 2. Vstupný tvar je špecifikovaný ako (X_train.shape[1], 1), kde X_train je trénovacia množina údajov a shape[1] predstavuje dĺžku sekvencie (počet stlačení klávesov). Parameter kernel_size=2 označuje, že konvolučné jadro bude posunuté o 2 pozície pozdĺž sekvencie. Voľba typu Conv1D je odôvodnená skutočnosťou, že vstupné údaje sú sekvenčné a predstavujú časovú sériu stlačení klávesov [5].

Skryté vrstvy siete Long Short Term Memory Network (LSTM): Po vstupnej vrstve nasledujú dve skryté vrstvy LSTM, každá s 32 neurónmi. Tieto vrstvy spracovávajú sekvenčné dáta a dokážu zachytiť dlhodobé závislosti medzi stlačeniami klávesov. Parameter return_sequences=True označuje, že tieto vrstvy vracajú celú sekvenciu aktivovaných neurónov, nielen konečný stav.

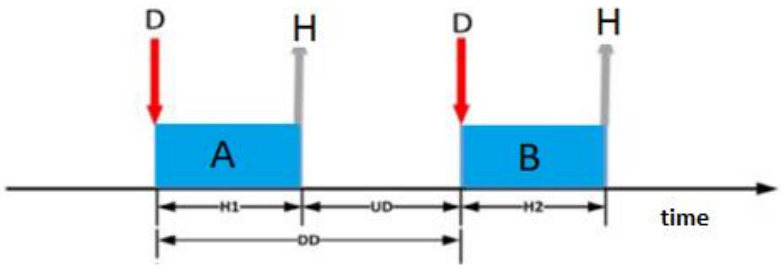
Vrstva Dropout (vrstva náhodného vypínania): Medzi dve vrstvy LSTM je vložená vrstva Dropout s mierou dropout 0,25. Táto vrstva slúži na zabránenie pretrénovania modelu a zlepšenie jeho zovšeobecnenia.

Výstupná vrstva: Po vrstvách LSTM nasleduje vrstva Flatten (splošťovacia vrstva), ktorá transformuje trojrozmerný výstup z vrstiev LSTM do dvojrozmerného tvaru. Následne sa v tomto príklade použije vrstva Dense so 4 neurónmi (plne prepojená vrstva) a keď potrebujeme mať v našom modeli vyškolených 15 ľudí, táto hodnota sa zmení na 15 a aktivuje sa softmax pre konečnú klasifikáciu do 4 kategórií. Parameter kernel_regularizer=l2(0.01) sa používa na reguláciu váh neurónov a zabránenie ich pretrénovaniu [6].

Predtým, ako môžeme použiť dáta na trénovanie modelu, musíme z nich extrahovať časové charakteristiky a uložiť ich do vektora. Z dát sme extrahovali 6 časových charakteristík. Všetkým

kľúčom je priradené ID číslo, ktoré bude tiež extrahované ako charakteristika podľa vytvorenej mapy kľúčov (Obr. 2). Toto sú časové charakteristiky:

- ID kľúča A
- ID kľúča B
- Trvanie stlačenia klávesu A, vypočítame, ako dlho sme držali prvý kláves ako H1
- Trvanie stlačenia klávesu B, dĺžku stlačenia klávesu B vypočítame aj ako H2
- Čas medzi stlačením tlačidla A a stlačením tlačidla B ako DD
- Čas medzi uvoľnením stlačeného tlačidla A a stlačením tlačidla B ako UD



Obr. 2 Časové charakteristiky dátového súboru

Tieto dáta sa potom uložia do vektora, čo sa robí pre každý pár po sebe nasledujúcich kľúčov. Tento vektor sa potom použije ako vstup do modelu (Obr. 3).

Obr. 3 Príklad vektora charakteristík

IV. Testovanie a Hodnotenie Presnosti Modelu

Na testovanie a nasadenie modelu identifikácie používateľa na základe písania na klávesnici bola vytvorená webová aplikácia. Webová aplikácia plní dve hlavné funkcie. Prvá časť aplikácie slúži na zhromažďovanie potrebných údajov od používateľa a druhá časť aplikácie slúži na demonštráciu natrénovaného modelu.

Keďže je potrebné získať čo najviac údajov o používateľovi, bolo zvolené vložiť na stránku 3 otázky, ktoré majú formu formulára. Tieto otázky majú textové pole, do ktorého môže používateľ napísať svoje odpovede na zadané otázky. Otázky boli navrhnuté tak, aby sa na ne dalo odpovedať dlhými vetami. Pri vypĺňaní týchto otázok by bolo pre väčšiu presnosť a kvalitu údajov najlepšie, keby používateľ počas rôznych relácií a odpovedania na otázky nemenil klávesnicu, ktorú používa, ani počítač.

Vyhodnotenie presnosti modelu posudzuje výkonnosť modelu dynamiky stlačení klávesov navrhnutého na identifikáciu používateľov so zameraním na presnosť predikcie používateľov v rôznych populáciách používateľov. Účinnosť modelu bola analyzovaná s použitím troch odlišných skupín používateľov: 14, 20 a 35 používateľov (Tabuľka 1).

- **14 používatelia**: Model dosiahol presnosť predikcie používateľov 79 %. To naznačuje relatívne vysokú úroveň presnosti pri identifikácii používateľov v rámci menšej populácie.
- **20 používateľov**: So zvýšenou používateľskou základňou sa presnosť znížila na 73 %. To naznačuje potenciálny pokles výkonu, pretože model má za úlohu rozlišovať medzi väčším počtom jednotlivcov.
- **35 používateľov**: Najväčšia skupina používateľov dosiahla najnižšiu presnosť 68 %. To ďalej posilňuje trend klesajúcej presnosti s rastúcou populáciou používateľov, čo zdôrazňuje výzvy spojené s udržiavaním presnosti pri rozsiahlejších nasadeniach.

Tabuľka 1
Presnosť modelu

Počet používateľov v modeli	Trénovací čas Presnosť predpovedí používateľov
14	79 %
20	73 %
35	68 %

Pozorovaný trend demonštruje inverzný vzťah medzi počtom používateľov a presnosťou predikcie modelu. S rastúcou populáciou používateľov sa schopnosť modelu presne rozlišovať medzi jednotlivcami znižuje. Je to pravdepodobne spôsobené zvýšenou podobnosťou vo vzoroch stlačení klávesov vo väčšej skupine, čo sťažuje modelu izoláciu jedinečných charakteristík používateľov.

Výsledky naznačujú, že výkon modelu je citlivý na veľkosť používateľskej populácie. Na zlepšenie škálovateľnosti modelu a zachovanie presnosti vo väčších skupinách používateľov je potrebný ďalší výskum. Potenciálne vylepšenia by mohli zahŕňať zdokonalenie techník extrakcie prvkov, preskúmanie sofistikovanejších algoritmov strojového učenia alebo začlenenie ďalších behaviorálnych biometrických údajov. Výsledky ukazujú, že hoci model má potenciál, je potrebné vynaložiť viac úsilia, aby bol použiteľný pre veľké používateľské základne. Model vykazuje dobrý výkon s malými skupinami používateľov.

Navrhovaný model založený na kombinácii konvolučnej a LSTM siete bol porovnaný s bežnými klasifikátormi ako SVM, Random Forest či GRU. Kým klasické modely (SVM, RF) vyžadujú manuálnu extrakciu príznakov a trpia obmedzenou schopnosťou pracovať so sekvenciami, modely typu GRU a LSTM dokážu efektívne zachytiť časové závislosti. V porovnaní s GRU ponúka LSTM lepšiu presnosť pri dlhších sekvenciách, aj keď za cenu vyššej výpočtovej náročnosti. Hybridné prístupy môžu dosahovať vyššiu presnosť, no vyžadujú zložitejšie spracovanie. Výhodou nášho riešenia je možnosť end-to-end učenia bez potreby manuálneho výberu čŕt a schopnosť spracovať voľný text s rôznou dĺžkou.

V. Záver

Táto práca skúmala životaschopnosť dynamiky stlačení kláves ako hardvérovo nezávislej a nenápadnej metódy autentifikácie používateľov v čoraz kritickejšom prostredí digitálnej bezpečnosti. Využitím jedinečných vzorcov písania jednotlivcov ponúka tento prístup potenciálne riešenie zraniteľností, ktoré sú vlastné tradičným metódam autentifikácie heslom a biometrickým metódam.

Výskum preukázal uskutočniteľnosť využitia modelu konvolučnej neurónovej siete (CNN) a rekurentnej neurónovej siete (RNN) na analýzu údajov o stlačeniach klávesov, so zameraním najmä na časové charakteristiky extrahované z po sebe nasledujúcich stlačení klávesov. Model, trénovaný na komplexnom súbore údajov Buffalo, priniesol sľubné výsledky, najmä v menších počtoch používateľov, pričom dosiahol mieru presnosti 79 % so 14 používateľmi. Štúdia však tiež odhalila jasný inverzný vzťah medzi veľkosťou populácie používateľov a presnosťou predikcie, pričom presnosť klesla na 68 %, keď sa používateľská základňa rozšírila na 35 jednotlivcov.

Zistenia zdôrazňujú potenciál dynamiky stlačení klávesov pre nepretržitú a nenápadnú autentifikáciu, ktorá ponúka cennú vrstvu zabezpečenia bez potreby špecializovaného hardvéru. Pozorovaný pokles presnosti u väčších skupín používateľov však zdôrazňuje potrebu ďalšieho výskumu a vývoja. Budúce úsilie by sa malo zamerať na zdokonaľovanie techník extrakcie prvkov, skúmanie pokročilých architektúr strojového učenia a riešenie výziev škálovateľnosti a odolnosti voči útokom nepriateľa. Okrem toho sú pre overenie dlhodobej účinnosti tohto prístupu kľúčové longitudinálne štúdie a testovanie s rôznymi demografickými skupinami používateľov.

Tento výskum v konečnom dôsledku prispieva k prebiehajúcemu skúmaniu behaviorálnej biometrie a jej potenciálu na zvýšenie digitálnej bezpečnosti v čoraz prepojenejšom svete. Hoci pretrvávajú výzvy, dynamika stlačení klávesov predstavuje sľubnú cestu k vývoju bezpečnejších a užívateľsky prívetivejších autentifikačných systémov.

Poďakovanie

Túto prácu čiastočne podporila Vedecká grantová agentúra Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky a Slovenská akadémia vied v rámci projektu VEGA 2/0165/21, Kultúrna a vzdelávacia grantová agentúra Ministerstva školstva, výskumu, vývoja a

mládeže Slovenskej republiky prostredníctvom projektu 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] N. Bussabong and T. Anusas-amornkul, "Enhanced Keystroke Dynamics Authentication Using Keystroke Vector Dissimilarity," *2023 15th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Chiang Mai, Thailand, 2023, pp. 223-228, doi: 10.1109/ICITEE59582.2023.10317643.

[2] B. Tural, Z. Örpek and S. Özmen, "Artificial Intelligence and Keystroke Dynamics: The Mysterious World of Personal Signatures," *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Istanbul, Turkiye, 2024, pp. 1-5, doi: 10.1109/HORA61326.2024.10550804.

[3] J. Arrigo, "Biometric Identification from Error Correction Behaviors Present in Keystroke Dynamics," *2024 IEEE MIT Undergraduate Research Technology Conference (URTC)*, Cambridge, MA, USA, 2024, pp. 1-4, doi: 10.1109/URTC65039.2024.10937520.

[4] H. M. C. K. B. Herath, K. G. C. Dulanga, N. V. D. Tharindu and G. U. Ganegoda, "Continuous User Authentication using Keystroke Dynamics for Touch Devices," *2022 2nd International Conference on Image Processing and Robotics (ICIPRob)*, Colombo, Sri Lanka, 2022, pp. 1-6, doi: 10.1109/ICIPRob54042.2022.9798728.

[5] R. Haluška *et al*., "Mobile Educational Application for Keystroke Dynamics Identification Systems," *2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA)*, Stary Smokovec, Slovakia, 2023, pp. 165-170, doi: 10.1109/ICETA61311.2023.10343749.

[6] A. Islam and S. B. Belhaouari, "Analysing keystroke dynamics using wavelet transforms," *2022 IEEE International Carnahan Conference on Security Technology (ICCST)*, Valeč u Hrotovic, Czech Republic, 2022, pp. 1-5, doi: 10.1109/ICCST52959.2022.9896483.

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@uke.sk

***Abstrakt* — Tento príspevok sa zaoberá možnosťami regulácie výkonu v ostrovnej prevádzke pre experimentálne účely. Jedná sa o model ostrovnej prevádzky, pomocou ktorej sa simulovali rôzne prevádzkové stavy a ich následná optimalizácia za účelom dosiahnutia požadovaných napätí v uzloch. Tento príspevok je rozdelený do niekoľkých častí, v ktorých sú popísané návrhy a modelovanie ostrovnej prevádzky a jej simulácií.**

***Kľúčové slová* — Ostrovná prevádzka, ATP-Draw, regulácia výkonu, fázové napätie, prevádzkový stav**

I. Úvod

V súčasnej dobe je ľudstvo čoraz viac závislejšie od spoľahlivosti a dostupnosti elektrickej energie, preto je nárast decentralizovaných a udržateľných zdrojov energie obrovský. Túto schopnosť pracovať samostatne zdieľa aj ostrovná prevádzka, preto je téma možností regulácie výkonu v ostrovnej prevádzke široká v oblasti elektroenergetiky. S rastúcou spotrebou elektrickej energie sa čoraz viac hľadajú novšie alternatívne zdroje, ktoré umožňujú efektívne využívať a skladovať elektrickú energiu. Prostredníctvom technologického vývoja sa viaceré podniky zameriavajú na ostrovnú prevádzku, ktorá ponúka niekoľko výhod v oblasti riadenia, stability alebo regulácie elektrickej energie. Aj napriek tomu, ostrovná prevádzka nie je ideálna, a prináša isté nevýhody, ktoré môžu značne ovplyvniť stabilitu siete.

II. Ostrovná prevádzka

Ostrovná prevádzka tzv. „mikrosieť“ je systém, ktorý je oddelený od predradenej distribučnej siete. Na základe priameho pripojenia do odberného elektrického zariadenia na konkrétnom odbernom mieste poskytuje danému zariadeniu spoľahlivé napájanie. Pri tomto spôsobe výroby elektrickej energie je odberné elektrické zariadenie napájané vlastným zdrojom elektrickej energie, ako napríklad batérie, diesel-generátory a iné zariadenia. Preto je potrebné poznamenať, že v ostrovnej prevádzke je nutné neustále udržiavať úroveň napätia a frekvenciu na požadovaných hodnotách. Taktiež je dôležité podotknúť, že v ostrovnej prevádzke bude napätie a frekvencia definované skôr pomocou záťaží, ako pomocou siete. Je to z dôvodu, že ostrovná prevádzka musí pokryť spôsobenú záťaž bez pomoci hlavnej siete. Z týchto poznatkov je možné definovať rôzne výhody, ako aj nevýhody ostrovnej prevádzky [3]:

- ***Výhody použitia ostrovnej prevádzky*** – Jednou z hlavných výhod je energetická nezávislosť, ktorá u ostrovnej prevádzky poskytuje úplnú kontrolu nad výrobou a spotrebou elektrickej energie. Týmto spôsobom sú odberatelia pripojení v ostrovnej prevádzky nezávislí od vonkajšej siete. Ďalšou z výhod je vyššia bezpečnosť a spoľahlivosť, pretože ostrovné prevádzky sú väčšinou konštruované pre núdzové stavy. Jednou z hlavných výhod je aj podpora lokálnych obnoviteľných zdrojov energie, ako napríklad využitie fotovoltických panelov, vodných elektrární alebo veterných fariem [4].
- ***Nevýhody použitia ostrovnej prevádzky*** – Jednou z hlavných nevýhod ostrovnej prevádzky sú vyššie počiatočné náklady na výstavbu, pretože pri výstavbe sa vyžaduje konkrétne zázemie na výrobu a skladovanie elektrickej energie, kde tieto vlastnosti plnia napríklad batérie alebo generátory. Ďalšou z nevýhod je obmedzená kapacita a stabilita systému, ktorá je menšia z pohľadu veľkosti samotnej ostrovnej prevádzky. Z tohto dôvodu môže počas špičky dôjsť k výpadkom alebo k zníženiu spotreby. Ďalšou z nevýhod je závislosť od obnoviteľných zdrojov energie, pretože stabilita prevádzky závisí od slnečného žiarenia pri

fotovoltických paneloch alebo od poveternostných podmienok pri veterných farmách (nefúka – menší výkon) [6]. Vzhľadom k uvedeným dobrým vlastnostiam ostrovnej prevádzky, ako sú napríklad energetická nezávislosť, spoľahlivosť a využitie obnoviteľných zdrojov energie, existujú aj riziká, na ktoré treba dohliadať, pretože môžu spôsobiť nestabilitu prevádzky alebo až úplný výpadok celej sústavy (blackout) [3].

III. PRINCÍPY REGULÁCIE VÝKONU

Elektrizačná sústava musí byť neustále udržiavaná v rovnováhe medzi výrobou a spotrebou elektrickej energie, pretože pri každom odklonení od tejto rovnováhy dochádza k zmene frekvencie, resp. napätia. Tieto zmeny môžu spôsobiť rôzne poruchy, ako aj znížiť spoľahlivosť dodávky elektrickej energie. Preto je regulácia výkonu v ostrovnej prevádzke dôležitým krokom pre udržanie bezpečnej a stabilnej prevádzky. Pre zabezpečenie tejto požiadavky sa využívajú rôzne mechanizmy a moderné technológie, ako aj využitie dostupných zdrojov (OZE), preto tieto princípy regulácie výkonu majú základ v riadení frekvencie či napätia [8], [9]:

- ***Regulácia frekvencie a činného výkonu*** – pri tomto spôsobe regulácie je dôležitou úlohou udržiavanie stabilnej frekvencie systému v určitom dovolenom pásme (v EÚ: f = 50 Hz), kde odchýlka tohto kvalitatívneho ukazovateľa môže dosahovať maximálne ±0,02 Hz. Pre udržanie frekvencie v požadovaných hraniciach je potrebné zabezpečiť rovnováhu medzi výrobou a spotrebou. Z tohto pohľadu môžu nastať 2 prípady odchýlky frekvencie [5]:
  - V prípade vyššej výroby ako dopytu sa frekvencia zvýši
  - V prípade vyššieho dopytu ako výroby sa frekvencia zníži

  Udržiavanie frekvencie v rámci povolených medzí je dôležité aj z pohľadu citlivých spotrebičov, ktoré sú pripojené v elektrizačnej sústave, ako napríklad bubnové mlyny a iné obrábacie stroje [5].
- ***Regulácia napätia a jalového výkonu*** – v tomto prípade regulácie je napätie považované za lokálny kvalitatívny ukazovateľ, ktorý je ovplyvňovaný v celom prenose až po koncový spotrebič. V prípade regulácie musí byť napätie rovnako ako frekvencia neustále udržiavané v rámci povolených medzí. K zmenám napätia môže dôjsť pri náhlych zmenách záťaže alebo výrobnej kapacity. K zvýšeniu stability napätia prispievajú transformátory alebo napäťové regulátory, no k riadeniu napätia je potrebné zabezpečiť dostatočnú rezervu jalového výkonu. To sa zabezpečuje požiadavkami na jalový výkon generátorov zapojených do ES alebo prevádzkou kondenzátorov, či tlmiviek.

  Technické zariadenia zabezpečujúce dostatočnú rezervu jalového výkonu [7]:
  - *Transformátory s odbočkami* – tieto transformátory dokážu zabezpečiť reguláciu napätia prepínaním odbočiek pod zaťažením.
  - *Regulácia budenia synchrónneho generátora* – prostredníctvom pôsobenia ich budenia ovplyvňujú napätia v sústave.
  - *Kompenzačné zariadenia (STATCOM)* – v tomto prípade ide o moderné elektronické systémy, ktoré sa využívajú na rýchlu kompenzáciu jalového výkonu.
  - *Kondenzátorové batérie a kompenzačné tlmivky* – tieto zariadenia slúžia na stabilizáciu napätia v prenosovej sústave.

Z uvedených vlastností je regulácia napätia dôležitou úlohou pri prenose elektrickej energie na veľké vzdialenosti, ale aj v oblasti distribučných sústav, kde OZE predstavujú veľký podiel zdrojov energie [7].

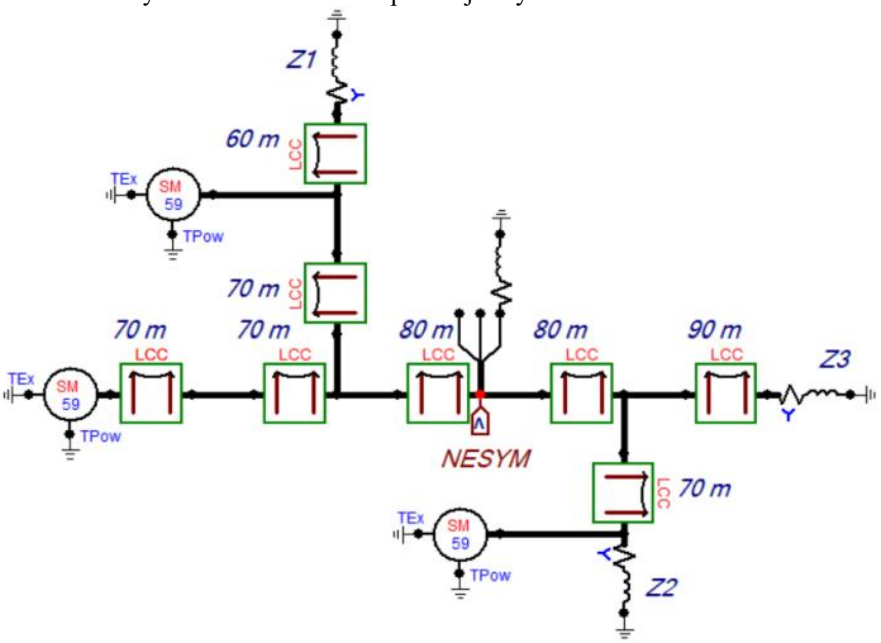
- ***Primárna, sekundárna a terciárna regulácia výkonu*** – z pohľadu regulácie výkonu ide o zabezpečenie stabilnej prevádzky sústavy a delí sa do troch úrovní [8], [9]:
  - *Primárna regulácia výkonu* – pôsobí pri zmenách frekvencie vplyvom porúch do niekoľkých sekúnd. Jej cieľom je pomocou turbínových regulátorov stabilizácia frekvencie, no nevráti ju späť na nominálnu úroveň (ostáva Δf).
  - *Sekundárna regulácia výkonu* – pôsobí od niekoľkých sekúnd a svoju činnosť ukončí do 15 minút a vyrovnáva odchýlky frekvencie spôsobené primárnou reguláciou na nominálnu úroveň.
  - *Terciárna regulácia výkonu* – optimalizuje prevádzku a zabezpečuje dostatočnú rezervu sekundárnej regulácie. Z hľadiska riadenia môže byť automatická alebo manuálna na pokyn dispečera.

Pri regulácii sa kladie dôraz hlavne na tieto princípy regulácie výkonu, pretože zabezpečujú rovnováhu medzi výrobou a spotrebou, ako aj efektívne využívanie zdrojov energie. Existujú aj iné princípy regulácie, ako napríklad: zálohovanie, riadenie zaťaženia, ochrana systému, monitorovanie alebo automatizácia. Vplyvom týchto princípov regulácie sa docieli efektívna, flexibilná a bezpečná prevádzka sústavy aj pri náročných podmienkach [8], [9].

IV. SIMULÁCIA POUŽITÍM GRAFICKÉHO PROGRAMU ATP-DRAW

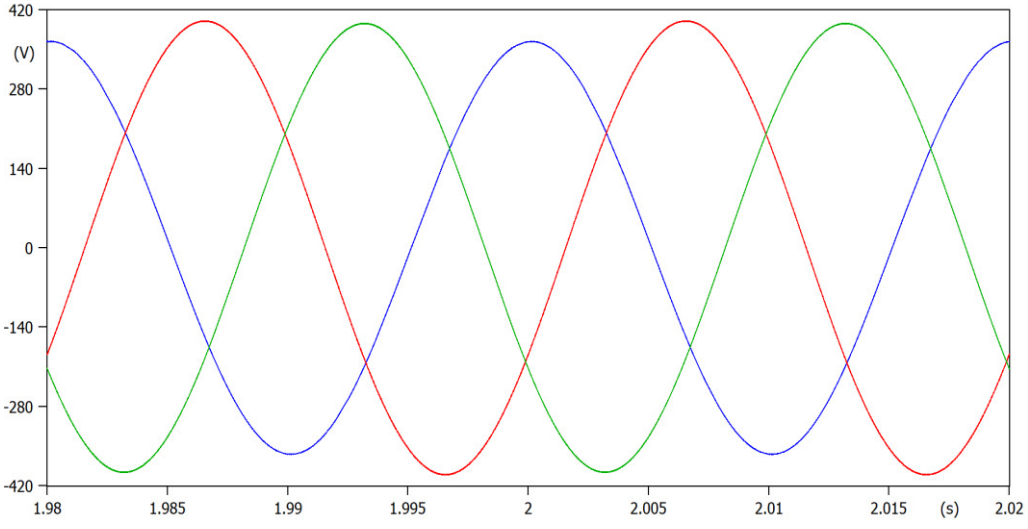
Simulácia rôznych prevádzkových stavov ostrovnej prevádzky sa uskutočňovala prostredníctvom grafického programu ATP-Draw, ktorý slúži na modelovanie rôznych elektrických obvodov a ich následnou simuláciou pre získanie požadovaných údajov. Okrem modelovania rôznych elektrických obvodov sa môže tento program využiť na analýzu prechodných javov, na návrh a testovanie ochranných systémov alebo pre štúdium harmonických a vyšších harmonických. Na orientáciu a modelovanie v programe využíva užívateľ počítačovú myš, kde pre namodelovanie elektrických obvodov slúži široká ponuka elektrických komponentov. V tomto programe je taktiež zakomponovaná funkcia PlotXY, pomocou ktorej si užívateľ dokáže vykresliť graf, z ktorého sa následne môžu odčítať údaje potrebné pre výpočty a porovnania výsledkov. V tomto prípade sa pracovalo v poslednej verzii programu 7.6, kde sa namodelovala sieť (ostrovná prevádzka) v rôznych prevádzkových stavoch [2]:

- Prevádzkový stav A – simulácia výpadku generátora z prevádzky
- Prevádzkový stav B – simulácia spínania záťaží/odberateľov
- Prevádzkový stav C – simulácia napäťovej nesymetrie v sieti



Obr. 1 Schéma ostrovnej prevádzky s jednofázovou záťažou RL (Prevádzkový stav C)

Na Obr. 1 je vyobrazená namodelovaná schéma ostrovnej prevádzky, ktorá sa zaoberala simuláciou napäťovej nesymetrie spôsobenej pripojením jednofázovej záťaže RL do siete, teda prevádzkový stav C. Pripojenie danej jednofázovej záťaže RL spôsobilo v ostrovnej prevádzke napäťovú nesymetriu, ktorá spôsobila v jednej z troch fáz úbytok napätia. Tento nežiadúci prechodný jav je možné vidieť na nasledujúcom priebehu napätia (obr. 2).



Obr. 2 Priebehy napätia v čase po pripojení jednofázovej záťaže RL

Po odčítaní hodnôt fázových napätí prostredníctvom kurzora z priebehu na Obr. 2, vykazovalo napätie vo fáze L1 úbytok o veľkosti skoro 35 V, kým napätia vo zvyšných fázach L2 a L3 neboli ovplyvnené a po odmeraní dosahovali požadovanú hodnotu okolo 400 V. V dôsledku tejto nesymetrie napätia nemusí daná jednofázová záťaž RL spĺňať požadované limity, ktoré udáva norma STN EN 50160. V rámci normy STN EN 50160 je obsiahnutá aj samotná nesymetria napätia, pretože ide o dôležitý kvalitatívny parameter elektrickej energie. Daná norma určuje maximálne povolené úrovne napäťovej nesymetrie do 2% nominálneho napätia, preto je potrebné po vypočítaní napäťovej nesymetrie posúdiť a porovnať tieto výsledky [1].

V. Analýza ukazovateľov kvality elektrickej energie

V nasledujúcej tabuľke sú zapísané fázové napätia, ktoré pred pripojením jednofázovej záťaže boli skoro identické, a po pripojení vykazovala jedna z troch fáz pomerne veľkú zmenu. Dané napätia boli odmerané pomocou priebehu vyobrazenom na obr. 2, pomocou ktorých sa následne vypočíta napäťová nesymetria.

Tabuľka 1
Napäťová nesymetria vplyvom pripojenia jednofázového zaťaženia RL

Fázové napätia pred pripojením jednofázovej RL záťaže			
Pred pripojením	U_{f1} [V]	U_{f2} [V]	U_{f3} [V]
	399,26	397,49	397,66
Fázové napätia po pripojením jednofázovej RL záťaže			
Po pripojení	U_{f1} [V]	U_{f2} [V]	U_{f3} [V]
	364,34	400,38	399,11

Prostredníctvom zapísaných hodnôt v tabuľke sa pomocou vzorca na výpočet nesymetrie napätia určí, o akú percentuálnu hodnotu nesymetrie napätia medzi jednotlivými fázami došlo. Najskôr je potrebné zohľadniť priemernú hodnotu fázového napätia pomocou nasledovného vzorca:

$$U_{pr} = \frac{U_{f1} + U_{f2} + U_{f3}}{3} = \frac{364{,}34 + 400{,}38 + 399{,}11}{3} = 387{,}94 \text{ [V]}$$

Po zohľadnení priemernej hodnoty fázového napätia po pripojení jednofázového zdroja energie do trojfázovej sústavy sa môžu vypočítať pre jednotlivé fázy najväčšie odchýlky od priemernej hodnoty fázového napätia podľa nasledujúceho vzorca:

$$\Delta L1 = U_{f1} - U_{pr} = 364{,}34 - 399{,}26 = 34{,}92 \text{ [V]}$$
$$\Delta L2 = U_{f2} - U_{pr} = 400{,}38 - 397{,}49 = 2{,}89 \text{ [V]}$$
$$\Delta L3 = U_{f3} - U_{pr} = 399{,}11 - 397{,}66 = 1{,}45 \text{ [V]}$$

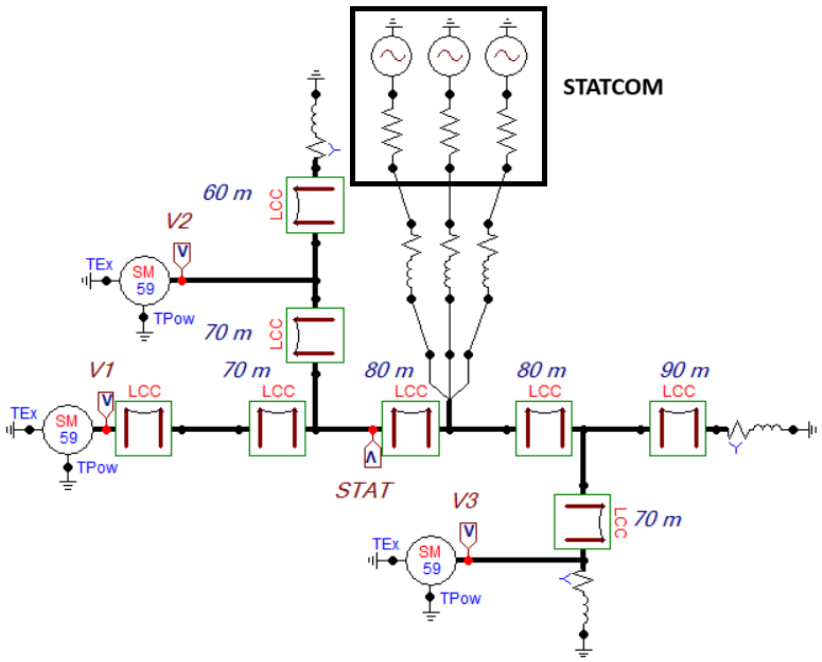
Po tomto kroku ostáva dopočítať napäťovú nesymetriu v trojfázovej sústave po pripojení jednofázovej záťaže RL pomocou výsledných hodnôt z predošlých výpočtov:

$$\text{Napäťová nesymetria} = \frac{([\Delta L1]; [\Delta L2]; [\Delta L3])}{U_{pr}} \cdot 100 = \frac{([34{,}92]; [2{,}89]; [1{,}45])}{387{,}94} \cdot 100$$
$$= \mathbf{9}; \mathbf{0,75}; \mathbf{0,37} \text{ [\%, V, V, V]}$$

Z týchto výsledkov je možné konštatovať, že napätie vo fáze L1 nevyhovuje povoleným limitom v rámci normy STN EN 50160 pre napäťovú nesymetriu, a preto je potrebné zabezpečiť optimalizáciu tejto siete za účelom zvýšenia napätia vo fáze L1:

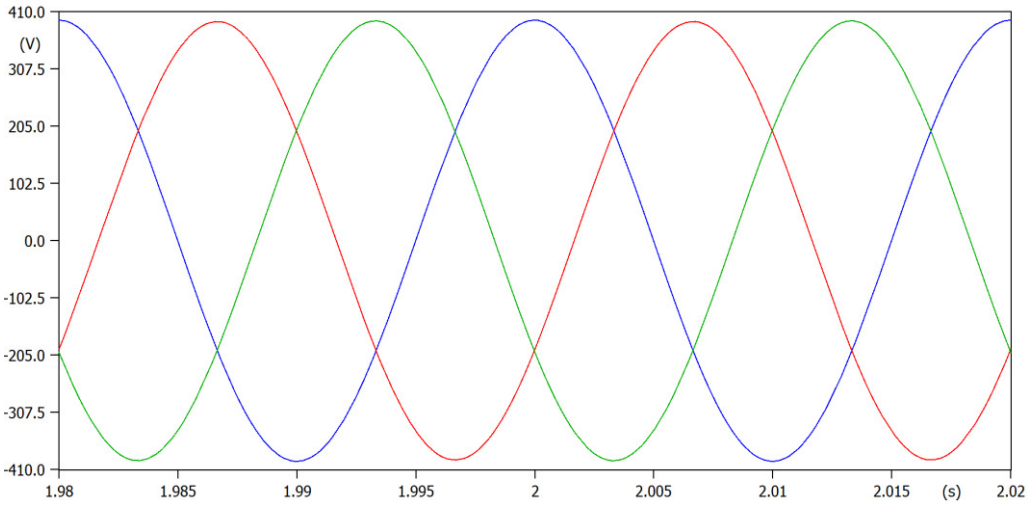
- L1 – 9 % > 2 % → nevyhovuje! (nad rámec limitov)
- L2 – 0,75 % < 2 % → v rámci limitov
- L3 – 0,37 % < 2 % → v rámci limitov

Pre tento prípad sa využil princíp statického synchrónneho kompenzátora (STATCOM), ktorého činnosť spočíva v poskytovaní alebo absorbovaní jalového prúdu za účelom regulácie napätia v mieste pripojenia do elektrickej siete. Výhodou tohto systému je rýchla odozva a prevádzková flexibilita, či možná korekcia účinníka alebo jalového výkonu. Pre tento princíp STATCOMU sa použijú na reguláciu napätia zdroje energie ako pasívny filter, ktoré prerozdelia výkon a rovnomerne rozložia napätia v jednotlivých fázach. Na obr. 3 je vyobrazená namodelovaná schéma ostrovnej prevádzky s aktívnym filtrom, teda časť zariadenia STATCOM.



Obr. 3 Namodelovaná ostrovná prevádzka so STATCOM zariadením

Po namodelovaní daného zapojenia sa v systéme rovnomerne rozložilo napätie v jednotlivých fázach, ktoré vykazovali minimálne rozdiely. Tieto menšie rozdiely mohli byť spôsobné nerovnomerným zaťažením danej ostrovnej prevádzky.



Obr. 4 Priebehy napätí v čase po pripojení kompenzačného zariadenia

Z daných priebehov napätia v čase je možné pozorovať rovnomerné rozloženie napätí v jednotlivých fázach, ktoré sa prostredníctvom kurzora odčítali a zapísali do nižšie uvedenej tabuľky. Pomocou týchto hodnôt fázových napätí sa vypočítala prostredníctvom vzorca na napäťovú nesymetriu (uvedená vyššie) percentuálna hodnota vzniknutej nesymetrie medzi jednotlivými fázami, kde konečná hodnota tejto nesymetrie bola v medziach STN EN 50160.

Tabuľka 2
Fázové napätia po vykompenzovaní napäťovej nesymetrie

Fázové napätia po pripojení STATCOM zariadenia			
Po pripojení	U_{f1} [V]	U_{f2} [V]	U_{f3} [V]
	394,9	392,45	393,77

- L1 – 0,3 % < 2 % → v rámci limitov
- L2 – 0,32 % < 2 % → v rámci limitov
- L3 – 0,02 % < 2 % → v rámci limitov

VI. ZÁVER

Cieľom tohto príspevku bol návrh regulácie výkonu v ostrovnej prevádzke, ktorá sa namodelovala prostredníctvom grafického programu ATP-Draw. Tento systém ostrovnej prevádzky bol napájaný z viacerých strán a tvoril NN sieť, pomocou ktorej sa simulovali rôzne prevádzkové stavy. Spôsobené nežiadúce prechodné javy v týchto stavoch bolo potrebné eliminovať správne navrhnutou a implementovanou reguláciou výkonu.

POĎAKOVANIE

Tento príspevok vznikol vďaka podpore Vedeckej grantovej agentúre Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporeného grantom VEGA 1/0532/25 a VEGA 1/0627/24.

ZOZNAM POUŽITEJ LITERATÚRY

[1] VSD, „*Technické podmienky prevádzkovateľa distribučnej sústavy*“, [online]. [cit. 2025-03-11]. Dostupné na internete: <https://www.vsds.sk/mdoc/dso.B6000.A/doc/20140101.pdf>.
[2] ATPDraw Home. [online]. [cit. 2024-10-27]. Dostupné na internete: <https://www.atpdraw.net/>.
[3] Islanded Mode – prehľad Témy ScienceDirect. [online]. [cit. 2024-10-13]. Dostupné na internete: <https://www.sciencedirect.com/topics/engineering/islanded-mode>.
[4] Power Grids Unplugged: How Islanding is Changing Autonomous Energy. [online]. [cit. 2025-02-21]. Dostupné na internete: <https://www.smpnet.tech/post/power-grids-unplugged-how-islanding-is-changing-autonomous-energy>.
[5] Prednaska Regulacia P A F. [online]. [cit. 2025-03-07]. Dostupné na internete: <https://www.scribd.com/document/707336167/Prednaska-regulacia-P-a-f>.
[6] The Benefits and Challenges of Microgrids. In TRC Companies [online]. [cit. 2025-02-21]. Dostupné na internete: <https://www.trccompanies.com/insights/benefits-and-challenges-of-microgrids/>.
[7] www.elektrotechnika1.estranky.sk - Elektroenergetika - Regulácia napätia a jalového výkonu v elektroenergetike. [online]. [cit. 2025-03-07]. Dostupné na internete: <https://elektrotechnika1.estranky.sk/clanky/elektroenergetika/regulacia-napatia-a-jaloveho-vykonu-v-elektroenergetike.html>.
[8] M. Kolcun, V. Griger, „*Riadenie prevádzky elektrizačnej sústavy*“, Mercury - Smékal, Košice, 2003, 288 strán, ISBN 80-89061-57-5.
[9] M. Kolcun, V. Griger, Ľ. Beňa, J. Rusnák, „*Prevádzka elektrizačnej sústavy*“, Košice TU, 2007, 306 strán, ISBN 978-80-8073-837-2.
[10] D. Čorňák, „*Možnosti regulácie výkonu v ostrovnej prevádzke*“, Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025.

Návrh funkčného modelu Stirlingovho motora

Matej HIRKO, Dušan MEDVEĎ

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

matej.hirko@student.tuke.sk, dusan.medved@tuke.sk

***Abstrakt* — Tento príspevok sa venuje návrhu a experimentálnemu overeniu funkčného modelu Stirlingovho motora modifikácie typu gama, so zameraním na analýzu vplyvu rôznych prevádzkových a konštrukčných parametrov na jeho činnosť. Cieľom príspevku bolo nielen zostrojiť funkčný model, ale aj detailne preskúmať vplyv rôznych druhov mazív, spôsobu ohrevu a intenzity horenia na štartovanie, stabilitu a účinnosť motora, ako aj schopnosť generovania elektrického napätia prostredníctvom jednoduchého generátora. Merania boli realizované pri rôznych štartovacích podmienkach (studený/horúci stav), pričom sa sledovali teploty jednotlivých komponentov, doba prevádzky a výstupné elektrické napätie. Získané výsledky poukázali na dôležitosť optimalizácie parametrov ako sú viskozita maziva, tepelná izolácia chladiča, rozloženie teplotných polí a štartovacia pozícia piestov. Identifikované boli viaceré konštrukčné a prevádzkové nedostatky, ktoré viedli k samovoľnému zastaveniu motora, zníženej účinnosti či nestabilnej prevádzke.**

***Kľúčové slová* — Stirlingov motor, chladič, model, návrh Stirlingovho motora, pracovný piest**

I. Úvod

Stirlingov motor si v posledných rokoch opäť získava pozornosť odbornej verejnosti, najmä vďaka svojej schopnosti efektívneho využívania rôznych zdrojov tepla, čím reaguje na rastúci dopyt po ekologických a energeticky udržateľných technológiách. Keďže ide o motor, ktorý premieňa tepelnú energiu na mechanickú, dochádza k jeho širokej možnosti využitia, kde jeho prevádzka môže byť zabezpečená skoro akýmkoľvek zdrojom tepla. Jeho veľkou výhodou je aj schopnosť pracovať s nízkoteplotnými zdrojmi, kde sa prevažne využíva odpadové teplo alebo slnečná energia. Z elektroenergetického hľadiska predstavuje Stirlingov motor perspektívne využitie v rámci kogeneračných jednotiek, kde dochádza ku kombinovanej výrobe elektrickej energie a tepla. Použitie Stirlingovho motora prispieva k zvýšeniu energetickej účinnosti v podmienkach, kde sú tradičné technológie z technických alebo ekonomických dôvodov nevhodné [1].

II. Analýza výsledkov nameraných hodnôt pri použití kachľovej pece typu Petrík ako zdroja tepla

Cieľom analýzy prevádzkových parametrov bolo zistiť, aký vplyv má intenzita horenia na vplyv teploty zdroja, na čas prevádzky a teplotu jednotlivých komponentov modelu Stirlingovho motora. Skúmaný bol aj potrebný rozdiel teplôt medzi pretláčacím valcom a chladičom pre potreby spustenia motora, taktiež aj teplotné podmienky, pri ktorých došlo k samovoľnému zastaveniu a aká mohla byť príčina tohto zastavenia.

Ďalším skúmaným faktorom bol vplyv typu použitého maziva na mazanie pracovného valca s piestom, a jeho dôsledky na efektivitu chodu motora.

Súčasťou meraní bolo aj overenie schopnosti modelu generovať elektrickú energiu. Pre účely overenia schopnosti výroby elektrickej energie bol použitý malý jednosmerný motorček, ktorý slúžil ako generátor. Motorček bolo potrebné mechanicky spojiť s kľukovým hriadeľom a jeho výstupné vodiče boli pripojené k multimetru, kde sa sledovalo maximálne generované jednosmerné napätie.

Merania č. 1–3 (Tab. 1) prebiehali hneď za sebou pričom meranie č.1 prebiehalo zo studeného stavu, naopak merania č. 2 a 3 prebiehali z horúceho stavu vplyvom samovoľného zastavenia. Ako mazivo bol v týchto meraniach použitý olej na mazanie reťaze motorovej píly, ktorý sa preukázal ako vhodná voľba pre mazanie pracovného valca. Veľkou nevýhodou tohto maziva

bola jeho nízka viskozita, v dôsledku čoho dochádzalo k jeho rýchlemu stekaniu z mazaných plôch a potrebe častejšieho dopĺňania maziva. Následkom stekania maziva a vplyvom netesností medzi pracovným valcom a pracovným piestom dochádzalo k poklesu tlaku v pracovnom priestore motora, čo znemožnilo vytváranie potrebných tlakových rozdielov pre jeho činnosť následkom čoho dochádzalo k samovoľným zastaveniam v spomenutých meraniach. Prvé meranie prebiehalo zo studeného stavu, čo zapríčinilo väčší rozdiel teplôt medzi pretláčacím valcom a chladičom. Po rozhorení paliva v kachľovej peci po približne 10 minútach sa na liatinovú platňu kachľovej pece umiestnil model Stirlingovho motora, ktorého nábeh bol približne za 90 sekúnd. Nábeh Stirlingovho motora bol realizovaný manuálnym roztočením zotrvačníka, pričom bolo potrebné tento pokus na nábeh opakovať v približne 10 sekundových intervaloch až kým nedošlo k stabilnému rozbehu motora. Pri rozbehu boli pomocou termokamery zaznamenané teploty pretláčacieho valca, chladiča a pracovného valca pričom sa zaznamenával aj čas prevádzky a maximálne vzniknuté elektrické napätie. Po samovoľnom zastavení sa opäť zaznamenali teploty komponentov a bol realizovaný opätovný nábeh, ale tentokrát z horúceho stavu. Doba prevádzky sa oproti prvému meraniu zvýšila vplyvom vyššej teploty na strane ohrevu pretláčacieho valca, čo viedlo k zvýšeniu tlaku počas expanzie pracovného plynu, čo malo za následok väčšiu silu pôsobiacu na piest. Väčšia sila, ktorá pôsobí na piest generuje väčší krútiaci moment na hriadeli, čím sa zotrvačník rýchlejšie roztočí a tým aj vzniknuté napätie je vyššie ako v predchádzajúcom prípade. Po opätovnom zastavení a následnom spustení bolo možné uviesť motor do prevádzky len na 10 sekúnd, vplyvom príliš vysokej teploty chladiča.

Merania č. 4–7 boli realizované pri štarte zo studeného stavu, pričom meranie č. 8 prebiehalo z horúceho stavu vplyvom poruchy, kde došlo k uvoľneniu závlačky v mieste spoja pretláčacieho piestu s ojnicou. Po uvoľnení závlačky bolo potrebné motor nasilu zastaviť a rýchlo zasunúť závlačku na svoje miesto. Po danom úkone bolo možné dokončiť dané meranie. Ako mazivo sa v uvedených meraniach použil motorový olej 5W-40, ktorý sa preukázal ako ideálna voľba pre mazanie pracovného valca. Jedinou nevýhodou tohto maziva bola aj v tomto prípade jeho nízka viskozita, v dôsledku čoho dochádzalo k jeho rýchlemu stekaniu z mazaných plôch a potrebe častejšieho dopĺňania maziva. Napriek tomu mazivo zabezpečovalo ideálne podmienky pre chod pracovného valca, čím nedochádzalo k stratám a následnému poklesu tlaku v pracovnom priestore, vďaka čomu nedošlo k samovoľnému zastaveniu motora. V dôsledku toho došlo k výraznému zlepšeniu prevádzkových parametrov, najmä k rapídnemu nárastu času prevádzky motora zo 60 sekúnd až na 230 sekúnd, taktiež aj k citeľnému zvýšeniu generovaného napätia z 230 mV až na 300 mV. Priaznivý vplyv na stabilitu a trvanie prevádzky motora bol dosiahnutý optimalizáciou horenia v kachľovej peci, konkrétne znížením prívodu vzduchu k horeniu. Tým sa predišlo príliš prudkému horeniu, čo malo za následok postupné a miernejšie ohrievanie liatinovej platne, pričom nárast teploty jednotlivých komponentov bol oveľa pomalší ako v prípade predošlých 3 meraní [3].

V tabuľke 2 je uvedený prepočet nárastu teploty chladiča a pretláčacieho valca za 10 sekúnd, na základe ktorého bolo možné určiť vplyv rýchlosti ohrevu týchto časti pri danom tepelnom zdroji pri rôznych intenzitách horenia. Na základe prepočítaných hodnôt uvedených v tabuľke možno konštatovať, že pri optimalizácií horenia za pomoci zníženia prívodu vzduchu k horeniu dochádza k výrazne pomalšiemu nárastu teploty jednotlivých komponentov Stirlingovho motora. Regulácia ohrevu má priaznivý vplyv na stabilitu a dĺžku prevádzky. Porovnaním merania č. 4 s meraniami č. 1–3 je viditeľný výrazný rozdiel v rýchlosti ohrevu [2]. Na rozdiel od predchádzajúcich meraní č. 1–3, kde došlo k zvýšeniu teploty chladiča o 3,88°C za 10 sekúnd, pri optimálnom horení v meraní č. 4 bol nárast len 1,52°C za 10 sekúnd. Tento viac ako dvojnásobný rozdiel v rýchlosti ohrevu sa priamo prejavil aj v takmer trojnásobnom predĺžení celkového času prevádzky motora, čím sa potvrdil aj vplyv intenzity horenia na celkovú efektivitu a účinnosť motora.

V meraní č. 4 (Tab. 1) možno pozorovať aj vplyv okolitej teploty motora a sálavého tepla prenášaného z tepelného zdroja. Následkom týchto vplyvov nastal výrazný rozdiel medzi teplotou pracovného valca pri spustení a následnom zastavení, pričom tento rozdiel predstavoval až 12°C. Ide o značný rozdiel oproti ostatným meraniam, kde sa rozdiel týchto teplôt pohyboval v rozmedzí 0 až 5°C. Spomínané negatívne vplyvy môžu spôsobovať nežiaduce prehrievanie komponentov, ktoré nie sú v priamom kontakte so zdrojom tepla. Týka sa to hlavne chladiča a pracovného valca. Zvýšená okolitá teplota chladiča prispieva k zníženiu účinnosti chladenia, čo sa negatívne prejavuje skrátením času prevádzky [4].

Merania č. 5 a 6 prebiehali za identických podmienok, kde sa následnou analýzou priebehu ohrevu zistilo, že rýchlosť nárastu teploty chladiča bola vyššia ako pri pretláčacom valci, čo sa vyskytlo len v uvedených meraniach (Tab. 2). Tento jav mohol byť spôsobený už vyššie spomínanou vyššou teplotou sálavého tepla a teploty okolia. Tieto merania boli uskutočnené po

vychladnutí motora po predchádzajúcom meraní. Z porovnania týchto dvoch meraní vychádza, že aj napriek minimálnemu rozdielu v teplote spustenia došlo pri meraní č. 5 k skoršiemu zastaveniu pri oveľa nižších teplotách ako v prípade merania č. 6, kde bol zároveň nameraný o 40 sekúnd dlhší čas prevádzky. Meranie č. 6 prebiehalo pri najnižšej teplote spustenia chladiča spomedzi meraní č. 4 až 8. V danom meraní bolo vidieť postupný plynulý nábeh motora, ktorý nastal po približne 90 sekundách od položenia modelu na zdroj tepla. Dosiahnuté otáčky sa postupne zvyšovali úmerne s časom až kým sa neustálili, pričom pri ustálení bolo dosiahnuté najvyššie namerané napätie, ktoré predstavovalo maximálnu nameranú hodnotu spomedzi všetkých meraní a to konkrétne 300 mV. Zároveň v danom meraní došlo aj k najväčšiemu rozdielu medzi teplotou spustenia chladiča a pretláčacieho valca, ktorý predstavoval až 72°C. Tento výrazný teplotný rozdiel viedol k vyššej efektivite a účinnosti motora, vďaka čomu došlo k zvýšeniu krútiaceho momentu na kľukovom hriadeli a už k spomínanému dosiahnutiu maximálneho napätia [5].

Meranie č. 7 bolo realizované zo studeného stavu podobne ako predošlé merania, avšak počas chodu došlo k poruche na mechanizme v mieste spojenia pretláčacieho piestu s ojnicou za pomoci závlačky. Po uvoľnení závlačky muselo dôjsť k úplnému zastaveniu motora. Po vykonaní rýchlej opravy sa motor opäť umiestnil na zdroj tepla, pričom spustenie tentokrát prebiehalo z horúceho stavu, čo v Tab. 1 zodpovedá meraniu č. 8. Na základe priebehu merania č. 7 môžeme predpokladať, že ak by nedošlo k danej poruche, čas prevádzky by bol pravdepodobne výrazne vyšší [6]. S veľkou pravdepodobnosťou by dokonca mohlo dôjsť až k prekročeniu maximálnej nameranej hodnoty času prevádzky dosiahnutej v rámci merania č. 4. Po porovnaní nameraných prevádzkových hodnôt medzi meraním č. 4 a 7 sa ukázalo, že teploty spustenia jednotlivých komponentov boli takmer úplne identické, čo nasvedčuje aj schopnosť opakovateľnosti merania pri identických podmienkach s rovnakými výstupnými parametrami.

Meranie č. 9 prebiehalo zo studeného stavu. Ako mazivo sa v tomto meraní použil silikónový olej s viskozitou 1100 cSt (centistokov), ktorý sa nepreukázal ako vhodná voľba pre mazanie pracovného valca. Veľkou nevýhodou tohto maziva bola jeho vysoká viskozita, ktorá obmedzovala vykonávanie voľného pohybu piestu vo valci, čo viedlo k zvýšeniu trenia a tým aj k spomaleniu jeho pohybu počas cyklu. Zvýšené trenie malo za následok zníženie účinnosti motora a skrátenie času prevádzky. Z tohto hľadiska silikónový olej s takouto vysokou viskozitou nie je vhodný na mazanie rýchlo sa pohybujúcich častí, ako v tomto prípade pracovného piestu. Po krátkej dobe prevádzky bolo zistené, že silikónový olej nedokáže zabezpečiť konzistentné mazanie po celom obvode a dĺžke pohybu pracovného piestu vo valci, čo viedlo k lokálnym suchým miestam, kde dochádzalo k únikom pracovného plynu. Následkom únikov pracovného plynu dochádzalo k poklesu tlaku v pracovnom priestore motora čo znemožnilo vytváranie potrebných tlakových rozdielov pre jeho činnosť, následkom čoho došlo k samovoľnému zastaveniu. Na samotné samovoľné zastavenie mohla mať vplyv aj teplota zdroja tepla, keďže po prepočte v Tab. 2 je viditeľný rapídny nárast teploty chladiča a pracovného valca za 10 sekúnd prevádzky oproti ostatným experimentálnym meraniam [7].

Porovnaním merania č. 9 s meraniami č.1–3 je viditeľný výrazný rozdiel v rýchlosti ohrevu. Na rozdiel od predchádzajúcich meraní č. 1–3, kde došlo k zvýšeniu teploty chladiča o 3,88°C za 10 sekúnd a teploty pretláčacieho valca o 4,71°C za 10 sekúnd, pri veľmi silnej intenzite horenia v meraní č. 9 bol nárast teploty chladiča na 4°C za 10 sekúnd, ale došlo k rapídnemu nárastu teploty pretláčacieho valca na hodnotu 15°C za 10 sekúnd, čo predstavuje až viac ako trojnásobne zvýšenie nárastu teploty pretláčacieho valca oproti meraniu č. 1–3. Dokonca pri porovnaní s meraním č. 4, kde nárast teploty chladiča predstavuje 1,52°C/10 s a teploty pretláčacieho valca 2,74°C/10 s tento rozdiel je ešte vyšší. V prípade chladiča ide o takmer trojnásobne zvýšenie, ale pri pretláčacom valci došlo až skoro k šesťnásobnému rozdielu oproti meraniu č. 4. Tieto obrovské rozdiely v rýchlosti ohrevu sa negatívne prejavili aj na celkový čas prevádzky motora, čím sa opäť potvrdil aj vplyv intenzity horenia na celkovú efektivitu a účinnosť motora. V meraní č. 9 možno pozorovať aj vplyv okolitej teploty motora a sálavého tepla prenášaného z tepelného zdroja. Následkom týchto vplyvov nastal výrazný rozdiel medzi teplotou pracovného valca pri spustení oproti ostatným meraniam, kde v tomto prípade už pri spustení pracovný valec dosahoval teplotu 28°C. Ide o značný rozdiel oproti ostatným meraniam, kde sa teplota pracovného valca pri spustení pohybovala v rozmedzí 15–24°C. Spomínané negatívne vplyvy môžu spôsobovať nežiaduce prehrievanie komponentov, ktoré nie sú v priamom kontakte so zdrojom tepla, týka sa to hlavne chladiča a pracovného valca [8]. Zvýšená okolitá teplota chladiča prispieva k zníženiu účinnosti chladenia, čo sa negatívne prejavuje skrátením času prevádzky. Napriek týmto skutočnostiam najvýraznejší negatívny vplyv na celkovú prevádzku bol spôsobený zlou voľbou maziva a príliš vysokou intenzitou ohrevu od zdroja tepla [9].

Tabuľka 1
Namerané prevádzkové teploty pri použití kachľovej pece typu Petrík

Číslo merania	Teplota spustenia (chladič) [°C]	Teplota zastavenia (chladič) [°C]	Teplota spustenia (pretláčací valec) [°C]	Teplota zastavenia (pretláčací valec) [°C]	Teplota spustenia (pracovný valec) [°C]	Teplota zastavenia (pracovný valec) [°C]	ΔT (pretláčací valec -chladič-) [°C]	Čas prevádzky [s]	Maximálne namerané napätie [mV]
1.	39	44	121	128	15	15	82	15	193
2.	48	65	131	148	15	16	83	60	230
3.	68	72	155	161	16	17	87	10	151
4.	60	95	112	175	20	32	52	230	280
5.	50	75	116	135	18	22	66	160	267
6.	45	90	117	155	23	28	72	200	300
7.	60	73	110	119	21	25	50	100	245
8.	73	80	119	135	24	25	46	25	195
9.	59	63	108	123	28	28	49	10	115
Priemerné hodnoty	55,78	73	121	142,11	20	23,11	65,22	90	219,55

V nasledujúcej tabuľke 2 je uvedený prepočet nárastu teploty chladiča a pretláčacieho valca na čas 10 sekúnd. Tento prepočet bolo potrebné analyzovať z hľadiska intenzity ohrevu tepelného zdroja počas jednotlivých meraní. Na základe tejto analýzy bol zistený zásadný vplyv intenzity horenia na celkový chod a prevádzku motora, ale aj negatívne javy vznikajúce pri príliš silnom ohreve, kde došlo k veľkému pôsobeniu sálavého tepla a teploty okolia. Všetky zistenia zo spomínanej analýzy boli použité pri analýze výsledkov experimentálnych meraní uvedených z tabuľky 1 [10].

Tabuľka 2
Prepočet nárastu teploty komponentov za čas (k=10 sekúnd)

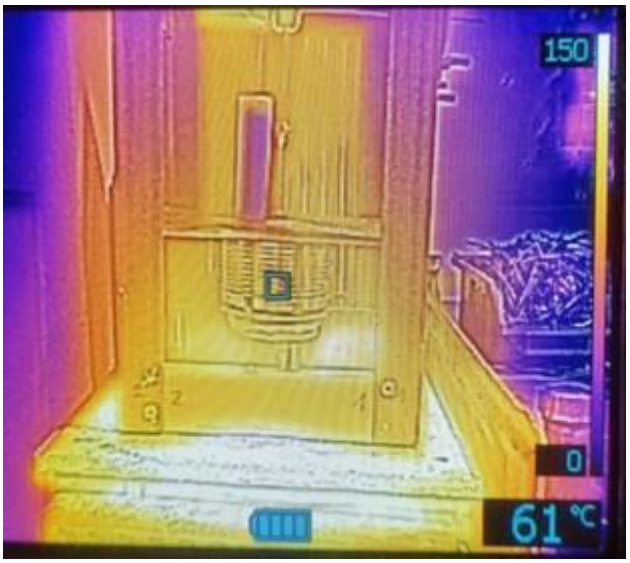
Číslo merania	Teplota spustenia (chladič) [°C]	Teplota zastavenia (chladič) [°C]	Teplota spustenia (pretláčací valec) [°C]	Teplota zastavenia (pretláčací valec) [°C]	ΔT (zastavenie chladič – spustenie chladič) $T_{chladič}$ [°C]	ΔT (zastavenie pretláčací valec - spustenie pretláčací valec) $T_{pretl.val.}$ [°C]	Prepočet nárastu teploty za 10 sekúnd (chladič) X_1 [°C/10 s]	Prepočet nárastu teploty za 10 sekúnd (pretláčací valec) X_2 [°C/10 s]	Čas prevádzky t [s]
1-3	39	72	121	161	33	40	3,88	4,71	85
4.	60	95	112	175	35	63	1,52	2,74	230
5.	50	75	116	135	25	19	1,56	1,19	160
6.	45	90	117	155	45	38	2,25	1,9	200
7-8	60	80	110	135	20	25	1,6	2	125
9.	59	63	108	123	4	15	4	15	10

Vzorce na prepočet nárastu teploty za čas 10 sekúnd:

$$x_1 = \frac{k \,.T_{\text{chladič}}}{t} = \frac{10\ s\ .33°C}{85\ s} = 3{,}88\ °\text{C} / 10\ \text{s} \quad (1)$$

$$x_2 = \frac{k \,.T_{\text{pretl.val.}}}{t} = \frac{10\ s\ .40°C}{85\ s} = 4{,}71\ °\text{C} / 10\ \text{s} \quad (2)$$

Obr. 1 Meranie teploty pracovného valca



Obr. 2 Meranie teploty chladiča

Obr. 3 Meranie pretláčacieho valca spolu aj s teplotou liatinovej platne

Obr. 4 Zostrojený model Stirlingovho motora modifikácie γ

Na základe vyššie uvedených nameraných výsledkov je možné uviesť možné využitie Stirlingovho motora, pretože ich hlavnou výhodou je schopnosť využívať nízkopotenciálne alebo odpadové teplo. Táto vlastnosť ich predurčuje na použitie v malých kogeneračných jednotkách alebo autonómnych energetických systémoch. Navrhnutý model však poukazuje aj na potrebu presného inžinierskeho spracovania a regulácie prevádzkových podmienok pre dosiahnutie stabilnej a spoľahlivej činnosti.

Do budúcnosti odporúčame realizovať merania v laboratórnych podmienkach s možnosťou kontinuálneho zberu dát (otáčky, tlak, prúd, napätie, teplotné polia) a doplniť model o sofistikovanejšie chladenie (kvapalinové alebo pomocou Peltierovho článku), ako aj upraviť mechanické parametre modelu s cieľom zvýšiť jeho spoľahlivosť, efektivitu a energetický výstup.

III. Záver

Po analýze výsledkov vykonaných experimentálnych meraní sa zistilo viacero nedostatkov, ktorých súvis má výrazný vplyv na funkciu Stirlingovho motora. Za najvýraznejší konštrukčný nedostatok sa ukázala nepresnosť pri výrobe pracovného piestu, ktorá vznikla počas obrábania. V dôsledku nadmerného ubratia materiálu došlo k zmenšeniu priemeru piestu viac než bolo potrebné. Vzniknutý problém sa prejavil počas vykonávania meraní, keď sa v dôsledku nadmernej vôle medzi pracovným piestom a stenami valca vyskytovali nežiaduce úniky pracovného plynu, čo malo za následok zníženie účinnosti motora.

Úniky v menšej miere boli spozorované aj v časti vymedzovača, kde k únikom dochádzalo medzi telom vymedzovača a nerezového drôtu, na ktorom sa nachádzal pretláčací piest. K ďalšiemu výraznému úniku došlo v mieste spoja pracovného valca s nerezovým plechom v dôsledku koncentrácie chemoprénového lepidla s použitými mazivami alebo technickým benzínom, kde sa pod vrstvou lepidla vytvorili malé vzduchové bublinky, ktoré po prasknutí prepúšťali pracovný plyn.

Na základe zistených skutočností je odporúčané pri obrábaní pracovného piestu zvoliť sústruh, kde dochádza k rovnomernému úbytku materiálu, poprípade zostrojiť viac prototypov pracovných piestov. V časti vymedzovača sa odporúča zmenšiť vôľu medzi nerezovým drôtom a stenami vymedzovača, resp. použiť tesniacu vložku. Pri zvolenom type lepidla je potrebné zistiť, s akými látkami nemôže dôjsť ku kontaktu v častiach lepených spojov.

Ďalší výrazný nedostatok bol pozorovaný v časti chladiča, kde dochádzalo k jeho nahrievaniu, čím strácal schopnosť plniť si svoju funkciu, čo spôsobilo zníženie času a schopnosti prevádzky. Zásadný vplyv na ohrev chladiča mal aj tepelný zdroj, kde vplyvom sálavého tepla a teploty okolia dochádzalo k jeho prebytočnému nahrievaniu. Ako sa neskôr ukázalo, príliš vysoká teplota zdroja tepla má negatívny vplyv na celkovú prevádzku motora.

Na elimináciu vzniknutých nedostatkov sa odporúča optimalizovať, prípadne regulovať intenzitu ohrevu tepelného zdroja. V časti chladiča sa odporúča miesto chladiaceho média, ktorým bol v tomto prípade vzduch, použiť alternatívu vo forme hélia alebo dusíka, prípadne použiť kvapalinové chladenie alebo Peltierov článok. Ďalším odporúčaním je aj použitie tepelnej izolácie alebo tepelného tienenia medzi studenou a horúcou časťou motora, čím sa eliminuje vplyv ohrevu chladiča vo forme sálavého tepla.

V rámci jednotlivých meraní bol spozorovaný aj vplyv trenia v časti pretláčacieho valca, kde dochádzalo ku kontaktu pretláčacieho piestu so stenami valca. Príčinou bolo nadmerné navinutie oceľovej vlny na časť pretláčacieho piestu. Možnou príčinou bolo aj mierne vychýlenie pretláčacieho piestu počas vykonávania pohybu. Toto vychýlenie mohlo byť spôsobené nepresným vycentrovaním kľukového hriadeľa, ku ktorému došlo pri ručnom ohýbaní pri výrobe, kde nebolo možné zabezpečiť dokonalú presnosť. Ako sa neskôr ukázalo, aj vplyv použitého maziva má vplyv na trenie. Pri použití silikónového oleja s vysokou viskozitou dochádzalo k obzvlášť vysokému treniu, ktoré sa prejavovalo aj brzdením pohybu.

Pre elimináciu týchto problémov sa odporúča navinúť na pretláčací piest menej oceľovej vlny o jednu otáčku. Pri výrobe kľukového hriadeľa je vhodné použiť prípravok určený na ohýbanie a aplikovať vhodný typ maziva s nízkou viskozitou.

Posledným významným faktorom ovplyvňujúcim činnosť motora mohla byť aj štartovacia pozícia pretláčacieho piestu. V danom prípade bola štartovacia pozícia nastavená až 3 cm nad úplným spodkom pretláčacieho valca, čím vznikol nevyužitý objem v spodnej časti valca. Tento konštrukčný nedostatok viedol k nedostatočnému stlačeniu plynu v studenej časti motora, čo spôsobilo zníženie potrebného teplotného rozdielu nevyhnutného na efektívnu expanziu pracovného plynu. V konečnom dôsledku tým došlo k poklesu výkonu a účinnosti motora.

Na základe týchto zistení je potrebné brať vysoký ohľad na správne nastavenie rozsahu pohybu pretláčacieho valca, čím sa docieli nárast výkonu a aj celkovej účinnosti motora.

Najdôležitejším faktorom pri realizácii experimentálnych meraní sú vhodné podmienky a kalibrované meracie prístroje. Pre dosiahnutie presnejších výsledkov sa odporúča dané merania vykonávať za rovnakých podmienok a v rovnakých časových intervaloch, ideálne v laboratórnom prostredí, kde je možné využiť senzory na nepretržité zaznamenávanie teplôt jednotlivých komponentov, snímanie otáčok kľukového hriadeľa ako aj prúdu, napätia a vzniknutého tlaku pracovného plynu vo valci. Pomocou presných výsledkov je následne možné vyhotoviť presné analýzy aj s grafickými závislosťami.

POĎAKOVANIE

Tento príspevok vznikol vďaka podpore Vedeckej grantovej agentúre Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporeného grantom VEGA 1/0532/25 a VEGA 1/0627/24.

ZOZNAM POUŽITEJ LITERATÚRY

[1] Interesting Engineering, „*The Everlasting, Nearly Emission Free Stirling Engine*“. [online] [cit. 2024-11-01]. Dostupné na internete: <https://interestingengineering.com/transportation/the-everlasting-nearly-emission-free-stirling-engine>.
[2] Stirling Ultracold, „*Stirling Engine: 200+ Year History*“. [online] [cit. 2024-10-31]. Dostupné na internete: <https://www.stirlingultracold.com/heritage/>.
[3] Kontax Stirling Engines. [online] [cit. 2024-10-31]. Dostupné na internete: <https://www.stirlingengines.co.uk/History.asp>.
[4] Zelené správy.cz, „*Stirlingův motor – Historie, princip a jeho využití při kogeneraci*“. [online] [cit. 2024-10-30]. Dostupné na internete:<https://www.zelenezpravy.cz/stirlinguv-motor-%E2%80%93-historie-princip-a-jeho-vyuziti-pri-kogeneraci/>.
[5] Swedish Stirling. [online] [cit.2024-10-30]. Dostupné na internete:<https://swedishstirling.com/>.
[6] Autoride, „*Stirlingov motor: Ako funguje a aké má výhody?*“ [online] [cit. 2024-11-01]. Dostupné na internete:<https://autoride.sk/stirlingov-motor-princip-prace-vyhody>.
[7] Ekofond Plus, „*Stirlingov motor na zemný plyn*“. [online] [cit. 2024-10-29]. Dostupné na internete:<http://platforma.ekofondplus.sk/moderne-vyucovanie/zdroje-energie/stirlingov-motor-na-zemny-plyn>.
[8] Coolenergy. [online] [cit. 2024-10-31]. Dostupné na internete:<https://coolenergy.com/>.
[9] Sciencedirect, „*Technological development in the Stirling cycle engines*“. 2006. [online] [cit. 2024-11-01]. Dostupné na:<https://www.sciencedirect.com/science/article/abs/pii/S1364032106000906?via%3Dihub>.
[10] M. Hirko, „Návrh funkčného modelu Stirlingovho motora“, Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025.

Vplyv elektromobility na mestskú elektrizačnú sieť a infraštruktúru

Pavlo HRYHORIEV, Dušan MEDVEĎ

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

pavlo.hryhoriev@student.tuke.sk, dusan.medved@tuke.sk

***Abstrakt* — Tento článok prezentuje vplyv elektrického vozidla na elektrickú sieť počas jeho nabíjania. Zameriava sa najmä na analýzu problémov, ktoré vznikajú v elektrickej sieti v dôsledku pripojenia elektromobilu. Na modelovanie vplyvu bol vytvorený simulačný model v prostredí Simulink. V článku sa uvažuje s dvomi scenármi nabíjania elektrického vozidla (jednofázovým a trojfázovým). V prvom prípade sa skúma správanie domácej elektrickej siete (400 V) pri rôznych hodnotách nabíjacieho prúdu. V druhom prípade sa analyzoval vplyv na distribučnú sieť (22 kV) a domácu sieť počas trojfázového nabíjania. V každom z týchto prípadov sa osobitne skúmala kvalita siete s filtrom a bez filtra. Výsledky modelovania umožňujú vyvodiť závery o optimalizácii procesu nabíjania elektrických vozidiel s cieľom znížiť negatívny vplyv na elektrickú sieť.**

***Kľúčové slová* —hybridný systém, elektromobil, fotovoltaické panely, batéria, nabíjanie, regulátor, usmerňovač, striedač, menič, elektromobilita, nabíjacia stanica, nabíjací bod, nabíjacia infraštruktúra**

I. Úvod

Súčasný celosvetový vývoj dopravného priemyslu poukazuje na výrazný nárast záujmu o elektrifikáciu kolesovej dopravy. Hlavnými dôvodmi tohto javu sú potreba znížiť emisie skleníkových plynov, boj proti klimatickým zmenám a potreba znížiť závislosť od fosílnych palív, ktorých zásoby sú obmedzené. Elektrické vozidlá sa vďaka svojej ekologickosti, energetickej účinnosti a schopnosti integrácie s modernými technológiami stali sľubnou alternatívou k tradičným vozidlám so spaľovacími motormi.

Spolu s rozvojom elektrických vozidiel však vzniká nový problém, vplyv elektrických vozidiel na elektrickú sieť. Nie je žiadnym tajomstvom, že počet elektrických vozidiel každoročne rastie. S týmto rastom sa spája potreba rozšíriť energetickú infraštruktúru vrátane nabíjacích systémov.

Nárast počtu nabíjacích staníc nevyhnutne vedie k nárastu spotreby elektrickej energie, a teda k potrebe dodatočnej výroby elektrickej energie. Táto výroba by samozrejme mala byť založená na obnoviteľných zdrojoch, pretože prechod na elektrickú dopravu nemá zmysel, ak sa energia pre ňu vyrába zo zdrojov, ktoré sú škodlivé pre životné prostredie.

V tejto súvislosti sa v tomto článku uvažuje s modelom napájania elektrického vozidla elektrickou energiou vyrobenou z fotovoltických panelov. Vďaka tomuto prístupu je proces nabíjania šetrnejší k životnému prostrediu.

Osobitná pozornosť sa venuje problému harmonického skreslenia napätia, ktoré sa vyskytuje v sieti počas procesu nabíjania. V praktickej časti článku je navrhnuté riešenie tohto problému pomocou filtrov. V modeli sa tiež analyzuje niekoľko rôznych scenárov nabíjania s cieľom porovnať ich vplyv na sieť.

II. Stav nabíjacej infraštruktúry na Slovensku

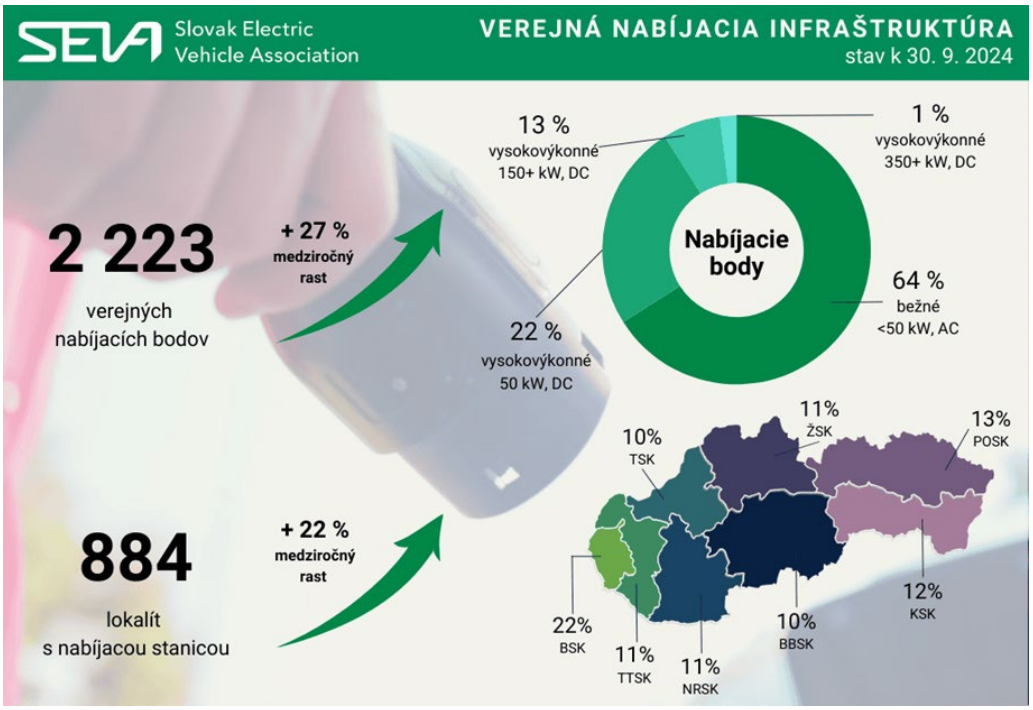
Ku koncu septembra 2024 bolo na Slovensku 2 223 verejných nabíjacích miest na 884 miestach s celkovým inštalovaným výkonom 107 562 kW. Rozdelenie podľa typu nabíjačky [1]:

- Nabíjacie body na striedavý prúd s výkonom do 22 kW: 1 421 jednotiek [2]
- DC rýchlonabíjačky do 50 kW: 421 nabíjačiek: 496 jednotiek [3]
- DC ultrarýchle nabíjacie stanice s výkonom nad 150 kW: 278 jednotiek [3]
- DC nabíjacie stanice s výkonom nad 350 kW: 26 jednotiek [2]

Napriek nárastu počtu nabíjacích staníc sa tempo výstavby spomalilo z dôvodu nízkeho využívania a očakávania financovania z Plánu obnovy a odolnosti.

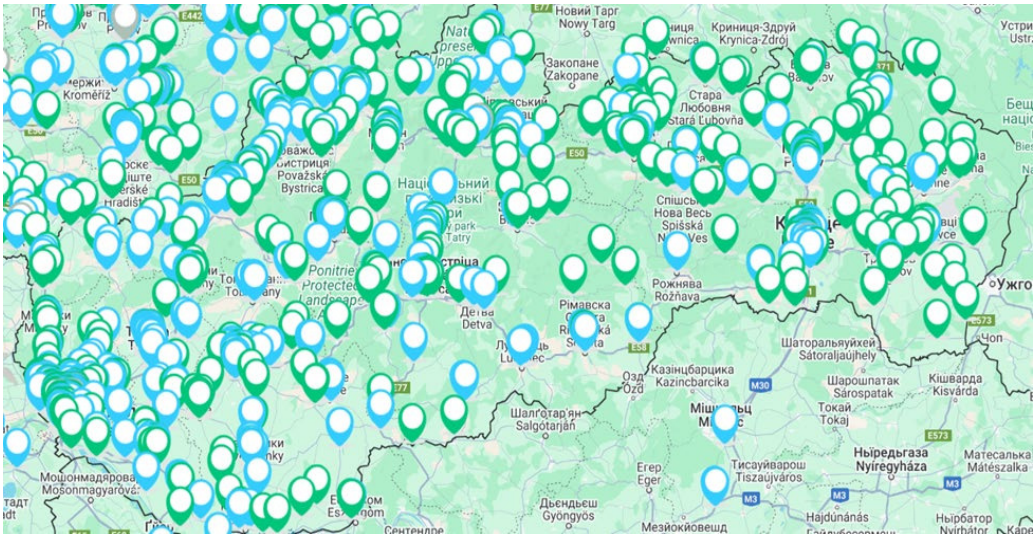
Výstavba verejných nabíjacích staníc pre elektromobily zaznamenala na Slovensku v prvom

polroku 2024 nárast, hoci z dvoch výziev z Plánu obnovy a rozvoja vidieka na podporu výstavby nebola zatiaľ nainštalovaná ani jedna. Hoci tempo výstavby nabíjacích staníc, v drvivej väčšine súkromnými spoločnosťami, je povzbudzujúce, pre efektívny rozvoj elektromobility je potrebná reálnejšia podpora zo strany štátu.



Obr. 1 Stav nabíjacej infraštruktúry na Slovensku [4]

Súhrnné informácie o nabíjacej infraštruktúre, ktoré spoločnosť SEVA zverejňuje štvrťročne, zachytávajú len počet verejných nabíjacích staníc. Je potrebné zdôrazniť, že v priemere sa viac ako 80 % nabíjania uskutočňuje na neverejných staniciach, t.j. v súkromných garážach, firmách, kancelárskych budovách alebo skladoch.



Obr. 2 Oficiálna mapa nabíjacích staníc na Slovensku [5]

III. Vplyv integrácie siete elektromobilov na energetickú sieť

Vplyv integrácie elektromobilu do siete možno klasifikovať na negatívne a pozitívne aspekty. Elektromobily predstavujú významné výzvy pre energetické podniky. Nadmerná integrácia elektromobilov do distribučnej siete ovplyvňuje stabilitu distribučnej sústavy. Tento nepriaznivý efekt je spôsobený zmenou profilu zaťaženia, nerovnováhou napätia a frekvencie, nadmerným zavádzaním harmonických zložiek napätia a stratami výkonu, ako je uvedené v tabuľke 1. Zhoršenie kvality elektrickej energie, špičkové zaťaženie a problémy s reguláciou výkonu môžu

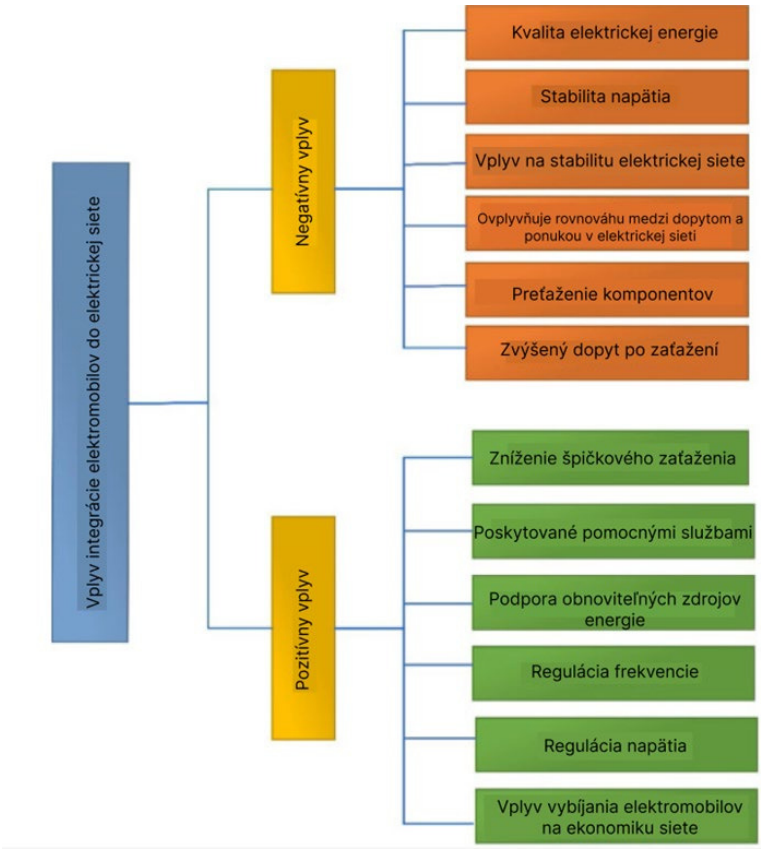
vzniknúť v dôsledku nadmerného prenikania elektromobilov do siete. Tieto problémy je možné riešiť pomocou pokročilých metód riadenia napájania. V tabuľke 2 je zhrnutý pozitívny vplyv integrácie elektromobilov. [6]

Tabuľka 1
Negatívne dôsledky integrácie elektromobilov do elektrickej siete

Dôsledky	**Stručné informácie**
Vplyv na kvalitu elektrickej energie	▪ Vplyv elektrickej siete na sieť V2G (Vehicle-to-grid) sa hodnotí predovšetkým prostredníctvom sledovania charakteru zdroja napájania ▪ Neistota spojená s nabíjaním elektromobilu vedie k viacerým problémom, ako sú prepätie v sieti, zhoršenie kvality elektrickej energie, zvýšené poškodenie vedenia a poruchy s vysokým prúdom ▪ Vďaka vysokofrekvenčným meničom, ktoré premieňajú striedavé napätie na jednosmerné napätie na nabíjanie elektromobilov, je sieť vystavená harmonickým, čo vedie k nízkej kvalite elektrickej energie ▪ Distribučné transformátory sa stávajú preťaženými v dôsledku harmonických, čo vedie k skráteniu ich životnosti
Vplyv na stabilitu napätia	▪ Nestabilita napájacieho systému je spôsobená nelineárnymi zaťaženiami elektromobilov, ktoré spotrebúvajú veľké množstvo energie počas krátkych časových období. ▪ Elektrické dopravné prostriedky zvyšujú pravdepodobnosť porúch v napájacom systéme a dlhšie sa vracajú do stabilného stavu ▪ Primerane riadený systém EVGI (electric vehicle grid integration) môže zvýšiť spoľahlivosť siete
Vplyv na stabilitu elektrických sietí	▪ Do systému je potrebné pridať nové elektromobily spolu s vyhodnotením systému zbernice IEEE (štandardy) 3 s využitím nabíjacích staníc alebo bez nich. Pred pridaním nových zaťažení elektromobilov do elektrického systému, ktorý bol nestabilný, je potrebné vykonať analýzu stability. ▪ Elektrická sieť sa analyzuje z hľadiska stability malého signálu. Elektromobil sa považuje za konštantné zaťaženie a impedanciu ▪ EV, modelované ako konštantné výkonové zaťaženie, sa ukázalo ako nestabilné, preto sa na riadenie EV, modelovaného ako konštantné impedančné zaťaženie, používa inteligentný riadiaci algoritmus
Vplyv na rovnováhu medzi dopytom a ponukou v elektrickej sieti	▪ Harmonické zložky sa zavádzajú do sieťového systému nabíjacími stanicami pre elektromobily. Tieto harmonické sa do siete dostávajú vďaka výkonovo-elektronickému meniču. Architektúra nabíjania tak ovplyvňuje kolísanie celkového harmonického skreslenia. ▪ Zmenlivosť harmonického skreslenia sa pohybuje od 7 % do 99 % pri nabíjacích zariadeniach pre elektromobily ▪ Výskumníci zistili, že úroveň penetrácie elektromobilov možno zvýšiť pomocou kondenzátorových batérií bez vyvolania porúch napäťových harmonických
Preťaženie komponentov	▪ Mimoriadne veľké množstvo EVGI (electric vehicle grid integration) vytvára alebo prenáša požiadavku na dodatočné zaťaženie ▪ Niektoré existujúce komponenty energetického systému nie sú navrhnuté na dodatočné zaťaženie, čo vedie k preťaženiu komponentov a skráteniu životnosti transformátora
Zvýšené zaťaženie	▪ Celkovo možno pridať 1000 TWh dodatočných zaťažení (nárast o 25 % v porovnaní so súčasnými úrovňami). ▪ Platby v špičke rastú v dôsledku nekontrolovaného nabíjania elektromobilov, čo výrazne zaťažuje verejné služby

Tabuľka 2
Pozitívne dôsledky integrácie elektromobilov do elektrickej siete

Dôsledky	**Stručné informácie**
Zníženie špičkového zaťaženia	▪ Využívanie časovo diferencovaných taríf (TOU – time-of-use) a koordinovaných stratégií nabíjania robí sieť energeticky efektívnejšou bez zvyšovania výrobnej kapacity ▪ Niekoľko obojsmerných elektromobilov pripojených do siete sa tu využíva ako súčasť priemyselnej elektrickej siete. Systémy riadenia batérií môžu ukladať obmedzenia týkajúce sa typu a rozsahu nabíjacích a vybíjacích signálov v závislosti od spôsobu používania. To negatívne ovplyvňuje prenikanie elektromobilov do sieťového systému ▪ Za praktických okolností možno znížiť 96 % špičkovej spotreby v energetickom systéme. Kedykoľvek má batéria elektromobilu veľkú kapacitu, výrazne profituje zo zníženia špičkovej vnútornej spotreby a dopytu zo siete. V konečnom dôsledku to môže viesť k lepšiemu využívaniu energetického systému
Poskytuje sa pomocnými službami	▪ Pomocné služby pre výrobu elektrickej energie zabezpečujú spoľahlivosť, dopyt, dodávku a stabilitu siete ▪ Využitím technológie V2G je možné poskytovať pomocné služby, čím sa zlepšuje stabilita elektrickej siete
Podpora obnoviteľných zdrojov energie	▪ Prevádzka elektromobilov ako akumulátorov energie môže znížiť neistotu obnoviteľnej energie ▪ Využitie elektromobilov ako vyrovnávacieho prvku obnoviteľnej energie znižuje emisie a môže priniesť úsporu peňazí
Regulácia frekvencie	▪ Odchýlka frekvencie siete sa koriguje reguláciou frekvencie ▪ Vďaka technológii V2G je možné zaťaženie vyvážiť prostredníctvom regulácie frekvencie siete ▪ Integrácia elektromobilov môže taktiež pomôcť minimalizovať výkyvy frekvencie spôsobené náhlymi zmenami zaťaženia, keď sú elektromobily integrované do siete
Regulácia napätia	▪ Regulácia napätia má zásadný význam pre zabezpečenie stability dopytu a ponuky jalového výkonu. Elektromobily môžu rýchlo reagovať na regulačné signály, keďže sú autonómne riadené samotným vozidlom ▪ Elektromobily autonómne spracovávajú regulačné signály, čo im umožňuje rýchlo reagovať na bezpečnostné signály. Nabíjanie elektromobilu sa preruší, keď napätie systému klesne pod určitú úroveň, a obnoví sa, keď napätie systému dosiahne stanovenú úroveň



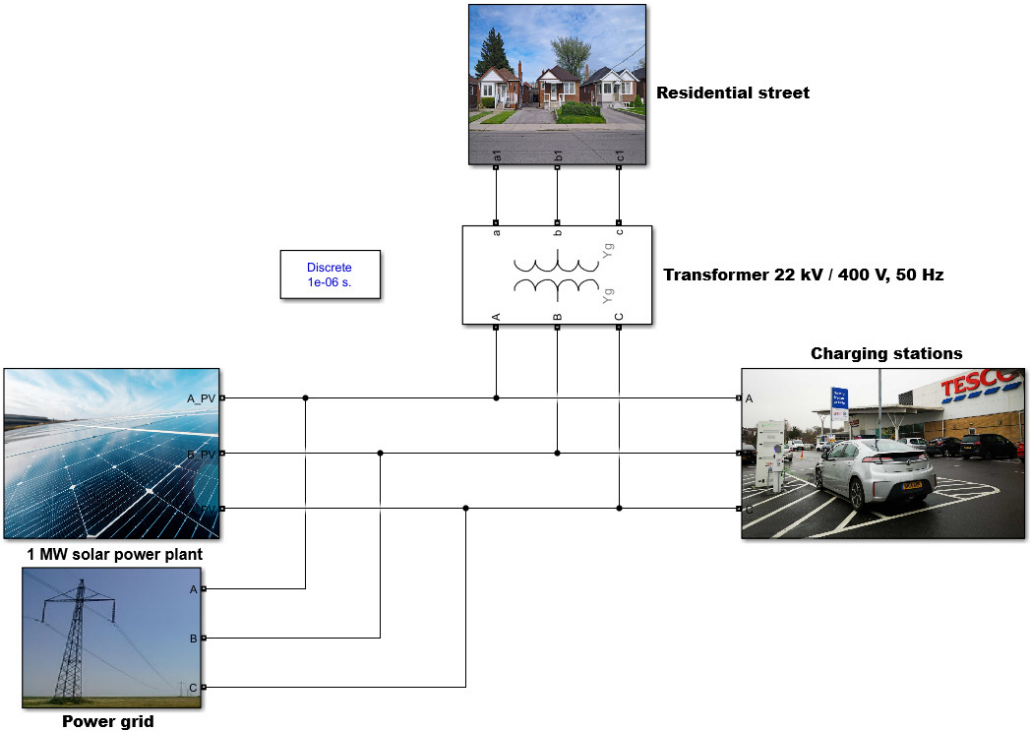
Obr. 3 Vplyv integrácie siete elektromobilov na energetickú sieť [6]

IV. CELKOVÝ MODEL A JEHO POPIS

Pre účely tohto článku bolo vytvorené modelové prostredie, ktoré simuluje distribuovanú elektrickú sieť so zohľadnením rôznych zdrojov výroby, transformátorov, typov zaťaženia, ako aj pripojenia elektromobilov. V tomto modeli je realizované pripojenie jednofázového aj trojfázového elektromobilu. Ich nabíjanie sa skúma v dvoch podmienkach, a to v rezidenčnej oblasti a na nabíjacej stanici pri supermarkete.

Súčasťou modelu sú dva zdroje fotovoltickej elektrárne: prvý s výkonom 1 MW a druhý s výkonom 100 kW. Na zníženie napätia z vysokonapäťovej úrovne na úroveň vhodnú pre spotrebiteľov boli použité dva transformátory s parametrami 22 kV/400 V. Okrem toho bol v modeli použitý zdroj typu Three-Phase Source, ktorý predstavuje pripojenie k hlavnej elektrickej sieti (Power Grid) a zabezpečuje energetickú rovnováhu v prípade nedostatku energie vyrobenej z fotovoltickej elektrárne.

Schematické zobrazenie vytvoreného systému je uvedené na obr. 4. Táto štruktúra zobrazuje hlavné prvky, ako sú zdroje výroby elektrickej energie, znižovací transformátor, rezidenčnú oblasť, supermarket s nabíjacími stanicami, ako aj samotné elektromobily.



Obr. 4 Schéma modelu elektrickej siete na výskum procesov nabíjania elektromobilov.

V. ZHODNOTENIE VÝSLEDKOV

Skúmané boli dva základné scenáre: jednofázové a trojfázové nabíjanie. V oboch prípadoch bola analyzovaná kvalita elektrickej energie, výskyt harmonických a zaťaženie siete, pričom sa porovnávali výsledky pri použití a nepoužití filtra LCL. Výsledky sú ilustrované v grafoch, ktoré zobrazujú priebehy napätia, prúdu a spektrálne rozbory prostredníctvom FFT [7].

A. Nabíjanie elektromobilu z jednofázovej domácej siete

V prvom scenári bolo vozidlo pripojené na jednofázovú domácu sieť s menovitým napätím 400 V. Nabíjací prúd sa menil v rozsahu 0 A – 16 A v závislosti od dostupnosti výkonu z fotovoltických panelov. Simulácie ukázali, že v momente pripojenia vozidla do siete dochádza k náhlemu zvýšeniu zaťaženia a poklesu napätia na príslušnej fáze. Tento jav bol sprevádzaný výskytom vyšších harmonických spôsobených činnosťou výkonových meničov.

Bez použitia filtra LCL bola úroveň harmonického skreslenia (THD) v miestach pripojenia výrazne nad úrovňami stanovenými v norme STN EN 50160, pričom dominovali 5., 7. a 11. harmonická. Po zaradení filtra LCL do obvodu sa hodnota THD znížila v priemere o 40–60 %, čo umožnilo splnenie štandardných limitov kvality napájania.

B. Trojfázové nabíjanie a jeho vplyv na distribučnú sieť

Druhý scenár simuloval nabíjanie elektromobilu z trojfázovej siete, teda typické pre verejné nabíjacie stanice. Nabíjací prúd bol nastavený na hodnotu 30 A a výstupné napätie na úrovni DC prípojnice bolo 800 V. Analýza ukázala, že pri trojfázovom nabíjaní je zaťaženie rovnomernejšie rozložené a výskyt harmonických nižší v porovnaní s jednofázovým scenárom.

Napriek tomu boli aj v tomto prípade zaznamenané skreslenia napätia a prúdu, najmä v oblasti znižovania napätia medzi 22 kV a 400 V. Použitie filtra LCL malo opäť významný pozitívny účinok, kde THD pokleslo pod 5 % a priebehy napätia aj prúdu sa výrazne vyhladili. Zaznamenané boli tiež pozitívne efekty na stabilitu napätia v celej sieti.

C. Porovnanie scenárov a význam filtrov

Porovnaním oboch scenárov možno konštatovať, že úroveň harmonických závisí nielen od nabíjacieho prúdu a typu siete, ale aj od topológie zapojenia a typu použitej filtrácie. V nižšej napäťovej hladine (0,4 kV) bolo skreslenie výraznejšie než na úrovni 22 kV, pretože to súvisí s vyššími prúdmi a nižšou impedanciou siete.

Filter typu LCL sa v oboch prípadoch preukázal ako efektívny nástroj na potlačenie harmonických. V jednofázovom režime bol jeho prínos najvýraznejší, pretože záťaž bola nelineárna a asymetrická. V trojfázovom režime zabezpečil stabilnejší priebeh všetkých veličín.

D. Praktické odporúčania pre návrh siete

Na základe simulácií môžeme odporučiť nasledovné opatrenia:

- Pri zvyšujúcom sa počte elektromobilov v mestských oblastiach je potrebné už vo fáze návrhu siete počítať so záťažovým profilom zohľadňujúcim súčasné aj budúce nabíjacie kapacity.
- Zabezpečenie rovnomerného rozdelenia záťaže medzi fázy znižuje riziko lokálneho preťaženia a poklesu napätia.
- Filtrovanie harmonických, najmä pomocou pasívnych filtrov LCL, významne prispieva k splneniu požiadaviek na kvalitu elektrickej energie.
- Pri rozsiahlejších implementáciách je vhodné zaviesť aj riadenie záťaže (load management) na úrovni nabíjacích staníc.

POĎAKOVANIE

Tento príspevok vznikol vďaka podpore Vedeckej grantovej agentúre Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporeného grantom VEGA 1/0532/25 a VEGA 1/0627/24.

ZOZNAM POUŽITEJ LITERATÚRY

[1] Trend, "*Posledné v Európe. Slovensko nie je dostatočne pripravené na prechod k elektromobilite*", [online]. Bratislava: TREND, 2024 [cit. 2025-05-18]. Dostupné na internete: <https://www.trend.sk/spravy/posledne-europe-slovensko-nie-je-dostatocne-pripravene-prechod-elektromobilite >.

[2] Mojelektromobil.sk, „*Nabíjacie stanice pre elektromobily – september 2024*“, [online]. 2024 [cit. 2025-05-18]. Dostupné na: <https://www.mojelektromobil.sk/nabijacie-stanice-pre-elektromobily-september-2024/>.

[3] Seva, „*Verejná nabíjacia infraštruktúra – Q3 2024*“, [online]. Bratislava: Slovenská asociácia pre elektromobilitu, 2024 [cit. 2025-05-18]. Dostupné na: <https://seva.sk/download/seva_verejna-nabijacia-infrastruktura_q3-2024/>.

[4] Mojelektromobil.sk, „*Nabíjacie stanice pre elektromobily na Slovensku: Počty a výkony k septembru 2024*“, [online]. 2024 [cit. 2025-04-09].Dostupné na internete: <https://www.mojelektromobil.sk/nabijacie-stanice-pre-elektromobily-september-2024/>.

[5] Seva, „*Slovenská asociácia pre elektromobilitu: Mapa verejných nabíjacích staníc na Slovensku*“, [online]. [cit. 2025-04-01].Dostupne na internete <https://nabijame.sk/>.

[6] M. S. Mastoi, et. al., „*An in-depth analysis of electric vehicle charging station infrastructure, policy implications, and future trend*“. In: Energy Reports, 2022, roč. 8, november, s. 11504-11529. ISSN 2352-4847.

[7] P. Hryhoriev, „*Vplyv elektromobility na mestskú elektrizačnú sieť a infraštruktúru*“, Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025.

Dynamické riadenie záťaže v inteligentných sieťach

Yurii HRYHORIEV, Dušan MEDVEĎ

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

yurii.hryhoriev@student.tuke.sk, dusan.medved@tuke.sk

***Abstrakt* — Tento článok sa venuje analýze a modelovaniu dynamického riadenia zaťaženia v nízkonapäťovej distribučnej sieti s využitím systému akumulácie energie (BESS). Hlavným cieľom je navrhnúť a overiť riadiaci algoritmus, ktorý umožňuje efektívne využívať BESS na stabilizáciu siete počas denného kolísania zaťaženia. V prostredí MATLAB/Simulink bol vytvorený model inteligentnej siete, ktorý integruje distribučnú sieť, dynamické odbery, batériový systém a adaptívny regulátor. Model bol testovaný v štyroch scenároch: bez BESS a s BESS počas letného aj zimného režimu, vrátane prípadu s dodatočným nabíjaním elektromobilu. Výsledky simulácií potvrdili, že použitie BESS výrazne prispieva k zníženiu špičkového zaťaženia, stabilizácii napätia a zvýšeniu energetickej účinnosti. Navrhnutý model je vhodný ako nástroj pre ďalší výskum v oblasti optimalizácie riadenia spotreby a akumulácie energie v podmienkach inteligentných distribučných sietí.**

***Klúčové slová* — inteligentná sieť, smart grid, dynamické riadenie, distribučná sieť, systém skladovania energie, Simulink, BESS**

I. Úvod

Rozvoj energetickej siete založený na koncepcii inteligentnej siete a intelektualizácii siete sa vo svete považuje za hlavný spôsob modernizácie energetického sektora. V súčasnosti existuje požiadavka spoločnosti na zabezpečenie neprerušovanej dodávky elektrickej energie, vhodnej spoľahlivosti a kvality elektrickej energie. Výrazne rastie aj vyrobená energia z obnoviteľných zdrojov energie, a to tak malých zariadení (do 1 MW), ako aj veľkých elektrární.

Jedným z riešení, ako zabezpečiť stabilnú prevádzku elektrickej siete v súvislosti s rastúcim podielom energie z obnoviteľných zdrojov, je používanie batériových systémov skladovania energie (BESS). Umožňujú akumulovať prebytočnú energiu vyrobenú v priaznivých obdobiach a dodávať ju spotrebiteľom počas hodín špičkového zaťaženia alebo nedostatku výroby. Účinnosť BESS však vo veľkej miere závisí od správneho nastavenia riadiacich algoritmov, ktoré si zasa vyžadujú presné modelovanie režimov systému a prechodných javov.

V súčasnosti sa na vytváranie a testovanie takýchto algoritmov široko používajú matematické modely implementované v softvérových prostrediach, ako je MATLAB/Simulink. Používanie takýchto modelov umožňuje projektantom analyzovať správanie sa systémov skladovania energie v rôznych scenároch prevádzky siete. Zložitosť takýchto modelov a potreba veľkého počtu vstupných parametrov však môže skomplikovať ich implementáciu v praxi. Jedným z možných riešení je vytvorenie zjednodušených modelov, ktoré však stále presne odrážajú hlavne prevádzkové režimy.

Keďže výroba elektrickej energie fotovoltickými elektrárňami je viac-menej nestabilná, systémy skladovania energie sa stávajú dôležitým atribútom systémov zásobovania energiou. Je potrebné uviesť, že elektrická energia nazbieraná počas dňa sa dodáva do siete v čase najväčšej spotreby, keď energia vyrobená vo fotovoltickej elektrárni nestačí na pokrytie potrieb spotrebiteľov. Je potrebné vziať do úvahy skutočnosť, že systém skladovania energie nie je len batéria, pretože známi svetoví výrobcovia vkladajú do tohto pojmu trochu iný význam. V tomto prípade sa spotrebiteľovi ponúka komplexný prístup, ktorý zahŕňa úložné zariadenia a softvérové riešenia na zabezpečenie kontroly ich stavu a optimálneho rozloženia zaťaženia.

II. Riadenie dopytu

Pod pojmom Demand Side Management (DSM) sa rozumie realizácia opatrení na zvyšovanie energetickej efektívnosti a riadenie výkonu na strane spotrebiteľa s cieľom celkovej

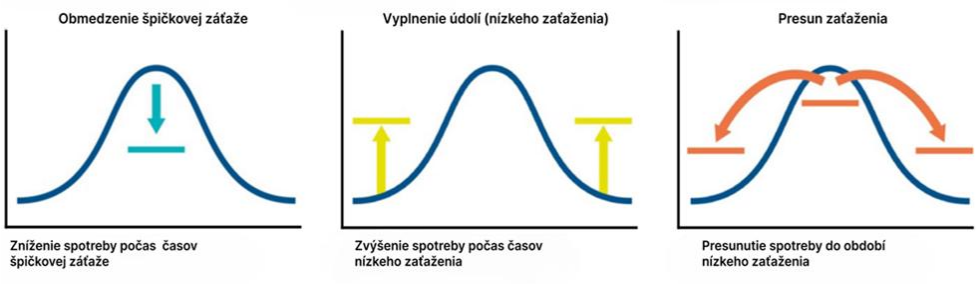
optimalizácie príslušného energetického systému [1]. To sa dosahuje využívaním flexibilného zaťaženia spotrebiteľov (tzv. „flexibility“) v prospech elektrizačnej sústavy.

Spotrebiteľ tak môže poskytovať systémové služby na zabezpečenie stabilnej prevádzky elektrickej siete alebo prispievať ku kompenzácii výkyvov vo výrobe elektriny z obnoviteľných zdrojov. To sa môže uskutočňovať napríklad prostredníctvom poskytovania regulačnej energie.

Regulačná energia je potrebná vždy, keď dostupný výkon v sieti nezodpovedá skutočnej spotrebe. V prípade nedostatku výkonu je potrebné dodať do siete viac elektrickej energie alebo znížiť spotrebu v určitých bodoch, v tomto prípade hovoríme o pozitívnej regulačnej energii.

Negatívna regulačná energia je potrebná vtedy, keď ponuka prevyšuje dopyt. V takom prípade je potrebné znížiť výkon generátorov alebo ich úplne odpojiť od siete, alebo zvýšiť spotrebu elektriny na strane odberateľov [2].

DSM teda predstavuje cielené riadenie spotreby elektrickej energie na strane odberateľa. Vo všeobecnosti možno rozlíšiť rôzne typy presunu zaťaženia, ktoré sú znázornené na obr. 1.



Obr. 1 Typy presunu zaťaženia v rámci DSM

Najdôležitejšou podmienkou pre flexibilnú prevádzku v rámci Demand Side Management sú medzizásobníky, ktoré dokážu kompenzovať časové posuny v rámci technologického procesu. Ako takéto zásobníky môžu byť použité napríklad tepelné alebo prečerpávacie zariadenia, ako aj akumulátorové batérie [3].

Systém akumulácie energie na báze batérií, Battery Energy Storage System (BESS), ktorý je predmetom tohto príspevku, má zabezpečiť práve takýto režim prevádzky. V rámci virtuálnej mikrosiete so zapojením inteligentných sieťových algoritmov by mal fungovať ako elektrochemický zásobník energie, poskytujúci flexibilitu pre potreby elektrizačnej sústavy. Toto sa realizuje v závislosti od výkonu a kapacity BESS.

A. Účel a scenáre využitia systémov na ukladanie energie

Systémy na ukladanie energie (ESS) zohrávajú čoraz dôležitejšiu úlohu v moderných elektroenergetických sústavách. Ich hlavný účel spočíva v ukladaní elektrickej energie na neskoršie použitie, čo zvyšuje flexibilitu, spoľahlivosť a odolnosť celej energetickej infraštruktúry.

Zavádzanie batériových systémov umožňuje:

- vyrovnávanie výroby a spotreby elektriny v čase,
- znižovanie zaťaženia siete počas špičiek,
- integráciu OZE (obnoviteľných zdrojov energie),
- poskytovanie systémových služieb pre prevádzkovateľov distribučných a prenosových sústav.

Hlavné funkcie ESS v európskom kontexte možno rozdeliť do týchto kategórií [4], [5]:

1. Systémové služby (Grid Services):
   - regulácia frekvencie,
   - podpora napätia,
   - black start, obnovenie napájania po výpadku,
   - odloženie investícií do rozšírenia siete.
2. Vyrovnávanie zaťaženia a znižovanie špičiek (Load Shifting & Peak Shaving):
   - ukladanie energie počas nízkej spotreby a využitie počas špičiek,
   - zníženie potrebného pripojeného výkonu.
3. Integrácia OZE (Renewable Integration):
   - kompenzácia nestability fotovoltických a veterných zdrojov,
   - minimalizácia strát výroby,
   - maximalizácia vlastnej spotreby vyrobenej energie z OZE.
4. Tarifná optimalizácia (Energy Arbitrage):

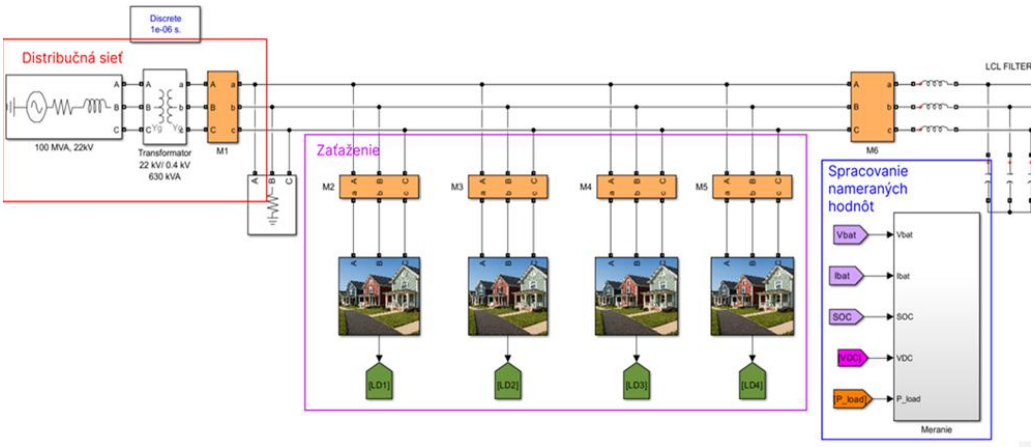
- nákup lacnej elektriny počas nízkych taríf a jej využitie alebo predaj počas vysokých taríf.

5. Záložné napájanie a odolnosť (Backup & Resilience):
   - zabezpečenie neprerušovaného napájania pre kritické odberné miesta,
   - autonómna prevádzka (off-grid) pri výpadku centrálnej siete.

Systém na ukladanie energie je multifunkčné zariadenie. Každé ESS, pri vhodnej konfigurácii, dizajne a riadiacom systéme, je schopné plniť široké spektrum funkcií, ktoré môžu patriť do jednej alebo viacerých vyššie uvedených kategórií.

III. ARCHITEKTÚRA MODELU

Vyvinutý model v prostredí Simulink (Simscape Power Systems) predstavuje fragment inteligentnej elektrickej siete (Smart Grid) s dynamicky sa meniacim zaťažením a integrovaným systémom ukladania energie. Hlavným cieľom vytvorenia tohto modelu je demonštrácia interakcie medzi rôznymi prvkami siete, najmä medzi zdrojom napájania, spotrebičmi a batériovým systémom, v podmienkach meniaceho sa denného zaťaženia. To umožňuje analyzovať efektívnosť využitia systému BESS na vyvažovanie energie v sústave.



Obr. 2 Schematická štruktúra zdroja napájania, transformátora, zaťaženia a bloku zberu údajov v modeli realizovanom v prostredí Simulink

Model pozostáva z týchto hlavných podsystémov:

- **Distribučná sieť**

Zdroj napájania simuluje trojfázový systém s napätím 22 kV a výkonom 100 MVA. Napätie je znížené pomocou výkonového transformátora na úroveň 0,4 kV, čo je štandardné nízke napätie pre domácich spotrebiteľov.

- **Zaťaženie**

Na výstup sekundárneho vinutia výkonového transformátora sú pripojené štyri bloky trojfázového zaťaženia (LD1–LD4), pričom každý z nich simuluje skupinu štyroch obytných domov. Celkovo ide o 16 domov. Zaťaženie je dynamické, pretože jeho veľkosť sa mení v čase a napodobňuje typickú dennú krivku spotreby elektrickej energie. To vytvára podmienky na využitie systému ukladania energie počas špičky.

- **Systém ukladania energie (BESS)**

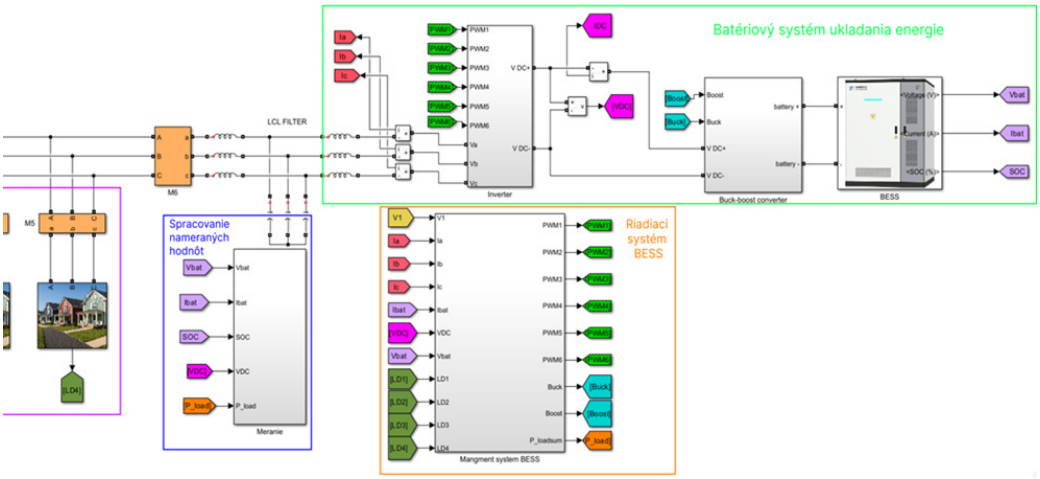
Batériový systém je pripojený paralelne k zaťaženiu. Skladá sa z akumulátora (DC zdroj), buck-boost meniča na nabíjanie a vybíjanie, striedača (invertora) na konverziu na striedavý prúd a LCL filtra na zníženie harmonických zložiek. Jeho hlavnou úlohou je dodávať energiu do siete počas špičkového zaťaženia a nabíjať sa v čase nízkeho dopytu.

- **Riadiaci systém BESS**

Je implementovaný ako samostatný logický blok (vo forme MATLAB Function a logických schém). Tento systém zabezpečuje monitorovanie stavu batérie (SoC, napätie, prúd), analýzu aktuálnej úrovne zaťaženia a prepínanie prevádzkových režimov: nabíjanie, vybíjanie alebo pohotovostný režim. Okrem toho riadi správnu činnosť striedača a buck-boost meniča. Plní funkciu adaptívneho regulátora na základe vstupných údajov v reálnom čase.

- **Merací a zberný systém**

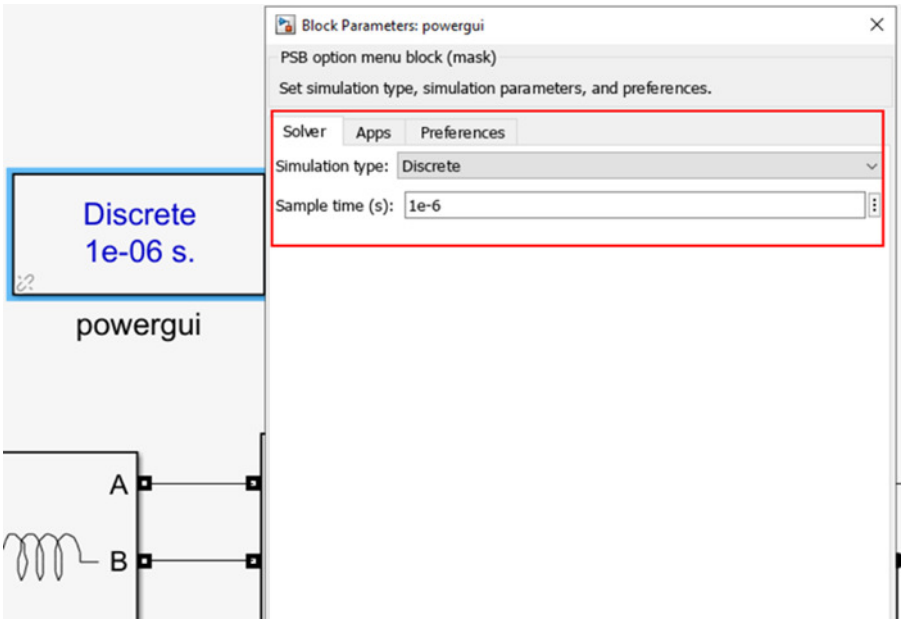
Obsahuje bloky na meranie vybraných parametrov: aktívneho výkonu, prúdu, napätia a stavu nabitia batérie (SoC). Údaje sa odosielajú do vizualizačného bloku a ukladajú sa na ďalšiu analýzu vo formáte *.mat.



Obr. 3 Schéma batériového systému ukladania energie (BESS) s riadiacimi prvkami a striedačom v modeli realizovanom v prostredí Simulink

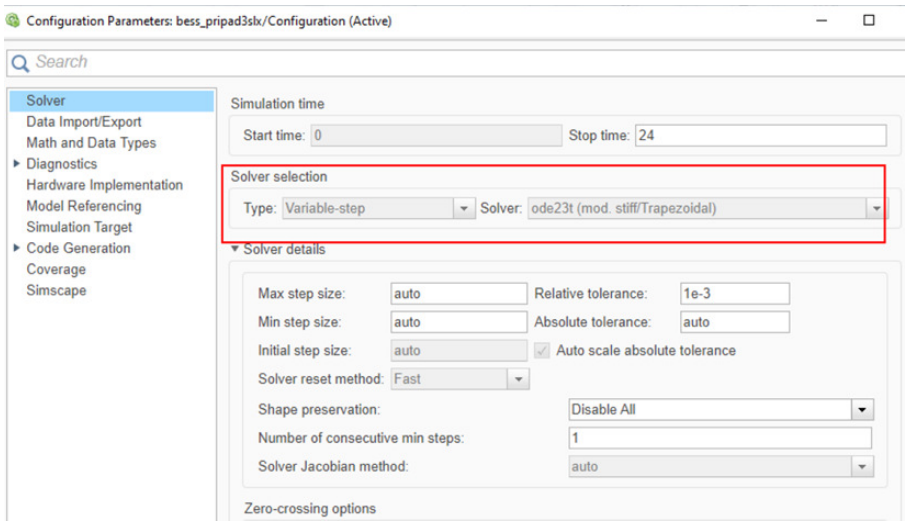
Každý z komponentov v modeli interaguje s ostatnými v rámci jednotného systému. Na zabezpečenie správnej činnosti všetkých blokov a presného modelovania bol v bloku Powergui nastavený typ simulácie na „Discrete“ a hodnota vzorkovacej periódy (Sample time) na 1e–6 s (obr. 4).

Toto nastavenie umožňuje vykonávať výpočty s vysokou presnosťou, čo je mimoriadne dôležité pri riadení energetických procesov v krátkych časových intervaloch.



Obr. 4 Nastavenie bloku Powergui

Parameter „Stop time“ bol nastavený na hodnotu 24, čo simuluje celodennú prevádzku systému. Okrem toho bol v okne „Configuration Parameters“ ako metóda riešenia diferenciálnych rovníc zvolený algoritmus „ode23t“ (modifikovaná tuhá/trapézová metóda) (obr. 5). Takáto konfigurácia umožňuje dosiahnuť stabilitu modelu aj pri zložitých prechodových javoch a zároveň presne vyhodnocovať vplyv zmeny zaťaženia a parametrov systému na celkovú stabilitu, energetické straty a účinnosť prevádzky akumulátorového systému.



Obr. 5 Nastavenie simulačných parametrov v okne Configuration Parameters

IV. ZÁVER

Cieľom tohto článku bolo analyzovať možnosti dynamického riadenia záťaže v nízkonapäťovej distribučnej sieti s využitím systému akumulácie energie typu BESS. Vytvorený simulačný model v prostredí MATLAB/Simulink umožnil sledovať interakcie medzi zdrojmi napájania, odberateľmi a akumulačným systémom v podmienkach premenlivého zaťaženia počas celého dňa.

Simulácie preukázali, že implementácia BESS vedie k výraznému zníženiu špičkového zaťaženia siete, stabilizácii napätia a vyrovnávaniu rozdielov medzi okamžitou výrobou a spotrebou energie. V prípade zimného režimu s vysokou záťažou, vrátane nabíjania elektromobilu, bol systém schopný udržať parametre napájania v súlade s normou STN EN 50160. Riadiaci algoritmus navyše efektívne reagoval na rýchle zmeny zaťaženia prepínaním medzi režimami nabíjania, vybíjania a pohotovostným režimom.

Získané výsledky potvrdzujú, že využitie batériových úložísk so zodpovedajúcim riadiacim systémom predstavuje efektívny nástroj na zabezpečenie kvality a stability napájania v inteligentných sieťach. V budúcnosti odporúčame rozšíriť model o viacero paralelných BESS modulov, zahrnúť predikciu zaťaženia pomocou metód umelej inteligencie a implementovať tarifné riadenie zohľadňujúce aktuálne trhové ceny elektrickej energie.

POĎAKOVANIE

Tento príspevok vznikol vďaka podpore Vedeckej grantovej agentúre Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporeného grantom VEGA 1/0532/25 a VEGA 1/0627/24.

ZOZNAM POUŽITEJ LITERATÚRY

[1] Deutsche energie-agentur (dena), „*A comparative Analysis and Simulation of DSM and Energy Efficiency in Chinese and German Industry*", [online]. Berlin: dena, 2022. [cit. 2025-04-23].Dostupné na internete: <https://www.dena.de/fileadmin/dena/Dokumente/Projektportrait/Projektarchiv/Entrans/Analysis_DSM_and_Energy_Efficiency_in_Chinese_and_German_Industry_EN.pdf>.

[2] Next Kraftwerke GMBH, „*Regelenergie & Regelleistung - was ist das?*" [online]. Köln: Next Kraftwerke, 2024. [cit. 2025-04-23]. Dostupné na internete: <https://www.next-kraftwerke.de/wissen/regelenergie>.

[3] International energy agency, „*Demand response*" [online]. Paris: IEA, 2024. [cit. 2025-04-19]. Dostupné na internete: <https://www.iea.org/energy-system/energy-efficiency-and-demand/demand-response>.

[4] European Commission, „*Energy storage*", [online]. Brussels: European Commission, 2025. [cit. 2025-05-04]. Dostupné na internete: <https://energy.ec.europa.eu/topics/research-and-technology/energy-storage_en>.

[5] ATP journal, „*Batérie nie sú jediný systém na uskladnenie energie*", [online]. Bratislava: ATP Journal, 2024. [cit. 2025-05-04]. Dostupné na internete: <https://www.atpjournal.sk/rubriky/prehladove-clanky/baterie-nie-su-jediny-system-na-uskladnenie-energie.html?page_id=41279>.

[6] Y. Hryhoriev, „*Dynamické riadenie záťaže v inteligentných sieťach*", Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025.

Optimalizácia ukladania elektrickej energie pre off-grid systém rodinného domu

Karina SAVENKO, Dušan MEDVEĎ

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

karina.savenko@student.tuke.sk, dusan.medved@tuke.sk

***Abstrakt* — Článok sa zaoberá návrhom a optimalizáciou systému skladovania elektrickej energie pre ostrovný (off-grid) fotovoltický systém určený pre rodinný dom v meste Košice. Cieľom bolo porovnať technicko-ekonomické parametre dvoch typov batériových technológií, olovených a lítium-iónových akumulátorov, s dôrazom na životnosť, hlboké vybitie (DoD), počiatočné investičné náklady (CAPEX), celkové náklady počas životnosti (NPC) a cenu vyrobenej energie (LCOE). Simulácie boli realizované v softvéri HOMER Pro pri rôznych úrovniach dennej spotreby energie a za rôznych klimatických podmienok. Výsledky ukazujú, že lítium-iónové batérie sú výhodnejšie pre domácnosti so strednou až vyššou spotrebou, zatiaľ čo olovené batérie môžu byť ekonomickejšie pri nízkej spotrebe. Článok prináša odporúčania pre výber optimálneho riešenia pre domácnosti mimo dosahu distribučnej siete.**

***Kľúčové slová* — Off-grid systém, skladovanie energie, HOMER Pro, akumulácia energie**

I. Úvod

V súčasnom svete je elektrická energia jedným z najdôležitejších faktorov zabezpečujúcich komfortný život. Jej dodávka k spotrebiteľom sa realizuje prostredníctvom energetického systému, ktorý zahŕňa prvky, ako sú distribučné a prenosové siete.

Avšak v mnohých menej rozvinutých krajinách, ako aj vo vzdialených oblastiach, ktoré nie sú pripojené k centralizovaným sieťam, vznikajú problémy so spoľahlivým zásobovaním elektrickou energiou. Na riešenie tohto problému boli vyvinuté autonómne energetické systémy, ktoré premieňajú obnoviteľnú energiu z okolia na elektrickú energiu a ukladajú ju do rôznych typov úložísk na neskoršie použitie. Podľa klimatických podmienok a špecifík regiónu môžu mať tieto systémy rôzne konfigurácie.

V posledných rokoch získavajú systémy akumulácie a skladovania elektrickej energie v praxi čoraz väčšiu popularitu. Rýchle rozširovanie obnoviteľných zdrojov energie sa stalo katalyzátorom rozvoja technológií na ukladanie energie. To umožňuje významné zmeny v štruktúre energetických systémov, znižuje využívanie fosílnych zdrojov a zmierňuje negatívny vplyv na životné prostredie.

Medzi najrozšírenejšie typy systémov akumulácie energie na svete patria vodné akumulujúce elektrárne, ktoré tvoria približne 96 % celkovej kapacity. Zároveň však rastie význam iných typov akumulátorov, najmä batériových systémov, ktoré môžu byť inštalované priamo pri spotrebiteľovi alebo na podstaniciach. Tieto systémy predstavujú flexibilný prvok v riadení parametrov elektrickej siete.

II. Analýza rôznych typov úložísk energie

Elektrická energia je najvhodnejšou formou energie na premenu na iné druhy energie. Základným princípom skladovania energie je jej transformácia na iné formy. Keď je spotreba elektrickej energie nižšia ako dodávka, je potrebné ju uskladniť. Pod pojmom skladovanie elektrickej energie rozumieme premenu elektrickej energie na iný druh energie a jej využitie v čase, keď je spotreba elektrickej energie vyššia ako jej dodávka.

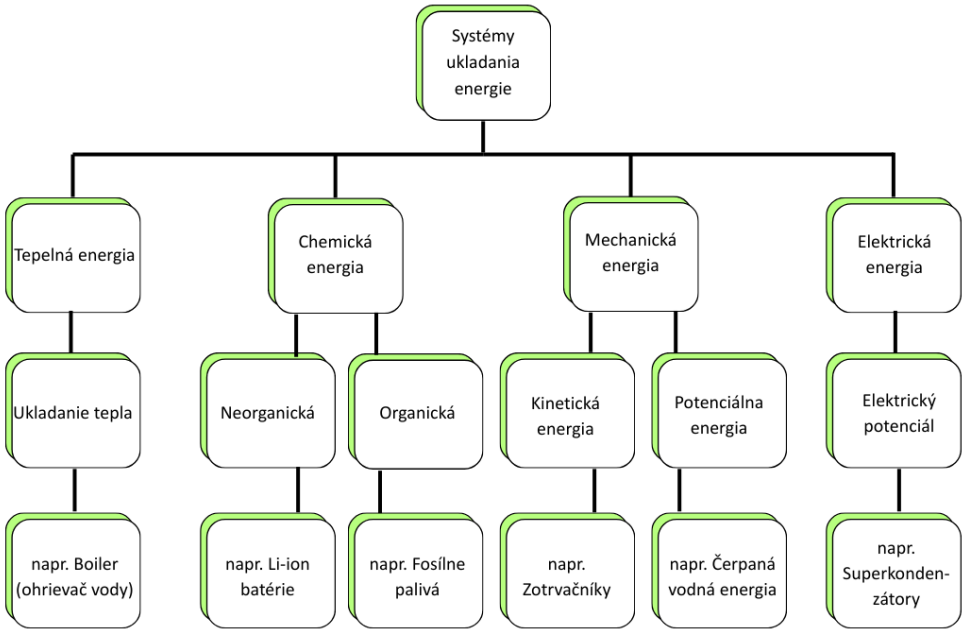
Celkovo môžeme energiu meniť na päť základných foriem:

- elektrochemickú;
- mechanickú;

- elektrickú;
- chemickú;
- tepelnú.

V závislosti od formy energie existujú zodpovedajúce metódy jej ukladania. K najrozšírenejším metódam skladovania energie patria (obr. 1) [5]:

- prečerpávacia vodná nádrž,
- tepelná premena,
- systémy akumulácie energie so stlačeným vzduchom,
- prietokové batérie (flow batérie) na akumuláciu energie,
- chemické skladovanie energie,
- zotrvačník,
- vysokoteplotné supravodivé (HTS) magnety,
- skladovanie energie v superkondenzátoroch.



Obr. 1 Klasifikácia systémov na uchovávanie energie podľa typu energie(vrátane príkladov)

Mechanická premena

Pri mechanickom skladovaní energie dochádza k premene elektrickej energie na potenciálnu alebo kinetickú energiu. Pri tomto type skladovania sa elektrická energia transformuje na formu skladovateľného energetického toku (nie elektriny), aby sa neskôr mohla opäť použiť na výrobu elektriny. Prebytočná elektrická energia z obnoviteľných zdrojov sa najčastejšie využíva na pohon mechanického systému.

Prečerpávacia vodná nádrž

Keď je v systéme prebytok elektrickej energie, využíva sa na čerpanie vody zo spodnej nádrže do hornej nádrže, takže sa energia akumuluje vo forme potenciálnej energie vody. Keď je potrebné vyrábať elektrickú energiu, voda sa vypúšťa z nádrže cez turbínu, ktorá je prepojená s generátorom, takže dochádza k výrobe elektrickej energie. V tomto procese sa potenciálna energia premieňa na mechanickú. Hlavnými výhodami tohto systému sú jeho relatívna jednoduchosť z prevádzkového hľadiska, nízke nároky na zložité technológie a účinnosť, ktorá dosahuje až 85 %. Hlavnými nevýhodami sú vysoké investičné náklady a potreba vhodných geografických podmienok.

Zotrvačník

Základom tohto princípu je premena energie na kinetickú. Prebytočná elektrická energia sa využíva na pohon motora, ktorý roztočí zotrvačník niekoľko-tisíckrát otáčok za minútu. Zotrvačník predstavuje rotujúcu hmotu, ktorá je držaná pomocou magnetických ložísk a uzavretá v kryte, aby sa znížili straty spôsobené vysokými otáčkami. Vďaka svojej schopnosti rýchlo reagovať na zmeny v sústave sa využíva na reguláciu frekvencie. K hlavným výhodám zotrvačníka patria aj jednoduchosť technickej údržby, vysoká ekologickosť a životnosť (približne 10 rokov).

Systémy akumulácie energie so stlačeným vzduchom

Princíp fungovania je založený na využití potenciálnej energie získanej stláčaním plynu. Plyn

je uložený pod tlakom v nepriepustnej a objemnej dutine, kde sa môže uchovávať dlhší čas. Keď je potrebné dodať energiu, plyn sa uvoľní z úložiska a nasmeruje na vzduchovú turbínu, kde sa jej kinetická energia premení na mechanickú energiu.

Elektrochemická premena

Princíp spočíva v premene chemickej energie, ktorú majú aktívne materiály, na elektrickú energiu pomocou elektrochemických reakcií. Pre tento typ skladovania elektrickej energie sa najčastejšie využívajú akumulátory. V súčasnosti existujú rôzne typy batérií s kapacitou od niekoľkých wattov po tisíce kilowattov. Vďaka svojmu širokému spektru výroby sú veľmi dostupné.

Olovená batéria

Hlavné výhody olovených batérií sú ich dobrá spoľahlivosť, nízka cena, dostupnosť olova, dlhý životný cyklus a elektrochemická účinnosť. Vďaka týmto vlastnostiam majú tieto batérie široké spektrum použitia. V závislosti od aplikácie sa vytvárajú rôzne konštrukcie batérií. Batéria sa skladá z hlavných častí, ako sú elektródy, separátory, ventilácia a iné.

Elektródy pozostávajú z mriežok a aktívnej hmoty. Mriežka kladnej elektródy musí byť odolná voči korózii, pretože korózia znižuje mechanickú pevnosť a elektrickú vodivosť. Na výrobu mriežok sa používajú zliatiny Pb-Ca alebo Pb-Sb s prídavkami, ako sú Sn, Cd a Se, ktoré sa tiež používajú na zlepšenie koróznej odolnosti.

Na výrobu aktívneho materiálu sa využíva oxid olovnatý (PbO), ktorý sa mení na červenohnedý oxid olovnatý (PbO_2) na kladnej elektróde a na šedé vodivé olovo (Pb) na zápornej elektróde.

Na oddelenie záporných a kladných elektród sa používajú separátory. Úlohou separátora je zabezpečenie elektrickej izolácie medzi zápornou a kladnou doskou, udržanie aktívnych materiálov v tesnej blízkosti s mriežkou, udržanie platní v predpísanej polohe (pôsobí ako podpora) a zabránenie skratom. Separátor môže byť vyrobený z drevenej dyhy, syntetických materiálov, gumy, polyetylénu, skla alebo celulózy. Elektrolyt predstavuje vodný roztok H_2SO_4, ktorý má hustotu 1,22–1,28 g/cm³ [1].

Li-ion batéria

V 70. rokoch 20. storočia sa začal vývoj lítiovo-iónových batérií s vysokou hustotou energie. Tieto batérie sú pri nabíjaní bezpečnejšie, pretože neobsahujú kovové lítium. Vo všeobecnosti sa lítiové batérie skladajú z elektród a organického elektrolytu. Na výrobu katódy sa používa lítium-metal oxid ($LiCoO_2$), zatiaľ čo anóda je grafitovo-uhlíková elektróda. Elektrolyt môže byť buď nevodný roztok organického rozpúšťadla s rozpustnou lítiovou soľou, alebo pevný polymér.

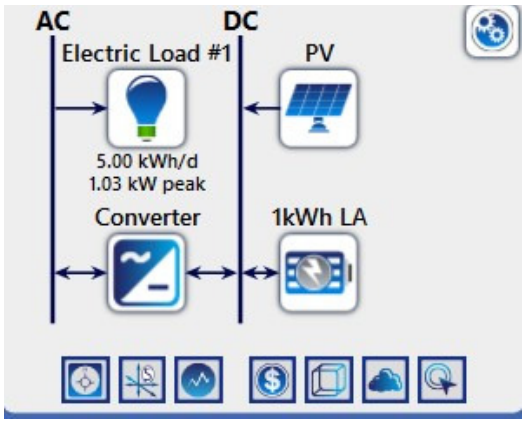
Aby sa zabránilo nadmernému zahrievaniu a taveniu tradičnej anódy, bol zavedený uhlíkový článok ako anóda (Li_xC_6), pretože lítiový karbón má vyšší bod tavenia (180 °C). Počas nabíjania sú lítiové ióny uvoľňované z oxidu lítneho a absorbované uhlíkovou anódou (LiC_6). Počas vybíjania prebieha opačný proces [2].

Tepelná premena

Princíp tepelnej premeny spočíva v tom, že počas prebytku elektrickej energie v energetickom systéme sa teplo transportuje do akumulačného média. Keď je potrebné vyrobiť elektrickú energiu, teplo sa z tohto média odovzdáva a môže byť použité na vykurovanie alebo chladenie budov, prípadne na výrobu elektriny. Výhodami použitia tohto systému sú zvýšenie celkovej účinnosti a lepšia spoľahlivosť. V praxi sa používajú rôzne typy tepelných akumulácií v závislosti od dĺžky skladovania: krátkodobé, strednodobé alebo dlhodobé. Tepelné úložiská (thermal energy storage – TES) sa delia na vysokoteplotné a nízkoteplotné. Za nízkoteplotné TES považujeme také, ktoré fungujú pri teplotách nižších ako 120 °C [3].

III. MODELOVANIE SYSTÉMU PRE SKLADOVANIE ENERGIE

Navrhnutý systém sa skladá z elektrickej záťaže, fotovoltických panelov, meniča a batériového úložiska (obr. 2).



Obr. 2 Schéma zapojenia

Pri riešení úlohy bol jedným z významných faktorov elektrické zaťaženie odberateľov. Ako vstupné údaje o elektrickom zaťažení sa vo väčšine prípadov pri optimalizácii používajú priemerné hodinové grafy elektrického zaťaženia pre každý mesiac v roku. Lepšie je však používať denné alebo skutočne namerané údaje o zmene činného výkonu v hodinových alebo 15 až 30-minútových intervaloch, ktoré poskytne viac informácií o rýchlosti zmeny elektrického zaťaženia a umožní identifikovať špičkové výkyvy. Najskôr je potrebné určiť dennú spotrebu elektrickej energie pre rodinný dom. Pre tento účel bol zvolený v programe typ záťaže „Residential“ a mesiac s vrcholovou záťažou je január, pretože v dôsledku sezónnych zmien sa deň skracuje, čo vedie k vyššej spotrebe elektrickej energie. Program podľa nastaveného regiónu (v tomto prípade Košice) vygeneruje priemerné hodnoty a zobrazí denný profil spotreby energie. Najvyššia spotreba sa zaznamenáva v rozmedzí od 17:00 do 20:00 hodiny (obr. 3). Priemerná celková hodnota spotreby elektrickej energie je 11,25 kWh/deň.

Obr. 3 Denná záťaž v prostredí programu Homer

Pre simuláciu systému bol pomocou programu HOMER Pro navrhnutý model ostrovného (off-grid) systému pre rodinný dom v Košiciach, ktorý využíva akumulátory na ukladanie energie. Charakteristiky porovnania systémov [4]:

- Ako priemerná denná spotreba elektrickej energie ovplyvňuje systém: 5 (kWh/deň), 10 (kWh/deň), 15 (kWh/deň), 20 (kWh/deň),

- Vplyv sezónnych zmien a okolia na systém. Znižovací faktor fotovoltických panelov: pre zimu je približne 70 %, zníženie hodnoty je spôsobené poveternostnými podmienkami: možné zamrznutie panelov, zasneženie, panely nemajú automatizované zariadenie na sledovanie polohy slnka. Znižovací faktor počas teplých období roka (jar, leto, jeseň) sa pohybuje približne na úrovni 85 %.
- Vplyv typu akumulátora na systém: Li-ion a olovené batérie (Lead-Acid).

Tabuľka 1
Výsledky simulácií pre každý z jednotlivých systémov

číslo simulácie	Typ batérie	Znižovací faktor (%)	Denná spotreba (kWh/deň)	CAPEX (euro)	NPC (euro)	LCOE (euro)	Počet batérií (ks)
1	Lead-Acid	85	11,25	40134	46344	1,15	55
2		70	11,25	41204	46488	1,16	60
3		85	5	10688	21817	0,926	35
4			10	20041	41426	0,879	77
5			15	43088	66774	0,944	115
6			20	57263	85052	0,902	155
1	Li-Ion	85	11,25	34596	46601	0,878	51
2		70	11,25	38535	51925	0,979	51
3		85	5	11257	19089	0,81	20
4			10	21250	42861	0,909	30
5			15	53183	70331	0,994	50
6			20	72160	93243	0,989	65

Aby systém mohol zabezpečiť energiu pre dom počas nepriaznivých poveternostných podmienok bolo potrebné zvýšiť počet akumulátorov na uchovanie energie a jej využitie počas dní autonómie. Pri porovnaní dvoch typov akumulátorov je potrebné na zabezpečenie rodinného domu rovnakým množstvom energie použiť viac olovených akumulátorov ako lítium-iónových. Toto súvisí s hĺbkou vybitia (DoD). Na rozdiel od olovených akumulátorov majú lítium-iónové akumulátory výhodu, pretože ich hĺbka vybitia je približne 80 %, zatiaľ čo u olovených je len okolo 50 %. Zníženie znižovacieho faktora vedie k zvýšeniu potrebného počtu batérií, čo následne zvyšuje celkové kapitálové a prevádzkové náklady.

Významnú úlohu v konečných výpočtoch zohráva aj denná spotreba energie: vyššia spotreba zlepšuje návratnosť systému vďaka úspore z rozsahu, takže sa znižuje celková cena elektrickej energie (LCOE). Oloveno-kyselinové batérie preukázali nižšie počiatočné náklady v niektorých scenároch, najmä pri nižšej spotrebe (5–10 kWh/deň). Avšak ich nižšia prípustná hĺbka vybitia (DoD), kratšia životnosť a potreba väčšieho počtu kusov (35–155 batérií) viedli k zvýšeným prevádzkovým nákladom. V najlepšom prípade dosiahli LCOE hodnotu 0,879 €/kWh, ale bolo potrebné nainštalovať až 77 akumulátorov, čo výrazne ovplyvňuje údržbu, priestorové nároky a spoľahlivosť systému.

Lítium-iónové batérie, hoci majú v niektorých prípadoch vyššie vstupné náklady, ponúkajú vyššiu účinnosť, väčšiu hĺbku vybitia (až 85 %), dlhšiu životnosť a menší počet potrebných jednotiek (od 20 do 65 batérií). Najlepší výsledok bol dosiahnutý pri dennej spotrebe 5 kWh, kde NPC dosiahla iba 19089 €, a LCOE bola na úrovni 0,81 €/kWh, čo je najnižšia hodnota zo všetkých simulovaných variantov. V tomto prípade bolo potrebných len 20 batérií, čo výrazne zjednodušuje údržbu a znižuje riziko porúch. V rozsahu dennej spotreby 5–10 kWh/deň majú Li-Ion batérie výrazne nižšie hodnoty LCOE v porovnaní s batériami typu Lead-Acid. Najnižšia hodnota LCOE pre Li-Ion je dosiahnutá pri spotrebe 5 kWh/deň (približne 0,81 €/kWh) [5].

IV. ZÁVER

V článku bol predstavený simulačný návrh systému akumulácie energie pre off-grid domácnosť s cieľom identifikovať technicky a ekonomicky optimálny typ batériového úložiska. Využitím softvéru HOMER Pro boli porovnané dve hlavné technológie batérií, olovené (Lead-Acid) a lítium-iónové (Li-Ion), v rôznych scenároch spotreby energie a klimatických podmienkach typických pre región Košíc.

Simulácie ukázali, že pri nízkej spotrebe (5 až 10 kWh/deň) môžu byť olovené batérie výhodné

vďaka nižším vstupným nákladom, avšak ich kratšia životnosť, menšia hĺbka vybitia a väčšie priestorové nároky znižujú ich celkovú efektivitu. Naopak, lítium-iónové batérie vykázali výrazne lepšie parametre LCOE a NPC pri vyššej spotrebe, pričom umožňujú efektívnejšie využitie uložených energetických zásob vďaka vyššej DoD a nižšiemu počtu potrebných kusov.

Z praktického hľadiska sa ukázalo, že lítium-iónové batérie predstavujú vhodnejšiu voľbu pre domácnosti s vyššími nárokmi na spoľahlivosť, priestorové obmedzenia a údržbu, najmä ak je prioritou dlhodobá ekonomická návratnosť. Výsledky článku potvrdzujú význam modelovania v procese návrhu hybridných systémov a ukazujú, že výber typu batérie musí vychádzať nielen z investičných nákladov, ale aj z prevádzkových a environmentálnych aspektov.

POĎAKOVANIE

Tento príspevok vznikol vďaka podpore Vedeckej grantovej agentúre Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporeného grantom VEGA 1/0532/25 a VEGA 1/0627/24.

ZOZNAM POUŽITEJ LITERATÚRY

[1] D. M. Vance, „*Development of a Methodology for Sizing Photovoltaics and Energy Storage for Islanded Communities*“. Purdue University, 2018. ProQuest Dissertations & Theses, No. 30505614. [online] [cit. 13.05.2025]. Dostupné na: https://www.proquest.com/docview/2838331160

[2] Bluebird Solar, „*Rozdiel medzi on-grid a off-grid solárnymi systémami*“, [online] [cit. 13.05.2025]. Dostupné na: https://bluebirdsolar.com/blogs/all/differencebetween-on-grid-and-off-grid-solar-system

[3] N. M. Kumar, et. al., „*On-Grid Solar Photovoltaic System: Components, Design Considerations, and Case Study*“, IEEE Xplore, 2018. [online] [cit. 14.05.2025]. Dostupné na: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8442403

[4] M. A. Eltawil, Z. Zhao, „*Grid-connected photovoltaic power systems: Technical and potential problems-A review*“, Renewable and Sustainable Energy Reviews, 2010, č. 14, s. 112-129. ISSN 1364-0321.

[5] K. Savenko, „*Optimalizácia energetického skladovania pre off-grid systémy*“, Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025.

Návrh zariadenia pre výrobu vodíka II

[1]*Vladimír SZOMOSI,* [2]*Dušan MEDVEĎ*

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

[1]vladimir.szomosi@outlook.sk, [2]dusan.medved@tuke.sk

***Abstrakt* — Tento článok sa zaoberá návrhom a konštrukciou elektrolyzéra schopného efektívnej výroby vodíka elektrochemickým procesom, konkrétnejšie finálnou fázou sústreďujúcou sa na laboratórne merania a vyhodnotenie výsledkov. Nadväzuje na článok Návrh zariadenia pre výrobu vodíka a článok, ktorý sa zaoberal možnosťami využitia elektrolyzéra pre elektroenergetické účely, s cieľom dosiahnuť optimálne podmienky pre elektrolýzu. Víziou je vytvorenie elektrolyzéra, ktorý by mohol slúžiť na získavanie vodíka určeného na akumuláciu nezužitkovanej elektrickej energie.**

***Kľúčové slová* — vodík, elektrolýza, kyslík, elektrolyzér, elektródy, nikel, KOH, nulová medzera**

I. Úvod

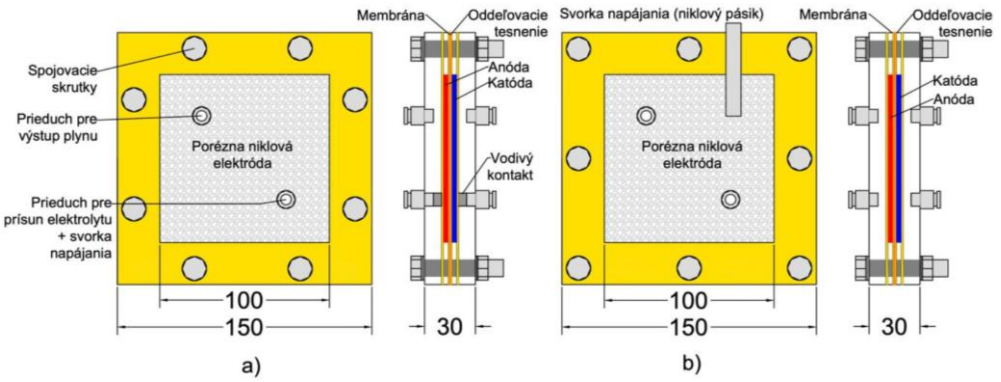
Prezentovaný článok sa opiera o výskum zaoberajúci sa výrobou alkalického elektrolyzéra slúžiaceho na výrobu Brownovho plynu (HHO) a predchádzajúci výskum zameraný na počiatočnú fázu návrhu a konštrukcie alkalického elektrolyzéra konštruovaného metódou nulovej medzery. Z hľadiska praktického použitia a bezpečnosti pri manipulácií je zaujímavejšie vyprodukovať čistý vodík (HHO plyn je nestabilný a ťažko využiteľný v praktickej aplikácií) priamo elektrochemickou premenou. V tejto finálnej praktickej fáze je cieľom odskúšanie funkčného skonštruovaného modelu na základe predloženého návrhu a následne po úspešných skúškach realizácia laboratórnych meraní s cieľom zistenia vybraných parametrov a stanovenia účinnosti zariadenia [5][6].

II. Elektrolýza technológiou nulovej medzery

Spojením výhod PEM (nulová medzera medzi membránou a elektródami) a alkalickej elektrolýzy (lacné materiály) možno dosiahnuť vysokú účinnosť aj pri nižších konštrukčných nákladoch na MEA (Membrane Electrode Assembly).

Technológia alkalického elektrolyzéra konštruovaného metódou nulovej medzery je založená na porozite elektród (anódy a katódy), ktorými bude obložená membrána prepúšťajúca hydroxidové ióny. Vzniknutá medzera medzi elektródami je o hrúbke membrány a zníži tak odpor elektrolytu. Redukcia vodíka následne prebieha vo vnútri poréznej elektródy a núti plynové bublinky vznikajúce v poréznej štruktúre vystupovať skrze prestupy elektrolyzéra [2][5].

Elektrolyzér produkujúci vodík je navrhnutý na základe predchádzajúceho výskumu. Ide o suché vyhotovenie s jedným elektródovým zväzkom MEA. Elektródy sú oddelené membránou ZIFRON PERL UTP 500, určenou pre alkalickú elektrolýzu, ktorú poskytla spoločnosť Agfa-Gevaert. Konštrukcia z priehľadného akrylátu umožňuje vizuálne sledovať tvorbu plynových bubliniek na elektródach. Elektrolyzér produkuje vodík (katóda) a kyslík (anóda), ktoré sú odvádzané skrze bočné portály [1][2][3][5][6].



Obr. 1 (a) Finálna verzia funkčného elektrolyzéra, (b) Prototyp elektrolyzéra [4]

III. ELEKTROLYZÉR

Problémom zisteným pri skúške prototypu bol únik skúšobnej látky pred samotným zavedením elektrolytu v dôsledku nesprávneho rozloženia väzobných skrutiek, čo bolo spôsobené nedostatočným utesnením konštrukcie. Riešením bola úprava rozloženia skrutiek podľa vzoru HHO elektrolyzéra. Ako neefektívne sa ukázalo aj pôvodné riešenie napájania elektród pomocou niklového pásika, ktorý taktiež spôsobil únik skúšobnej látky. To predstavovalo vážny problém, keďže nespoľahlivé napájanie znemožňuje prevádzku. Nové riešenie využíva rýchlospojky z pochrómovanej ocele, ktoré pôvodne slúžili na prívod elektrolytu a odvod plynu. Ich závitová časť bola predĺžená, čím sa sprostredkovalo spoľahlivé vodivé spojenie s elektródami. Toto riešenie sa pri skúške osvedčilo pri minimálnom prechodovom odpore. Na Obr. 1 je prezentované porovnanie verzií elektrolyzérov. Obr. 1 (a) znázorňuje finálnu verziu, (b) znázorňuje nedokonalý prototyp. Viditeľná je aj zmena rozloženia skrutiek [4].

IV. FUNKČNÉ SKÚŠKY, UVEDENIE DO PREVÁDZKY A POSÚDENIE PRODUKOVANÉHO VODÍKA

Cieľom laboratórneho merania je overiť správnu funkciu skonštruovaného elektrolyzéra, a to najmä porovnaním reálnej produkcie vodíka s vypočítanými teoretickými hodnotami. Skúška bola realizovaná jednoduchým testom výstupných plynov vznietením cudzím zdrojom, pri ktorom sa sledujú akustické a vizuálne prejavy charakteristické pre vodík. Tento kvalitatívny test síce nepotvrdí presnú čistotu plynu, ale poskytne rýchlu indíciu, či skutočne ide o vodík. Skúška taktiež napomôže identifikovať možné konštrukčné nedostatky alebo riziká, ktoré by mohli ovplyvniť dlhodobú prevádzku elektrolyzéra. Pre kvantitatívne vyhodnotenie bude elektrolyzér napájaný jednosmerným laboratórnym zdrojom s vopred definovanými hodnotami napätia a prúdu. Odmeraná bude aj aktuálna produkcia vodíka v litroch za minútu, pričom sa očakáva, že zvyšovanie napájacieho prúdu bude viesť k úmerne vyššej produkcii vodíka, ako aj k zvyšovaniu teploty elektrolytu. Pri zvýšení napätia zdroja síce dochádza k nárastu produkcie vodíka, no zároveň sa znižuje účinnosť celého procesu v dôsledku tepelných strát. Takéto merania umožnia presnejšie vyhodnotiť celkovú efektivitu pri rôznych pracovných podmienkach [4].

A. Vyhodnotenie skúšky

V počiatočnej fáze prípravy elektrolytu bolo nutné dbať na bezpečnosť, keďže vznikalo veľké množstvo tepla. Použitý bol odolný polypropylénový kontajner a chemicky odolné rukavice. Po zavedení elektrolytu do systému sa objavili drobné netesnosti v okolí skrutiek a závitov, ktoré sa vyriešili dotiahnutím a použitím gumových O-krúžkov. Počas skúšky bola pozorovaná zvýšená produkcia kyslíka, čo viedlo k podozreniu na chybný kontakt medzi rýchlospojkou a elektródou. V mieste sa potvrdil prechodový odpor, ktorý spôsoboval prehrievanie a degradáciu niklovej peny. Riešením bolo privarenie predlženej závitovej časti a zavedenie niklovej pásky na rozšírenie kontaktnej plochy s elektródou. Po opätovnom zložení sa produkcia stabilizovala. Plyny boli napokon testované zapálením – vodík vykázal očakávaný zvukový prejav, kyslík nevykazoval prejav žiaden, čo potvrdilo správnu funkciu separačnej membrány [4].

B. Vzťahy použité pri jednotlivých meraniach

1) Výpočet relatívneho rozdielu medzi vypočítaným a odmeraným objemom

$$\Delta_{rel} = \frac{|V_{odmer.} - V_{vyp.}|}{\frac{V_{odmer.} + V_{vyp.}}{2}} \cdot 100 \ [\%] \tag{1}$$

kde $V_{\mathrm{odmer.}}$ je odmeraný objem vygenerovaného vodíka
$V_{\mathrm{vyp.}}$ je výpočet predpokladaného objemu vygenerovaného vodíka.

2) Výpočet spotreby elektrickej energie

$$E = U_{\mathrm{z}} \cdot I_{\mathrm{p}} \cdot t \ [\mathrm{J}] \tag{2}$$

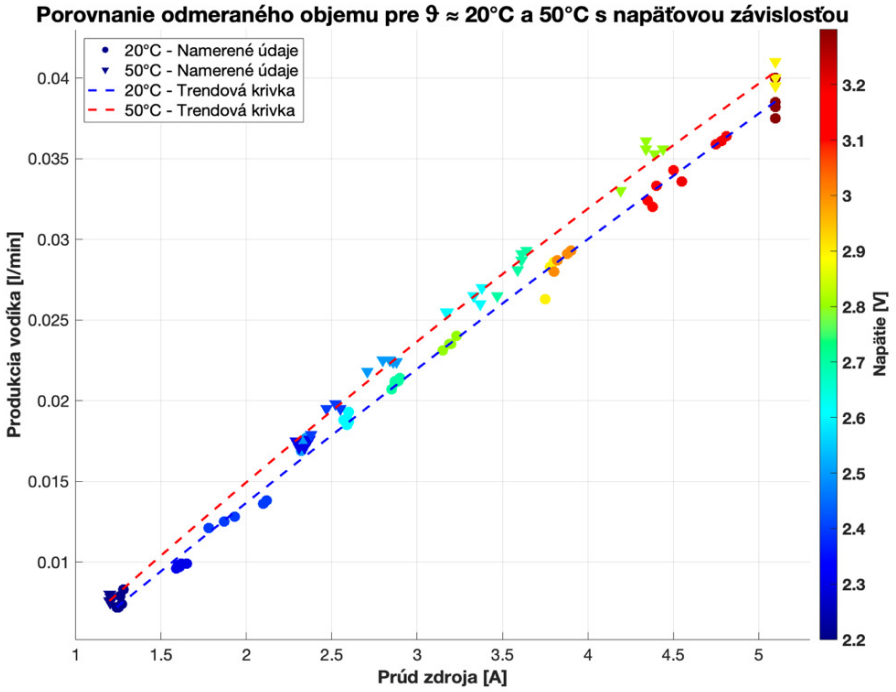
kde U_{z} je napätie dodané do redoxového procesu,
I_{p} je priemerný prúd odoberaný zo zdroja,
t je časový úsek merania produkcie vodíka.

3) Výpočet teoretickej účinnosti zariadenia pre výrobu vodíka

$$\eta = \frac{n \cdot \Delta_{\mathrm{R}} H \cdot V}{U \cdot I \cdot t} \cdot 100 \ [\%] \tag{3}$$

kde n je počet molov vodíka ≈ 0,0416 mol,
$\Delta_R H$ je reakčná entalpia pre elektrolýzu vody 286 kJ/mol,
V je odmerané množstvo vygenerovaného objemu vodíka,
U je napätie dodané do procesu elektrolýzy,
I je priemerný prúd odoberaný zo zdroja pri elektrochemickom procese,
t je časový úsek merania produkcie vodíka 60 s.

C. *Grafické závislosti, meranie č.1*



Obr. 2 Porovnanie odmeranej produkcie vodíka pre $\vartheta \approx 20\,°C$ a $50\,°C$ [4]

D. *Vybrané tabuľky meraní produkcie vodíka a vypočítanej teoretickej účinnosti zariadenia*

Tabuľka 1
Meranie produkcie vodíka pri $\vartheta \approx 20\,°C$ a $U_z = 2{,}5\,V$

č.m	I_p [A]	$V_{odmer.}$ [l/min]	$V_{vyp.}$[l/min]	Spotreba E [J]	Δ_{rel} [%]
1.	2,33	0,0171	0,0174	204,4	1,89
2.	2,32	0,0169	0,0174	205,9	2,64
3.	2,34	0,0174	0,0175	201,7	0,58
4.	2,33	0,0171	0,0174	204,4	1,89
5.	2,35	0,0177	0,0176	199,2	0,71

Tabuľka 2
Meranie produkcie vodíka pri $\vartheta \approx 50\,°C$ a $U_z = 2{,}4\,V$

č.m	I_p [A]	$V_{odmer.}$ [l/min]	$V_{vyp.}$[l/min]	Spotreba E [J]	Δ_{rel} [%]
1.	2,38	0,0179	0,0196	191,5	9,18
2.	2,52	0,0198	0,0208	183,3	4,81
3.	2,47	0,0195	0,0204	182,4	4,33
4.	2,47	0,0195	0,0204	182,4	4,33
5.	2,55	0,0195	0,0210	188,3	7,52

Tabuľka 3
Výpočet teoretickej účinnosti zariadenia bez použitia núteného obehu $\vartheta \approx 20\,°C$

č.m	I_p [A]	U_z[V]	$V_{odmer.priemer}$ [l/min]	η [%]
1.	1,26	2,2	0,0076	54,37
2.	1,614	2,3	0,0098	52,34
3.	1,96	2,4	0,013	54,63
4.	2,334	2,5	0,0172	58,59

Tabuľka 4
Výpočet teoretickej účinnosti zariadenia bez použitia núteného obehu $\vartheta \approx 50\,°C$

č.m	I_p [A]	U_z[V]	$V_{odmer.priemer}$ [l/min]	η [%]
1.	1,202	2,2	0,0078	58,34
2.	2,338	2,3	0,0174	62,86
3.	2,487	2,4	0,0192	64,15

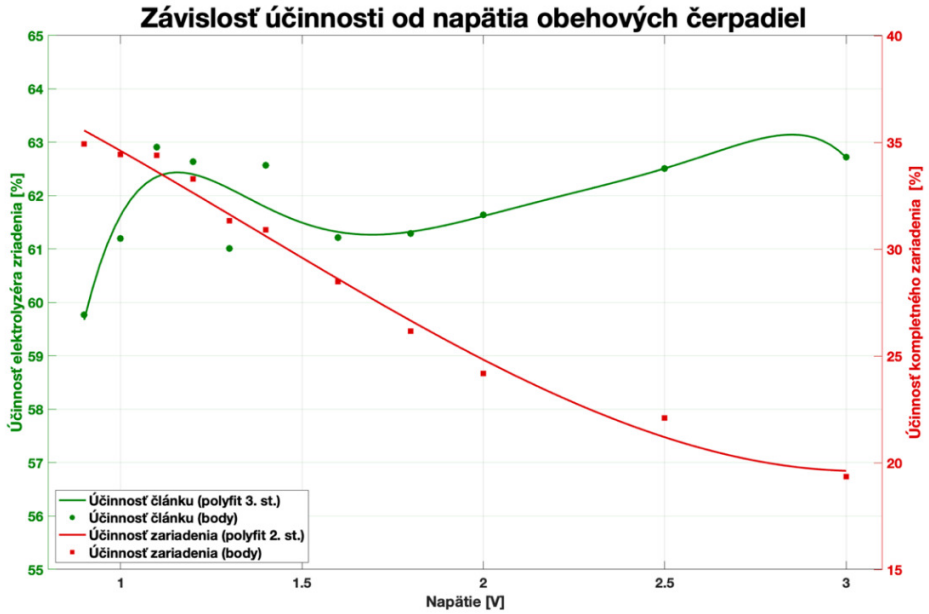
E. Vyhodnotenie a záver laboratórneho merania č.1

Prvé meranie bolo zamerané na zistenie reálnej produkcie vodíka a vyhodnotenie účinnosti elektrochemického procesu pri rôznych prevádzkových teplotách. Začiatok elektrolytickej reakcie bol pozorovaný pri napätí približne 1,49 V, čo potvrdzuje výpočtom získané termoneutrálne napätie 1,48 V pre alkalickú elektrolýzu. Produkcia vodíka rástla so zvyšujúcim sa napätím, avšak účinnosť samotného procesu s rastúcim napätím klesala, čo ukázalo, že táto cesta nie je energeticky výhodná. Porovnaním studeného a ohriateho elektrolytu sa potvrdilo, že predhriatie elektrolytu významne zvyšuje účinnosť zariadenia. Pri napätí 2,5 V bola pri studenom elektrolyte dosiahnutá účinnosť 58,59 %, zatiaľ čo pri ohriatom elektrolyte už pri napätí 2,4 V účinnosť vzrástla na 64,15 %, pričom rozdiel medzi teoretickým výpočtom a reálnou produkciou vodíka predstavoval len 7,34 % pri studenom a 5,77 % pri ohriatom elektrolyte. Pre porovnanie, starší HHO generátor vykazoval odchýlku až 45,33 %, čo potvrdzuje vyššiu kvalitu a presnosť meraní v aktuálnom systéme. Ohriaty elektrolyt umožnil dosiahnuť maximálny prúdový limit zdroja už pri 2,9 V oproti 3,3 V pri studenom, čím sa znížila celková spotreba energie na jednotku produkovaného vodíka. Tieto výsledky poukazujú na fakt, že najefektívnejším spôsobom zvyšovania produkcie je optimalizácia prevádzkových podmienok – najmä teploty a koncentrácie elektrolytu – nie samotné zvyšovanie napätia. Zariadenie je tak schopné produkovať čistý vodík s vysokou účinnosťou, pričom celý systém je možné škálovať pridaním ďalších MEA [4].

V. Experimentálne posúdenie vplyvu núteného obehu na produkciu vodíka

Zavedenie núteného obehu elektrolytu pomocou vákuových čerpadiel môže zlepšiť účinnosť a stabilitu alkalickej elektrolýzy. Prietok elektrolytu znižuje odpor článku, čím klesá potrebné napätie pri rovnakom odberanom prúde, zlepšuje taktiež iónovú vodivosť membrány a urýchľuje odvod plynových bublín z povrchu elektród, čím sa zvyšuje efektivita elektrochemickej reakcie. Príliš vysoký prietok však môže viesť k odplavovaniu OH^- iónov a vyššej spotrebe elektrolytu. Meranie preto skúma, či optimálny prietok znižuje spotrebu energie a zvyšuje produkciu vodíka, pričom hodnotí vplyv na celkovú energetickú bilanciu systému a efektivitu stabilizácie elektrolyzéra. Výsledky zároveň umožnia posúdiť, či sa prínos núteného obehu vyrovná dodatočnej spotrebe energie čerpadiel [4].

A. Grafické závislosti, meranie č.2



Obr. 3 Závislosť účinnosti od napätia obehových čerpadiel [4]

B. Vybrané tabuľky meraní posúdenia vplyvu núteného obehu na produkciu vodíka

Tabuľka 5
Meranie produkcie vodíka pri $U_z = 2{,}5\ V$, nútený obeh $U_{č.} = 1{,}1\ V, I_{č.} \approx 0{,}63\ A$

č.m	I [A]	$V_{\text{odmer.}}$ [l/min]	$V_{\text{vyp.}}$ [l/min]	Δ_{rel} [%]	$E_{\text{čl.}}$ [J]	$E_{\text{če.}}$ [J]	E [J]
1.	2,33	0,0181	0,0174	3,79	193	22,97	216,1
2.		0,0192		10,69	182,0	21,66	203,7
3.		0,019		8,64	183,9	21,88	205,83
4.		0,018		3,24	194,2	23,1	217,3
5.		0,0181		3,79	193,1	22,97	216,1

Tabuľka 1
Výpočet teoretickej účinnosti zariadenia s použitím núteného obehu $\vartheta \approx 20\ °C$

č. Tab	I [A]	V_p [l/min]	$\eta_{čl.}$[%]	η [%]	E_p [J]
25.	2,317	0,01746	59,77	34,93	217,48
26.	2,325	0,01794	61,2	34,44	217,74
27.	2,33	0,01848	62,91	34,39	211,41
28.	2,33	0,0184	62,64	33,29	215,54
29.	2,335	0,01796	61,01	31,33	223,6
30.	2,335	0,01842	62,57	30,91	222,12
31.	2,34	0,01806	61,22	28,47	231,4
32.	2,35	0,01816	61,29	26,17	240,7
33.	2,355	0,0183	61,64	24,18	248,1
34.	2,36	0,0186	62,51	22,09	269,3
35.	2,36	0,01866	62,72	19,36	297,96

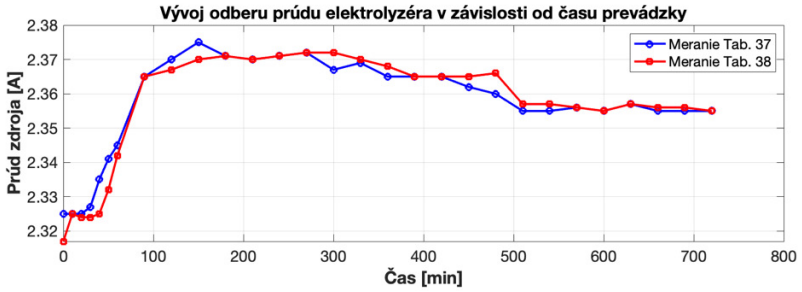
C. Vyhodnotenie a záver laboratórneho merania č.2

Po zavedení núteného obehu elektrolytu boli vykonané merania pri napätí $U_z = 2{,}5$ V, pričom produkcia vodíka bola sledovaná pri jedenástich rôznych napätiach na obehových čerpadlách (0,9 – 3 V). Pre každé napätie bolo zaznamenaných 5 meraní (spolu 55 meraní). Teplota elektrolytu bola stabilná (20 ± 0,2 °C) a zariadenie nevykazovalo žiadne netesnosti. Výsledky potvrdili stabilizáciu odoberaného prúdu a zníženie fluktuácií vďaka cirkulácii elektrolytu. Rýchlejší odvod plynových bublín viedol k nárastu prúdovej zaťažiteľnosti článku, ale aj k zvýšeniu spotreby energie, najmä pri vyšších prietokoch. Najvyššiu účinnosť (34,39 %) dosiahlo zariadenie pri $U_č = 1{,}1$ V a $I_č \approx 0{,}63$ A. V porovnaní s prevádzkou bez čerpadiel bola účinnosť článku vyššia o 3,18 %, avšak celková účinnosť zariadenia klesla priemerne o 29,54 % kvôli spotrebe čerpadiel. Z energetického hľadiska sa preto zavedenie núteného obehu javí ako neefektívne. Napriek tomu má cirkulácia elektrolytu prevádzkové výhody – znižuje kolísanie hladiny, zabezpečuje plynulý vývin vodíka a znižuje výskyt nárazovej produkcie plynu [4].

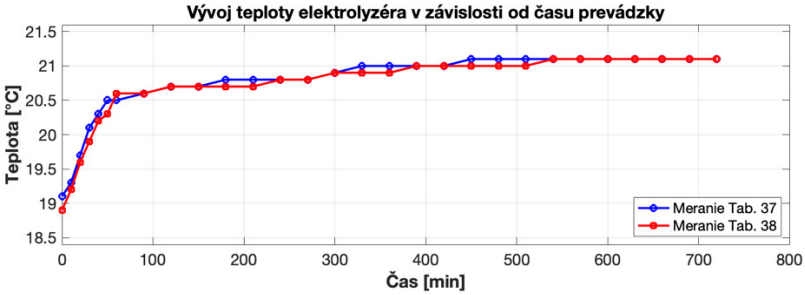
VI. Posúdenie vlastností a produkcie zariadenia v závislosti od času prevádzky

Cieľom merania je analyzovať vplyv dlhodobej prevádzky elektrolyzéra na jeho výkonové parametre, najmä na produkciu vodíka, prúdovú zaťažiteľnosť a účinnosť. Na rozdiel od HHO generátora, kde sa v dôsledku lineárne rastúcej teploty vplyvom prevádzky zvyšovala produkcia plynu, je cieľom tohto merania pozorovať správanie zariadenia s poréznymi niklovými elektródami pri zavedenom nútenom obehu. Vzhľadom na pomalšie zahrievanie článku vplyvom teplotnej bilancie a odlišné fyzikálne vlastnosti použitých materiálov sa predpokladá stabilnejší prúdový odber bez výrazného teplotného nárastu. Meraním je sledovaný priebeh produkcie vodíka, vývoj prúdu a teploty počas predĺženej prevádzky pri optimálnych podmienkach stanovených v predchádzajúcich kapitolách ($U_z = 2{,}5$ V, $U_č = 1{,}1$ V). Cieľom je posúdiť dlhodobú stabilitu zariadenia a identifikovať možné odchýlky v jeho výkone, ktoré by mohli ovplyvniť jeho účinnosť pri reálnej aplikácii [4].

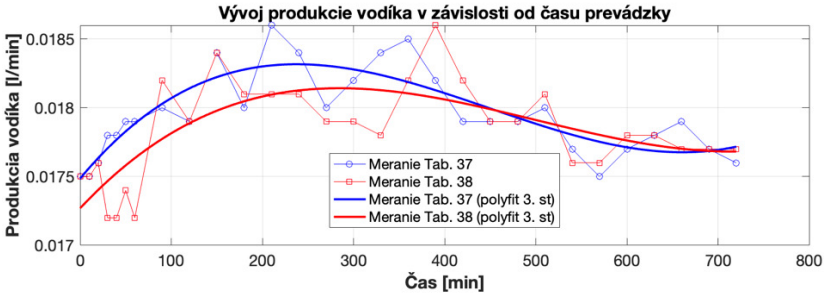
A. Grafické závislosti, meranie č.3



Obr. 4 Vývoj odberu prúdu elektrolyzéra v závislosti od času prevádzky [4]



Obr. 5 Vývoj teploty elektrolyzéra v závislosti od času prevádzky [4]



Obr. 6 Vývoj produkcie vodíka v závislosti od času prevádzky [4]

B. Vyhodnotenie a záver laboratórneho merania č.3

Cieľom merania bolo simulovať reálnu prevádzku zariadenia s núteným obehom a sledovať vývoj jeho technických parametrov, predovšetkým teploty elektrolytu, prúdu a produkcie vodíka. Pôvodne plánovaná jednodňová prevádzka bola predčasne ukončená kvôli dosiahnutiu stabilného teplotného režimu, následne bola meraná opakovane s rovnakými podmienkami. Zariadenie pracovalo pri optimálnom napätí $U_z = 2{,}5$ V, pričom boli zaznamenané merania v 30-minútových intervaloch. Priemerná účinnosť dosiahla 33,09 % pri prvom meraní a 32,88 % pri druhom, čo potvrdzuje stabilnú produkciu vodíka aj po dlhšom čase. Počas prvej hodiny stúpla teplota elektrolytu približne o 1,4 °C na konečných 21,1 °C a následne sa ustálila. Mierna oxidácia bola pozorovaná len na anóde, bez ďalšieho progresu a bez negatívneho vplyvu na produkciu vodíka. Prúd kopíroval teplotný vývoj a po približne dvoch hodinách prešiel do stacionárneho režimu s ustálenou hodnotou 2,355 A. Rastúca teplota spôsobila zvýšenie vodivosti elektrolytu z pôvodných ≈ 7,61 S·m^2/mol na ≈ 8,4 S·m^2/mol, čím sa znížila spotreba energie na jednotku vyprodukovaného plynu. Vývoj produkcie vodíka (Obr. 6) zodpovedal vývoju prúdu (Obr. 4) a potvrdil dynamickú stabilitu systému. Nútený obeh spolu s poréznou štruktúrou elektród prispeli k stabilizácii teploty a chladiacemu efektu. Z hľadiska technickej spôsobilosti preukázal elektrolyzér schopnosť dlhodobej prevádzky bez výrazných zmien či poškodenia [4].

VII. ZÁVER

Cieľom tohto článku bolo prezentovať návrh pre zostrojenie funkčného alkalického elektrolyzéra s nulovou medzerou pre výrobu čistého vodíka. Zariadenie preukázalo stabilnú prevádzku, primeranú účinnosť a vhodnosť na ďalší výskum. Merania potvrdili závislosť účinnosti od teplotného koeficientu elektrolytu a prúdového zaťaženia. Zavedenie núteného obehu zabezpečilo stabilný chod, aj keď s poklesom celkovej energetickej účinnosti. Zariadenie je technicky spôsobilé na dlhodobejšiu prevádzku.

POĎAKOVANIE

Tento príspevok vznikol vďaka podpore Vedeckej grantovej agentúre Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporeného grantom VEGA 1/0532/25 a VEGA 1/0627/24.

REFERENCIE

[1] Agfa-Gevaert NV Zirfon PERL UTP 500. In [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] R. W. Phillips, C. Dunnill, „*Zero gap alkaline electrolysis cell design for renewable energy storage as hydrogen gas*“. In RSC Advances [online]. 2016. Vol. 6, no. 102, s. 100643–100651. [cit. 2025-05-17]. Dostupné na internete: <https://pubs.rsc.org/en/content/articlelanding/2016/ra/c6ra22242k>.

[3] B. Sudarmanta, et al., „*Application of dray cell HHO gas generator with pulse width modulation on sinjai spark ignition engine performance*“. In: IJRET: International Journal of Research in Engineering and Technology. 2016. .ISSN 2319-1163.

[4] V. Szomosi, „*Návrh zariadenia pre výrobu vodíka*“, Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025.

[5] V. Szomosi, D. Medveď, „*Možnosti využitia elektrolyzéra pre elektroenergetické účely*“. In Electrical Engineering and Informatics [online]. Faculty of Electrical Engineering and Informatics Technical University of Košice Letná 9, 040 01 Košice, Slovak Republics. 486. ISBN 978-80-553-4407-2. Dostupné na internete: <https://eei.fei.tuke.sk/wp-content/uploads/2024/03/EEI_14.pdf>.

[6] V. Szomosi, D. Medveď, „*Návrh zariadenia pre výrobu vodíka*“. In Electrical Engineering and Informatics [online]. Faculty of Electrical Engineering and Informatics Technical University of Košice Letná 9, 040 01 Košice, Slovak Republics. 524. ISBN 978-80-553-4682-3. Dostupné na internete: <https://eei.fei.tuke.sk/wp-content/uploads/2024/07/EEI_15.pdf>.

Úloha prosumera vo flexibilnom riadení siete a jeho vplyv na energetickú bezpečnosť

Artem VOZDRAHANOV, Dušan MEDVEĎ

Katedra elektroenergetiky, Fakulta elektrotechniky a informatiky,
Technická univerzita Košice, Slovenská republika

artem.vozdrahanov@student.tuke.sk, dusan.medved@tuke.sk

***Abstract* — Predkladaný článok sa zaoberá prosumermi, t.j. spotrebiteľmi, ktorí sú schopní nielen odoberať energiu zo siete, ale ju v prípade potreby aj dodávať späť, v kontexte ich vplyvu na bezpečnosť a flexibilitu siete. Článok prezentuje výsledky simulácie malej časti distribučnej siete, v ktorej sú pripojení nielen bežní spotrebitelia, ale aj prosumeri vybavení batériami. Pomocou týchto batérií, ktoré sa nabíjajú v noci počas zníženého zaťaženia siete a dodávajú energiu počas špičkového odberu, môžu prosumeri ovplyvniť flexibilitu a stabilitu siete. Rozvoj tejto technológie má veľký potenciál a môže výrazne ovplyvniť zaužívané spôsoby riadenia distribučnej siete.**

***Kľúčové slová* — prosumer, batériové úložisko, Simulink, flexibilita, bezpečnosť**

I. Úvod

V súčasnosti sa čoraz väčší dôraz kladie na prechod k decentralizovaným a nízkouhlíkovým energetickým systémom. S rastúcim podielom obnoviteľných zdrojov energie sa zároveň zvyšuje potreba flexibility v riadení elektrizačnej sústavy. Jedným z dôležitých prvkov tejto transformácie je rozvoj aktívnych účastníkov trhu s elektrinou, tzv. prosumerov, ktorí dokážu nielen energiu spotrebovávať, ale aj vyrábať a uskladňovať.

Úloha prosumerov v distribučnej sieti naberá na význame najmä v súvislosti s ich schopnosťou ovplyvňovať lokálne toky energie a prispievať k stabilite siete prostredníctvom poskytovania flexibility. Prosumeri môžu aktívne reagovať na požiadavky siete napríklad úpravou svojej spotreby, dodávkou energie z batérií, či znížením odberu v čase špičky. Takéto správanie má potenciál zvýšiť energetickú bezpečnosť a znížiť potrebu externých zásahov do systému.

V tomto článku sme sa zamerali na simuláciu časti distribučnej siete pozostávajúcej zo štyroch domov, z ktorých dva predstavujú klasických pasívnych odberateľov a dva sú prosumeri vybavení batériovým úložiskom. Model bol vytvorený v prostredí MATLAB/Simulink, ktorý umožňuje presné modelovanie elektrických obvodov, dynamických systémov a interakcií medzi komponentmi siete.

II. Charakteristika prosumera a jeho úloha v energetike

Termín prosumer vznikol spojením slov producer (výrobca) a consumer (spotrebiteľ) a označuje účastníka elektrickej siete, ktorý je schopný energiu nielen spotrebovávať, ale aj vyrábať a v prípade potreby uskladňovať. Vďaka rozšíreniu technológií, ako sú fotovoltické panely, batériové úložiská a systémy riadenia spotreby, sa prosumeri stávajú dôležitým článkom modernej energetiky [1].

Prosumeri zohrávajú významnú úlohu v decentralizácii výroby elektrickej energie a prispievajú k znižovaniu strát v distribučných sieťach. Okrem environmentálnych benefitov však poskytujú aj systémové služby, napríklad demand-side response, vyhladzovanie špičiek v zaťažení, či dodávku energie počas výpadkov. Vďaka batériovým systémom môžu prosumeri energiu akumulovať v čase prebytku a neskôr ju spätne dodávať do siete, takže zvyšujú jej flexibilitu [2].

S rastúcim počtom prosumerov však narastá aj potreba ich inteligentného riadenia. Distribučné siete, ktoré boli pôvodne navrhnuté pre jednosmerný tok energie, sa musia prispôsobiť novým podmienkam. Riadenie distribučných systémov v reálnom čase, lokálne rozhodovanie a predikcia správania prosumerov sa stávajú dôležitými oblasťami výskumu a vývoja [3].

Okrem technických aspektov je dôležitý aj vplyv prosumerov na energetickú bezpečnosť.

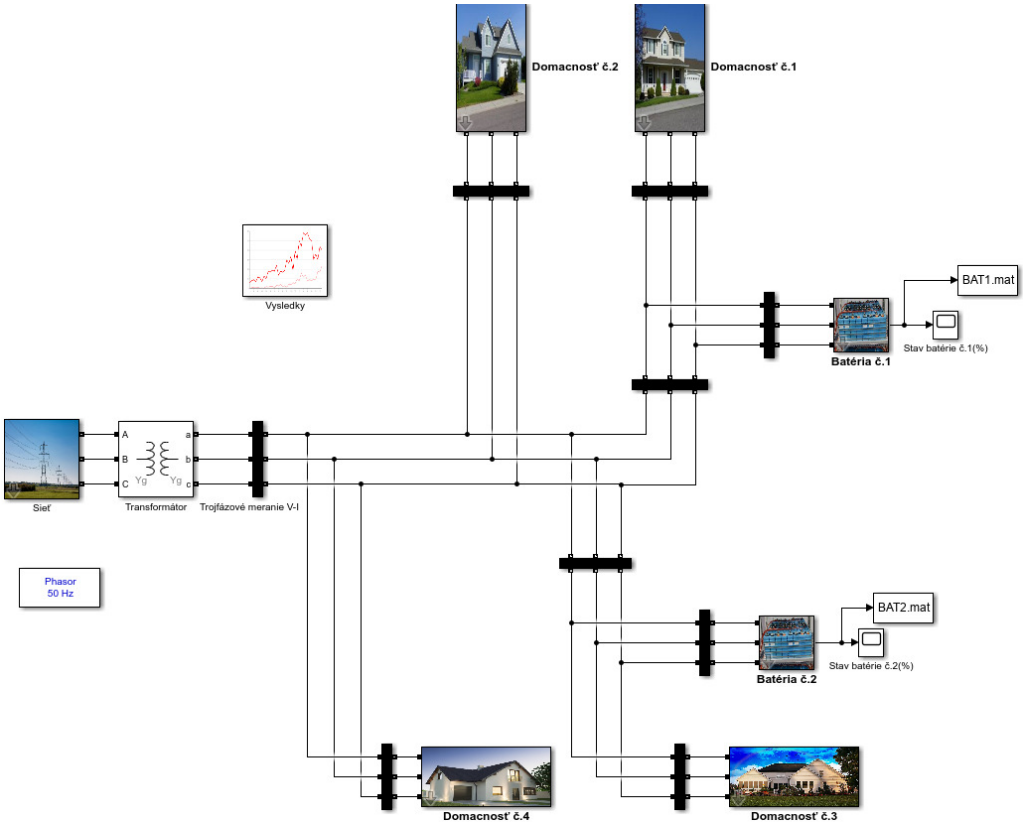
Decentralizovaná výroba znižuje závislosť od centrálnych zdrojov a zahraničných dodávok palív. Príkladom dobrej praxe sú krajiny ako Nemecko alebo Holandsko, kde sa prosumeri podieľajú na poskytovaní rezervných kapacít a integrácii obnoviteľných zdrojov do siete [4].

III. Popis modelu simulácie v prostredí Simulink

Na analýzu vplyvu prosumerov na distribučnú sieť bol vytvorený simulačný model v prostredí MATLAB/Simulink, konkrétne pomocou balíka Simscape Electrical, ktorý umožňuje detailné modelovanie elektrických obvodov s časovo premenlivými prvkami.

V simulačnom prostredí Simulink bol vytvorený model, ktorý reprezentuje úsek distribučnej siete s pripojenými štyrmi domácnosťami ako odberateľmi elektrickej energie. Dve z týchto domácností vystupujú ako štandardní spotrebitelia, ktorých energetické nároky sa menia v priebehu dňa a odrážajú sezónne vplyvy. Zvyšné dve domácnosti fungujú ako prosumeri, teda spotrebujú elektrickú energiu, no zároveň ju môžu v prípade potreby spätne dodávať do distribučnej sústavy.

Celý model pozostáva z komponentov, uvedených na obr. 1.

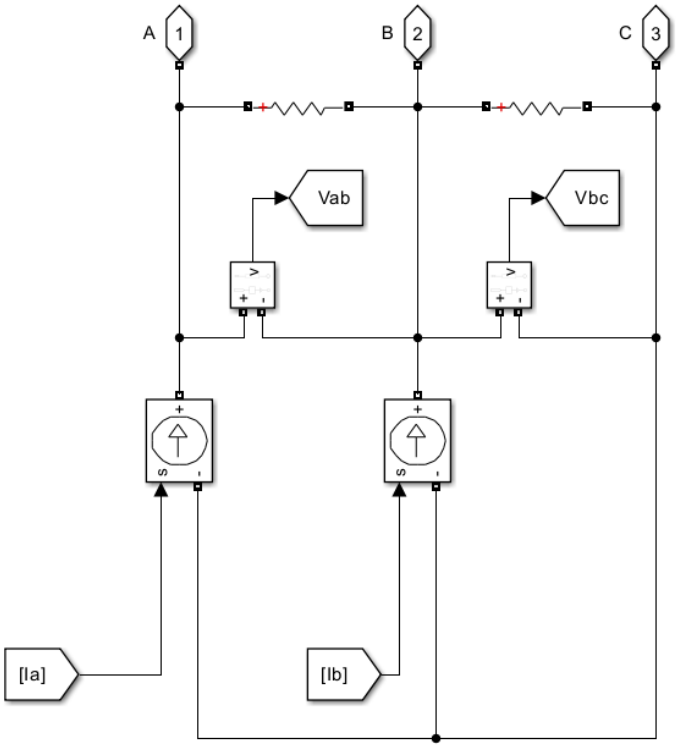


Obr. 1 Schéma modelovanej časti obvodu

Na obr. 2 je znázornené zapojenie dynamickej záťaže v rámci trojfázového simulačného modelu, ktorá predstavuje odber jednotlivých domácností. Ide o model spotreby elektrickej energie, ktorá sa mení v čase, t.j. každých 30 minút počas simulácie, na základe preddefinovaného profilu odberu. Takýto prístup umožňuje presne simulovať správanie domácnosti s rôznorodým dopytom počas dňa.

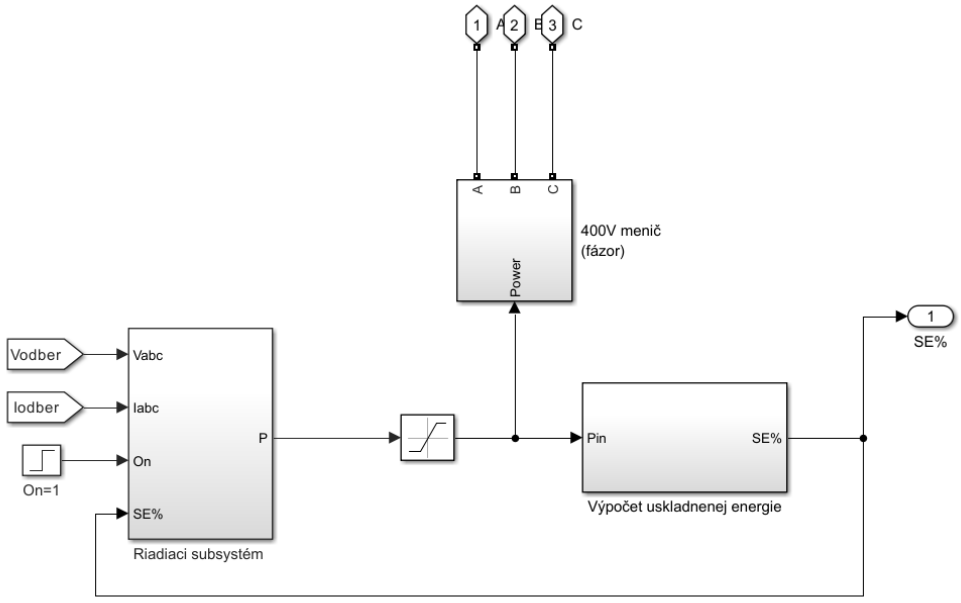
Systém pozostáva z troch fáz (A, B a C), pričom medzi jednotlivými fázami sú umiestnené napäťové meracie bloky, ktoré zaznamenávajú aktuálne napätia medzi vodičmi. Tieto hodnoty napätia slúžia ako vstup do výpočtu potrebného odberového prúdu. Výpočty sú realizované pomocou matematických blokov, ktoré určujú okamžité hodnoty prúdov, generovaných následne prúdovými zdrojmi v jednotlivých fázach.

Výhodou tohto prístupu je, že umožňuje simulovať ľubovoľný časovo závislý profil spotreby. Na rozdiel od pasívnych impedančných záťaží, ktoré reagujú na napätie podľa Ohmového zákona, tento model umožňuje zadávať výkonové profily, ktoré sú následne transformované do hodnoty prúdu v reálnom čase podľa aktuálneho napätia v systéme. Takýto typ modelovania je veľmi vhodný na testovanie reakcie systému na flexibilnú záťaž, ako aj na analýzu možností riadenia a akumulácie energie v reálnych domácnostiach alebo v malých priemyselných prevádzkach.



Obr. 2 Schéma zapojenia dynamickej záťaže

Dôležitou súčasťou simulačného modelu je batériový úložný systém, ktorý slúži na akumuláciu prebytočnej energie v čase nízkej spotreby a jej dodávku v čase špičky. V tomto modeli bol úložný systém implementovaný pomocou 400 V meniča pripojeného k trojfázovej sieti, ako je znázornené na obr. 3.



Obr. 3 Schéma zapojenia bloku batérie

Základ modelu tvorí blok „Riadiaci subsystém“, ktorý získava informácie o okamžitých napätiach a prúdoch vo fázach a vypočítava aktuálny výkon P dodaný alebo odoberaný z batérie. Tento výkon je následne upravený podľa stavu riadiaceho signálu „On“ a podľa aktuálneho stavu nabitia úložiska (SE%).

Z riadiaceho subsystému ide upravený výkon do bloku meniča, ktorý zabezpečuje výmenu energie medzi batériovým systémom a elektrickou sieťou. Fázorový menič 400 V je nastavený

tak, aby umožňoval obojsmerný tok energie v závislosti od riadiaceho vstupu.

Vstupný výkon na vstupe do meniča (Pin) sa prenáša do bloku „Výpočet uskladnenej energie“, ktorý integruje výkon v čase a tým sleduje aktuálnu úroveň stavu nabitia batérie v percentách (SE%). Tento údaj sa spätne posiela do riadiaceho subsystému, ktorý na základe neho rozhoduje, či je možné pokračovať v nabíjaní alebo vybíjaní úložiska.

Logika riadenia je založená na nasledujúcich pravidlách:

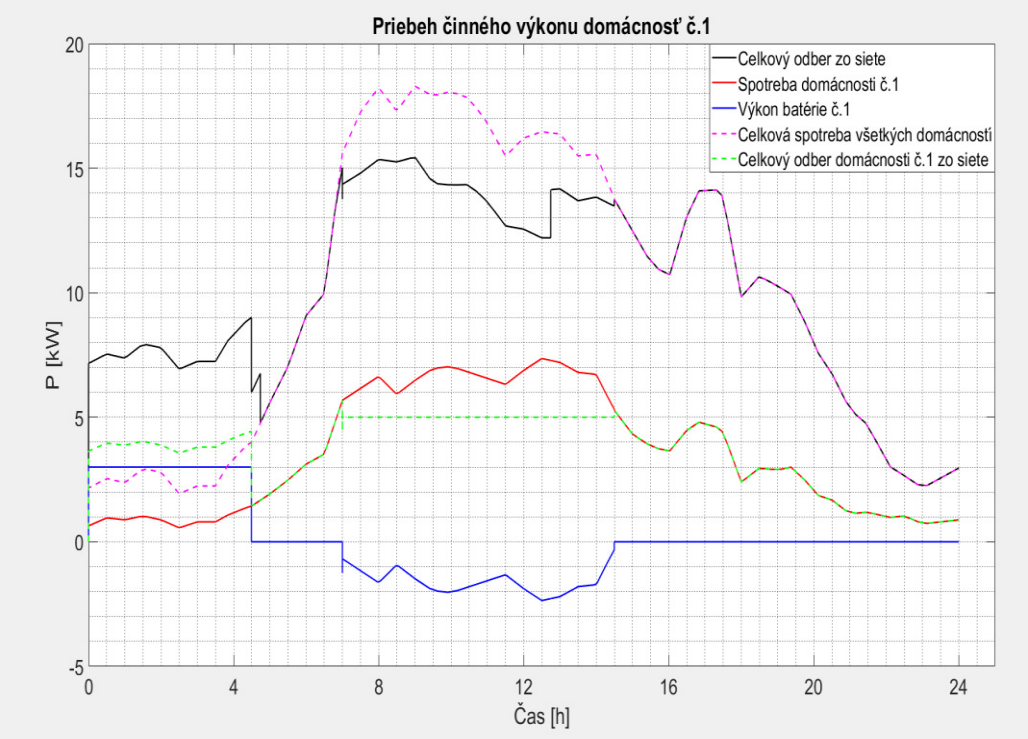
- Nabíjanie je povolené, ak je dostatok prebytočnej energie a stav nabitia je pod 100 %.
- Vybíjanie je aktivované v prípade, že spotreba prevyšuje dostupný výkon zo siete a stav nabitia je vyšší než minimálny prah (20%).

Takto navrhnutý model umožňuje realistické a flexibilné testovanie rôznych scenárov riadenia batérie v domácnosti, ktorá sa správa ako aktívny účastník trhu s elektrinou, teda prosumer.

IV. VÝSLEDKY SIMULÁCIÍ

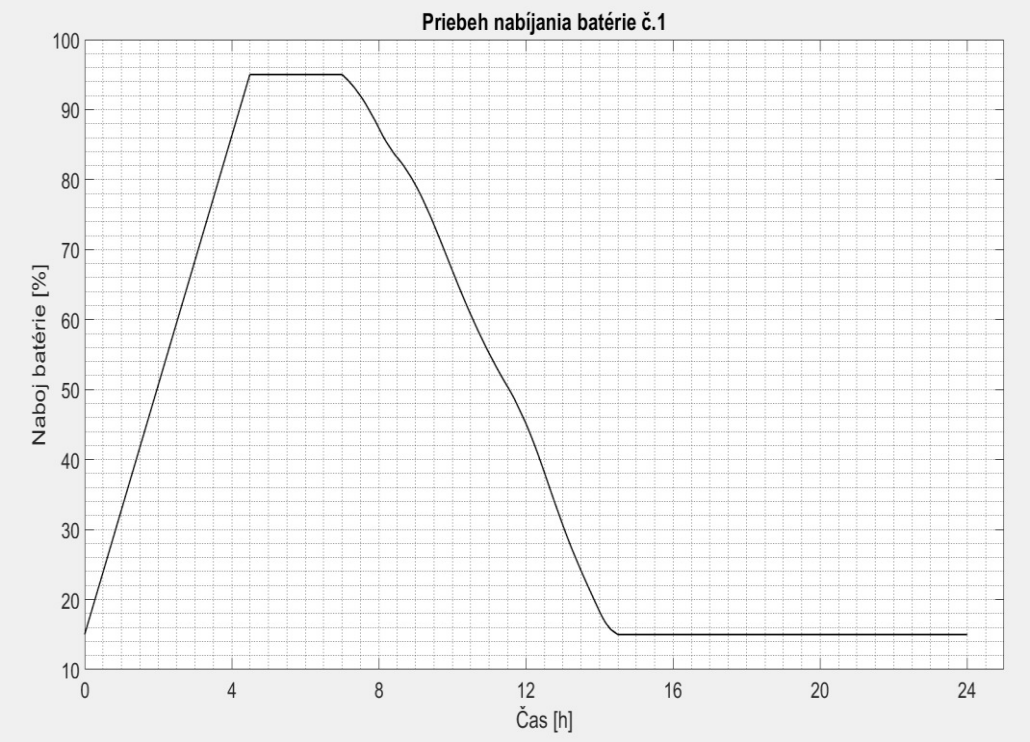
Simulácie boli vykonané pre dva hlavné prevádzkové režimy a to letný profil s nižšou spotrebou energie a zimný profil s vyšším energetickým zaťažením. Každý profil pokrýva obdobie 24 hodín, pričom zaťažovací profil sa mení každých 30 minút podľa preddefinovaného harmonogramu domácnosti.

Na obr. 4 je znázornený priebeh spotreby elektrickej energie počas dňa v zime. Je viditeľné, že najvyššie odbery nastávajú v ranných a popoludňajších hodinách. V nočných hodinách je spotreba minimálna.



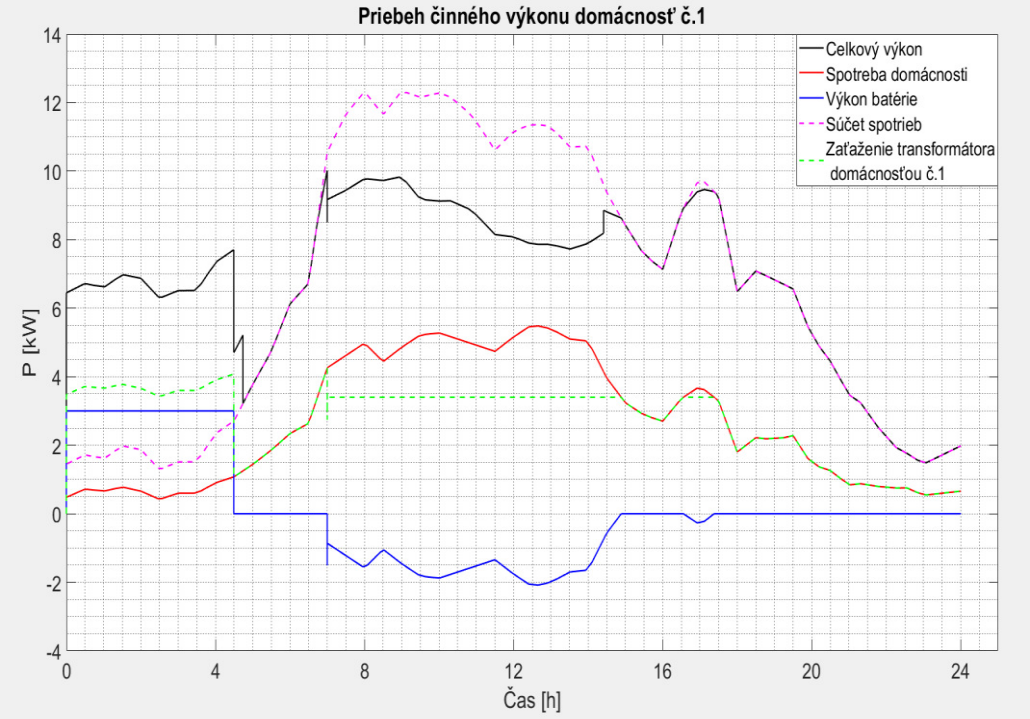
Obr. 4 Priebeh činného výkonu pre domácnosť č.1 počas zimného obdobia

Na obr. 5 je zobrazený stav nabitia batériového úložiska (SoC – State of Charge) v priebehu simulácie. Počas období s nízkou spotrebou sa batéria nabíja, zatiaľ čo počas špičkového zaťaženia dochádza k jej vybíjaniu. Tým sa znižuje maximálny výkon odoberaný zo siete, čo prispieva k rovnomernejšiemu zaťaženiu distribučného systému.



Obr. 5 Zmena stavu nabitia batérie (State of Charge – SoC) počas zimného obdobia

Počas letného obdobia bola celková spotreba domácnosti znížená, keďže nebolo potrebné vykurovanie. Spotreba elektrickej energie bola koncentrovaná najmä v doobedňajších a popoludňajších hodinách (obr. 6). V noci a skoro ráno bola spotreba minimálna.



Obr. 6 Priebeh činného výkonu pre domácnosť č.1 počas letného obdobia

Dynamický model v tomto prípade ukázal, že batéria bola schopná pokryť krátkodobé špičky v spotrebe, čím sa znížil okamžitý odber zo siete. Priebežné nabíjanie batérie prebiehalo v čase minimálneho zaťaženia, a to najmä počas noci a skorého rána, keď bol výkonový profil systému priaznivý. Letný režim potvrdil, že aj pri nízkej základnej spotrebe môže vhodné využitie úložiska prispieť k zníženiu zaťažovacích špičiek a tým zlepšiť stabilitu siete.

V. Záver

Tento článok bol venovaný vplyvu prosumerov na flexibilitu a bezpečnosť distribučnej siete. Simulácie ukázali, že používanie batérií prosumerov pomáha znižovať špičkové zaťaženie transformátora a vyrovnávať denný profil spotreby, čo zlepšuje stabilitu siete. V letnom období batérie dokážu pokryť väčšinu špičkovej záťaže, v zime je odporúčané zvýšiť ich kapacitu.

Prosumeri takisto prispievajú k službám flexibility podľa platnej legislatívy, čím znižujú potrebu spúšťania dodatočných zdrojov a zvyšujú bezpečnosť distribučnej siete. Aktuálne však počet prosumerov a kapacita batérií nie sú v ES dostatočné na výrazný vplyv na celú sieť.

Poďakovanie

Tento príspevok vznikol vďaka podpore Vedeckej grantovej agentúre Ministerstva školstva, vedy, výskumu a športu SR a SAV, podporeného grantom VEGA 1/0532/25 a VEGA 1/0627/24.

Zoznam použitej literatúry

[1] M. R. Atallah, L. A. Galea, „*A generative literature analysis on the rise of prosumers, Renewable and Sustainable Energy Reviews*“, [online], [cit.: 14.5.2025]. Elsevier, 2024. Dostupné: https://www.sciencedirect.com/science/article/pii/S0957178724000924

[2] A. Galván, M. M. Galán, A. Gómez-Expósito, „*Prosumer flexibility as an enabler for ecosystem value co-creation, Applied Energy*“, [online], [cit.: 14.5.2025]. Elsevier, 2025. Dostupné: https://www.sciencedirect.com/science/article/pii/S0306261925005446

[3] Clou Global, „*The Rising Role of Prosumers in Grid Flexibility*“, CLOU Electronics Co., Ltd., 2024. [online], [cit.: 14.5.2025]. Dostupné: https://clouglobal.com/the-rising-role-of-prosumers-in-grid-flexibility

[4] SolarPower Europe, „*Grid Guardians – Prosumers boosting flexibility and cutting energy waste, SolarPower Europe*“, [online], [cit.: 14.5.2025]. Dostupné: https://www.solarpowereurope.org/features/grid-guardians-prosumers-boosting-flexibility-and-cutting-energy-waste

[5] A. Vozdrahanov, „*Úloha prosumera vo flexibilnom riadení siete a jeho vplyv na energetickú bezpečnosť*“, Bakalárska práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2025.

Návrh zariadenia pre zapovanie

[1]*Milan LACKO,* [2]*Viliam JAKUBEC*

[1,2] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzity v Košiciach, Slovenská republika

[1]milan.lacko@tuke.sk, [2]viliam.jakubec@student.tuke.sk

***Abstrakt* — Článok opisuje návrh a konštrukciu zariadenia pre zapovanie, ktoré patrí medzi biorezonančné terapie. Zariadenie generuje nízkonapäťové impulzy rôznej frekvencie pomocou digitálneho signálového procesora (DSP). Zariadenie obsahuje LCD rozhranie, možnosť nastavenia frekvencie v rozsahu od 1 Hz do 1 MHz a používateľskú pamäť na ukladanie frekvenčných procedúr. V článku je opísaný návrh elektronických a riadiacich obvodov, celkový dizajn zariadenia s kompletnou elektrickou schémou zapojenia.**

***Kľúčové slová* — zapovanie, frekvencia, signálový mikrokontrolér, displej, PWM, I2C**

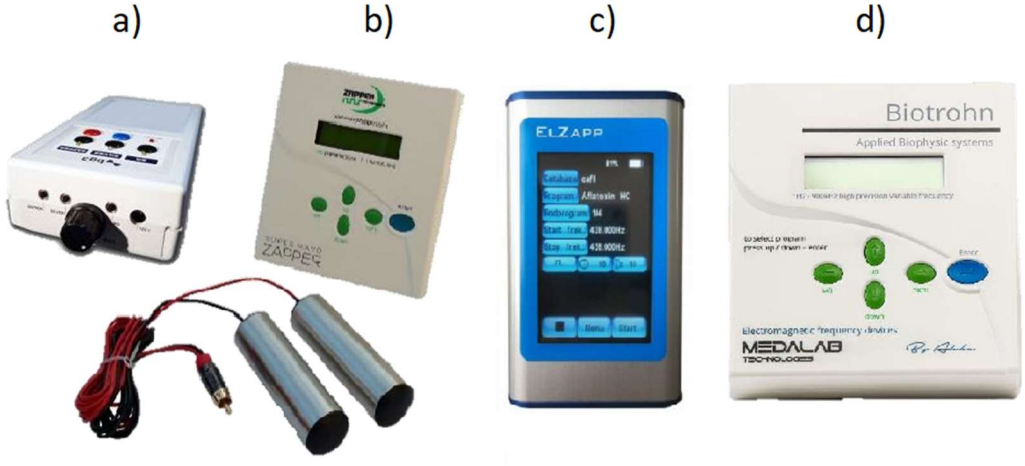
I. Úvod

Biorezonancia, metóda alternatívnej medicíny, pracuje s hypotézou, že každý organizmus, čiže aj patogén, má svoju rezonančnú frekvenciu. Ak telo vystavíme tejto frekvencii je údajne možné ovplyvniť prítomnosť patogénov v tele. Jedno z mnohých zariadení, ktoré túto terapiu umožňuje, sa nazýva „zapper"- z angl, slova zapping, čo znamená ničenie. Cieľom je navrhnúť cenovo dostupný zapper s parametrami porovnateľnými s komerčne dostupnými zariadeniami.

Biorezonančná terapia k liečbe primárne nepoužíva lieky ani iné chemické látky. Základným predpokladom biorezonancie je, že každý organizmus (živý, aj neživý) má vlastnú rezonančnú frekvenciu. Liečba pacienta spočíva v odstránení nežiaducich organizmov z tela pomocou elektromagnetických vĺn s rezonančnou frekvenciou daných nežiadúcich organizmov. Biorezonančnú terapiu vynašli v roku 1977 Franz Morell a Erich Rasche. Svoj objav uviedli na trh pod názvom „MORA–Therapie". Názov bol odvodený z ich priezvisk (MOrell a RAsche). [1]

A. Zapovanie ako podoblasť biorezonancie

Zapovanie je činnosť, pri ktorej je organizmus vystavený pulzujúcemu kladnému napätiu s rezonančnou frekvenciou daného organizmu. Zapper generuje jednosmerné pulzujúce napätie s požadovanou frekvenciou. Rozsahy frekvencií sa pohybujú od jednotiek hertzov až po stovky kilohertzov. Presnosť frekvencie je kľúčová pre úspešnosť liečby a pohybuje sa v stotinách ojedinele až tisícinách hertzov.



Obr. 1 Existujúce zappery a) Multifunkčný nízkofrekvenčný generátor BG3 [7] b) Super Ravo Zapper [8] c) ElZap [9] d) Mikrofrekvenčný generátor Biotrohn [10]

Úroveň napätia sa pohybuje v jednotkách voltov, čo je pre človeka bezpečné napätie. Pulzujúce napätie sa dostane do tela pacienta pomocou dvoch elektród, ktoré počas terapie drží v rukách. Elektródy sú vyrobené z dobre vodivého kovového materiálu a majú rôzny elektrický potenciál. Alternatívou môžu byť dve elektródy s kovovou podložkou. V takomto prípade majú elektródy rovnaký potenciál a kovová podložka, na ktorej pacient stojí bosý, tvorí druhý potenciál.

Spojením pacienta s elektródami sa elektrický obvod uzatvorí a telom pacienta začne pretekať jednosmerný pulzujúci prúd o veľkosti jednotiek μA. Tento prúd pre pacienta nie je nebezpečný ani bolestivý, vo väčšine prípadov ho ani necíti. Liečba pozostáva z niekoľko minút trvajúcich procedúr, medzi ktorými je potrebná prestávka. Počas prestávky pacient odpočíva a zapper nepoužíva. Najlepšie výsledky sa podľa doktorky Clark dosahujú pri troch 7 minútových procedúrach striedaných dvoma približne 20 minútovými prestávkami pri napätí 5 – 10 V. Dr. Clark uvádza, že zapperom sa dajú liečiť bolesti hlavy, chrbtice, brucha, chrípka, horúčka, bradavice, herpes, ale aj závislosti na nikotíne, alkohole, čiastočne aj HIV či zmierniť príznaky rakoviny. [2] Príklady existujúcich zaperov od rôznych výrobcov sú uvedené na Obr. 1

II. NÁVRH ZARIADENIA PRE ZAPOVANIE

Nami navrhnuté zariadenie pre zapovanie by malo spĺňať požiadavky ako sú malá hmotnosť a rozmery, široký frekvenčný rozsah (minimálne od 1 Hz do 1 MHz) s presnosťou na tisíciny Hz, funkcia plynulej zmeny frekvencie od stanovenej dolnej hodnoty k hornej hodnote, jednoduché používateľské rozhranie (LCD displej a tlačidlá, zvukový signál), možnosť uchovania frekvencií používateľov v pamäti, regulácia výstupného napätia, celková cena zariadenia by mala byť nižšia ako podobných komerčných zariadení.

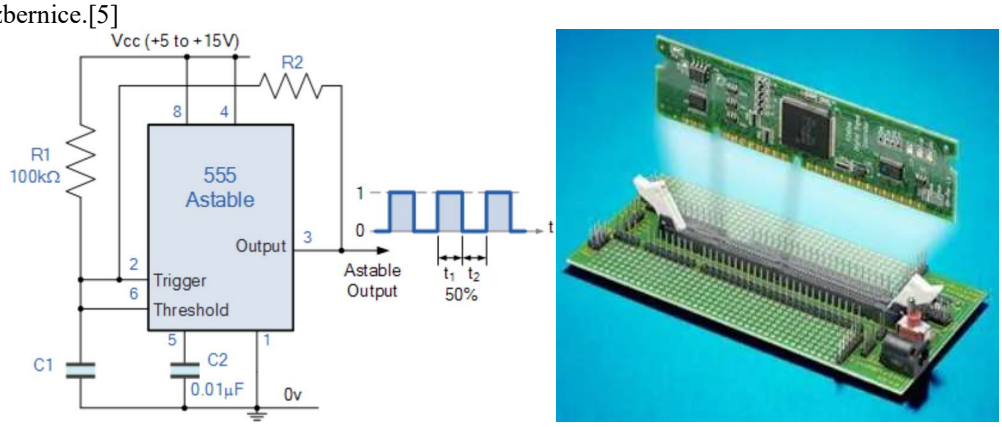
Po elektronickej stránke je zariadenie pre zapovanie v najjednoduchšej forme frekvenčný generátor. Prvým krokom pri návrhu nášho zappera bude zvoliť frekvenčný generátor, ktorý by svojím rozsahom a presnosťou vyhovoval stanoveným požiadavkám.

A. Analógové obvody

Prvou voľbou pri výbere frekvenčného generátora je použiť analógový obvod s časovačom TE555. Časovač TE555 by bol zapojený ako astabilný multivibrátor, ktorý by generoval obdĺžnikový signál, zapojenie uvedené na Obr. 2a. Väčšinou sa takéto zapojenie používa pre napájanie s nastaviteľnou striedou pri konštantnej frekvencii. Malou zmenou zapojenia je možné docieliť konštantnú 50 % striedu s nastaviteľnou frekvenciou.[4]

Ako ďalší spôsob generovania frekvencie môžeme spomenúť napäťovo — frekvenčný menič. Pripojením referenčného napätia a zmenou vstupného napätia sa mení výstupná frekvencia. [3]

V ďalšom bádaní v oblasti elektronických obvodov pre generáciu frekvencie nemá zmysel pokračovať vzhľadom na to, že naše zariadenie má byť vybavené používateľským rozhraním (displej), a budeme potrebovať mikrokontrolér. Preto sme sa rozhodli vybrať taký mikrokontrolér, ktorý má príslušné periférie pre generovanie frekvencií s požadovanou presnosťou a rozsahom. Pre spracovávanie, snímanie a ovládanie s cieľom zlepšenia chodu v uzavretej slučke v reálnom čase. Procesor F28335 (Obr. 2b) pracuje s frekvenciou až 150 MHz. má 32-bitový časovač so 16-bitovou deličkou. Tým bude zaručený požadovaný frekvenčný rozsah a presnosť frekvencie pri generovaní. Mikrokontrolér F28335 bude tiež obsluhovať LCD displej a tlačidlá. Komunikácia medzi mikrokontrolérom a LCD displejom bude pomocou I2C zbernice.[5]



Obr. 2a) Astabilný multivibrátor s časovačom 555 [4] b)Mikrokontrolér TMS320F28335 od fy Texas Instruments [11]

B. TMS320F28335 TEXAS INSTRUMENTS

Ide o mikrokontrolér rady C2000 od spoločnosti TEXAS INSTRUMENTS, ktorý je

optimalizovaný pre spracovávanie, snímanie a ovládanie s cieľom zlepšenia chodu v uzavretej slučke v reálnom čase. Procesor F28335 pracuje s frekvenciou až 150 MHz a má 32-bitový časovač so 16-bitovou deličkou. Tým bude zaručený požadovaný frekvenčný rozsah a presnosť frekvencie pri generovaní. Mikrokontrolér F28335 bude tiež obsluhovať LCD displej a tlačidlá. Komunikácia medzi mikrokontrolérom a LCD displejom bude pomocou I2C zbernice. [5]

C. Časovače mikrokontroléra TMS320F28335

Mikrokontrolér F28335 má tri časovače. My z týchto troch časovačov budeme používať časovač CPU-Timer 0 a CPU-Timer 1. Časovačom CPU-Timer 0 budeme generovať frekvenciu. Pre zvolenú frekvenciu mikrokontrolér vypočíta a nastaví deličku a odpovedajúcu periódu. Delička je reprezentovaná dvoma 8-bitovými registrami TDDRH a TDDR (spolu tvoria jeden 16-bitový register). Perióda je reprezentovaná dvoma 16-bitovými registrami PRDH a PRD (spolu tvoria jeden 32-bitový register). Hodnota z registra deličky sa nakopíruje do registrov počítadla deličky PSCH a PSC. Hodnota z registra periódy sa nakopíruje do registrov počítadla časovača TIMH a TIM. V týchto registroch mikrokontrolér počíta, vždy smerom nadol. Po spustení časovača sa bude s každým strojovým krokom dekrementovať perióda, až kým neprejde nulou. Vtedy sa hodnota v deličke raz dekrementuje a do registra počítadla časovača sa nakopíruje hodnota z registra časovača. Celý proces sa opakuje, kým register počítadla deličky neprejde nulou. Vtedy sa vygeneruje príznak prerušenia, ktorý procesor zaznamená a vznikne žiadosť o prerušenie, ktorá sa obslúži. V obsluhe prerušenia časovača Timer 0 budeme prepínať (negovať aktuálnu logickú hodnotu, prepisovať opačnou hodnotou) jeden z výstupov mikrokontroléra, ktorým bude spínaný tranzistor T1. Časovač Timer-1 bude slúžiť ako pomocný časovač počas generovania frekvencie. Jednou z požiadaviek na naše zariadenie je funkcia plynulej zmeny frekvencie. Zapper začne zapovať na dolnej hodnote frekvencie a s pravidelným prírastkom bude svoju frekvenciu zvyšovať. Časovač Timer-1 bude zabezpečovať v pravidelných intervaloch zvyšovanie frekvencie o spomínaný prírastok.

D. Vstupy a výstupy mikrokontroléra TMS320F28335

Mikrokontrolér je vybavený až 88 univerzálnymi vstupno-výstupnými pinmi. Každý jeden z nich je možné použiť až štyrmi rôznymi spôsobmi. Okrem predvolenej vstupno-výstupnej funkcie môžu slúžiť napríklad perifériám I2C, SPI, CAN, ePWM, eCAP... My budeme potrebovať šesť vstupných pinov pre tlačidlá a štyri výstupné piny, jeden pre LED diódu, jeden pre spínanie tranzistora T1 pripojenom na pine 00 a dva budú slúžiť pre I2C zbernicu (SDA na pine 32 a SCL na pine 33). LED dióda bude slúžiť pre indikáciu a bude pripojená k pinu 09. Keď bude zariadenie zapovať, LED dióda bude blikať. Funkcie tlačidiel budú: vľavo (pin 17), dole (pin 27), hore (pin 25), vpravo (pin 14), zrušiť (pin 13), potvrdiť (pin 11). Pre nastavenie tlačidiel a LED diódy musíme najprv nastaviť dané piny ako GPIO. Ďalej musíme nastaviť smer toku signálu (vstup/výstup) v registri GPADIR (nami používané piny sú z registrov skupiny A, okrem nich sú ešte skupiny B a C) a pre tlačidlá povoliť vnútorné pull-up rezistory v registri GPAPUD. Pull-up rezistor slúži na udržanie logickej hodnoty na logickej jednotke (high) na pine v stave nečinnosti. Stlačením tlačidla sa logická hodnota pinu zmení na logickú nulu (low). Aktuálnu hodnotu pinu zistíme prečítaním hodnoty v registri GPADAT.

Ako opatrenie pred neplatným prečítaním hodnoty pinu je možné nastaviť tzv. kvalifikačné vzorky. Procesor pred určením, že hodnota pinu sa zmenila bude musieť prečítať niekoľkokrát po sebe tú istú hodnotu. My nastavíme šesť kvalifikačných vzoriek v registroch GPAQSEL1 a GPAQSEL2 a vzorkovaciu periódu nastavíme v registri GPACTRL podľa časovania systému. Výstupnú hodnotu pinu meníme zápisom do registrov GPASET (nastavíme do log. jednotky), GPACLEAR (nastavíme do log. nuly) a GPATOGGLE (nová hodnota je negáciou predchádzajúcej).

III. FUNKCIE ZOSTROJENÉHO ZARIADENIA PRE ZAPOVANIE

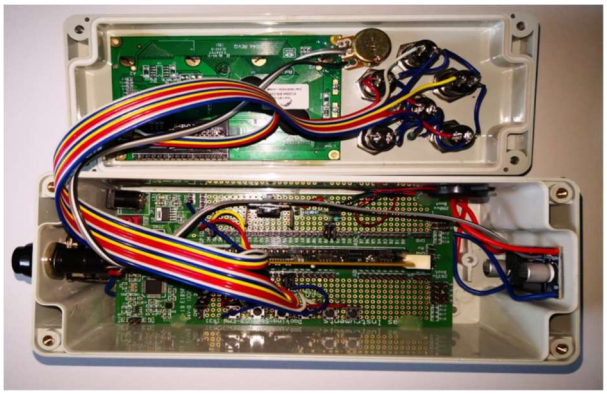
A. Napájanie zariadenia pre zapovanie

Mikrokontrolér F28335 potrebuje napájacie napätie 5 V. Výstupný obvod, ku ktorému bude pacient cez elektródy pripojený sme dimenzovali na 15 V. Zároveň požadujeme, aby bolo možné znížiť veľkosť tohto napätia. Mohli by sme mikrokontrolér a výstupný obvod napájať každý samostatne. Rozhodli sme sa napájať ich spoločne. Prívodom do zappera je výstup sieťového adaptéra 15 V DC 1 A. Toto napätie je priamo pripojené k vstupu DC-DC meniča, pomocou ktorého budeme vedieť meniť veľkosť napätia pre samotné zapovanie. Napájanie mikrokontroléra zabezpečuje lineárny stabilizátor napätia L7805. Vstupom tohto stabilizátora je napätie 15 V a výstupom napätie 5 V pre napájanie mikrokontroléra a displeja..

B. Nastavenie výstupného napätia zariadenia pre zapovanie

V kapitole III sme si stanovili požiadavky, ktoré by naše zariadenie malo spĺňať. Jednou z nich bola možnosť plynule meniť výstupné napätie v obvode zappera. Jednoduchým riešením by bolo použitie potenciometra zapojeného ako napäťový delič. Takéto zapojenie by však bolo energeticky nehospodárne. Preto sme sa rozhodli použiť jednosmerný znižovací (step-down) DC - DC menič LM2596. Menič LM2596 je veľmi rozšírený, najmä kvôli svojím malým rozmerom, nízkej cene, vysokej účinnosti (až 92 %), širokému rozsahu vstupného (3,2 – 40 V) a výstupného (1,25 – 35 V) napätia s malým šumom. Keďže sme zvolili znižovací menič, môžeme výstupné napätie v obvode regulovať od nuly po hodnotu napájacieho napätia. Zvolený menič je znázornený na Obr. 3. [6]. Kompletne zapojené zariadenie pre zapovanie osadene v krabičke spolu s displejom, tlačidlami a samotnou doskou s mikrokontrolérom pri pohľade do vnútra je znázornené na Obr. 4.

Obr. 3 LM2596 step-down menič DC-DC [6]



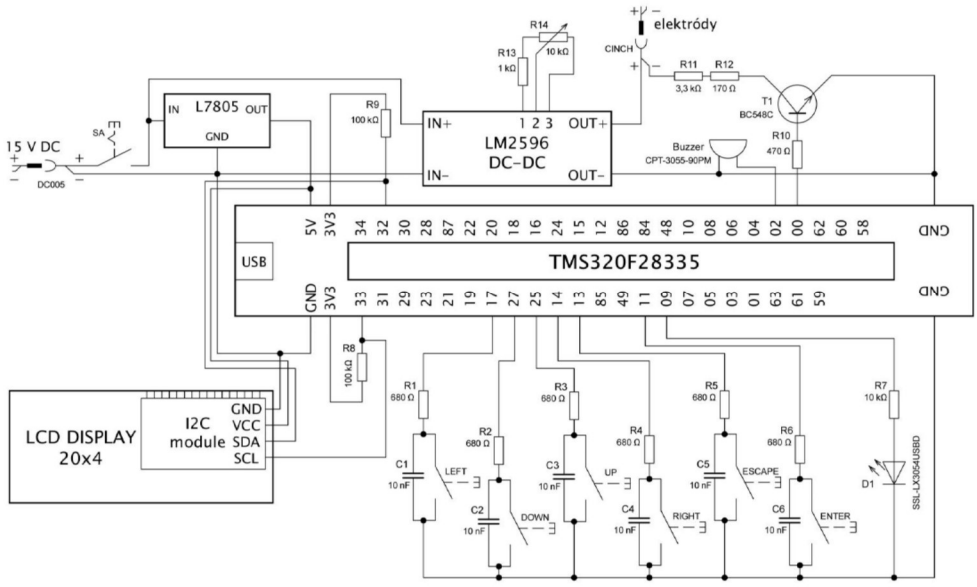
Obr. 4 Realizované kompletné zapojenie zariadenia

C. Zvukové rozhranie

Naše zariadenie je vybavené displejom, ktorý zabezpečuje vizuálnu komunikáciu s používateľom. Avšak, ak sa používateľ na displej nepozerá, komunikácia je prerušená. Tento problém sme sa rozhodli vyriešiť pridaním zvukovej komunikácie, presnejšie piezoelektrickým bzučiakom. Tie možno deliť na pasívne a aktívne. Aktívne bzučiaky majú vnútorný oscilátor, ktorý po pripojení na konštantné napätie rezonuje konštantnou frekvenciou a tým vydáva jeden tón. Pasívne bzučiaky neobsahujú oscilačný zdroj a vyžadujú vonkajší zdroj frekvencie. Piezoelektrický kryštál bzučiaka kmitá v závislosti na dodanej frekvencii. Zmenou dodávanej frekvencie môžeme generovať jednotlivé hudobné tóny a jednoduché melódie. V našom zariadení sme použili pasívny bzučiak, pretože takýto bzučiak sa riadi PWM signálom. Generovanie PWM signálu mikrokontrolérom sme už raz použili pri spínaní tranzistora T1, takže použiť ďalší PWM výstup nebol problém. Vytvorili sme vlastnú knižnicu funkcií pre generovanie jednotlivých tónov (frekvencie tónov z [6]) a vytvorili sme pomocné funkcie, ktoré v stanovených časoch generujú tóny, čím vzniká hudba. Jednoduché melódie používateľovi oznamujú začiatky a konce procedúr a prestávky medzi procedúrami. Pri zapnutí zariadenia sa tiež ozve krátka melódia.

IV. ZAPOJENIE ZARIADENIA

Zapnutím sa spustí mikrokontrolér a rozsvieti LCD displej. Užívateľ s pomocou tlačidiel ovláda kurzor displeja. Môže tak nastavovať hodnoty frekvencií, časov procedúr a prestávok medzi procedúrami, spúšťať samotné procedúry, pozastavovať či predčasne ich ukončovať. Pred spustením procedúry pacient pripojí elektródy k zapperu. Ku konektoru v zappere je pripojený výstup DC-DC meniča v sérii s bipolárnym tranzistorom T1. Báza tranzistora T1 je pripojená na pin 00, ktorý je spínaný požadovanou frekvenciou. Zopnutím tranzistora T1 sa na pripojených elektródach objaví napätie meniča. Dotknutím sa elektród pacient uzatvorí elektrický obvod a obvodom začne pretekať elektrický prúd. Veľkosť prúdu kvôli skratu v obvode je obmedzená rezistorom. Aby prúd tečúci obvodom nebol pri dotyku elektród bolestivý (citlivejší ľudia), tak sme pre obmedzenie pridali výstupný rezistor. Tým sa prúd v obvode zníži na nebolestivú hodnotu. Na elektródy sme použili medené rúrky, ktoré sme vyleštili a zbavili od vrstvy oxidu a pripojili k nim kábel, ktorý je možné pripojiť k výstupu zappera. Celková schéma zapojenia je na Obr. 5 a výsledné hotové zariadenie je na Obr. 6.



Obr. 5 Celková schéma zapojenia zapovacieho zariadenia

Obr. 6 Výsledné zhotovené zariadenia pre zapovanie

V. ZÁVER

Navrhnuté zariadenie pre zapovanie demonštruje praktickú realizáciu nízkonapäťového generátora pulzných signálov s variabilnou frekvenciou založeného na mikrokontroléri TMS320F28335. Hardvér aj softvér boli vyvinuté s dôrazom na presnosť výstupnej frekvencie (až na tisíciny hertzu), jednoduchosť používateľského ovládania a flexibilitu v nastavovaní procedúr. Vďaka digitálnemu riadeniu a použitiu DC-DC meniča LM2596 je možné presne kontrolovať výstupné napätie, čo je kľúčové z pohľadu účinnosti terapie ale tiež v kontexte pocitu užívateľa.

Z technického hľadiska zariadenie funguje spoľahlivo a je schopné plniť deklarované parametre. Mikrokontrolér zvláda plynulé zvyšovanie frekvencie, generovanie PWM signálu na viacero výstupov a zároveň obsluhuje vstupy z tlačidiel, LCD rozhranie aj zvukovú signalizáciu. Táto integrovaná funkcionalita ukazuje, že aj v podmienkach nízkonákladového vývoja je možné

vytvoriť relatívne komplexný systém, ktorý konkuruje komerčne dostupným riešeniam.

V budúcnosti je možné rozšíriť zariadenie o napájanie z batérie, čím by vzniklo prenosné zariadenie vhodné na použitie kdekoľvek s nulovou obavou o bezpečnosť pacienta. Taktiež je možné uvažovať o doplnení diaľkového ovládania cez Bluetooth alebo Wi-Fi, doplniť spätnoväzbový systém na meranie prúdu prechádzajúceho cez pacienta, či prepojiť zariadenie s mobilnou aplikáciou umožňujúcou sledovanie histórie terapií.

Záverom možno povedať, že vytvorené zariadenie je spoľahlivý a funkčný prototyp, ktorý slúži ako výborný základ pre ďalší vývoj. Kým biorezonancia ostáva kontroverznou oblasťou na hranici medzi technológiou a pseudovedeckým presvedčením, samotná technická práca ukazuje, že dizajn a realizácia frekvenčných terapeutických zariadení sú hodnotným cvičením v oblasti embedded systémov, digitálneho spracovania signálu a používateľského rozhrania.

ACKNOWLEDGMENT

Táto práca bola podporená z projektov APVV-18-0436, VEGA 1/0363/23 a KEGA 059TUKE-4/2024.

REFERENCES

[1] MORA Academie: The MORA - Bioresonance Therapy: The MORA Therapy. [online] Zug: MORA Akademie, 2008. [cit. 2023-03-03]. Dostupné na internete: http://mora-akademie.org/mora-therapie.html?&L=1.

[2] CLARK, Hulda Regehr: Revoluce v léčení všech nemocí. Zapper Technology, 2011. 620 s. ISBN: 978-80-254-5881-5.

[3] Burr-Brown Products from TEXAS INSTRUMENTS: High-Frequency VOLTAGE-TO-FREQUENCY CONVERTER: SBVS021A. [online]. Dallas, Texas: TEXAS INSTRUMENTS, október 1988, apríl 2007 [cit. 2023-03-11]. Dostupné na internete: https://www.ti.com/lit/ds/sbvs021a/sbvs021a.pdf.

[4] TEXAS INSTRUMENTS: Technical Reference Manual: SPRUI07. [online]. Dallas, Texas: TEXAS INSTRUMENTS, marec 2020 [cit. 2023- 0302]. Dostupné na internete:https://www.ti.com/lit/ug/sprui07/sprui07.pdf.

[5] Techfun: LM2596 step-down menič DC-DC. [online]. [cit. 2025-06-01]. Dostupné na internete: https://techfun.sk/produkt/lm2596-step-down-menic-dc-dc/.

[6] SUITS, Bryan H.: Physics of Music – Notes: Tuning. [online]. Michigan: Michigan Technological University, 1998, 2023 [cit. 2023-04-17]. Dostupné na internete: https://pages.mtu.edu/~suits/notefreqs.html.

[7] BIOpotraviny RAJ: Multifunkčný nízkofrekvenčný generátor BG3. [online]. [cit. 2023-05-14]. Dostupné na internete: https://biopotravinyraj.sk/nizkofrekvencny-generator-bg3.

[8] ZAPPER CENTRUM: SUPER RAVO ZAPPER. [online]. [cit. 2023-05-14]. Dostupné na internete: https://www.zapper-centrum.cz/super-ravo-zapper/super-ravo-zapper-2/.

[9] ZAPPER CENTRUM: ElZapp. [online]. [cit. 2023-05-14]. Dostupné na internete: https://www.zapper-centrum.cz/produkty/elzapp/.

[10] MEDALAB: mikrofrekvenčný generátor Biotrohn. [online]. [cit. 2023-05-14]. Dostupné na internete: https://www.medalab.com/sk/biotrohn/.

[11] Mouser electronics, [cit. 5.6.2025]. Dostupné na internete: https://www.mouser.fi/ProductDetail/Texas-Instruments/TMDSDOCK28335?qs=N6lMO %252BDQdCoi3mhoqayZXw%3D%3D

Optimálne vytváranie vektorových reprezentácií vzhľadom na sémantiku slovenského jazyka

[1]Kristián Sopkovič, [2]Eva Kupcová, [3]Matúš Pleva

[1,2,3]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 9, 04001 Košice, Slovenská republika

[1]kristian.sopkovic@tuke.sk, [2]eva.kupcova@tuke.sk, [3]matus.pleva@tuke.sk

Abstrakt **– Vektorové reprezentácie slov a viet hrajú zásadnú úlohu v súčasnom spracovaní prirodzeného jazyka (z angl. *Natural Language Processing, NLP*). Kvalita týchto reprezentácií má priamy dopad na úspešnosť mnohých úloh v oblasti NLP. Tento článok sa sústreďuje na optimalizáciu procesu vytvárania vektorových reprezentácií, pričom kladie dôraz na zachytenie sémantiky slovenského jazyka. Predstavujeme prístup, ktorý využíva jemné doladenie (z angl. *fine-tuning*) metódy Sentence-BERT (SBERT) na slovenskom strojovo preloženom jazykovom korpuse SNLI s cieľom posúdiť sémantickú podobnosť viet. Následne ukazujeme, ako sa naučené reprezentácie dajú využiť pri úlohe sémantického vyhľadávania dopytov k slovenským Wikipédia dokumentom na slovenskej databáze SK-QuAD, pričom hodnotíme výsledky pomocou metrík ako MAP a nDCG. Naše experimenty potvrdzujú, že jemne doladené modely SBERT majú potenciál efektívne a kvantifikovateľne zachytiť sémantické vzťahy v slovenčine, čo vedie k zlepšeniu výkonu v následných úlohách.**

Kľúčové slová – **NLP, vektorové reprezentácie, sémantická podobnosť, Sentence-BERT, SNLI, SK-QuAD**

I. Úvod

V posledných rokoch sa schopnosť počítačov porozumieť a spracovať ľudský jazyk výrazne zlepšila, najmä vďaka pokrokom v oblasti hlbokého učenia a vývoja rozsiahlych jazykových modelov. Kľúčovým krokom v tomto procese je premena textu na numerické vektorové reprezentácie, ktoré zachytávajú sémantický význam slov, fráz a celých viet. Kvalita týchto reprezentácií má priamy dopad na výkon širokého spektra úloh spracovania prirodzeného jazyka (z angl. *Natural Language Processing, NLP*), ako sú klasifikácia textu, rozpoznávanie pomenovaných entít, strojový preklad, sémantická podobnosť a vyhľadávanie informácií [1].

Tradičné prístupy k vytváraniu vektorových reprezentácií, ako Word2Vec [2] a GloVe [3], produkujú statické vektory pre každé slovo. Tieto prístupy však nezohľadňujú kontextuálnu povahu jazyka, kde sa význam slova môže meniť v závislosti od jeho okolia. S príchodom transformátorovej architektúry, ako sú modely BERT, došlo k prevratu v schopnosti modelovať kontextuálne významy slov. Model BERT generuje dynamické vektorové reprezentácie, ktoré berú do úvahy kontext každého slova v rámci vety, čím zvyšujú presnosť mnohých úloh spracovania prirodzeného jazyka [4].

Aj keď modely ako BERT dosahujú vynikajúce výsledky v rôznych úlohách NLP, ich priame využitie pri úlohách sémantickej podobnosti viet môže byť náročné na výpočtové zdroje, pretože si vyžaduje porovnávanie všetkých párov tokenov medzi dvoma vetami. Sentence-BERT (SBERT) je modifikácia architektúry BERT, navrhnutá na efektívne vytváranie vektorových reprezentácií celých viet. SBERT aplikuje rôzne stratégie na spracovanie výstupov BERT (pooling) a trénovacie ciele (ako siamské a tripletové siete) na generovanie kvalitných vektorových reprezentácií viet, ktoré sú sémanticky zmysluplné a vhodné na porovnávanie viet [5].

Motiváciou tejto štúdie je preskúmať optimálne metódy vytvárania vektorových reprezentácií so zameraním na sémantiku slovenského jazyka. Hoci existuje množstvo výskumov zameraných na anglický jazyk, menej pozornosti sa venovalo špecifickým výzvam a príležitostiam, ktoré sú spojené so slovenským jazykom, ktorý má zložitú morfológiu a syntaktickú štruktúru. V tomto

článku sa zameriame na metódu SBERT a jej prispôsobenie pre slovenský jazyk. Budeme skúmať dopad jemného doladenia SBERT modelu na strojovo preložnej databáze SNLI zameranej na sémantickú podobnosť dvojíc viet a hodnotiť kvalitu naučených reprezentácií v úlohe vyhľadávania na slovenskej časti Wikipédie označenú na úlohu odpovedania na otázky - databáze SK-QuAD [6].

II. Súvisiaca práca

V posledných rokoch si aplikácia siamských architektúr, najmä v súvislosti rozšírených modelov BERT, získala významnú pozornosť v úlohách spracovania prirodzeného jazyka (NLP). Siamská architektúra BERT ponúka zaujímavý prístup k extrakcii kvalitných vektorových reprezentácií viet, obzvlášť pre úlohy, ktoré zahŕbajú sémantickú podobnosť a implikáciu. Jedna významná štúdia od Laviho a kol. [7] predstavuje conSultantBERT, jemne doladený siamský model Sentence-BERT navrhnutý na porovnávanie pracovných ponúk s životopismi. Využitím databázy s viac ako 270 000 pármi táto štúdia preukázala efektívnosť modelu pri zachytávaní sémantických nuáns v pároch pracovných ponúk a uchádzačov o zamestnanie, pričom výrazne prekonala existujúce systémy.

Ďalším príkladom použitia siamských architektúr je práca Khatryho a kol. [8], ktorí implementovali model Sentence-BERT v rámci svojej metodológie nazvanej Target Similarity Tuning (TST). Tento prístup vylepšil zarovnanie vektorových reprezentácií viet doladením kosínusových podobností, aby sa efektívne zhodovali s priradenými úsekmi kódu. Ich zistenia zvýrazňujú, ako môže zlepšenie reprezentácie menej rozšírených programovacích jazykov prostredníctvom takto systematického prístupu obohatiť praktické aplikácie v programovacích prostrediach. Tento potenciál je obzvlášť dôležitý pre úlohy, ktoré si vyžadujú pochopenie sémantických vzťahov, čím naznačuje silnú závislosť pre jazyky, ktoré presahujú angličtinu a bežne používané programovacie jazyky.

V samostatnej štúdii Choi a kol. [9] preskúmali účinnosť modelov založených na BERT, vrátane ALBERT, pri hodnotení sémantickej textovej podobnosti (STS) a inferencii prirodzeného jazyka (NLI). Výsledky zdôrazňujú dôležitosť jemného doladenia modelov ako Sentence-BERT na dosiahnutie kvalitných vektorových reprezentácií pre nasledujúce úlohy, pričom naznačujú, že inicializácie vrstiev predtrénovaného modelu BERT môžu ovplyvniť výsledky učenia v rôznych aplikáciách NLP. To ukazuje na širšiu relevantnosť efektívnych tréningových metód pri optimalizácii vektorových reprezentácií viet pre konkrétne úlohy, ako sú tie, ktoré využívajú nuansy jazykových charakteristík.

Ďalej štúdie od Choi a kol. [9] skúmali rôzne metodológie na získavanie vektorových reprezentácií viet pomocou BERT a jeho variantov. Ich zistenia podporujú rôzne signály supervízie a metódy kontrastného učenia, pričom zdôrazňujú, ako môžu zlepšiť kvalitu reprezentácií viet. Tieto práce celkovo poukazujú na vyvíjajúce sa stratégie v rámci využitia siamských architektúr a ich aplikáciu na úlohy, ako je identifikácia parafráz, čo vedie k výrazným zlepšeniam v hodnotiacich metrikách naprieč rôznymi databázami.

Celkovo integrácia siamských architektúr do modelov založených na BERT má potenciál zlepšiť hodnotenie sémantickej podobnosti, čím otvára cestu pre pokrok v jazykovom spracovaní naprieč rôznymi jazykovými kontextmi. Naša práca sa zhoduje s týmito trendmi, keď aplikujeme prístup jemného doladenia prispôsobený slovenskému jazyku, čím ilustrujeme prispôsobavenie a účinnosť siamského modelu Sentence-BERT pri zvládaní jedinečných jazykových vlastností slovenčiny a prispievame k medzinárodnej diskusii o vektorových reprezentáciách viet v spracovaní jazyka pre slovenský jazyk.

III. Metodológia

V tejto časti článku podrobne opisujeme metodológiu trénovania a vyhodnocovania vektorových reprezentácií pre slovenský jazyk. Náš prístup pozostáva z dvoch hlavných fáz:

1) jemné doladenie modelu Sentence-BERT (SBERT) na databáze sémantickej podobnosti,
2) vyhodnotenie získaných reprezentácií v úlohe sémantického vyhľadávania.

A. Jemné doladenie SBERT na sémantickú podobnosť

Na jemné doladenie modelu SBERT sme použili predtrénovaný model slovakbert a taktiež sme experiment zopakovali aj pre multilinguálne modely ako napríklad mBERT. Tieto modely boli dotrénované na slovenskej verzii strojovo preloženej databázy SNLI (Stanford Natural Language Inference). Databáza SNLI je rozsiahly súbor údajov navrhnutý pre úlohu Natural Language Inference (NLI), v ktorej model určuje vzťahy medzi dvojicou viet (premisa a hypotéza). Obsahuje stovky tisíc párov viet anotovaných jednou z troch tried: entailment (implikácia), contradiction (kontradikcia) alebo neutral (neutrálny). Strojový preklad tejto databázy do slovenčiny nám umožnil získať trénovaciu množinu pre úlohu sémantickej podobnosti v slovenskom jazyku. Použili

sme SNLI na tréning modelov, avšak pre kvalitnejšie hodnotenie (evaluáciu) sémantickej podobnosti sme využili databázu STSB (Semantic Textual Similarity Benchmark). STSB je rozsiahly benchmark, ktorý obsahuje páry viet anotované so spojitým skóre podobnosti (od 0 do 5), čo umožňuje presnejšie posúdenie sémantickej podobnosti v porovnaní s kategorickými triedami v SNLI. Štruktúru dát znázornujeme v tabuľke 1.

Tabuľka 1
Porovnanie štruktúr databáz SNLI, STS-B a SK-QuAD

Vlastnosť	**SNLI**	**STS-B**	**SK-QuAD**
Úloha	Natural Language Inference	Semantic Textual Similarity	Otázka-Odpoveď (QA)
Vstup	Premisa + hypotéza	Dve vety	Kontext + otázka
Výstup	Trieda (entailment, neutral, contradiction)	Reálne číslo (0–5)	Textová odpoveď z kontextu
Jazyk	Slovenčina (strojovo preložená)	Slovenčina (strojovo preložená)	Slovenčina
Počet vzoriek	∼570k	∼8.6k	∼80k
Dátové polia	`premise`, `hypothesis`, `label`	`sentence1`, `sentence2`, `score`	`context`, `question`, `answers`
Zdroj	SNLI (Stanford)	GLUE Benchmark	TUKE KEMT

Počas trénovania sme optimalizovali model SBERT s cieľom minimalizovať stratu pre správne označené páry viet. Model bol trénovaný pomocou stratovej funkcie (z angl. *Multiple Negative Ranking Loss, MNRL*). Táto stratová funkcia je vhodná pre úlohy, kde chceme rozlíšiť medzi relevantnými a irelevantnými dokumentami.

B. Vyhodnotenie na slovenskom vyhľadávaní

Pre vyhodnotenie kvality naučených vektorových reprezentácií v úlohe vyhľadávania sme použili slovenskú databázu časti wikipédie SK-QuAD. SK-QuAD (Slovak Question Answering Dataset) je slovenský súbor údajov pre úlohu otázok a odpovedí, ktorý obsahuje približne 80 000 párov otázok a odpovedí extrahovaných zo slovenskej Wikipédie. Ku každej otázke je priradený kontext (časť textu z Wikipédie), z ktorého je možné vygenerovať odpoveď.

S cieľom prispôsobiť model SBERT pre úlohu vyhľadávania, boli vygenerované trojice dát (query, relevantný dokument, irelevantný dokument). Relevantné dokumenty boli priamo spojené s odpoveďami na otázky, zatiaľ čo irelevantné dokumenty boli vybrané náhodne, čo sa však vo všeobecnosti neodporúča, nakoľko to môže viesť k falošne pozitívnym či negatívnym výsledkom.

Na zhodnotenie kvality a relevantnosti výsledkov vyhľadávania boli aplikované štandardné metriky určené pre vyhľadávacie systémy: **priemerná presnosť'** (z anglického *Mean Average Precision, MAP*) a **normalizovaný diskontovaný kumulatívny zisk** (z anglického *Normalized Discounted Cumulative Gain, NDCG*).Tieto metriky nám umožňujú kvantifikovať presnosť a relevantnosť vrátených výsledkov vyhľadávania. Medzi ďalšie metriky patrí: **Presnosť'** (Precision), ktorá meria podiel relevantných dokumentov spomedzi vybraných, **Návratnosť'** (Recall), ktorá meria podiel relevantných dokumentov, ktoré boli nájdené, a **Skóre F1**, ktoré je harmonickým priemerom presnosti a návratnosti.

C. Experimentálne nastavenie

Všetky experimenty boli vykonané s použitím knižnice SentenceTransformers. Na trénovanie bola použitý následovný hardvér a trénovacie parametre

- Grafická karta NVIDIA A6000 (48GB VRAM)
- Rýchlosť učenia (learning rate): 2e-5
- Počet iterácií databázy použitej na trénovanie (epochs): 1
- Počet "warmup"krokov pred začiatkom trénovania: 0.1

Trénovanie bolo vzhľadom na použitú architektúru modelov veľmi rýchle a približná doba trénovania jedného modelu bola okolo 1 hodiny strojového času.

IV. Dosiahnuté výsledky a diskusia

V tejto kapitole prezentujeme a analyzujeme výsledky experimentov zameraných na trénovanie a vyhodnocovanie vektorových reprezentácií viet pre slovenský jazyk. Cieľom týchto experimentov je overiť efektivitu navrhnutej metodológie a kvalitu získaných reprezentácií v rôznych úlohách založených na sémantickej podobnosti.

Tabuľka 2
Vyhodnotenie výsledkov sémantickej podobnosti dvojíc viet STS-B-SK

Názov modelu	Stratová funkcia	Pearsonová korelácia	Spearmanová korelácia
slovakbert	SOFTMAX	0.59	0.61
bert-uncased-multi	SOFTMAX	0.55	0.55
bert-cased-multi	SOFTMAX	0.48	0.50
slovakbert	MNLR	0.70	0.70
bert-uncased-multi	MNLR	0.69	0.67
bert-cased-multi	MNLR	0.68	0.67

Tabuľka zobrazuje výsledky vyhodnotenia sémantickej podobnosti dvojíc viet na datasete STS-B-SK pre rôzne modely a stratové funkcie (trénovanie prebiehalo na datasete SNLI - strojovo preložený do slovenčiny).

A. *Vyhodnotenie sémantickej podobnosti*

Článok prezentuje vyhodnotenie modelov Sentence-BERT (SBERT) pre sémantické porovnávanie dvojíc viet, trénovaných na databáze SNLI-SK a testovaných na STS-B-SK.Výsledky tohto vyhodnotenia sú uvedené v Tabuľke 2.

Výsledky v Tabuľke 2 sú vyjadrené pomocou dvoch korelačných koeficientov: Pearsonovej a Spearmanovej korelačnej miery. Tieto miery kvantifikujú silu a smer vzťahu medzi predpovedanými hodnotami sémantickej podobnosti modelov a skutočnými (manuálne ohodnotenými) hodnotami z datasetu STS-B-SK.

- **Pearsonova korelácia** meria lineárny vzťah medzi dvoma premennými. Vypočíta sa ako kovariancia dvoch premenných, vydelená súčinom ich štandardných odchýlok. Hodnota Pearsonovej korelácie sa pohybuje v rozmedzí od -1 do +1. Hodnota +1 indikuje dokonalú pozitívnu koreláciu (ak jedna premenná rastie, druhá tiež), -1 indikuje dokonalú negatívnu koreláciu (ak jedna premenná rastie, druhá klesá) a 0 indikuje žiadnu lineárnu koreláciu.
- **Spearmanova korelácia** meria monotónny vzťah medzi dvoma premennými, čo znamená, že meria, ako dobre sa vzťah medzi premennými dá vyjadriť pomocou monotónnej funkcie. Vypočíta sa na základe poradia hodnôt premenných, a nie ich presných hodnôt. Rovnako ako Pearsonova korelácia, aj Spearmanova korelácia sa pohybuje v rozmedzí od -1 do +1 a interpretuje sa rovnako. Spearmanova korelácia je robustnejšia voči odľahlým hodnotám (outliers) ako Pearsonova korelácia.

B. *Vyhodnotenie vyhľadávania*

Neskôr sa v práci diskutuje aj o výkonnosti modelov na úlohe slovenského vyhľadávania na Wikipédii (SK-QuAD) pre modely SBERT natrénované na SNLI databáze. Výsledky pre siamsku architektúru sú uvedené v Tabuľke 3 [10].

Tabuľka 3
Výkonnosť SBERT modelov na slovenskom vyhľadávaní (siamská architektúra)

Model	Accuracy @10	Recall @10	Precision @10	NDCG @10	MAP @100
Slovakbert-mnlr-snli	0.34	0.34	0.11	0.36	0.29
Bert-uncased-snli-mnlr (base)	0.34	0.34	0.11	0.37	0.28
Bert-cased-snli-mnlr (base)	0.35	0.35	0.12	0.37	0.29

Tabuľka zobrazuje výkonnosť SBERT modelov na slovenskom vyhľadávaní (siamská architektúra) s metrikami Accuracy, Recall, Precision, NDCG a MAP, pričom @10 je značenie pre top 10 dokumentov.

C. *Diskusia výsledkov*

Výsledky vyhodnotenia sémantickej podobnosti na datasete STS-B-SK (Tabuľka 2) jasne ukazujú, že použitie stratovej funkcie MNLR (Multiple Negative Ranking Loss) vedie k výrazne lepším výsledkom v porovnaní so stratovou funkciou SOFTMAX. Modely trénované s MNLR dosahujú vyššie hodnoty Pearsonovej a Spearmanovej korelácie, čo naznačuje lepšiu schopnosť zachytiť sémantickú podobnosť medzi dvojicami viet. Konkrétne, model *slovakbert* s MNLR dosiahol koreláciu 0.70 pre obe metriky, čo je podstatné zlepšenie oproti 0.59/0.61 s SOFTMAX. Podobné zlepšenia sú viditeľné aj pri multilinguálnych modeloch *bert-uncased-multi* a *bert-cased-multi*. Toto potvrdzuje, že MNLR je efektívnejšia stratová funkcia pre úlohy sémantickej podobnosti v kontexte SBERT modelov.

Pri vyhodnotení vyhľadávania na datasete SK-QuAD (Tabuľka 3) je zaujímavé, že rozdiely vo výkonnosti medzi jednotlivými SBERT modelmi (jemne doladenými na SNLI-SK) sú relatívne malé. Všetky modely dosahujú podobné hodnoty pre metriky Accuracy@10, Recall@10,

Precision@10, NDCG@10 a MAP@100. Model *bert-cased-snli-mnlr (base)* vykazuje mierne lepšie výsledky v Precision@10 (0.12) a NDCG@10 (0.37), zatiaľ čo MAP@100 je najvyššie pre *slovakbert-mnlr-snli* a *bert-cased-snli-mnlr (base)* (0.29). Tieto výsledky naznačujú, že jemné doladenie na datasete SNLI s MNLR je robustný prístup, ktorý vedie k stabilnému výkonu v úlohe sémantického vyhľadávania pre slovenský jazyk. Je dôležité poznamenať, že hoci hodnoty metrík môžu pôsobiť na prvý pohľad nízko, pre komplexnú úlohu vyhľadávania na rozsiahlej databáze, akou je Wikipédia, ide o sľubné výsledky, ktoré naznačujú potenciál na ďalšie zlepšovanie. Porovnanie s multilingválnymi modelmi naznačuje, že hoci slovakbert je špecificky trénovaný na slovenčine, multilinguálne modely majú vďaka svojej širšej predtrenovanej báze potenciál dosiahnuť komparovateľné, alebo dokonca mierne lepšie výsledky v niektorých metrikách.

Celkovo výsledky potvrdzujú, že jemné doladenie SBERT modelov na sémantickú podobnosť s použitím vhodnej stratovej funkcie je efektívny prístup k získavaniu kvalitných vektorových reprezentácií pre slovenský jazyk, ktoré sú využiteľné v úlohách sémantického vyhľadávania.

V. Záver a práca do budúcnosti

Článok sa zaoberá problematikou optimálneho vytvárania vektorových reprezentácií s dôrazom na sémantiku slovenského jazyka. Skúmali sme efektívnosť jemného doladenia modelu Sentence-BERT (SBERT) na slovenských databáze SNLI pre úlohu sémantickej podobnosti a následne sme vyhodnotili takto získané reprezentácie v úlohe sémantického vyhľadávania na databáze SK-QuAD.

Vykonané experimenty ukázali, že jemné doladenie modelu SBERT na SNLI vedie k zlepšeniu schopnosti modelu zachytiť sémantické vzťahy medzi slovenskými vetami, čo sa prejavilo vo vyšších hodnotách Pearsonovej a Spearmanovej korelácie. V úlohe sémantického vyhľadávania sme pozorovali, že modely trénované na SNLI dosahujú porovnateľné výsledky, pričom model *bert-cased-snli-mnlr (base)* vykazoval mierne lepšiu výkonnosť. Porovnanie s multilingválnymi modelmi ukázalo, že tieto modely majú potenciál dosiahnuť ešte lepšie výsledky v úlohe vyhľadávania.

Zistenia naznačujú, že jemné doladenie SBERT modelov je efektívny prístup k získavaniu kvalitných vektorových reprezentácií pre slovenský jazyk. Tieto reprezentácie môžu byť následne úspešne využité v rôznych NLP úlohách, vrátane sémantického vyhľadávania.

V budúcej práci sa chceme zamerať na rozsiahlejšie experimentovanie s rôznymi architektúrami SBERT a trénovacími stratégiami. Okrem toho by sme chceli preskúmať vplyv augmentácie dát a použitia pokročilejších metód výberu negatívnych vzoriek na výkonnosť vyhľadávania. Ďalším smerom výskumu by mohlo byť skúmanie efektívnosti týchto modelov v iných slovenských NLP úlohách, ako je napríklad klasifikácia textu alebo rozpoznávanie pomenovaných entít.

Poďakovanie

Táto práca bola podporená Kultúrnou a edukačnou grantovou agentúrou Ministerstva školstva, výskumu, vývoja a mládeže SR cez projekt KEGA 049TUKE-4/2024, a v rámci projektov podporených Agentúrou na podporu výskumu a vývoja APVV-22-0261 a APVV-22-0414.

Literatúra

[1] D. Chandrasekaran and V. Mago, "Evolution of semantic similarity—a survey," *ACM Computing Surveys*, vol. 54, pp. 1–37, 2021.

[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013. [Online]. Available: https://arxiv.org/abs/1301.3781

[3] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: https://aclanthology.org/D14-1162/

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423/

[5] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," 2019. [Online]. Available: https://arxiv.org/abs/1908.10084

[6] D. Hládek, J. Staš, J. Juhár, and T. Koctúr, "Slovak dataset for multilingual question answering," *IEEE Access*, vol. 11, pp. 32 869–32 881, 2023.

[7] N. Lavi, O. Medentsiy, and D. Graus, "conSultantBERT: Fine-tuned siamese sentence-bert for matching jobs and job seekers," in *Proceedings of the Workshop on Recommender Systems for Human Resources (RecSys in HR 2021)*, vol. 2967. Amsterdam, The Netherlands: CEUR-WS.org, October 2021, also available as arXiv preprint: arXiv:2109.06501. [Online]. Available: http://ceur-ws.org/Vol-2967/paper_8.pdf

[8] A. Khatry, S. Gulwani, P. Gupta, V. Le, M. Singh, A. Singha, and G. Verbruggen, "TSTR: Target similarity tuning meets the real world," in *Findings of the Association for Computational Linguistics: EMNLP 2023*. Singapore: Association for Computational Linguistics, December 2023, pp. 10 256–10 261. [Online]. Available: https://aclanthology.org/2023.findings-emnlp.687/

[9] H. Choi, J. Kim, S. Joe, and Y. Gwon, "Evaluation of bert and albert sentence embedding performance on downstream nlp tasks," January 2021, preprint. [Online]. Available: https://arxiv.org/abs/2101.10642
[10] K. Sopkovič, "Neurónové vyhľadávanie na základe sémantickej podobnosti vektorov," Diplomová práca, Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, Košice, Slovensko, 2024, bez DOI.

Cybersecurity Implications of Pickle-based Model Serialization in Containerized Machine Learning Environments

[1]*Erik KAJÁTI,* [2]*Ľubomír URBLÍK,* [3]*Veronika KOSIUK,*

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

[1]erik.kajati@tuke.sk, [2]lubomir.urblik@tuke.sk, [3]veronika.kosiuk@student.tuke.sk

Abstract **– The integration of machine learning models into containerized real-time data processing platforms presents substantial cybersecurity challenges, particularly due to the utilization of Python Pickle for model serialization. This paper comprehensively analyzes the security implications of a containerized machine learning system employing Docker containers, ZeroMQ communication, and Pickle-serialized models for classification, regression, and clustering tasks. Through meticulous security assessments, we identify critical vulnerabilities associated with Pickle deserialization, which can facilitate arbitrary code execution, privilege escalation, and system compromise. Our analysis underscores that while containerization offers isolation benefits, the combination of untrusted model serialization and inter-container communication engenders attack vectors capable of compromising entire distributed systems. Consequently, we propose a security framework incorporating input validation, sandboxing, alternative serialization methods, and monitoring strategies to effectively mitigate these risks while preserving system functionality and performance.**

Keywords **– containerization, machine learning security, pickle vulnerability, serialization attacks**

I. Introduction

The rapid adoption of containerized machine learning (ML) systems has revolutionized how organizations deploy and scale artificial intelligence applications. Modern ML platforms increasingly rely on microservices architectures where pre-trained models are packaged into isolated containers and communicate through message-passing protocols [1]. While this approach offers benefits in terms of scalability, maintainability, and deployment flexibility, it also introduces significant cybersecurity challenges that are often overlooked during system design and implementation.

Python Pickle, a built-in serialization library, has become a de facto standard for model persistence in the Python ecosystem due to its simplicity and comprehensive object serialization capabilities [2]. However, the convenience of Pickle comes with substantial security risks. The deserialization process can execute arbitrary Python code, making it a prime vector for remote code execution attacks when untrusted data is processed [3]. This vulnerability becomes particularly concerning in distributed systems where serialized objects are transmitted between services and potentially stored for extended periods.

The intersection of these technologies creates a complex threat landscape. Containerized ML systems that rely on Pickle serialization for model storage and distribution may unknowingly expose themselves to sophisticated attacks that can compromise not only individual containers but entire clusters. The problem is exacerbated by the fact that many developers are unaware of Pickle's security implications, treating serialized models as benign data files rather than potential attack vectors.

This paper presents a comprehensive security analysis of containerized ML environments that utilize Pickle-based model serialization. Our research is motivated by a real-world implementation that demonstrates the typical architecture patterns found in modern ML platforms: Docker containerization, ZeroMQ-based inter-service communication, and automated model loading from Pickle files [1]. Through systematic evaluation, we identify critical vulnerabilities and propose mitigation strategies that balance security requirements with operational needs.

II. Related Work

The security implications of object serialization have been extensively studied, with particular attention to Java and Python ecosystems. In [2], authors highlighted the dangers of Python Pickle deserialization at BlackHat 2011, demonstrating how malicious payloads could achieve arbitrary code execution. This seminal work established the foundation for understanding serialization-based attacks in Python environments.

Recent research has expanded on these findings, with authors [3] conducting a comprehensive analysis of 7,543 open-source Python projects, discovering that 36 projects implemented Pickle defense strategies, of which nine were incorrectly implemented. Their work reveals the widespread nature of Pickle vulnerabilities and the challenges in implementing effective countermeasures.

The Java ecosystem has faced similar challenges with deserialization attacks. Researchers have developed sophisticated techniques for detecting and exploiting Java deserialization vulnerabilities [4], [5]. In [6], authors proposed a bottom-up approach for detecting Java deserialization gadget chains, reducing search complexity from exponential to polynomial time. These techniques have inspired similar approaches for Python environments.

Container security has emerged as a critical research area as organizations increasingly adopt containerized architectures. Traditional security models often assume trusted code execution within containers, but the reality of complex supply chains and third-party dependencies introduces numerous attack vectors [7]. The isolation provided by containers can create a false sense of security, particularly when malicious code can execute within container boundaries.

ZeroMQ, while providing efficient message passing capabilities, introduces its own security considerations. The lack of built-in authentication and encryption mechanisms means that security must be implemented at the application layer, creating opportunities for misconfiguration and attack [8]. When combined with serialized data transmission, these vulnerabilities can be amplified.

Machine learning specific security research has primarily focused on adversarial attacks against model inference rather than deployment security [9]. However, the rise of MLOps and automated model deployment has created new attack surfaces that require comprehensive security analysis. The trusted nature of ML models in many organizations makes them attractive targets for supply chain attacks.

III. Materials and Methods

Our security analysis focuses on a containerized ML platform developed as part of a comprehensive study on ML integration into real-time data processing systems [1]. The target system represents a typical modern ML deployment architecture and serves as an excellent case study for identifying security vulnerabilities in production-like environments.

The analyzed system consists of six interconnected Docker containers: a data preprocessor, three ML model containers (classification, regression, clustering), a prediction saver, and a ZeroMQ proxy for message routing. Each ML model container loads a pre-trained model from a Pickle file during initialization and processes incoming data through ZeroMQ publish-subscribe communication patterns.

The microservices-based architecture depicted in Figure 1 emphasizes modularity, scalability, and clear separation of concerns. A local data storage structure contains directories for raw, cleaned, and predicted data. Raw data is stored, retrieved by the data preprocessor, validated, and cleaned. Cleaned data is stored back and published for downstream processing.

A ZeroMQ proxy acts as a central message broker, enabling asynchronous, topic-based message passing. Each microservice subscribes to or publishes messages relevant to its role. The data preprocessor publishes cleaned data, which is consumed by model containers for classification, regression, and clustering. Each container subscribes to the cleaned data topic, processes the input, and generates predictions.

Predictions are published back through the messaging system. The prediction saver subscribes to the predictions topic, ensuring reliable persistence of inference results in the local storage. This approach guarantees traceability and supports future analysis or auditing.

JSON serialization is used for payloads, ensuring interoperability and readability. Model objects are serialized using Pickle, allowing efficient storage and retrieval of trained models. The architecture's design principles—separation of concerns, loose coupling, and scalability—are realized through independent microservices and standardized, message-based communication. This structure facilitates easy maintenance, debugging, and system extension. The workflow ensures robust, maintainable, and scalable machine learning operations.

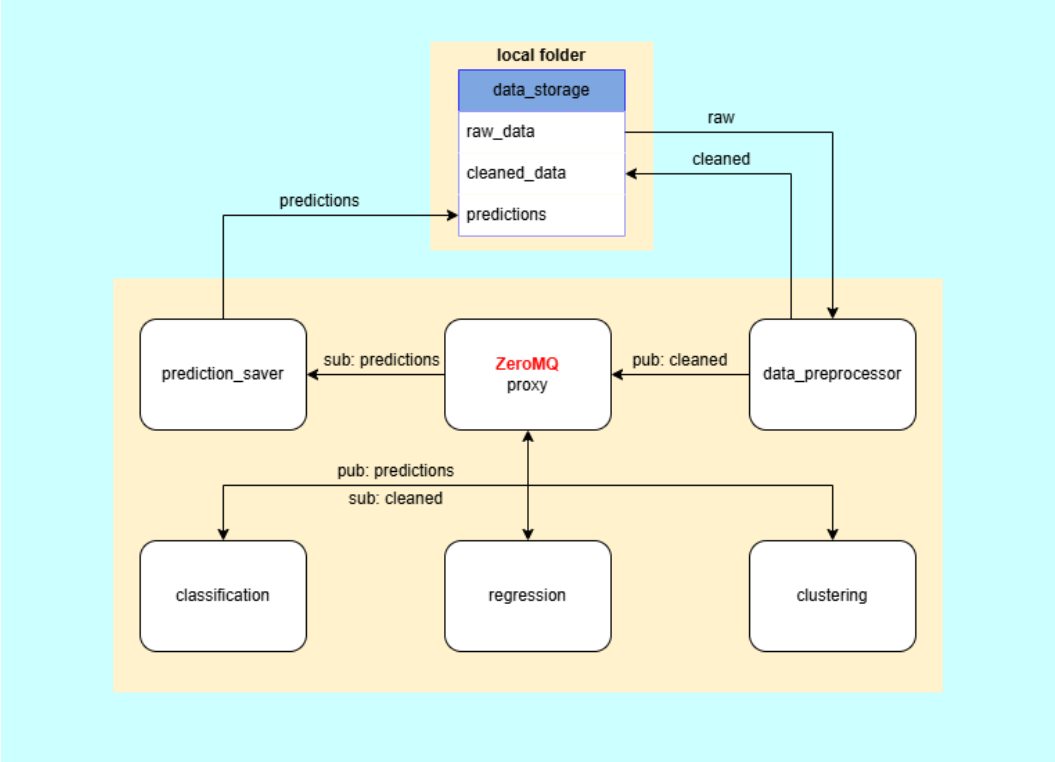


Fig. 1 Solution architecture.

IV. Results

Our security assessment revealed multiple critical vulnerabilities that could compromise the entire containerized ML system. The findings demonstrate that seemingly isolated security issues can combine to create severe attack scenarios with system-wide impact.

A. *Pickle Deserialization Vulnerabilities*

The most critical vulnerability stems from the system's reliance on Python Pickle for model serialization. During container initialization, each ML service automatically loads its corresponding model file using `pickle.load()` without any validation or sandboxing mechanisms. This creates an immediate arbitrary code execution vulnerability if an attacker can replace or modify model files.

Our analysis revealed that successful exploitation requires only write access to the model directory, which could be achieved through various vectors including compromised CI/CD pipelines, supply chain attacks, or container escape exploits. Once a malicious Pickle file is in place, the next container restart triggers automatic code execution with full container privileges.

B. *Inter-Container Communication Security*

The ZeroMQ-based communication system lacks authentication and encryption, creating multiple attack vectors for network-based adversaries. Our analysis identified that the proxy component accepts connections from any source without verification, allowing unauthorized clients to subscribe to sensitive data streams or inject malicious messages.

Message injection attacks proved particularly effective, as the JSON-based message format can be easily crafted to trigger unexpected behavior in receiving containers.

The publish-subscribe pattern amplifies these vulnerabilities by broadcasting sensitive data to all subscribers without access control. An attacker gaining access to one container can potentially monitor all inter-service communication, revealing business logic, model performance metrics, and potentially sensitive training data characteristics.

C. *Container Isolation Bypass*

Docker offers process and filesystem isolation, but our assessment identified mechanisms that circumvent container boundaries, posing substantial security vulnerabilities. While containerization isolates applications and dependencies, practical deployments frequently compromise these boundaries.

Bind mounts between containers and the host share directories or files, creating shared attack surfaces. Compromising a container with access to a bind mount enables attackers to manipulate or replace files on the host, impacting other containers that share the mount. This undermines isolation, as a breach can propagate to other containers via the shared filesystem. Improper permissions or inadequate access controls on these mounts increase the risk of unauthorized modifications or data exfiltration.

Network-based attacks effectively moved laterally between containers. Docker networking permits unrestricted communication between containers on the same network bridge, enabling attackers to scan, probe, and exploit vulnerabilities in neighboring containers with minimal resistance. We observed attackers escalating privileges, intercepting sensitive data, or disrupting critical inter-container communications. The absence of network segmentation or firewall rules amplifies exposure, allowing attackers to traverse the internal network without triggering external security controls.

D. Supply Chain Attack Vectors

The automated model loading mechanism creates significant supply chain attack opportunities. Our analysis revealed that the system trusts model files implicitly, making it vulnerable to attacks that compromise the model training or distribution pipeline.

Malicious actors can introduce backdoored models during the training process. These models appear to function normally during testing but contain hidden Pickle payloads that activate under specific conditions. Such attacks could remain undetected for extended periods while providing persistent access to production systems.

The Docker image build process also presents supply chain risks, particularly when using untrusted base images or downloading model files from external sources during container construction. These attack vectors bypass traditional security controls by embedding malicious content within trusted deployment artifacts.

V. Discussion

The identified vulnerabilities represent fundamental security design flaws rather than implementation bugs, making them particularly challenging to address through simple patches. The root cause lies in the inherent tension between operational convenience and security requirements in ML deployment pipelines.

A. Security Implications

The combination of Pickle deserialization vulnerabilities with containerized deployment creates a perfect storm for sophisticated attacks. Traditional network security controls fail to detect these attacks because they exploit legitimate application functionality. These attacks are persistent and stealthy, maintaining access across system updates, container restarts, and even migration. Malicious payloads bypass security monitoring systems designed to detect traditional malware due to the trusted nature of ML models in many organizations. Compromise of a single model container can quickly escalate to full system compromise due to the interconnected nature of modern ML systems. The shared data processing pipeline offers numerous opportunities for lateral movement and privilege escalation.

B. Mitigation Strategies

Addressing vulnerabilities in containerized ML systems requires a multi-layered approach. Input validation and sanitization are crucial, but modern serialization attacks make complete detection challenging. Alternative serialization formats like ONNX, TensorFlow SavedModel, or custom binary formats offer a long-term solution by eliminating arbitrary code execution risks. Sandboxing and containment limit attack impact, preventing full system compromise even with successful deserialization vulnerabilities. Network security controls, including authentication, encryption, and access control for inter-container communication, prevent lateral movement attempts. Implementing mutual TLS authentication for ZeroMQ connections addresses communication vulnerabilities.

C. Broader Implications for MLOps

These findings have significant implications for the broader MLOps ecosystem. Containerized ML deployment has outpaced security framework development, creating systemic vulnerabilities. Integrate security into ML pipeline design from the start. Collaborate between ML engineers, security teams, and infrastructure operators to develop secure-by-design architectures. Organizations rely on the trustworthiness of ML models. Scrutiny and verification processes are essential. Model provenance tracking, cryptographic signing, and runtime integrity monitoring should be standard practices in production ML deployments.

VI. Conclusion

Containerized ML systems face significant cybersecurity challenges requiring immediate attention. Python Pickle serialization vulnerabilities combined with containerized deployment architectures create attack vectors compromising entire ML platforms through sophisticated supply chain and deserialization attacks.

Traditional security approaches are insufficient for modern ML systems. The trusted nature of ML models and complexity of containerized deployments create unique attack surfaces requiring specialized defensive strategies. Organizations must evaluate their security posture and implement comprehensive mitigation strategies.

Pickle-based model serialization poses an security risk in production environments. Organizations should prioritize migration to secure alternatives like ONNX or TensorFlow SavedModel formats. Robust input validation, network security controls, and container hardening are essential for maintaining secure ML deployments.

Future research should focus on developing security frameworks for ML systems, including automated vulnerability detection tools, secure model distribution mechanisms, and runtime protection systems. The rapid evolution of ML deployment practices requires corresponding advancements in security methodologies to prevent widespread exploitation of these vulnerabilities.

The implications extend beyond individual systems to the broader AI ecosystem. As ML becomes critical to business operations, the security of ML deployment infrastructure is a strategic imperative. Organizations should invest in secure MLOps practices to protect systems and the broader digital ecosystem from sophisticated attacks targeting AI infrastructure.

Acknowledgment

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 17R05-04-V01-00004, Competence and Training Center for Cyber and Information Security TUKE.

References

[1] V. Kosiuk, "Aplikácia metód strojového učenia v kontajnerizovanom prostredí," Košice, Slovak Republic, 2025, bachelor's thesis, 46 pages, Supervisor: Ing. Erik Kajáti, PhD.

[2] M. Slaviero, "Sour pickles: A serialised exploitation guide," in *BlackHat USA*, Las Vegas, USA, 2011, security conference presentation on Python Pickle vulnerabilities.

[3] B. Alliance and S. Researchers, "Pain pickle: Bypassing python restricted unpickler for automatic exploit generation," *IEEE Security and Privacy Workshops*, December 2022, analysis of 7543 open-source Python projects for Pickle vulnerabilities.

[4] A. Yamani and C. S. Team, "Exploitation of the java deserialization vulnerability to access forgerock-openam server," *IEEE International Conference on Cyber Security and Resilience*, pp. 1–8, September 2023.

[5] L. Zhang, C. Wang, and M. Liu, "A vulnerability mining model of java json deserialization based on ast," *IEEE International Conference on Computer Science and Educational Informatization*, pp. 1–6, November 2020.

[6] H. Gu, X. Li, W. Zhang, and K. Chen, "Efficient detection of java deserialization gadget chains via bottom-up gadget search and dataflow-aided payload construction," *IEEE Transactions on Software Engineering*, May 2024, advanced techniques for detecting deserialization vulnerabilities.

[7] S. Sultan, I. Ahmad, and T. Dimitriou, "Container security: Issues, challenges, and the road ahead," *IEEE Access*, vol. 7, pp. 52 976–52 996, 2019.

[8] P. Hintjens, *ZeroMQ: Messaging for Many Applications*. O'Reilly Media, 2013, comprehensive guide to ZeroMQ messaging patterns and security considerations.

[9] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014, foundational work on adversarial attacks in machine learning.

Spätný vplyv LED žiaroviek na napájaciu sieť

[1]*Ján KAŇUCH,* [2]*Peter GIROVSKÝ,* [3]*Matej ONDERKO*
[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]matej.onderko@student.tuke.sk

***Abstrakt* — Tento príspevok sa zaoberá problematikou spätného vplyvu LED osvetlenia na napájaciu elektrickú sieť. Cieľom je analyzovať technické a fyzikálne aspekty LED svetelných zdrojov, ich vývoj, konštrukciu a hlavne ich interakciu s napájacím systémom. Osobitná pozornosť je venovaná nežiaducim javom, ktoré pri prevádzke LED žiaroviek vznikajú, ako sú vyššie harmonické, nízky účinník, elektromagnetické rušenie (EMI) a generovanie jalového výkonu. Príspevok objasňuje, ako LED technológia, hoci energeticky efektívna, môže pri hromadnom nasadení ovplyvniť kvalitu napätia v distribučnej sieti. Záver práce poukazuje na dôležitosť zohľadnenia týchto javov pri navrhovaní a výbere svetelných systémov, aby sa minimalizoval ich negatívny dopad na elektrickú sieť a zabezpečila čo najvyššia kvalita elektrickej energie.**

***Kľúčové slová* — LED osvetlenie, harmonické skreslenie, účinník, PFC, elektromagnetické rušenie.**

I. Úvod

LED osvetlenie sa v posledných rokoch stalo bežnou súčasťou domácností, kancelárií aj verejných priestorov. Hlavnými dôvodmi sú nízka spotreba energie, dlhá životnosť a šetrnosť k životnému prostrediu. Avšak je potrebné si uvedomiť aj fakt, že LED technológia môže mať aj nežiaduci vplyv na elektrickú sieť.

S rastúcim počtom LED žiaroviek v prevádzke sa čoraz častejšie objavujú technické problémy ako vyššie harmonické, znížený účinník či elektromagnetické rušenie. Tieto javy môžu spôsobovať zhoršenie kvality elektrickej energie, problémy pri prevádzke iných elektrických zariadení a zvýšené náklady na údržbu elektrických rozvodov.

Energia je základnou podmienkou rozvoja spoločnosti. Zvyšovanie životnej úrovne ľudí zákonite vedie k rastúcej spotrebe energie. Preto riešenie energetických problémov spočíva nielen v znižovaní spotreby, ale aj v odstraňovaní zbytočného plytvania energiou. Keďže energia nie je bezplatná, očakávame od úspor aj finančné výhody [1].

V minulosti sa pri návrhu svietidiel len málo prihliadalo na ich energetickú účinnosť. Výsledkom je, že v mnohých domácnostiach sa stále používajú zariadenia s vysokou spotrebou energie. S rastúcimi cenami energie je čoraz dôležitejšie nahradiť tieto staré svietidlá úspornejšími riešeniami. Straty spôsobuje aj napr. prach na svietidlách a silnejšia vrstva môže znížiť svetelný výkon až o polovicu [2].

Kvalitné osvetlenie by nemalo zabezpečiť len dostatočné množstvo svetla, ale aj správne rozloženie jasu, obmedzenie oslnenia, vhodnú farbu a verné podanie farieb. Malo by tiež zohľadňovať fyziologické a psychologické potreby pri zrakovej činnosti. Účinné osvetlenie by malo byť navrhnuté s ohľadom na jeho funkčnosť a energetickú úspornosť. V obytných priestoroch nie je vždy nutné osvetliť celý priestor – postačuje zamerať sa na konkrétne zóny, ako je pracovný stôl či miesto určené na čítanie. Z energetického hľadiska je vhodnejšie používať bodové osvetlenie, ktoré cielene osvetľuje požadovanú oblasť a zároveň znižuje celkovú spotrebu elektrickej energie.

Úspornosť osvetlenia závisí najmä od typu použiteľného svetelného zdroja. LED osvetlenie je výrazne najúspornejšie, keďže väčšinu energie premieňa priamo na svetlo. Zároveň má dlhú životnosť a nízku spotrebu. Hoci je obstarávacia cena LED svietidiel vyššia, tak celková prevádzka je ekonomicky výhodnejšia. Výber osvetlenia preto závisí od požiadaviek používateľa a dlhodobej návratnosti investície [3].

II. Vývoj LED žiaroviek a základné vlastnosti

LED žiarovky priniesli revolúciu do sveta osvetlenia vďaka svojej inovatívnej technológii a výnimočným vlastnostiam. V porovnaní s tradičnými žiarovkami ponúkajú mnoho výhod, ktoré

vedú k ich postupnému nahradeniu. Avšak majú aj niektoré nevýhody, na ktoré netreba zabúdať.

A. História LED diódy a žiarovky

Prakticky sa LED diódy začali využívať od roku 1962, pričom ich vynálezcom bol Nick Holonyak junior. Pre širokú verejnosť sa stali dostupnejšie až koncom 60. rokov. Zozačiatku sa LED diódy objavovali len v drahších zariadeniach, ako boli laboratórne nástroje či elektronické meracie prístroje. Postupne našli uplatnenie aj v televízoroch, rádiách a podobných technológiách. Ich využitie však obmedzoval nízky výkon, ktorý neumožňoval osvetlenie priestoru. Preto slúžili hlavne ako náhrada klasických indikátorov alebo súčasť sedem-segmentových displejov. Prvé LED diódy boli dostupné výlučne v červenej farbe [4].

Okolo roku 1971 sa začali objavovať LED diódy aj v iných farbách. Postupný technologický vývoj viedol k zvýšeniu ich svetelného výkonu, čo umožnilo ich využitie aj na osvetľovacie účely.

V 80. rokoch zažili LED diódy prvý výraznejší rozmach, keď sa stali dostupnejšími v rôznych farbách a tvaroch (Obr. 1).

Obr. 1 LED diódy v rôznych farbách a tvaroch.

Používali sa ako indikačné prvky v elektronických zariadeniach, ako brzdové svetlá v automobiloch, zobrazovanie alfanumerických znakov, alebo na tvorbu zobrazovacích panelov a reklamných tabúľ.

Prelomovým krokom bolo zavedenie prvých LED žiaroviek s bežnými päticami, ktoré umožnili ich použitie v štandardných svietidlách. Ich svietivosť však stále nebola dostatočne vysoká a farebné spektrum neumožňovalo plnohodnotnú náhradu klasických žiaroviek. Využívali sa prevažne v dekoratívnych a okrasných svietidlách. Výrobca osvetlenia Osram Sylvania predstavil prvú LED žiarovku so svetelným výkonom 1500 lúmenov, čo zodpovedá 100-wattovej žiarovke (Obr. 2).

Obr. 2 LED žiarovka fy OSRAM [5].

Na nasledujúcich obrázkoch môžeme vidieť v súčasnosti vyrábanú LED žiarovku 8W – G45/E27/SMD/4000K–ZLS829 (Obr. 3) a tú istú LED žiarovku v rozobratom stave s pohľadom na jej vnútornú elektroniku a LED diódy (Obr. 4).

Obr. 3 LED žiarovka 8W – G45/E27/SMD/4000K.

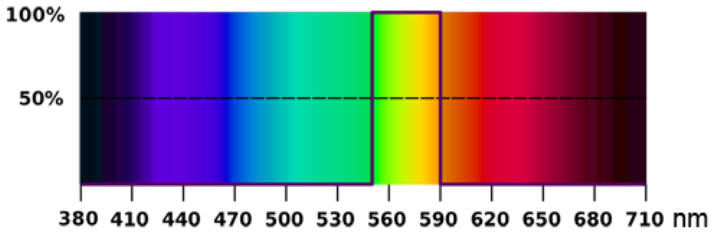
Obr. 4 LED žiarovka 8 W – G45/E27/SMD/4000 K v rozobratom stave.

Spôsob osvetlenia klasickými žiarovkami bol a je energeticky neefektívny, a preto sa od jeho používania upustilo. V Spojených štátoch amerických začalo postupné vyraďovanie žiaroviek v roku 2007 v súlade so zákonom o energetickej nezávislosti a bezpečnosti Energy Independance and Security Act. V Európskej únii bol predaj klasických žiaroviek zakázaný od roku 2012 a v USA vstúpil v roku 2023 do platnosti zákaz výroby a predaja týchto žiaroviek prijatý administratívou prezidenta Bidena [6].

B. Farby svetla LED diódy a žiarovky

Pomerne dlho trval vývoj LED diódy emitujúce modre a vysoko svietivé biele svetlo. Staršie biele LED používali kombináciu troch čipov na miešanie farieb, aby sa dosiahol efekt bieleho svetla. keďže priamy výstup bieleho svetla nie je možný, LED diódy využívajú luminofor. Niektoré LED emitujú modré svetlo, ktoré je na čipe transformované luminoforom na žlté, čím vzniká biele svetlo. Iné typy bielych LED využívajú ultrafialové žiarenie, ktoré sa tiež transformuje na biele svetlo pomocou luminoforu [7].

LED vyžarujú svetlo v rámci spektra (Obr. 5) , ktoré je definované maximálnou vlnovou dĺžkou (IPK), meranou v nanometroch. Táto vlnová dĺžka závisí od materiálu LED čipu, pričom výrobné odchýlky môžu byť až ±10 nm. Ľudské oko je najcitlivejšie na spektrum medzi 565 a 600 nm, preto je ľahšie rozoznať farby žltých a oranžových LED ako u iných farieb.



Obr. 5 Vlnové dĺžky farieb viditeľného svetla.

C. Svetelný výkon

Svetelný výkon LED závisí od rôznych faktorov, ako je typ čipu, zapuzdrenie a účinnosť doštičiek. Výrobcovia niekedy používajú pojmy ako "super-bright" alebo "ultra-bright", no tieto termíny nemajú jednotný priemyselný štandard. Svietivosť LED sa meria v mcd, čo sa líši od merania pri bežných žiarovkách s vláknom [7].

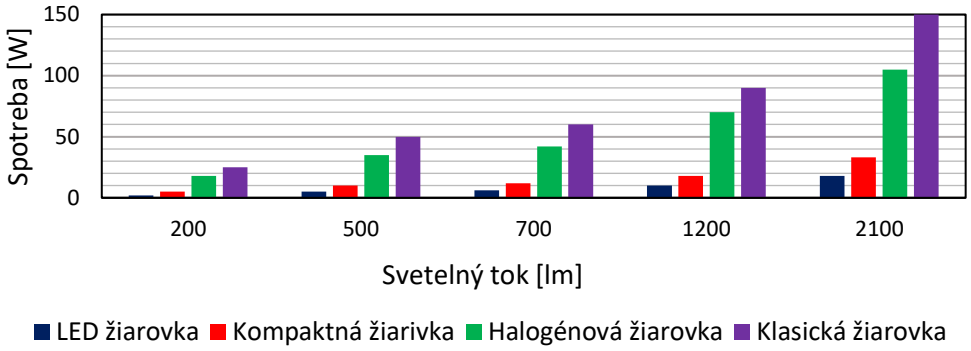
Svietivosť LED závisí od prúdu, ktorý prechádza čipom – čím väčší prúd, tým viac svetla sa vyprodukuje. LED sú navrhnuté pre prúd 20 mA, ale vyšší prúd zvyšuje produkciu tepla. Viac LED na čipe tiež zvyšuje tepelné zaťaženie. Vyššie napätie vedie k väčšiemu množstvu odpadového tepla. LED sú navrhnuté tak, aby si udržali dlhú životnosť pri dodržiavaní správnych prevádzkových podmienok.

Svetelný výkon je kľúčový pri výbere žiarovky, pretože určuje, či bude miestnosť dostatočne osvetlená. Tento výkon sa udáva v lúmenoch. Napríklad 40W žiarovka má svetelný výkon približne 420 lm. Pri nahrádzaní tradičnej žiarovky s LED alebo žiarivkou je dôležité skontrolovať svetelný tok v lúmenoch na obale alebo energetickom štítku [8].

Rôzne typy žiaroviek spotrebujú rôzne množstvo energie na vytvorenie rovnakého svetla. V tabuľke 1 je uvedené, koľko wattov potrebujú LED žiarovky, kompaktné žiarivky, halogénové a klasické žiarovky na dosiahnutie určitého svetelného výkonu. Na Obr. 6 je znázornené grafické porovnanie spotreby energie jednotlivých typov žiaroviek.

Tabuľka 1
Porovnanie spotreby rôznych typov žiaroviek [9].

Svetelný tok	*LED žiarovka*	*Kompaktná žiarovka*	*Halogénová žiarovka*	*Klasická žiarovka*
200 lm	2 W	5 W	18 W	25 W
500 lm	5 W	10 W	35 W	50 W
700 lm	6 W	12 W	42 W	60 W
1200 lm	10 W	18 W	70 W	90 W
2100 lm	18 W	33 W	105 W	150 W



Obr. 6 Spotreba jednotlivých typov žiaroviek pri rovnakom svetelnom toku.

D. Vyžarovací uhol

Vyžarovací uhol závisí od typu LED čipu, šošovky z epoxidovej živice a umiestnenia diód v svietidle. Svetelný tok z LED je smerový, takže pre vyšší výkon je svetlo zamerané do úzkeho lúča. Svetlo sa viac rozptyľuje do strán, čím väčší je tento uhol.

Zapuzdrenie LED slúži ako šošovka, ktorá zosilňuje svetlo vychádzajúce z čipu. Farba zapuzdrenia ovplyvňuje viditeľnosť vyžarovaného svetla. Difúzna úprava zabezpečuje rovnomerný rozptyl svetla, zatiaľ čo nedifúzne a priesvitné LED zariadenia produkujú jasnejšie svetlo s užším uhlom viditeľnosti.

Ako uhol vyžarovania sa najčastejšie počíta ten priestor, do ktorého dopadá aspoň 50 % z intenzity svetla, ktorým svetelný zdroj svieti kolmo. Ak by svetlo svietilo na všetky strany s veľkou intenzitou, bol by uhol vyžarovania 360°. LED osvetlenie je možné vyrobiť s vyžarovacím uhlom 15–360°.

E. Výhody a nevýhody LED žiaroviek

LED žiarovky sú v posledných rokoch čoraz populárnejšie, pretože v porovnaní s tradičnými žiarovkami ponúkajú mnoho výhod, ktoré vedú k ich postupnému nahradeniu. Avšak majú aj niektoré nevýhody, ktoré treba zvážiť. Základné výhody LED žiaroviek sú:

- Dlhá životnosť – LED žiarovky môžu vydržať až päťdesiat krát dlhšie než tradičné žiarovky, pričom ich životnosť dosahuje až päťdesiattisíc hodín. Okrem toho zvládajú až dvadsaťpäťtisíc cyklov zapnutia a vypnutia. Vďaka týmto vlastnostiam sú nielen ekonomickejšie, ale aj odolnejšie. Navyše, pri ich výrobe sa nepoužívajú škodlivé látky ako ortuť, čo má pozitívny vplyv na životné prostredie. [11]
- Bezpečnosť – Štandardná žiarovka produkuje 80% tepla a iba 20% svetla, zatiaľ čo LED diódy fungujú opačne. Preto sú bezpečnejšie a častejšie sa používajú na miestach, kde by nadmerné teplo mohlo byť problémom, ako napríklad v blízkosti textílií, postelí, tienidiel, nábytku alebo na miestach, kde chceme udržať nižšiu teplotu, ako sú špajze alebo chladničky. V domácnostiach je vhodné inštalovať LED žiarovky do svietidiel, ktoré sú ľahko prístupné, najmä v miestach, kde sa môžeme lampy náhodne dotknúť alebo v blízkosti materiálov ako drevo či papier.
- Úspora energie – Energetická účinnosť je jedným z hlavných dôvodov, prečo si ľudia vyberajú LED osvetlenie. LED žiarovky spotrebúvajú menej energie ako bežné osvetlenie, čím znižujú náklady na elektrinu. To je spôsobené tým, že LED diódy produkujú veľmi málo tepla a väčšina spotrebovanej energie sa premieňa na svetlo. Úspory na elektrine môžu byť niekoľko desiatok eur mesačne, v závislosti od typu žiaroviek a počtu svetiel.
- Cena – Mnohí ľudia vnímajú vyššiu cenu LED žiaroviek ako nevýhodu, no ide o jednorazový výdavok, ktorý sa z dlhodobého hľadiska oplatí. Klasické žiarovky majú životnosť približne dvetisíc hodín, zatiaľ čo LED osvetlenie vydrží až dvadsaťpäťkrát dlhšie, čo znamená, že ich bude treba meniť menej často a ušetríme peniaze v budúcnosti.

- Vplyv teploty – Hoci sú LED diódy bezpečné a energeticky úsporné, sú citlivé na vysoké teploty, čo môže viesť k zníženiu ich svietivosti. V bežných domácich podmienkach však takéto extrémne situácie nenastávajú. Pri vonkajšom použití je dôležité zabezpečiť dostatočnú ochranu svietidiel, pričom minimálna úroveň krytia by mala byť IP40, aby sa zabránilo vniknutiu vody a poškodeniu. Dôležitá je aj pravidelná údržba, ktorá je nevyhnutná pri LED osvetlení.
- LED žiarovky neobsahujú krehké prvky, ako sú napríklad sklenené tyčinky v klasických žiarovkách, takže nie sú náchylné na poškodenie pri ľahkých otrasoch alebo nárazoch. To znamená, že majú dlhšiu životnosť a sú odolnejšie. Táto odolnosť zlepšuje nielen ich bezpečnosť, ale aj ich trvanlivosť. LED žiarovky nevyžadujú tienidlá na ochranu, sú bezpečné samy o sebe a môžu slúžiť aj ako dekorácia.

Tu sú niektoré z hlavných nevýhod LED žiaroviek:

- Nedostatočný rozptyl svetla môže byť problematický, najmä pri doplnkovom osvetlení, Efektívnym riešením sú LED filamentové žiarovky, ktoré vďaka rovnomernému rozmiestneniu diód poskytujú vyvážené osvetlenie. Na zlepšenie rozptylu svetla je možné použiť napríklad mliečne alebo matné sklo. Moderným trendom sú aj dekoratívne COB LED diódy, ktoré vyžarujú svetlo po celej svojej dĺžke a zabezpečujú homogénnejšie osvetlenie.

III. Nepriaznivý vplyv LED žiaroviek na napájaciu sieť

Aby sa kvalita napätia vo verejnej elektrickej sieti zachovala aj v budúcnosti na vysokej úrovni, je nevyhnutné prijať určité opatrenia. Kvalita elektriny, podobne ako kvalita iných výrobkov, nezávisí len od samotnej výroby, ale vo veľkej miere aj od správania a rozhodnutí konečných užívateľov.

Zatiaľ čo v minulosti sa elektronické zariadenia navrhovali s veľkými rezervami a dôrazom na spoľahlivosť, dnešným cieľom výrobcov je ponúkať zariadenia čo najmenšie, ľahké a najmä cenovo dostupné. Tento trend je možný vďaka využitiu moderných polovodičových súčiastok a miniaturizovaných napájacích zdrojov.

Jedným z dôsledkov tohto vývoja je zaťaženie elektrickej siete vyššími harmonickými, hlavne nepárnymi, ktoré vznikajú kvôli čoraz častejšiemu používaniu elektroniky. V osvetlení k tomu výrazne prispievajú LED svietidlá, ktoré nahrádzajú klasické žiarovky. Ich napájacie obvody sa správajú ako nelineárne záťaže, čo spôsobuje deformáciu prúdu a vznik vyšších harmonických.

V posledných rokoch sa technológia LED stala preferovaným riešením osvetlenia v domácnostiach, priemysle a verejných priestoroch. Je široko prijímaný najmä pre svoju veľmi efektívnu spotrebu energie, dlhú životnosť a všade použiteľnosť. Okrem týchto výhod má LED osvetlenie samozrejme aj technické nevýhody, ktoré sa prejavujú najmä pri interakcii s napájacou sieťou [3].

A. Vyššie harmonické a LED žiarovky

Vyššie harmonické sú dnes považované za jeden z hlavných ukazovateľov kvality elektrickej energie. Spôsobujú deformáciu napätia a prúdu v sieti. Ich výskyt sa začal výraznejšie prejavovať od 70. rokov 20. storočia, kedy sa začali vo väčšej miere používať polovodičové súčiastky. Rozvoj moderných technológií, ako sú aj LED svietidlá, prispel k ďalšiemu nárastu týchto harmonických. Kým v prenosových sieťach býva skreslenie napätia väčšinou nižšie ako 1–2 %, na úrovni koncových odberateľov tieto hodnoty často rastú [10].

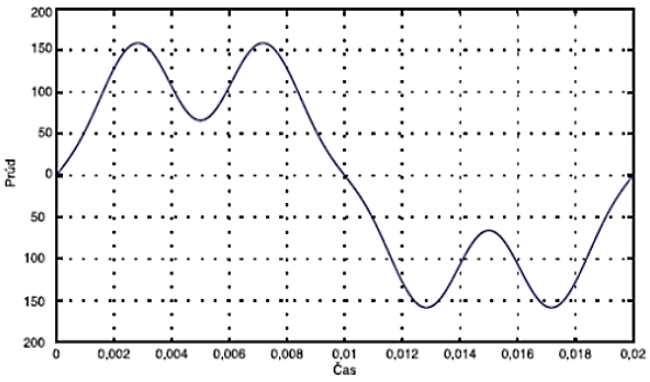
Súčtom viacerých harmonických zložiek, ktorých frekvencia je nepárnym násobkom 50 (napr. 150 Hz, 250 Hz), napríklad prvej, tretej a piatej harmonickej, vznikne výsledný priebeh, ktorý sa líši od ideálnej sínusovej krivky. Takýto signál obsahuje vyššie harmonické frekvencie a predstavuje skreslený, neharmonický priebeh elektrickej veličiny (Obr. 7). Harmonické skreslenie LED svietidiel môžu spôsobovať rôzne faktory:

- Elektronické faktory:

1. Nelinearita napájacích obvodov (LED driver): Väčšina LED svetiel obsahuje spínaný zdroj (napr. buck, boost alebo flyback konvertor). Tieto obvody neriadia prúd plynule, ale v rýchlych prepínacích cykloch, čím vznikajú vyššie harmonické zložky v prúde.

2. Usmerňovače a filtračné kondenzátory: Striedavé napätie sa najprv usmerní a následne vyhladí. Prúd tečie len v krátkych impulzoch pri špičkách sínusovky (keď je napätie vyššie ako napätie na kondenzátore), čo vytvára harmonické zložky.

3. Chýbajúce alebo nedostatočné korekcie účinníka (PFC): LED svetlá bez aktívnej korekcie účinníka (Power Factor Correction) majú nízky účinník a vysoké THD (Total Harmonic Distortion). Napr. lacné žiarovky môžu mať THD > 100%, čo značí veľmi skreslený prúd.



Obr. 7 Priebeh prúdu s obsahom vyšších harmonických.

• Fyzikálne faktory:

1. Prúdovo riadené LED diódy: LED diódy sú prúdovo závislé zariadenia. Ich charakteristika je nelineárna – malé zmeny napätia spôsobujú veľké zmeny prúdu. Táto nelinearita prispieva k vzniku vyšších harmonických, ak nie je dostatočne lineárne riadený prúd.

2. Teplotná závislosť LED a súčiastok: Zmeny teploty ovplyvňujú vodivosť polovodičových prvkov, čo mení ich správanie a môže spôsobovať kolísanie alebo dodatočné harmonické zložky.

B. Význam účinníka pri LED osvetlení

Vo všeobecnosti platí, že s rastúcim výkonom zariadenia rastie aj prúd a zároveň svetelný tok. Táto priama úmera však neplatí vždy, a práve odchýlky od tohto vzťahu súvisia s pojmom účinník a celkovou energetickou efektivitou LED zariadení [12].

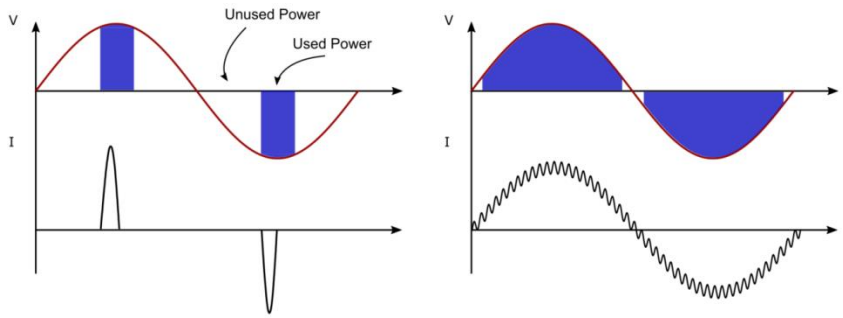
Pri prevádzke svetelných zdrojov elektrická energia odoberaná zo siete pozostáva z dvoch zložiek: činného výkonu, ktorý sa skutočne využíva na osvetlenie a jalového výkonu, ktorý predstavuje straty spôsobené reaktívnou záťažou.

Účinník má priamy vplyv na efektívnosť osvetľovacích zariadení. Čím nižšia je jeho hodnota, tým menej efektívne svietidlo premieňa zdanlivý výkon na činný výkon potrebný na osvetlenie. Svietidlá s nízkym účinníkom vytvárajú vyšší jalový výkon, ktorý zbytočne zaťažuje elektrickú sústavu bez prínosu pre samotnú funkciu zariadenia. V domácich inštaláciách účinník priamo neovplyvňuje cenu elektriny, keďže sa účtuje podľa spotreby v kWh. Nízky účinník však zvyšuje zdanlivý výkon, čo môže vyžadovať hrubšie vodiče a výkonnejšie ističe. Preto je potrebné zohľadniť účinník už pri návrhu elektroinštalácie. V komerčných a priemyselných objektoch sa cena elektriny často skladá zo spotreby a odberového maxima. Zariadenia s nízkym účinníkom, ako je osvetlenie, môže zvýšiť odberové maximum, čím rastú náklady. Celková cena zahŕňa obe tieto zložky. Z uvedeného vyplýva, že pri výbere svietidiel je vhodné zohľadniť účinník. Nízky účinník môže v domácnostiach vyžadovať väčšie dimenzovanie elektroinštalácie a v komerčných objektoch zvýšiť prevádzkové náklady na elektrinu.

C. Význam PFC v napájacích zdrojoch pre LED žiarovky

Korekcia účinníka (PFC – Power Factor Correction) je dôležitou funkciou napájacích zdrojov s konštantným napätím, ktoré sa používa pri LED osvetlení. Hlavnou funkciou je zvýšenie účinnosti zariadenia a minimalizácia jeho negatívneho vplyvu na elektrickú sieť. PFC zabezpečuje, aby prúd odoberaný zariadením bol vo väčšej miere vo fáze s napätím. V praxi mnohé elektrické zariadenia, vrátane LED napájacích zdrojov, odoberajú prúd, ktorý má nesínusový priebeh. Tento prúd môže spôsobovať nežiaduce javy v distribučnej sieti, ako je nadmerné zahrievanie vodičov alebo kolísanie napätia. Princíp spočíva v akumulácii energie v kondenzátore počas časti striedavého cyklu, keď je prúd vyšší ako napätie. Táto energia sa následne uvoľní v časti cyklu, keď je prúd nižší. Tým sa dosiahne plynulejší, sínusovejší priebeh prúdu, čo vedie k zlepšeniu účinníka [13].

Použitie korekcie účinníka prináša viacero výhod. Jednou z nich je zvýšená účinnosť napájacieho zdroja, pretože znižuje tepelné straty – účinnosť sa tak môže zlepšiť až o 10%. Zároveň sa znižuje prúd, ktorý zariadenie odoberá zo siete, čím sa znižuje jej zaťaženie a zlepšuje stabilita. V niektorých krajinách je tiež požiadavka, aby zaradenia spĺňali určité normy PFC, takže pomáha aj pri dodržiavaní predpisov. Na Obr. 8 je znázornený priebeh napätia a prúdu s PFC (Obr. 8a) a bez PFC (Obr. 8b).



Obr. 8 Priebeh napätia a prúdu s nízkym účinníkom bez PFC (a) a s korigovaným účinníkom a PFC (b) [14].

Korekcia účinníka v LED napájacích zdrojoch sa vykonáva dvoma spôsobmi – pasívnym a aktívnym. Pasívna PFC využíva jednoduché súčiastky, ako sú tlmivka a kondenzátor. Je lacnejšia, ale menej účinná. Aktívna PFC obsahuje zložitejší obvod so spínačom a riadením. JE síce drahšia, ale účinnejšia a dosahuje lepšie výsledky pri zlepšovaní účinníka. Pri výbere napájacieho zdroja pre LED osvetlenie je dôležité zohľadniť hodnotu účinníka. Pokiaľ je to možné, odporúča sa zvoliť zariadenia s aktívnou PFC, ktoré zabezpečí vyššiu efektívnosť a menšie zaťaženie elektrickej siete. Korekcie účinníka okrem lepšej účinnosti a menšieho zaťaženia siete pomáha aj predĺžiť životnosť LED svietidiel, pretože znižuje ich namáhanie. Celkovo je PFC dôležitou súčasťou napájacích zdrojov pre LED osvetlenie [15].

D. Jalový výkon a LED osvetlenie

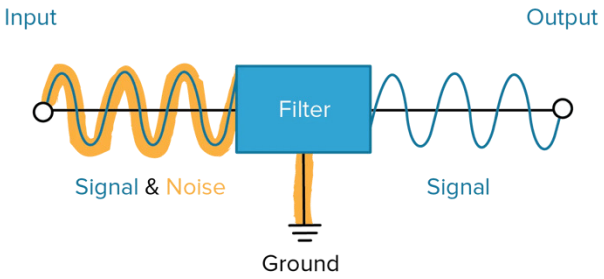
Z hľadiska prevádzky energetickej sústavy zohráva jalový výkon významnú úlohu. Je dôležitý najmä pre správnu funkciu indukčných zariadení a pomáha udržiavať stabilné napätie v sieti. Na druhej strane však spôsobuje zvýšenie prúdu v elektrickej sieti, čo má za následok vyššie energetické straty.

Moderné LED osvetlenie síce môže znížiť spotrebu elektrickej energie až o približne 70% avšak zároveň do elektrickej siete vnáša kapacitný jalový výkon. Pre svoju činnosť LED diódy vyžadujú napájací zdroj, ktorý generuje určitý objem jalového výkonu, pričom jeho množstvo závisí od kvality spracovania a použitých komponentov zo strany výrobcu. Kvalitné napájacie zdroje bývajú vybavené ochrannými obvodmi a prvkami na kompenzáciu jalového výkonu, pričom sa vyznačuje vysokým účinníkom cos φ. Naopak, svietidlá s nekvalitnými zdrojmi môžu produkovať výrazné množstvo jalového výkonu, čo môže viesť k tomu, že celkové náklady na elektrickú energiu neklesnú, ale dokonca vznikne vyššia spotreba než pri starších žiarovkových svetelných zdrojoch, ktoré jalový výkon neodoberajú. Z tohto dôvodu je pri výmene klasického osvetlenia za LED technológiu dôležité skontrolovať účinník cos φ, keďže výrazne ovplyvňuje výslednú výšku platieb za elektrinu [16].

E. Elektromagnetické rušenie

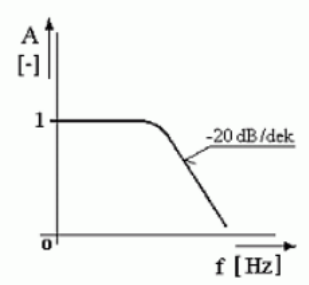
Elektromagnetické rušenie (EMI) vzniká, keď elektrický prúd prechádza vodičmi a súčiastkami a vytvára magnetické pole, ktoré sa šíri do okolia. Rozsah tohto vyžarovania závisí od veľkosti prúdu, napätia, frekvencie a konštrukcie zariadenia – napríklad od tienenia vodičov, typu krytu alebo uzemnenia.

Vo výkonovej elektronike sa používajú EMI filtre na potlačenie a odstránenie neželaného vysokofrekvenčného rušenia, ktoré vzniká pri spínaní v elektronických obvodoch. Keramické EMI filtre zabezpečujú, že rušivé vysoko frekvenčné signály majú možnosť prejsť cez nízko impedančnú cestu do zeme, zatiaľ čo užitočný napájací signál môže prechádzať obvodom bez útlmu (Obr. 9).



Obr. 9 Základný princíp fungovania EMI filtra.

EMI filter teda funguje ako dolno-priepustný filter (Obr. 10), ktorý zoslabuje vysoké frekvencie a bráni ich šíreniu v systéme [25].

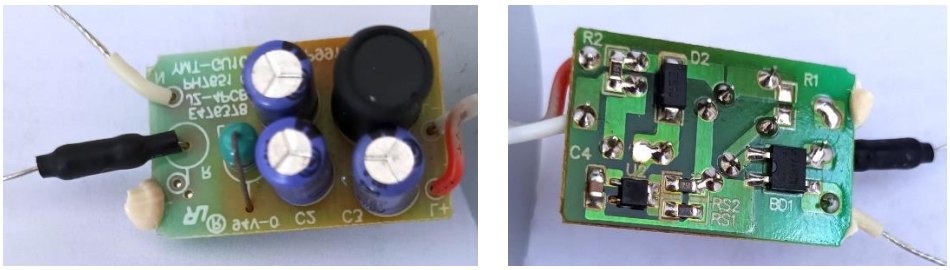


Obr. 10 Amplitúdovo-frekvenčná charakteristika filtru typu dolný priepust.

EMI filtre zvyšujú spoľahlivosť elektronických zariadení tým, že znižujú rušenie a chránia citlivé súčiastky pred poruchami alebo chybami. Ich použitie je bežné v napájacích zdrojoch, meničoch či LED osvetlení – teda v oblastiach, kde je potlačenie rušenia mimoriadne dôležité. Vďaka správnemu filtrovaniu môžu byť tieto zariadenia bezpečne nasadené aj v kritických aplikáciách, ako sú zdravotnícke prístroje alebo letecká technika, kde si poruchy nemožno dovoliť.

Na výrobu EMI filtrov sú vhodné keramické materiály vďaka dobrým elektrickým vlastnostiam. Používajú sa najmä viacvrstvové keramické kondenzátory, ktoré umožňujú účinne odfiltrovať vysokofrekvenčné rušenie. Tieto filtre sa zapájajú buď priamo na napájacie vodiče, alebo medzi zdroj a chránené zariadenie. Môžu byť súčasťou plošného spoja alebo zapojené samostatne, pričom ich počet a usporiadanie závisí od náročnosti konkrétnej aplikácie.

LED žiarovky obsahujú elektronický menič (Obr. 11), ktorý premieňa striedavé napätie (230 V AC) na jednosmerné (DC) vhodné pre LED diódy.



Obr. 11 AC-DC Menič LED žiarovky 10 W – 806 lm/6000 K.

Súčasťou tohto meniča je veľmi často spínaný zdroj, ktorý pracuje na vysokých frekvenciách (desiatky až stovky kHz). Toto vysokofrekvenčné spínanie spôsobuje prudké zmeny prúdu a napätia, čo generuje elektromagnetické rušenie. Ak nie je tento menič dobre odtienený alebo filtrovaný, elektromagnetické rušenie sa môže šíriť do okolia. Ako je všeobecne známe, tak veľmi lacné alebo nekvalitné LED žiarovky často nemajú dostatočné EMI filtre a preto spôsobujú vyššie elektromagnetické rušenie.

IV. ZÁVER

LED osvetlenie je dnes neoddeliteľnou súčasťou moderného spôsobu osvetľovania. Vďaka svojej vysokej účinnosti, nízkej spotrebe elektrickej energie a dlhej životnosti sa stalo prirodzenou náhradou za klasické svetelné zdroje. Hoci sú výhody LED technológie nepopierateľné, v článku sa poukázalo na fakt, že jej masové rozšírenie môže mať aj negatívny vplyv na kvalitu energie v napájacej sieti.

LED svietidlá predstavujú nelineárne zaťaženie, ktoré generuje vyššie harmonické prúdy, znižuje účinník a môže spôsobovať elektromagnetické rušenie. Tieto javy sa nepriaznivo prejavujú v celej sieti, nielen v mieste vzniku. Problémom sa stáva najmä ich hromadné používanie v obytných aj priemyselných objektoch, kde sa ich účinky navzájom sčítavajú.

Pri návrhu a používaní LED osvetlenia nestačí myslieť len na úsporu energie. Dôležité je zohľadniť aj to, ako tieto svetelné zdroje vplývajú na sieť. Aby sa predišlo problémom, je vhodné vyberať kvalitné LED žiarovky, ktoré majú dobrú korekciu účinníka (PFC), prípadne používať zariadenia s filtrom, ktoré znižujú EMI.

Vyššie harmonické prúdy, ktoré LED osvetlenie vytvára, neovplyvňujú len samotné svietidlo, ale môžu zasiahnuť aj ďalšie zariadenia pripojené v sieti. Preto je dôležité, aby projektanti, výrobcovia aj bežní používatelia vedeli o týchto javoch. Jedna žiarovka problém nespôsobí, no vo veľkom množstve sa účinky môžu sčítať a spôsobiť komplikácie v celom obvode.

POĎAKOVANIE

Táto práca bola podporená projektom vedeckej agentúry KEGA podľa zmluvy číslo: 032TUKE-4/2024.

LITERATÚRA

[1] Srdečný, K., Macholda, F., „Úspory energie v dome“. Praha: Grada Publishing, 2004. ISBN 80-247-0523-0
[2] Úsporná žiarovka. Najpoužívanejšie typy žiaroviek. . [online], [cit. 15.1.2025]. Dostupné na internete: <https://www.uspornaziarovka.sk najpo uzivanejsie-typy-ziaroviek/ >.
[3] Kowalska, A., Napierala, J., Staszewski, M., Adamczewski, M., „Effect of LED Lighting on Selected Quality Parameters of Electricity“. Sensors, 2023, Vol. 23, No. 3, ´pp. 1582-1589.
[4] Svet svietidiel. História a zaujímavosti o LED osvetlení. [online], [cit. 15.12.2024]. Dostupné na internete: <https://www.svet-svietidiel.sk/novinky/detail/historia-a-zaujimavosti-o-led-osvetleni/ >.
[5] Futurezone Newsletter. LED-Lampe: 100 Watt Licht mit 14 Watt Verbrauch. [online], [cit. 15.2.2022]. Dostupné na internete: <https://futurezone.at/produkte/led-lampe-100-watt-licht-mit-14-watt-verbrauch/24.567.356 >.
[6] Encyclopedia Britannica. LED. [online], [cit. 8.5.2025]. Dostupné na internete : < https://www.britannica.com/ technology/LED >.
[7] Ako LED diódy fungujú. . [online], [cit. 8.5.2025]. Dostupné na internete internete : < https://www.123led.sk/led-diody-technicke-udaje/ >.
[8] Ako sú označené LED žiarovky. [online], [cit. 13.12.2024. Dostupné na internete: < https://www.uspornaziarovka. sk/ako-su-oznacene-led-ziarovky/ >.
[9] Rozdiel medzi klasickými a LED žiarovkami. [online]. [cit. 16.4.2025], Dostupné na internete: https://revolo.sk/ rozdiel-medzi-klasickymi-a-led-ziarovkami-/?srsltid=AfmBOoplxyr2mqD8OU4ZVu4386eAW Texln0uV9KOBhaRgxyD9lOSgf-w/ >.
[10] Szathmáry, P., „Kvalita elektrickej energie“, PRO Vydavateľstvo, Banská Bystrica, 2003. ISBN: 80-89057-04-7.
[11] KANLUX. LED – plusy a mínusy. [online], [cit. 16.4.2025]. Dostupné na internete: https://www.kanlux.com/ sk /clanky/LED-plusy-minusy/ >.
[12] Power factor for LED light sources. [online], [cit. 5.5.2025]. Dostupné na internete: https://ansell-lighting. com/en/ articles/news/power-factor-for-led-light-sources/ >.
[13] Vplyv jalového výkonu na LED osvetľovacie zariadenia. [online], [cit. 7.5.2025], Dostupné na intetnete: < https://onninen.pl/sk/clanok/vplyv-jaloveho-vykonu-na-led-osvetjovacie-zariadenia>.
[14] Power Factor Correction (PFC) Explained. Monolithic Power Systems [online]. [cit 9.5.2025]. Dostupné na internete: < https://www.monolithicpower.com/en/learning/resources/power-factor-correction?srsltid =AfmBOor LO8ler7W4G9_fZL0wWRGjcP6X0VbHosWoDc4oC3PC8riiA1B5 >.
[15] Yang, Y., „What is PFC on an LED constant voltage power supply? „[online], [cit. 9.5.202]. Dostupné na internete: <https://www.linkedin.com/pulse/what-pfc-led-constant-voltage-power-supply-yanhong-yang-7oqyc>.
[16] Vplyv jalového výkonu na LED osvetľovacie zariadenia. [online], [cit. 9.5.202]. Dostupné na internete: https://onninen.pl/sk/clanok/vplyv-jaloveho-vykonu-na-led-osvetjovacie-zariadenia
[17] Yates, J. „What are EMI Filters?“, [online], [cit. 22.5.2025], Dostupné na internete: <https://blog.knowles capacitors.com/blog/what-are-emi-filters>

Creation of a Tool for Extracting Metadata from Student Technical Documentation Works

[1]Tomáš BUČEK,[2]Martin PACANOVSKÝ,[3]Ján GENČI

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

[1]tomas.bucek@tuke.sk, [2]martin.pacanovsky@student.tuke.sk, [3]jan.genci@tuke.sk

Abstract **– This paper deals with the topic of metadata extraction and analysis from text documents in various formats (DOC, DOCX, ODT, PDF). The aim of the paper is to design and implement a tool capable of automatically processing files and retrieving important information such as the creation date, last modification date, editing time, or number of pages. The work uses a combination of XML parsing techniques, binary file processing, and regular expressions in the Python programming language. The result is software capable of sorting and verifying documents in terms of their authenticity. This tool can be especially useful in academic environments, particularly for checking the origin and content of submitted documents.**

Keywords **– metadata, extraction, automatization**

I. Introduction

The authenticity of digital documents is a real concern in modern evaluation of student works. With the increasing volume of student submissions in electronic formats, teachers require efficient tools to validate not only content but also contextual metadata such as editing time, creation timestamps or their absence. This paper presents the development of a tool for metadata extraction and verification from student works, with the goal of automating and supporting document evaluation processes.

The tool was developed to support mass document analysis in courses such as Operating Systems at the Technical University of Košice. By analyzing internal metadata of document formats like DOCX, ODT, and PDF, the system identifies potentially suspicious documents that may indicate plagiarism or questionable authorship. The software is implemented in Python and processes both individual documents and ZIP archives typically submitted by students.

II. System Background

Digital documents typically contain more than just visible content. In many cases, they also include embedded metadata, which can be used for various analytical, verification, and forensic purposes. This section provides an overview of how metadata are represented in different document formats and why they are relevant in detecting anomalies or potential academic misconduct.

A. Metadata in Documents

Metadata are defined as data about data and they play a fundamental role in document management and verification. Common types of document metadata are author name, creation date, last modified timestamp, total editing time, page count, and file format. These values are mostly embedded automatically by processing software and can be viewed to retrieve extra information about a specific document [1].

Microsoft Word (DOCX), LibreOffice (ODT), and Adobe Acrobat (PDF) each utilize different metadata structures:

- **DOCX and ODT** - store metadata in XML format embedded in ZIP archives.
- **PDF** - uses binary-based metadata stored in an Info Dictionary.

Metadata like this can provide insight into how much time was invested into document creation, whether the file has been modified recently, or if it was copied from another source [2]. All of these information are beneficial for teachers for grading and examining student works [3].

B. Motivations and Use in University

Teachers often receive hundreds of assignments, especially in technical courses like Operating Systems. Manual verification of authorship or suspicious documents is inefficient and often inconsistent [4]. Students may attempt to reuse old documents or submit plagiarized work, and while plagiarism checkers focus on content similarity, metadata reveal structural anomalies that text-based methods may miss.

An example of a suspicious document is an assignment that was created months or even years before the current academic term, or one with an unusually short total editing time, such as only three minutes for a twelve-page paper. Other examples include files submitted in outdated formats like DOC, or submissions that are protected by encryption or lack basic metadata. The developed system addresses these concerns by performing batch metadata extraction and generating alerts based on predefined evaluation criteria.

C. Limitations of Existing Solutions

Online tools and mostly older applications claim to extract metadata, such as Metadata2Go and PDFYeah.

Such tools have, according to [5], several limitations. They do not support batch processing or compressed ZIP files, and they are not privacy-safe for handling sensitive student submissions. Furthermore, they lack detection logic for various anomalies, do not support alternative file formats such as ODT, and are often restricted by paywalls or usage limitations. File properties shown in Windows Explorer or within document editors are additionally limited by differing operating environments. Therefore, a new specialized system was developed to address these issues and integrate directly into university courses, effectively resolving the limitations of such tools [6].

III. System Requirements and Design Goals

A. Functional Requirements

The tool was designed with the following functional goals: It supports document formats including DOCX, ODT, and PDF. It enables the extraction of key metadata fields such as title, author, creation date, modification date, total edit time, page count, and encryption status. Additionally, it is capable of scanning ZIP archives and processing all nested documents. The system is designed to detect suspicious documents based on predefined evaluation rules and to generate color-coded summary reports in PDF format.

B. Non-functional Requirements

The system also must be simple to use within a terminal environment, ensure cross-platform compatibility for both Windows and Linux systems, and provide high performance, allowing the processing of hundreds of files in a single use.

C. Target Users and Environment

The main users are university techers of Operating Systems course working at TUKE. They are expected to use a terminal-based environment, to batch process submissions which are stored in structured ZIP archives downloaded from platforms such as Moodle.

The system is optimized for use explicitly in course Operating Systems at TUKE, but the modularity allows generalization to other institutions and education programs.

IV. System Architecture and Implementation

A. Overview of System Design

The developed system is implemented in the **Python** programming language, chosen for its simplicity and rich library options. The usage supports both single-file analysis and recursive batch processing of ZIP archives. The entire process does not alter the analyzed files (non-intrusive).

The program consists of several logically separated modules, each with clearly defined responsibility. All of the program files are managed by **"main.py"**. This separation allows for easy maintenance and potential future extensions, such as a graphical interface or additional file format extraction.

B. File Parsing Modules

Metadata extraction depends on the file format (PDF or DOCX/ODT). Each document type is handled by a specific parser module:

- **pdf_parser.py**: Extracts metadata by parsing Info Dictionary containing metadata of most PDF files.
- **doc_parser.py**: Parses DOCX, ODT and DOC files using zipfile and ElementTree to locate XML files and their properties.

Each parser is designed to extract the following metadata:

- Author and Title (folder and file name)
- Creation and Modification Dates
- Total Editing Time (DOCX, ODT and DOC)
- Estimated or actual Page Count
- Encryption or other restrictions

Estimation mechanisms are used when certain fields are missing. For example, page count may be estimated using paragraph count and average layout assumptions.

C. Document Evaluation Module

The **checker.py** module is responsible for rule-based evaluation of extracted metadata. Based on set rules, the system flags suspicious files using a simple traffic-light scheme:

- **Green**: Metadata appears correct (e.g., created recently, sufficient editing time).
- **Orange**: Metadata are missing or unusual, but not suspicious (e.g., missing editing time).
- **Red**: File has suspicious characteristics (e.g., old creation date, short editing time, encrypted content).

These rules are based on observation of real student submissions and document writing standards, such as 12 minutes to write a page.

D. Report Generation

The **report.py** module generates a color-coded PDF summary using the FPDF library, and saves it as ”report.pdf” in directory where the extraction occured. Generated report is also marked with a time stamp with date and time on which generation occured for administration purposes. Each entry contains the filename, extracted metadata, and an evaluation status.

The reports are optimized for print or digital archive and help instructors quickly identify problematic files and establish next steps.

V. Results Evaluation

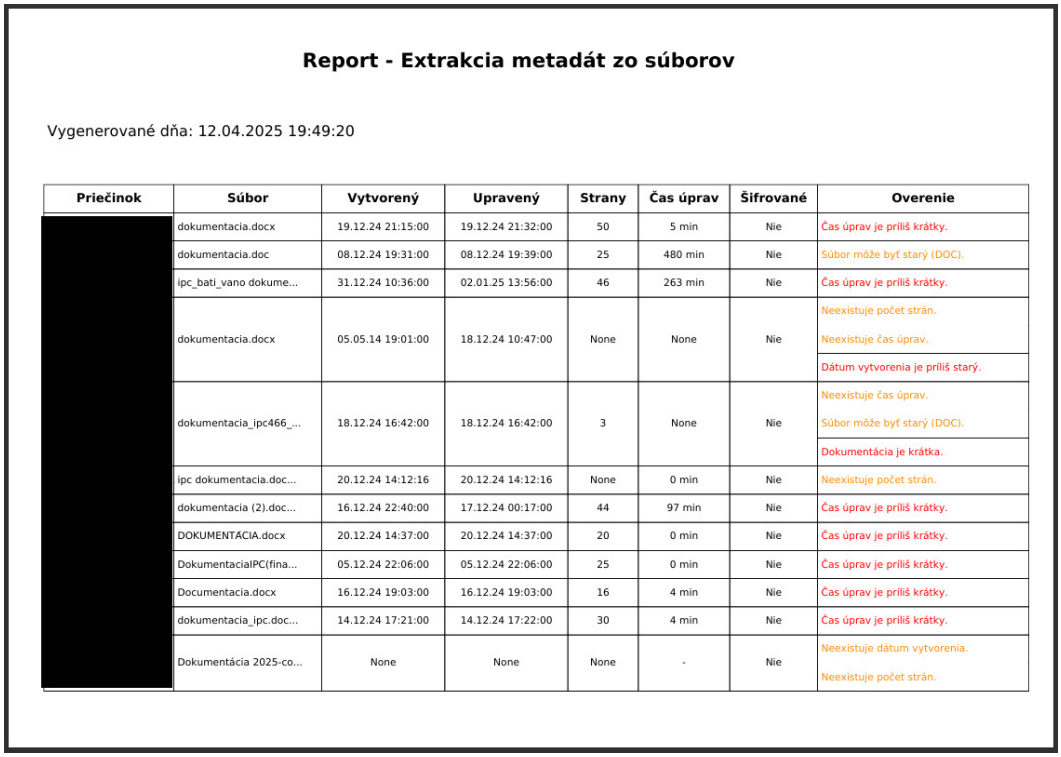
A. Testing the System

The system was tested using authentic student submissions collected during the 2014/2015 and 2024/2025 academic years. These samples were sourced from the Operating Systems course at the Technical University of Košice, which regularly receives a large volume of submissions in the form of ZIP archives. Each archive usually contains a single document (DOCX, ODT, or PDF), but the system will extract metadata even from multiple files in each archive.

The testing dataset (2014/2015) consisted of 127 ZIP files, each containing at least one student document for a total of 186 files, which is presented in Table 1. Results are presented in Tables 2 and 3. These documents were processed in batch mode. The tool automatically unzipped the content, located the documents, extracted relevant metadata, and applied evaluation rules. The resulting report was compared with manually verified outcomes to establish the reliability and performance of the tool.

Table 1
Overview of file metadata issues by file type

File Type	Number of Files	Missing Metadata	Faulty (%)
PDF	17	2	11.76
DOC/DOCX	151	19	12.58
ODT	18	5	27.78
All Files	**186**	**26**	**13.98**



Report - Extrakcia metadát zo súborov

Vygenerované dňa: 12.04.2025 19:49:20

Priečinok	Súbor	Vytvorený	Upravený	Strany	Čas úprav	Šifrované	Overenie
[redacted]	dokumentacia.docx	19.12.24 21:15:00	19.12.24 21:32:00	50	5 min	Nie	Čas úprav je príliš krátky.
	dokumentacia.doc	08.12.24 19:31:00	08.12.24 19:39:00	25	480 min	Nie	Súbor môže byť starý (DOC).
	ipc_bati_vano dokume...	31.12.24 10:36:00	02.01.25 13:56:00	46	263 min	Nie	Čas úprav je príliš krátky.
	dokumentacia.docx	05.05.14 19:01:00	18.12.24 10:47:00	None	None	Nie	Neexistuje počet strán. Neexistuje čas úprav. Dátum vytvorenia je príliš starý.
	dokumentacia_ipc466_...	18.12.24 16:42:00	18.12.24 16:42:00	3	None	Nie	Neexistuje čas úprav. Súbor môže byť starý (DOC). Dokumentácia je krátka.
	ipc dokumentacia.doc...	20.12.24 14:12:16	20.12.24 14:12:16	None	0 min	Nie	Neexistuje počet strán.
	dokumentacia (2).doc...	16.12.24 22:40:00	17.12.24 00:17:00	44	97 min	Nie	Čas úprav je príliš krátky.
	DOKUMENTACIA.docx	20.12.24 14:37:00	20.12.24 14:37:00	20	0 min	Nie	Čas úprav je príliš krátky.
	DokumentaciaIPC(fina...	05.12.24 22:06:00	05.12.24 22:06:00	25	0 min	Nie	Čas úprav je príliš krátky.
	Documentacia.docx	16.12.24 19:03:00	16.12.24 19:03:00	16	4 min	Nie	Čas úprav je príliš krátky.
	dokumentacia_ipc.doc...	14.12.24 17:21:00	14.12.24 17:22:00	30	4 min	Nie	Čas úprav je príliš krátky.
	Dokumentácia 2025-co...	None	None	None	-	Nie	Neexistuje dátum vytvorenia. Neexistuje počet strán.

Fig. 1 Example of generated report.pdf with extracted metadata and colored evaluation.

B. Results Summary

The extracted metadata and corresponding evaluation results were analyzed in terms of document format, metadata completeness, and detection accuracy.

Accuracy: The tool successfully flagged more than 95% of documents as suspicious. Most false negatives were due to minimal metadata availability (e.g., missing timestamps in PDFs) and unusually short total editing time for longer documents.

Processing Time: The average processing time for one batch was approximately 2 seconds. Batch processing of all 186 submissions with user input was completed in under 30 seconds, demonstrating high efficiency and scalability.

Format Support Results:

- **DOCX:** Provided the richest metadata set. Over 85% of files contained complete metadata, including editing time, creation/modification dates, and page count.
- **ODT:** Metadata availability varied significantly. Files created with LibreOffice preserved metadata well, but often lacked key fields.
- **PDF:** Metadata fields were frequently limited, 12% contained unreadable metadata, which triggered automatic red flags.

Table 2
Suspicion level in faulty files by type

File Type	Not Suspicious	Slightly Suspicious	Suspicious
PDF	0	2	0
DOC/DOCX	2	3	14
ODT	0	0	0
All Files	**2**	**5**	**14**

Table 3
Suspicion level across all submitted files

File Type	Not Suspicious	Slightly Suspicious	Suspicious
PDF	11	2	4
DOC/DOCX	38	15	98
ODT	3	1	14
All Files	**52**	**18**	**116**

C. Analysis of Suspicious Patterns

Suspicious files are the ones with:

- **Old creation dates:** Some documents had timestamps dated years before the semester started.
- **Unusually low edit time:** Several submissions showed under 5 minutes of total editing time for multi-page documents.
- **DOC format submissions:** Although outdated, some files were still submitted as DOC format, which lacks structured metadata and was automatically flagged orange color.
- **Short documetations:** Documents with really low page count (less than 10 pages) were marked red.
- **Encryption and protection:** Documents with password protection or inaccessible metadata sections were flagged red by default due to evaluation uncertainty.

VI. Conclusion

This paper presented a metadata extraction tool designed to improve the efficiency and integrity of student document evaluation in academic environments. Implemented in Python with support for major document formats and ZIP archives, the system enables instructors to detect suspicious files and anomalies related to document authorship and authenticity.

Test results confirmed its usefulness in real-world scenarios, highlighting its performance, accuracy, and simplicity. The tool represents a meaningful addition to the digital academic toolkit, complementing plagiarism detectors with structural metadata verification.

In addition to the tested course, the system has potential applications across various academic disciplines, including thesis verification, coursework review in humanities and engineering, or early detection of atypical document creation behavior. It can also be integrated into existing plagiarism detection workflows to enhance the robustness of academic integrity checks.

Future improvements include the addition of a graphical interface, support for cloud-based operations, and integration with learning management systems. The project also opens the door for further research in document forensics and academic integrity analytics.

References

[1] J. Schneider, A. Bernstein, J. Vom Brocke, K. Damevski, and D. C. Shepherd, "Detecting plagiarism based on the creation process," *arXiv preprint arXiv:1612.09183*, 2016. [Online]. Available: https://arxiv.org/abs/1612.09183

[2] C. Johnson and R. Davies, "Using digital forensic techniques to identify contract cheating: A case study," *Journal of Academic Ethics*, vol. 18, no. 1, pp. 105–113, 2020.

[3] e. a. Johnson, "Using digital forensics in higher education to detect academic misconduct," *International Journal for Educational Integrity*, vol. 18, no. 12, 2022.

[4] S. Edwards and C. Cramer, "Are we ready for large-scale use of plagiarism detection tools?" *Assessment Evaluation in Higher Education*, vol. 40, no. 5, pp. 698–714, 2015.

[5] P. Dawson, W. Sutherland-Smith, and J. Ricksen, "Using digital forensics in higher education to detect academic misconduct," *International Journal for Educational Integrity*, vol. 18, no. 1, pp. 1–17, 2022.

[6] Z. Boukhers and C. Yang, "Comparison of feature learning methods for metadata extraction from pdf scholarly documents," *arXiv preprint arXiv:2501.05082*, 2025. [Online]. Available: https://arxiv.org/abs/2501.05082

Návrh riadenia BLDC motora

[1]*Peter GIROVSKÝ* , [2]*Ján KAŇUCH,* [3]*Tadeáš KMECIK* , [4]*Pavol MACKANIČ*

[1, 2, 3, 4] Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]peter.girovsky@tuke.sk, [2]jan.kanuch@tuke.sk, [3]tadeas.kmecik@tuke.sk, [4]pavol.mackanic@student.tuke.sk

***Abstrakt* — V tomto článku je predstavený návrh riadenia BLDC motora. Návrh riadenia BLDC motora bol robený v programe MATLAB\Simulink pre zvolený typ motora. Následne toto riadenie bolo implementované a otestované na reálnom BLDC motore, pričom riadenie bolo realizované pomocou mikrokontroléra. V článku sú prezentované výsledky jednak simulačného návrhu ako aj praktickej realizácie riadenia BLDC motora.**

***Kľúčové slová* — BLDC motor, riadenie, mikrokontrolér, simulácia**

I. ÚVOD

BLDC motory patria v súčasnosti medzi jedny z najviac využívaných motorov v elektrických pohonoch vzhľadom ich vlastnostiam. Vyznačujú sa vysokou účinnosťou, nízkou hmotnosťou a dlhodobou spoľahlivosťou vďaka absencii mechanického komutátora a kefiek, čo eliminuje opotrebenie a znižuje požiadavky na údržbu. Pre tieto vlastnosti sú BLDC motory ideálnymi pre široké spektrum aplikácií, od priemyselných pohonov cez zdravotnícke prístroje až po moderné elektromobily.

Riadenie BLDC motorov je však technicky náročnejšie. Elektronická komutácia, ktorá nahrádza mechanický komutátor, vyžaduje presné riadenie pomocou algoritmov a senzorov na zisťovanie polohy rotora. Tieto riadiace metódy umožňujú optimalizovať výkon motora, minimalizovať energetické straty a zabezpečiť hladký chod pri rôznych záťažových podmienkach. [1][2][3]

II. BLDC MOTOR

BLDC (brushless direct current) je jednosmerným bezkefovým, bezkomutátorovým motorom s permanentnými magnetmi. Môže byť konštruovaný buď s vonkajším alebo s vnútorným umiestnením rotora. Vinutie je umiestnené na statore, pričom na rotore sa nachádzajú permanentné magnety. Tie môžu byť striedavo rozložené po povrchu alebo vo vnútri rotora. Vinutie statora môže byť jednofázové, dvojfázové alebo trojfázové, pričom vyšší počet fáz zaručuje plynulejšie prepínanie vinutí. To vedie k vyššej účinnosti, vyššiemu výkonu a tichšej prevádzke motora. [1][3]

Vzhľadom k absencii kief a komutátora sa BLDC motory, v porovnaní s komutátorovými motormi, vyznačujú viacerými výhodami:

- vyššou: spoľahlivosťou, životnosťou, účinnosťou, výkonom, otáčkami, robustnosťou
- nižšími nárokmi na údržbu a nižšou hlučnosťou
- možnosťou pracovať v nebezpečných alebo mimoriadne čistých prostrediach

Medzi nevýhody patrí vyššia obstarávacia cena a zložitejšie riadenie. Vďaka svojím nesporným výhodám nachádzajú BLDC motory široké uplatnenie v rôznych odvetviach ako napr. priemyselná automatizácia, robotika, automobilový priemysel, zdravotnícka technika či letectvo.

A. Matematický model BLDC motora

Medzi najčastejšie využívané patria trojfázové BLDC motory, ktoré môžeme popísať nasledujúcim matematickými rovnicami:

$$\begin{aligned} U_a &= i_a R_a + L_a \frac{di_a}{dt} + M_{ab} \frac{di_b}{dt} + M_{ac} \frac{di_c}{dt} + e_a \\ U_b &= i_b R_b + L_b \frac{di_b}{dt} + M_{ba} \frac{di_a}{dt} + M_{bc} \frac{di_c}{dt} + e_b \\ U_c &= i_c R_c + L_c \frac{di_c}{dt} + M_{cb} \frac{di_b}{dt} + M_{ca} \frac{di_a}{dt} + e_c \end{aligned} \quad (1)$$

kde R je elektrický odpor fázového vinutia, L – indukčnosť fázového vinutia, M – vzájomná

indukčnosť medzi fázovými vinutiami, i_a, i_b, i_c – prúdy tečúce jednotlivými fázami vinutia, U_a, U_b, U_c – fázové napätia vinutia. Ak uvažujeme symetrické fázové vinutia BLDC motora, bude platiť:

$$R_a = R_b = R_c = R \tag{2}$$

$$L_a = L_b = L_c = L \tag{3}$$

$$M_{ab} = M_{ac} = M_{bc} = M_{ba} = M_{ca} = M_{cb} = M \tag{4}$$

B. Riadenie a komutácia

BLDC motor neobsahuje komutátor, komutácia je zabezpečovaná elektronicky – tranzistorovým meničom. Ten v závislosti od polohy rotora, prepína prúd do jednotlivých cievok statorového vinutia. Cieľom je zabezpečiť stálu, vzájomnú kolmosť vektorov magnetického poľa statora a magnetického poľa permanentných magnetov rotora. Na zistenie presnej polohy rotora sa využívajú rôzne typy senzorov, napr. enkodéry, resolvery, reluktančné senzory alebo Hallove sondy. Použitie týchto snímačov zvyšuje presnosť riadenia a umožňuje bezchybné riadenie otáčok a momentu. V prípade dynamicky nenáročných pohonov sa môže využiť aj bezsenzorové riadenie založené na analýze meraného indukovaného napätia, ktoré umožňuje využitie jednoduchších, robustnejších a lacnejších BLDC motorov bez snímačov.

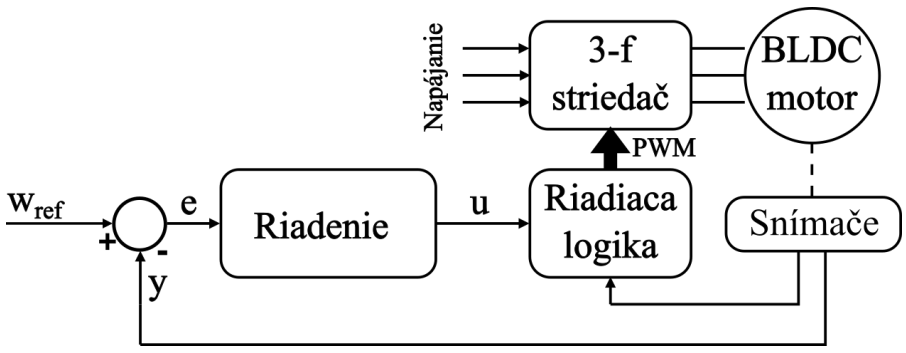
Pre riadenie sa zvyčajne využíva algoritmus 6 krokovej komutácie, ktorý počas jednej elektrickej otáčky motora, vygeneruje 6 rôznych vektorov napätia. Tie riadiaca jednotka generuje takým spôsobom, aby platila podmienka vzájomnej kolmosti magnetického poľa rotora a statora. Stredná hodnota napätia privádzaného na svorky statora, ktorou je možné ovládať otáčky motora, je riadená pulznou šírkovou moduláciou (PWM). [2][4]

III. Simulácia BLDC motora v programe MATLAB\Simulink

Pre návrh riadenia bol zvolený BLDC motor s nasledujúcimi parametrami uvedenými v Tabuľke 1.

Tabuľka 1
Parametre BLDC motora

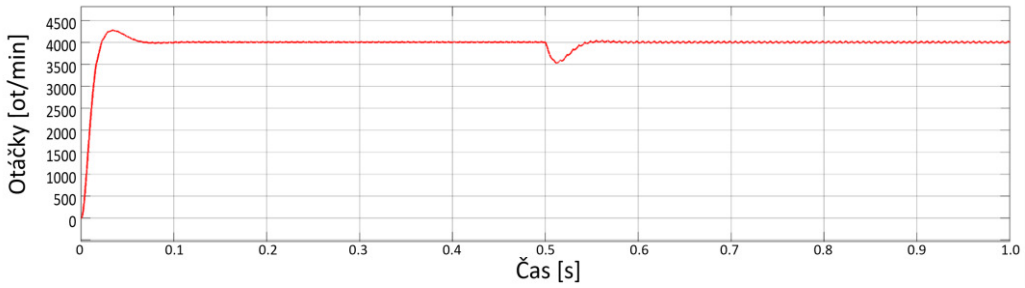
Symbol	Veličina	Hodnota
U_{DC}	Napájacie napätie	24 [V]
P_N	Nominálny výkon	53 [W]
M_N	Nominálny moment	0.127 [Nm]
n	Nominálne otáčky	4000 [ot/min]
R_S	Statorový odpor	0.8 [Ω]
L_S	Statorová indukčnosť	1.2 [mH]
J	Moment zotrvačnosti	4.805×10^{-9} [kg.m^2]
k_t	Momentová konštanta	0.04104 [Nm/A]



Obr. 1 Principiálna simulačná schéma pre BLDC motor

Na Obr.1 je zobrazená principiálna simulačná schéma s riadením BLDC motora. Pre potreby riadenia boli získavané hodnoty z Hallových snímačov a informácia o rýchlosti BLDC motora. Signály z Hallového snímača vstupujú do riadiacej logiky, kde sú spracované pomocou pravdivostných tabuliek a na základe nastavených pravidiel sa prepočítavajú impulzy, ktoré

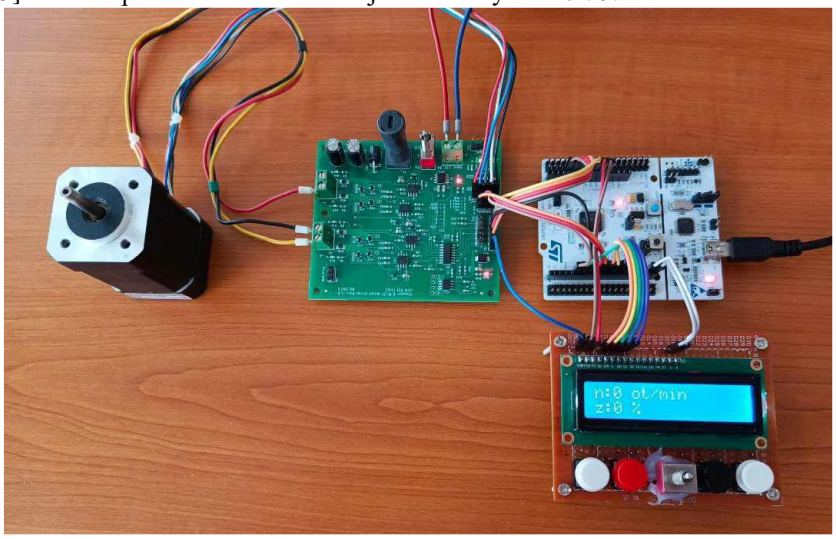
spínajú tranzistory 3-f striedača. Pre reguláciu rýchlosti BLDC motora bol použitý PI regulátor, hodnota skutočnej rýchlosti bola snímaná priamo na motore. Na Obr. 2 je zobrazený priebeh rýchlosti BLDC motora.



Obr. 2 Priebeh rýchlosti BLDC motora

IV. PRAKTICKÁ VERIFIKÁCIA RIADENIA BLDC MOTORA

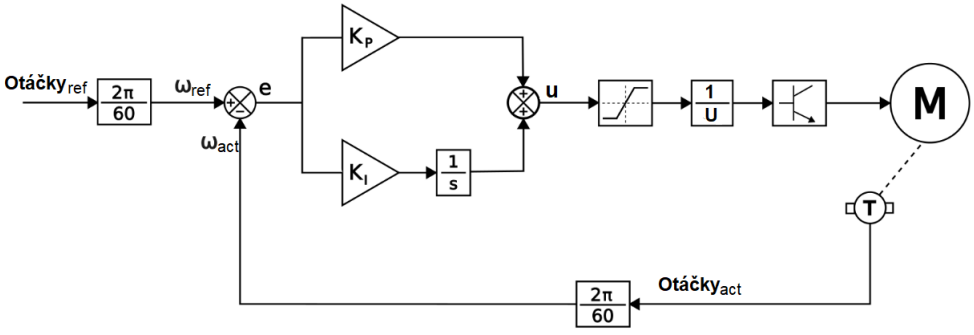
Praktická verifikácia riadenia bola robená na BLDC motore - BLY172S (parametre motora sú uvedené v tabuľke 1), pričom riadenie bolo robené pomocou mikrokontroléra STM32NUCLEO-F446R. [5] Hardvér pre testovanie riadenia je zobrazený na Obr. 3.



Obr. 3 Hardvér pre testovanie riadenia BLDC motora

A. *Riadenie motora s PI regulátorom rýchlosti*

Pri riadení BLDC motora pomocou PI regulátora rýchlosti bolo použité priame riadenie rýchlosti bez prúdového regulátora.



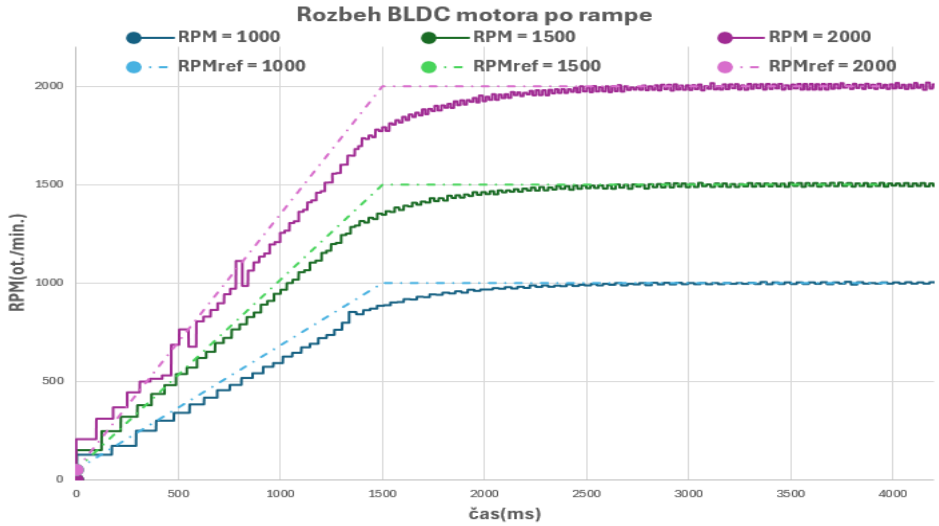
Obr. 4 Principiálna bloková schéma riadenia BLDC motora

Vstupom do regulátora rýchlosti bola žiadaná hodnota uhlovej rýchlosti a výstupom z PI regulátora bola hodnota napätia prepočítaná na striedu PWM, ktorou boli spínané tranzistory meniča. Principiálna bloková schéma regulačného obvodu sa nachádza na Obr. 4.

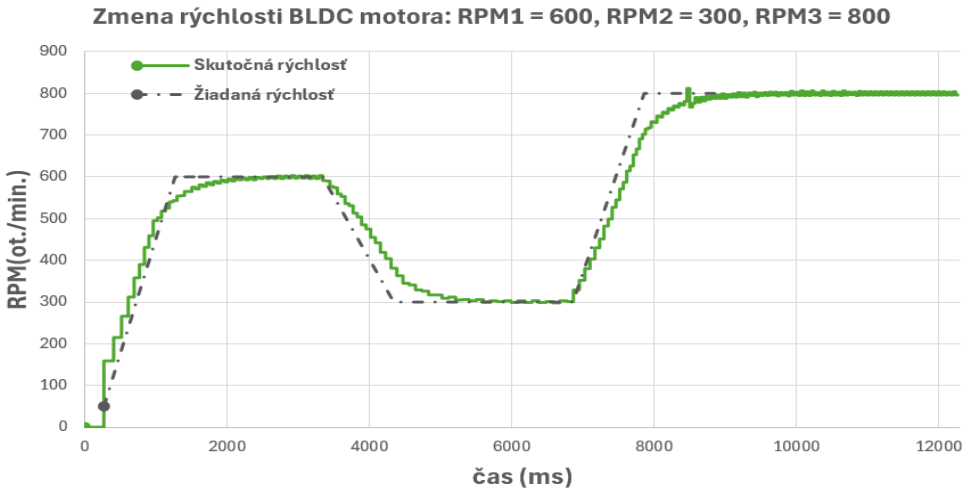
Pre výpočet parametrov regulátora rýchlosti bola použitá metóda KOM (Kritérium optimálneho modulu) a výpočtom podľa tejto metódy bola určená proporcionálna konštanta regulátora K_P=0,0017 a integračná konštanta regulátora K_I=0,00029.

Na Obr. 5 sú zobrazené priebehy rozbehu rýchlosti a na Obr. 6 je zobrazený priebeh rýchlosti

BLDC motora pre žiadanú hodnotu.



Obr. 5 Principiálna bloková schéma riadenia BLDC motora



Obr. 6 Principiálna bloková schéma riadenia BLDC motora

V. ZÁVER

Cieľom príspevku bolo predstaviť návrh riadenia BLDC motora a jeho praktickú verifikáciu. V príspevku bol predstavený matematický model BLDC motora, kde na základe tohto modelu bol zostavený simulačný model s riadením v programe MATLAB\Simulink. V rámci tejto simulácie boli otestované základné princípy riadenia pre jeho následnú praktickú verifikáciu. Pre verifikáciu riadenia boli vykonané viaceré merania ako napr. rozbeh motora bez rampy, s rampou, zmena referenčnej rýchlosti a pod. Výsledky nám ukázali, že navrhnutá regulácia zabezpečila stabilné a spoľahlivé riadenie BLDC motora.

POĎAKOVANIE

Tento príspevok vznikol s podporou projektu KEGA 032TUKE-4/2024. Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00018.

LITERATÚRA

[1] GIERAS, Jacek: Pernament magnet motor technology. USA: Taylor & Francis Group, LLC, 2010. 600 s. ISBN 978-1-4200-6440-7
[2] XIA, Changliang: Pernament Magnet Brushless DC motor drives and controls, China: John Wiley & Sons. 2012. 282 s. ISBN 978-1-118-18833-0
[3] HANSELMAN, Duane: Brushless Permanent Magnet Motor Design. USA: Magna Physics Publishing. 2006. 392 s. ISBN: 1-881855-15-5.
[4] NISE, Norman S. Control Systems engineering. Pomona: John Wiley & Sons. 2011. 929 s. ISBN 978-0-470-54756-4
[5] STMicroelectronics: NUCLEO-F446RE. [online]. 2025. Dostupné na internete <https://www.st.com/en/evaluation-tools/nucleo-f446re.html#featured_resources-0>

Implementácia inerciálnej jednotky pre zvýšenie bezpečnosti práce

Slavomír KARDOŠ, Dmytro BABITSKYI

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

slavomir.kardos@tuke.sk

***Abstrakt* — Inerciálne meracie jednotky (IMU) sa osvedčili ako všestranné snímacie prvky v širokej škále odvetví. Príspevok sa zameriava na využitie akcelerometrickej jednotky na zvýšenie bezpečnosti obsluhy rezacích zariadení. Téma je aktuálna vzhľadom na štatistiky nehôd, ktoré reportujú vážne zdravotné i smrteľné následky. Experimentálne merania boli vykonané s použitím 3-osového akcelerometra aplikovaného na reťazovú pílu, ktorá je vysokorizikovým zariadením. Reakčný čas takéhoto senzora je výrazne kratší ako reakčný čas požadovaný na ochranu obsluhy. Benefitom je teda možnosť zníženia rizika zranení v dôsledku spätného rázu, ktorý je najčastejšou príčinou týchto zranení.**

***Kľúčové slová* — akcelerometer, MEMS, motorová píla, bezpečnosť práce**

I. Úvod

Inerciálne meracie jednotky našli široké uplatnenie vďaka schopnosti reagovať na mechanický pohyb vrátane vibrácií. Vždy je to forma snímania zrýchlenia alebo rýchlosti otáčania [1][2], prípadne ich fúzia [3] a transformácia na požadované mechanické parametre. Ako ochranné prvky umožňujú predchádzať poškodeniu priemyselných alebo spotrebiteľských zariadení vrátane počítačov. Snímanie vibrácií priemyselných rotačných strojov a ich spektrálna analýza umožňuje identifikovať degradačné procesy a tým zabrániť rozsiahlejšiemu poškodeniu [4]. Signál IMU je predspracovaný, digitalizovaný a analyzovaný alebo zapojený do riadiaceho reťazca. V prípade monitorovacích alebo diagnostických jednotiek sa ich signál prenáša bezdrôtovo alebo sa ukladá na neskoršiu analýzu.

Pracovné obrábacie stroje ako reťazové píly sú užitočné, ale potenciálne nebezpečné zariadenia. Takéto výkonné zariadenia sa používajú na mechanické spracovanie materiálov, t. j. na úpravu ich tvaru a povrchových vlastností. Na dosiahnutie výsledku nástroj vykonáva lineárny alebo rotačný pohyb, je teda uvedený do stavu s určitou kinetickou energiou alebo určitým krútiacim momentom tak vysokým, že nesprávna manipulácia alebo porucha pracovného nástroja môže spôsobiť zranenie alebo fatálne následky.

Reťazová píla prenáša na zuby reťaze okamžitý výkon dostatočný na spracovanie materiálov na báze celulózy. Podobne uhlová brúska s rýchlosťou viac ako 10 000 otáčok za minútu a hmotnosťou kotúča 40 g má zotrvačnú silu ďaleko prevyšujúcu odpor živého tkaniva a tuhších častí živého organizmu. V nepredvídateľnom prípade má kontakt pracovného nástroja s mäkkým tkanivom devastačné následky, pretože je výrazne poddajnejšie. Ochranné prostriedky sú nevyhnutnosťou, ale nechránia celé telo. Preto aj skúsený pracovník môže utrpieť pracovný úraz, a to aj pri dodržiavaní predpísaných pracovných predpisov a používaní certifikovaného náradia.

Nedostatočné skúsenosti obsluhy a pracovné prostredie sú hlavnými faktormi, ktoré spôsobujú väčšinu nehôd. Aj skúsený operátor robí chyby - a ak ich robia profesionáli, ich pravdepodobnosť je vyššia u menej skúsených používateľov. Štatistiky amerického Úradu pre bezpečnosť a ochranu zdravia pri práci (OSHA) vykazujú 20 % fatálnosť pri zraneniach spôsobených spätným rázom brúsky alebo píly. Ostatných 80 % zranení je natoľko vážnych, že si vyžadujú hospitalizáciu. Podľa Centra pre kontrolu a prevenciu chorôb (CDC), nemocnice hlásia každý rok desiatky tisíc zranení a stovky úmrtí súvisiacich s motorovou pílou, pričom takmer všetkým sa dá predísť. Podľa OSHA sa približne 40 % hlásených nehôd s motorovou pílou stane na nohách, 35 % na ľavej ruke a zápästí a 1 % vedie k amputácii. Americká Komisia pre bezpečnosť spotrebiteľských výrobkov uviedla v priemere 110 stehov v súvislosti s poraneniami spôsobenými motorovou pílou. CDC odhaduje, že náklady na zdravotnú starostlivosť v súvislosti s takýmito zraneniami sú ročne najmenej 350 miliónov dolárov. A čo je vážnejšie, OSHA uvádza,

že ročne dochádza k približne 250 úmrtiam súvisiacim s motorovými pílami [5].

Reťazové píly sú zdrojom mechanických vibrácií a akustických signálov, ktorých úrovne výrobca deklaruje v technickom liste produktu na základe štandardných prevádzkových podmienok. Potenciálny vplyv takéhoto vibračného zaťaženia na obsluhu je uvedený vo viacerých štúdiách, napríklad v [6]. Normy EN a EN ISO sa zaoberajú požiadavkami a ochranou zdravia obsluhy pri práci s takýmito zariadeniami. V súvislosti s dlhodobým alebo profesionálnym používaním motorových píl sa spája niekoľko ochorení, jednoduchej alebo zložitej povahy, ako napríklad traumatické vazospastické ochorenie (T.V.D.). Už štúdia [7] uvádza, že charakteristická frekvencia vibračného signálu, ktorá je okolo 125 Hz, má vážny vplyv na nervové receptory obsluhy, pokiaľ ide o amplitúdy a trvanie zaťaženia.

Frekvencia vibračného signálu generovaného benzínovou a elektrickou pílou sa líši, pretože v benzínovej píle má voľnobeh charakter voľnobežných otáčok bez uzatvorenej spojky. Zvyšovaním otáčok sa spojka spája s pohonom reťaze píly, pričom otáčky sa úmerne menia so sýtením spaľovacej zmesi. Píla tak dosahuje otáčky okolo 12 000 ot./min pri plnom plyne, čo zodpovedá frekvencii prvej harmonickej okolo 200 Hz [8].

II. Definícia kritických udalostí

Povaha pracovného režimu zariadenia určuje typ a rozsah relevantného signálu. Na posúdenie je potrebné analyzovať povahu natívneho signálu, ako aj dynamiku pohybu v prípade nepredvídaných udalostí s potenciálom rizika zranenia. Na základe analýzy reportovaných rizikových udalostí s fatálnymi následkami boli identifikované dva relevantné typy.

Prvý typ, známy ako „spätný ráz“, predstavuje neočakávané uvoľnenie pracovného nástroja v dôsledku zachytenia kontaktnej časti rezného nástroja o obrábaný materiál, pričom obsluha nie je schopná kompenzovať ťahovú silu nástroja. Výsledkom je vymrštenie zariadenia rýchlosťou úmernou udelenej hybnosti, v kritickom prípade až do oblasti tela obsluhy. Keďže zariadenie je nesené na kĺboch obsluhy, jeho lineárny pohyb sa transformuje na rotačný pohyb smerom k ramenám, hrudníku alebo hlave obsluhy.

K spätnému rázu zvyčajne dochádza bez akýchkoľvek predchádzajúcich okolností, keď sa zuby reťaze píly zachytia o materiál alebo nerovný predmet (uzly v dreve, cudzie predmety) a otáčajú sa okolo špičky lišty, čo spôsobí, že lišta je vymrštená smerom k obsluhe. Hoci sú reťazové píly vybavené ochrannými zariadeniami určenými na minimalizáciu následkov spätného rázu, najviac zranení spôsobuje táto udalosť. Jej iniciovanie a trvanie je aj menej ako pätina sekundy, takže obsluha nemá žiadny reakčný čas [5][9].

Neočakávaná udalosť typu spätného rázu z hľadiska zrýchlenia pracovného zariadenia je reprezentovaná nízkofrekvenčným impulzom v desatinách hertzov, t. j. s trvaním kratším ako sekunda a zrýchlením v rozsahu low-g.

Druhý typ nebezpečnej udalosti je oveľa prozaickejší a neobvyklý pre skúsenú obsluhu, pretože mu predchádza neprofesionálna manipulácia s reťazovou pílou. Ide o vertikálny rez do horizontálne umiestneného materiálu, pri ktorom je spodná končatina v jednej línii s pracovným smerom pohybu motorovej píly. Neočakávané uvoľnenie po typickom intenzívnejšom reze z dôvodu lokálnej tuhosti materiálu má za následok kritické alebo fatálne následky.

III. Mechanická odozva

Zariadenie typu reťazovej píly je fyzikálny systém s vynúteným vibračným pohybom, čo je pohyb, ktorý vzniká v reakcii na kontinuálne budenie. Amplitúda signálu sa mení s časom harmonickou funkciou. Výsledný signál je komplex primárnych a sekundárnych zložiek mechanického vibračného signálu. Budiaca sila pôsobiaca na hmotnú štruktúru s jedným stupňom voľnosti (1-DOF) vedie k odozve so zmenou polohy x telesa s hmotnosťou m oproti počiatočnej polohe:

$$x = A \sin \omega_n t + B \cos \omega_n t + \frac{1}{k} \frac{F_0}{1-(\omega/\omega_n)^2} \tag{1}$$

kde A a B sú charakteristické koeficienty, $F = F_0 \sin \omega t$ je pôsobiaca sila a $\omega_n = \sqrt{k/m}$ je vlastná frekvencia hmotného prvku na pružinovom prvku s tuhosťou k [10].

Takéto zariadenie je systém obsahujúci množstvo mechanických prvkov, ktoré vedú mechanické vibrácie z aktuátora a samy sú sekundárnym zdrojom mechanických vibrácií, prevažne v akustickom frekvenčnom rozsahu. Celková štruktúra zariadenia potom generuje komplexný signál:

$$x(t) = A_0 + \sum_{i=0}^{0} A_i \cos(2\pi f_i t + \varphi_i) \tag{2}$$

kde A_0 predstavuje jednosmerný posun a $A_i \cos(2\pi f_i t + \varphi_i)$ individuálnu spektrálnu zložku harmonického signálu s frekvenciou f a fázovým posunom φ_i [11]. Zariadenia, ktoré sú vystavené

intenzívnemu vibračnému zaťaženiu, vyžadujú použitie materiálov a komponentov so zvýšenou mechanickou odolnosťou a podliehajú príslušným požiadavkám na testovanie [14][15].

Reťazová píla je systém s premenlivým operačným vektorom, pričom os hnacieho hriadeľa a otáčania reťaze je kolmá na rovinu reťaze. Akčný záber je v rámci plošného uhla, nie priestorového, v zmysle hybnosti reťaze. Preto je najrelevantnejšie snímanie v dvoch osiach, pričom laterálny pohyb má minimálny príspevok, aj to iba v prípade výrazne nehomogénneho spracovávaného materiálu. Ak sa akčný záber všetkých elementárnych kĺbov v súradnicovom systéme redukuje na súradnicové osi x, y, konkrétne na rotáciu okolo osi z, absolútna hodnota okamžitého zrýchlenia je:

$$|a| \cong \sqrt{a_x^2 + a_y^2} \ @ \ a_z \ll a_x;\ a_y \tag{3}$$

Keďže kapacitný akcelerometer je vystavený dynamickému aj statickému zrýchleniu, každá zložka vektora zrýchlenia, ktorá nie je kolmá na vektor gravitačného zrýchlenia, obsahuje okrem dynamickej zložky aj aditívnu statickú zložku zrýchlenia. Taktiež platí, že veľkosť absolútnej hodnoty zrýchlenia v prípade stacionárneho stavu je:

$$|a| = \sqrt{a_x^2 + a_y^2 + a_z^2} = 1\ g \ @ \ \overrightarrow{a_d} = 0 \tag{4}$$

pretože gravitačné zrýchlenie je konštantné, iba sa rozkladá medzi jednotlivé snímacie osi pri naklonení zariadenia [12].

Je vhodné poukázať na dominantné pohybové zložky signálu v jednotlivých častiach systému tvoreného kĺbovými segmentmi operátora a samotným zariadením. Počas oboch definovaných rizikových udalostí sú zapojené všetky tri kĺby obidvoch horných končatín s rôznou veľkosťou statických a dynamických zložiek zrýchlenia, ako aj mierou rotácie. Súvisí to s prirodzenou štruktúrou horných končatín, ktorých každý segment má niekoľko stupňov voľnosti pohybu (DOF), a preto ako celok tvorí komplexný pohybový systém.

Druhá časť systému, samotné zariadenie, je s operátorom spojená prostredníctvom dvoch bodov a kĺby sú tvorené zápästnými kĺbmi a frikčným kontaktom dlaní s úchopmi. Neočakávaný pohyb pracovného nástroja počas oboch rizikových udalostí je iniciovaný v smere udelenej hybnosti a v súlade s predchádzajúcimi úvahami je transformovaný na komplexný lineárno-rotačný pohyb.

IV. Experimentálna zostava

Na základe predchádzajúcich úvah je možné lokalizovať snímacie polohy s najvyššou mierou dynamickej zmeny počas rizikových udalostí, ale ďalšími faktormi sú technická realizovateľnosť a praktickosť. Lakťové kĺby vykazujú počas nich najvyššiu mieru rotačného pohybu, ale je nepraktické aplikovať snímací systém na telo obsluhy. Vhodnosť umiestnenia snímacej jednotky má zmysel na pracovné zariadenie.

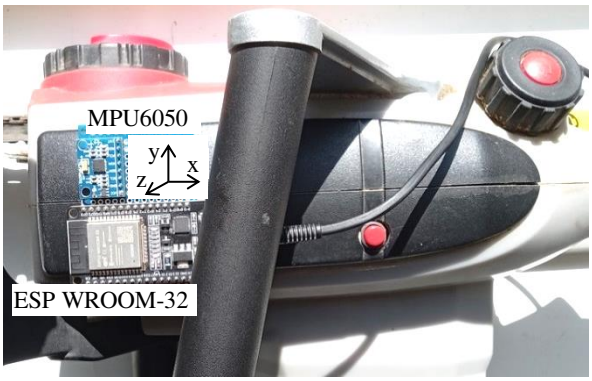
Vzhľadom na kontaktný bod zariadenia s opracovávaným materiálom dochádza počas spätného rázu k rotačnému pohybu v uvažovanej rovine x, y okolo osi z, ale obzvlášť významné sú lineárne zložky zrýchlenia a zložky náklonu. Smerom k telu zariadenia sa lineárne zrýchlenie znižuje v dôsledku klesajúcej obvodovej rýchlosti, ale ďalšie dve spomínané zložky zostávajú rovnaké kvôli zachovanému uhlu rotačného pohybu. Umiestnenie snímacej jednotky na konci reznej lišty je však problematické, optimálne je preto jej umiestnenie v oblasti tela zariadenia.

Na základe predpokladov bola navrhnutá a implementovaná experimentálna zostava snímacieho a vyhodnocovacieho systému, mechanicky prepojená s telom motorovej píly. Vzhľadom na orientáciu pracovného nástroja bola orientácia snímača stanovená rovnobežne s jeho osou z, so záporným zmyslom.

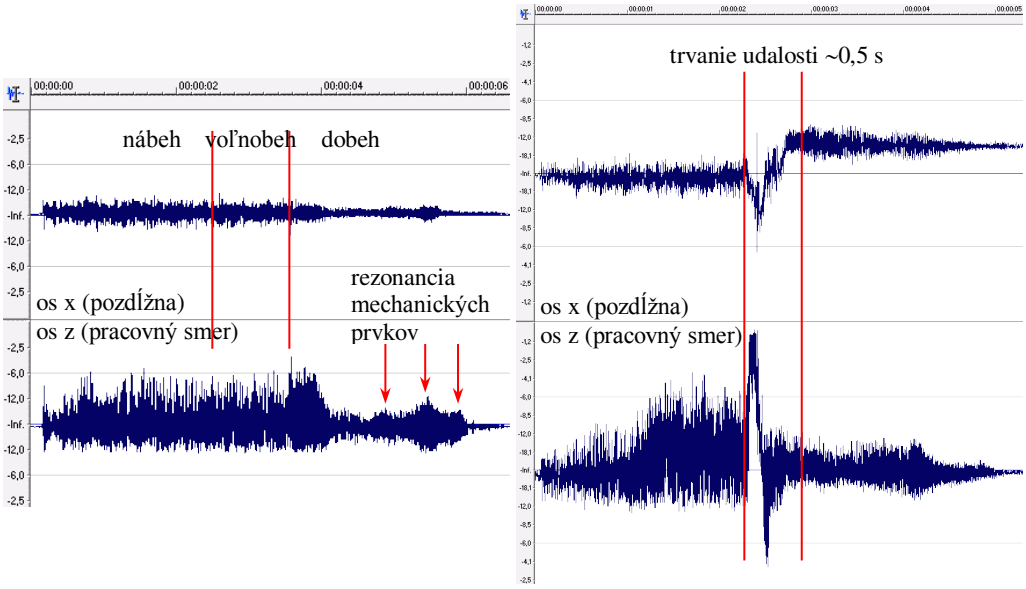
Testovací signál vykazoval vysoké amplitúdy charakteristických zložiek signálu, v ktorých je nízkofrekvenčná časť spektra výrazne potlačená. Na vyriešenie tohto problému bol použitý mechanický tlmič vo forme disku z polymérnej peny, ktorého vplyv na signál má charakter dolnopriepustného filtra (LPF). So zahrnutím mechanického LPF signál už vykazoval prijateľné vlastnosti vrátane relatívneho pomeru amplitúd harmonických zložiek v spektre. Napriek namáhaniu počas spracovania materiálu nedošlo v spektre k žiadnej významnej zmene. Naopak, signál zo simulácie spätného rázu vykazoval výrazné preskupenie spektra, ktoré možno pripísať deformácii niektorých mechanických prvkov s efektom podobným ako u strún a jazýčkov hudobných nástrojov.

Pre účely ďalších meraní bol zvolený akcelerometer s vyšším meracím rozsahom (±16 g), ktorý je súčasťou 6-DOF jednotky MPU6050 [13]. Bol umiestnený na nosnom substráte s mikrokontrolérom pre zber údajov (Obr. 1), pre potreby manipulácie na štyroch magnetických spojoch na mechanickom tlmiči. Keďže udalosť spätného rázu generuje signál vo frekvenčnom pásme rádovo Hz, šírka pásma je ďalej zúžená filtrovaním.

Testovací scenár pozostával zo spustenia, voľnobehu, spracovania a dobehu, pričom mechanický signál bol zaznamenaný na PC pomocou Python aplikácie pre príjem a spracovanie údajov. Náhľad testovacieho signálu v časovej doméne je znázornený na Obr. 2 s definovanými jednotlivými charakteristickými oblasťami.



Obr. 1. Umiestnenie senzorového modulu s mikrokontrolérom na tele reťazovej píly.



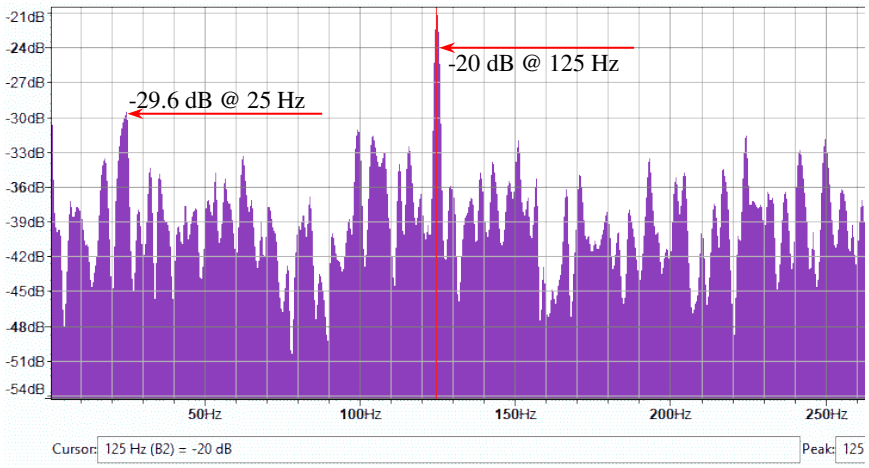
Obr. 2. Signálová odozva rozbehu, voľnobehu a dobehu zariadenia (vľavo) a rizikovej udalosti „kickback“ (vpravo).

Spektrálna analýza jednotlivých častí prevádzkového režimu zariadenia podrobnejšie poukazuje na vyššie uvedené vlastnosti, pričom vyššia zložitosť spektra počas rozbehu a dobehu je daná kontinuálnou zmenou prevádzkovej rýchlosti.

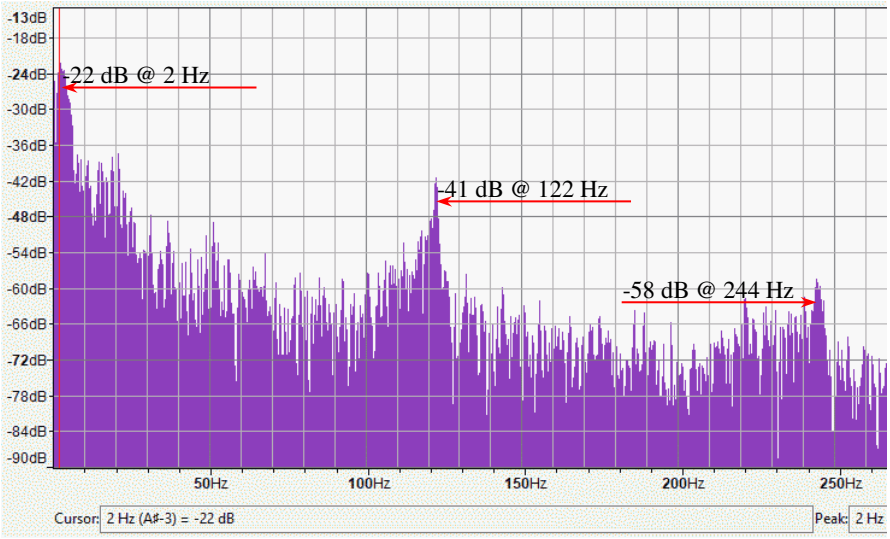
Prirodzenými prvkami signálu dobehu sú významné rezonančné frekvencie niektorých prvkov mechanickej zostavy systému (hlavné vrcholy pri -33 dB pri 161 Hz, -43 dB pri 80 Hz atď.). Prítomnosť viacerých mechanických prvkov pod vplyvom vibračného zaťaženia preto prirodzene prináša riziko naladenia na rezonančnú frekvenciu konkrétneho prvku. Jedná sa o podobný problém ako v avionike, kde však už predstavuje rizikový faktor.

Stav voľnobehu má jednoduchšie rozložené spektrum, pretože nedochádza k žiadnej kontinuálnej zmene rýchlosti, pričom charakteristické rezonančné frekvencie mechanických prvkov systému reťazovej píly sú jasnejšie viditeľné (Obr. 3), najmä -20 dB pri 125 Hz, -29,6 dB pri 25 Hz, -31,1 dB pri 100 Hz atď.

Spätný ráz je nepredvídaná udalosť, o to nebezpečnejšia pre obsluhu. Preto je obťažné vyvolať podmienky pre túto udalosť. Jej príprava bola vykonaná s veľkou starostlivosťou a čo najvyššou predvídateľnosťou. Signálna odozva je znázornená na Obr. 2 vpravo a naznačuje vysoký detekčný potenciál udalosti vďaka veľmi nízkej polohe jej signálu vo frekvenčnom spektre a dostatočne vysokým amplitúdam jeho zložiek. Spektrálna analýza indikuje frekvenčné vlastnosti signálu s udalosťou spätného rázu, ktorého poloha je vo frekvenčnom rozsahu jednotiek Hz, s dominantným vrcholom -22 dB pri 2 Hz (Obr. 4) a je veľmi dobre rozlíšiteľný frekvenciou i amplitúdou od voľnobežného i pracovného signálu zariadenia. Ďalšie významné zložky sú vo frekvenčnom spektre dostatočne vzdialené - 122 Hz a jej druhá harmonická 244 Hz.



Obr. 3. Spektrálne zložky signálovej odozvy počas voľnobehu



Obr. 4. Spektrálne zložky signálovej odozvy riadeného spätného rázu

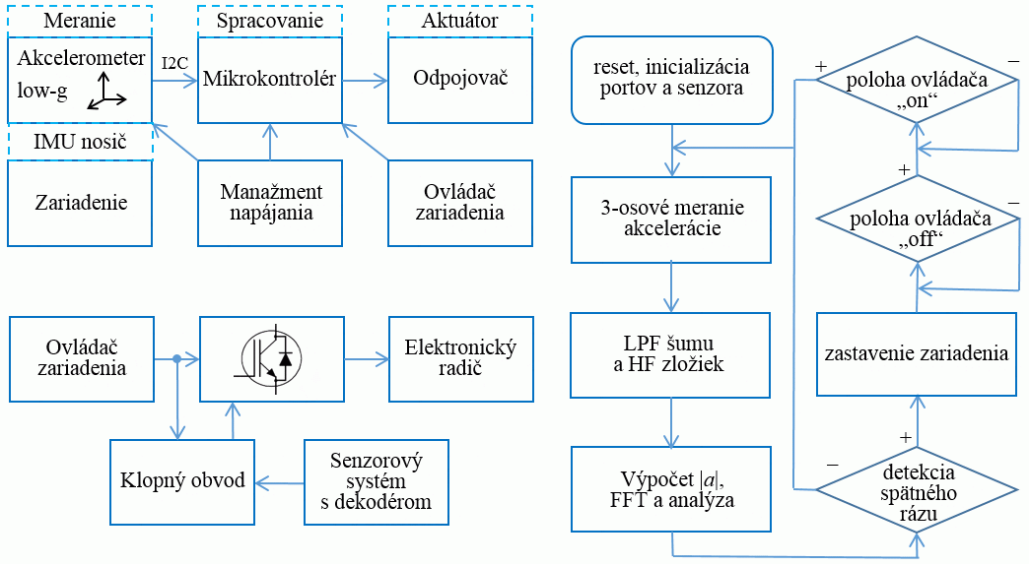
V. Implementácia systému

Ako je uvedené v teoretickom rozbore, neočakávaný spätný ráz má rýchly priebeh, s trvaním kratším ako je reakčný čas obsluhy, daný aplikovaným krútiacim momentom. Ak má byť preventívny systém účinný, musí prevziať túto funkciu a konať v dostatočne krátkom čase predtým, ako pohybujúca sa reťaz zraní obsluhu. Konštrukčne je reťazová píla elektromechanický systém, ktorého primárnym zdrojom energie je benzín alebo elektrická energia. Prvý typ je typický voľnobehom bez uzavretej spojky. V oboch prípadoch sú však hlavný a bezpečnostný spínač prvky, ktorých vypnutie zabezpečí okamžité zastavenie reťaze. Preto je v ceste týchto prvkov potrebné zahrnúť elektronický odpojovač ovládaný logikou riadenou detekčným snímacím systémom, ako je znázornené v blokovej schéme na Obr. 5 vľavo hore a vo vývojovom diagrame funkcií na Obr. 5 vpravo. Logika odpojovača s klopným obvodom na Obr. 5 vľavo dole zabezpečuje, že aj po reakcii akcelerometra zostane zariadenie v bezpečnostnom stave vypnutia, ktorý je možné resetovať uvoľnením hlavného spínača a jeho opätovným zapnutím. Je dôvodné zahrnúť bezpečnostný časovač, ktorý nedovolí náhodnému zapnutiu zariadenia, kým neuplynie bezpečnostný interval.

VI. Záver

Nosnou témou príspevku je zvýšenie bezpečnosti práce s reznými zariadeniami, keďže štatistiky reportujú vysokú mieru nehôd, vrátane fatálnych následkov, a to aj v profesionálnej praxi. Keďže ide o dynamický systém, skúmaná bola možnosť použitia trojosového akcelerometra na snímanie kritických udalostí, ako je spätný ráz (kickback). Bolo realizovaných niekoľko scenárov merania, ktoré predstavujú rozbeh, voľnobeh, opracovanie materiálu a spätný ráz. Odozva signálu vykazovala charakteristické oblasti v časovej doméne a spektrálna analýza preukázala zložitosť mechanickej zostavy s množstvom rezonančných frekvencií jednotlivých mechanických prvkov. Keďže odozva signálu udalostí typu spätného rázu je dobre rozlíšiteľná

vo frekvenčnom spektre i amplitúdami, po spracovaní signálu filtrovaním a frekvenčnou transformáciou je možné vykonať opatrenia v riadiacom systéme zariadenia, a tým zvýšiť jeho bezpečnosť predchádzaním následkov.



Obr. 5. Bloková schéma reťazovej píly s implementovanou elektronikou na snímanie a analýzu spätného rázu (vľavo hore), riadiaca logika núdzového zastavenia reťazovej píly (vľavo dole), vývojový diagram snímania údajov a analýzy spätného rázu (vpravo).

Poďakovanie

Príspevok bol vypracovaný s podporou projektov KEGA č. 011TUKE-4/2023, VEGA č. 1/0049/24 a č. 1/0278/24. Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. "09I05-03-V02-00002“. Študent Dmytro Babitskyi sa na príspevku podieľal v rámci riešenia svojej bakalárskej práce.

Literatúra

[1] M. Looney, ‘An Introduction to MEMS Vibration Monitoring. ADI Analog Dialogue’. Analog Devices, Inc., Jun. 2014.
[2] M. Looney, ‘MEMS Vibration Monitoring: From Acceleration to Velocity’. Analog Devices, Inc., Jun. 2017.
[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] B. Campagnie, ‘Resource Library Technical Articles: Choose the Right Accelerometer for Predictive Maintenance’. Analog Devices, Inc., 2019.
[5] R. Gast, ‘Working Safely with a Chainsaw’. Adirondack Almanack, Oct. 2023. [Online]. Available: https://www.adirondackalmanack. com/2023/10/working-safely-with-a-chainsaw.html.
[6] W. Rukat, B. Jakubek, R. Barczewski, and K. Grochalski, ‘Identification of operating mode of a petrol chainsaw based on short-time parametrization and analysis of vibro-acoustic signals’, *Applied Acoustics*, vol. 192, p. 108704, Apr. 2022, doi: 10.1016/j.apacoust.2022.108704.
[7] J. Hyvärinen, I. Pyykkö, and S. Sundberg, ‘Vibration frequencies and amplitudes in the ætiology of traumatic vasospastic disease’, *The Lancet*, vol. 301, no. 7807, pp. 791–794, 04/1973, doi: 10.1016/S0140-6736(73)90598-9.
[8] M. G. Matache, M. Munteanu, D. N. Dumitru, and M. Epure, ‘Evaluation of hand transmitted chainsaw vibrations during wood cutting’, *E3S Web Conf.*, vol. 180, p. 03013, 2020, doi: 10.1051/e3sconf/202018003013.
[9] A. Dąbrowski, ‘Kickback risk of portable chainsaws while cutting wood of different properties: laboratory tests and deductions’, *International Journal of Occupational Safety and Ergonomics*, vol. 21, no. 4, pp. 512–523, Oct. 2015, doi: 10.1080/10803548.2015.1095547.
[10] R. E. Blake, ‘Basic vibration Theory’. Cooper Union, 02/2005. Available: https://engfac.cooper.edu/pages/tzavelis/uploads/Vibration%20Theory.pdf
[11] D. Preis, *SBE Broadcast Engineering Handbook. Chapter: Signals and Spectra*. The Society of Broadcast Engineers, 2005. [Online]. Available: https://sbe.org/handbook/fundamentals/Audio/Audio-Signals_and_Spectra.pdf
[12] M. Pedley, ‘Tilt Sensing Using a Three-Axis Accelerometer’. Freescale Semiconductor, Inc., Mar. 2013.
[13] InvenSense Inc., ‘MPU-6000 and MPU-6050 Product Specification Revision 3.4’. Aug. 2013.

Porovnanie smerovacích algoritmov v MANET pomocou metód posilňovaného učenia

[1]Jozef BADÁR, [2]Ján PAPAJ, [3]Dávid MAČEK

[1 2 3]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]jozef.badar@tuke.sk, [2]jan.papaj@tuke.sk, [3]david.macek@student.tuke.sk

Abstrakt **– Tento článok predstavuje porovnanie smerovacieho protokolu AODV s dvoma prístupmi posilňovaného učenia — Monte Carlo a Deep Q-Learning — v prostredí mobilných ad hoc sietí (MANET). Simulácie boli vykonané pomocou nástroja NS-3 v rôznych konfiguráciách siete a hustotách uzlov. Výkonnosť bola hodnotená na základe pomeru doručených paketov (PDR), priepustnosti a priemerného oneskorenia. Výsledky ukázali, že Deep Q-Learning stabilne prekonáva tradičný AODV protokol aj metódu Monte Carlo vo všetkých metrikách. Zistenia potvrdzujú potenciál techník umelej inteligencie pre adaptívne smerovanie v dynamických sieťových prostrediach.**

Kľúčové slová **– mobilné ad hoc siete, smerovanie, AODV, posilňované učenie, Monte Carlo, Deep Q-Learning, NS-3, umelá inteligencia**

I. Úvod

Mobilné ad hoc siete (MANET) predstavujú dôležitý smer vývoja v oblasti bezdrôtovej komunikácie, najmä v prostrediach bez pevnej infraštruktúry. Vďaka svojej decentralizovanej povahe a schopnosti autonómne sa organizovať nachádzajú uplatnenie v oblasti krízového riadenia, vojenských operácií či inteligentnej dopravy [1]. Dynamická topológia MANET však výrazne komplikuje smerovanie paketov, čím sa zvyšuje náročnosť na použité smerovacie protokoly [2].

Tradičné algoritmy ako Ad hoc On-Demand Distance Vector (AODV) často zlyhávajú v prostrediach s vysokou mobilitou alebo preťažením [3]. V posledných rokoch sa preto pozornosť výskumu sústreďuje na využitie umelej inteligencie, najmä metód posilňovaného učenia (RL), ako nástroja na zlepšenie smerovacích rozhodnutí. Tento článok sa zaoberá porovnaním výkonnosti klasického protokolu AODV a dvoch prístupov založených na RL – Monte Carlo (MC) a Deep Q-Learning (DQL) – v simulovanom prostredí Network Simulator 3 (NS-3) [4], [5].

Hlavným cieľom je vyhodnotiť výkonnosť jednotlivých riešení na základe metrík *Packet Delivery Ratio* (PDR), priepustnosti a priemerného oneskorenia v rôznych scenároch a posúdiť, či RL algoritmy dokážu prekonať tradičné smerovanie.

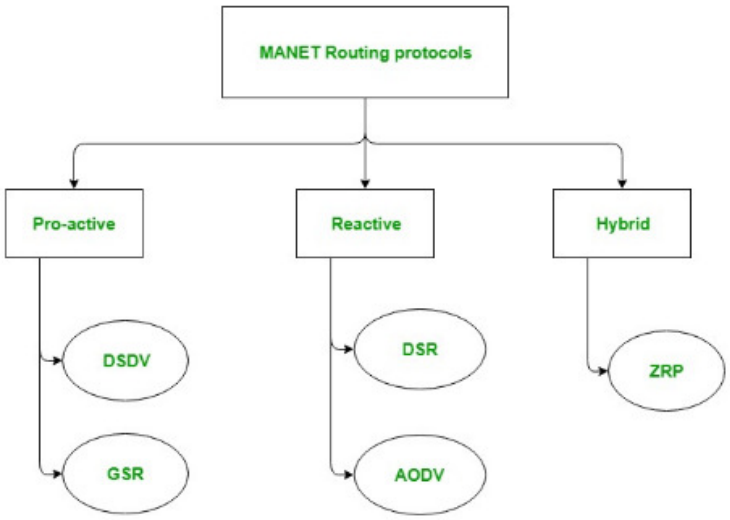
II. Súvisiaca práca a pozadie

MANET sú dynamicky sa meniace bezdrôtové siete, kde každý uzol funguje ako smerovač a prenáša dáta medzi ďalšími uzlami bez potreby centrálnej infraštruktúry. Vďaka ich flexibilite sa využívajú najmä v scenároch, kde je nemožné zabezpečiť stabilnú sieťovú infraštruktúru [1], [2].

Jedným z najpoužívanejších smerovacích protokolov v MANET je AODV, ktorý vytvára trasy len v prípade potreby. Jeho výhodou je nízka réžia, no často trpí oneskoreniami pri vyhľadávaní ciest a znižujúcou sa spoľahlivosťou pri vysokej mobilite uzlov [3]. Prehľad hlavných smerovacích protokolov v MANET je na Obr. 1.

V posledných rokoch sa objavujú prístupy, ktoré využívajú techniky strojového učenia, konkrétne RL, na inteligentné rozhodovanie o smerovaní. V práci [6] autori sumarizujú potenciál RL v oblasti sieťovej optimalizácie a smerovania. Algoritmy ako Monte Carlo a DQL sa ukazujú ako vhodné pre prostredia s neustále sa meniacou topológiou vďaka schopnosti adaptívne sa učiť z prostredia [7], [5], [8].

V práci porovnávame výkonnosť protokolu AODV s dvoma RL prístupmi v rovnakých podmienkach pomocou simulačného nástroja NS-3, ktorý umožňuje detailné testovanie v kontrolovanom a opakovateľnom prostredí.



Obr. 1 Smerovacie protokoly MANET

III. METODIKA

Na účely experimentálneho porovnania bol použitý diskrétny simulačný nástroj NS-3, ktorý poskytuje realistické modely bezdrôtových sietí, možnosť programovania vlastných protokolov v jazykoch C++ a Python, ako aj vizualizáciu pomocou nástroja NetAnim [9].

Simulácie boli realizované v dvoch scenároch:

- **Scenár 1**: menšia oblasť (100 m × 100 m) s nižším počtom uzlov (10, 20, 30), podrobné parametre sú uvedené v Tab. 1.
- **Scenár 2**: väčšia oblasť (150 m × 150 m) s vyšším počtom uzlov (50, 60, 70), podrobné parametre sú uvedené v Tab. 2.

Tabuľka 1
Vstupné parametre simulácie č. 1

Parameter	**Hodnota**
Počet uzlov	10, 20, 30
Veľkosť oblasti	100 m × 100 m
Veľkosť paketu	512 bajtov
Prenosová rýchlosť	512 kbps
Čas simulácie	100 sekúnd
Mobilný model	RandomWalk2dMobilityModel
WIFI štandard	IEEE 802.11n
Rýchlosť uzlov	3–5 m/s

Tabuľka 2
Vstupné parametre simulácie č. 2

Parameter	**Hodnota**
Počet uzlov	50, 60, 70
Veľkosť oblasti	150 m × 150 m
Veľkosť paketu	512 bajtov
Prenosová rýchlosť	512 kbps
Čas simulácie	100 sekúnd
Mobilný model	RandomWalk2dMobilityModel
WIFI štandard	IEEE 802.11n
Rýchlosť uzlov	3–5 m/s

Všetky scenáre používali rovnaký model mobility (`RandomWalk2dMobilityModel`) s rýchlosťou uzlov 3–5 m/s a štandard Institute of Electrical and Electronics Engineers (IEEE) 802.11n. Dĺžka simulácie bola 100 sekúnd. Prúd dát bol generovaný pomocou User Datagram Protocol (UDP) aplikácie s veľkosťou paketu 512 bajtov a prenosovou rýchlosťou 512 kbps.

AODV

Tradičný reaktívny protokol AODV bol použitý vo forme natívnej implementácie dostupnej v NS-3. Trasy sa vytvárali na požiadanie pomocou správ Route Request (RREQ) a Route Reply

(RREP).

Monte Carlo

RL agent na báze Monte Carlo používa jednoduchú neurónovú sieť a aktualizuje hodnoty Q až po ukončení každej epizódy. Počas prenosu paketu sa zaznamenávajú prechody (stav, akcia, odmena), ktoré sa spätne spracujú na konci. Politika výberu je *epsilon-greedy*.

Deep Q-Learning

Algoritmus DQL využíva dvojicu sietí (online a cieľovú) a metódu *experience replay*. Stav reprezentuje kombinácia vzdialeností k susedom, zaťaženie uzlov a ich rýchlosti. Aktualizácia prebieha priamo počas simulácie na základe Bellmanovej rovnice [4]. Odmena závisí od počtu skokov, úspešnosti prenosu a zaťaženia trasy.

Použité metriky

Výkon smerovacích algoritmov bol hodnotený pomocou troch hlavných metrík:

- **Packet Delivery Ratio (PDR)** – pomer úspešne doručených paketov,
- **Priepustnosť (Throughput)** – množstvo prenesených dát za časovú jednotku,
- **Priemerné oneskorenie (Average Delay)** – čas medzi odoslaním a doručením paketu.

IV. Výsledky a diskusia

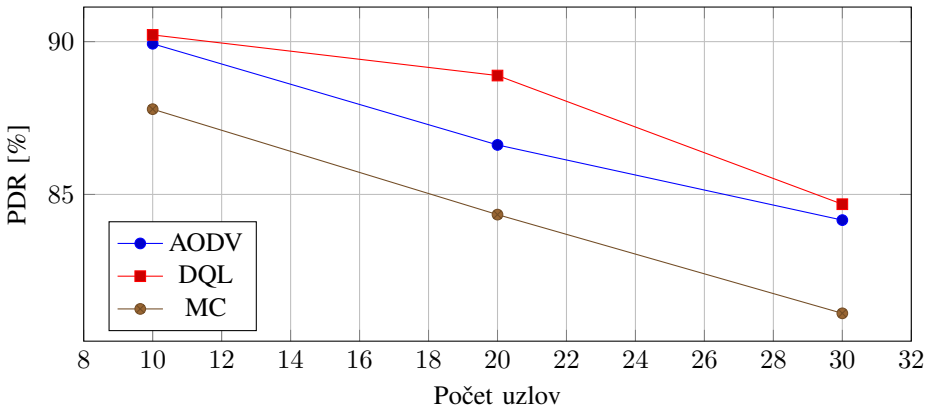
Výkonnosť smerovacích algoritmov bola hodnotená pomocou troch metrík: PDR, priepustnosť a priemerné oneskorenie. Simulácie prebiehali v dvoch odlišných topológiách a pre tri rôzne počty uzlov v každej z nich. Výsledky sú prezentované v nasledovných tabuľkách a grafickej vizualizácii.

Pomer doručených paketov (PDR)

Hodnoty PDR sú uvedené v tabuľke 3 pre simuláciu č. 1 a v tabuľke 4 pre simuláciu č. 2. Zodpovedajúce vizualizácie sú zobrazené na grafoch 2 a 3.

Tabuľka 3
Výsledné PDR v % pre simuláciu č. 1

Počet uzlov	AODV	DQL	MC
10	89.93	90.22	87.79
20	86.62	88.89	84.34
30	84.16	84.68	81.11

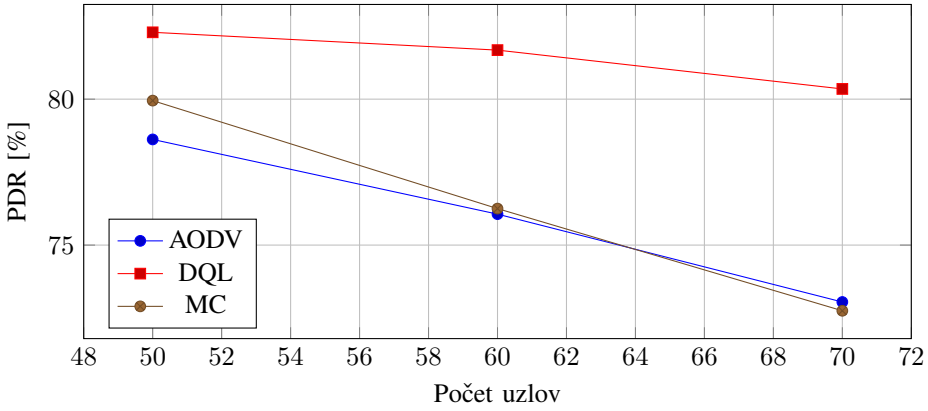


Obr. 2 Porovnanie PDR v simulácii č. 1

Výsledky ukazujú, že algoritmus DQL prekonal ostatné prístupy vo všetkých scenároch a počtoch uzlov. Najvyšší rozdiel bol zaznamenaný pri väčšej sieti so 70 uzlami (tab. 4), kde DQL dosiahol o viac ako 7 % vyšší PDR než AODV. S rastúcim počtom uzlov výkonnosť AODV aj MC výrazne klesá, zatiaľ čo DQL si udržiava stabilnú spoľahlivosť doručovania. To naznačuje lepšiu škálovateľnosť a adaptabilitu DQL v dynamickom prostredí.

Tabuľka 4
Výsledné PDR v % pre simuláciu č. 2

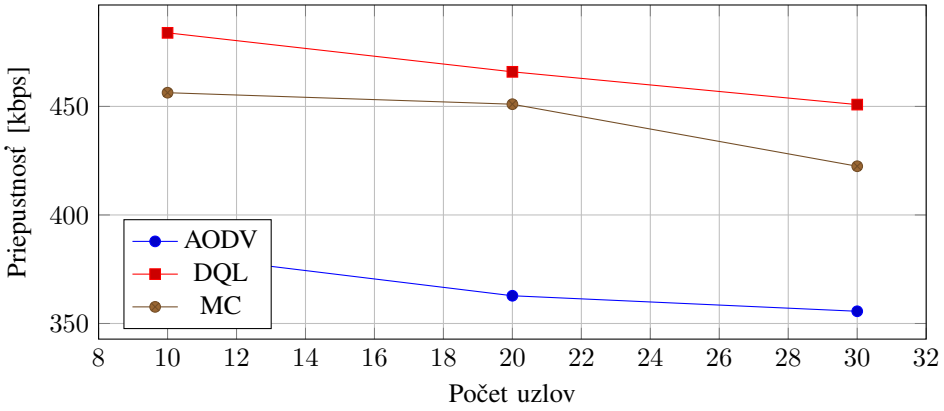
Počet uzlov	AODV	DQL	MC
50	78.62	82.29	79.95
60	76.06	81.68	76.25
70	73.05	80.35	72.75



Obr. 3 Porovnanie PDR v simulácii č. 2

Tabuľka 5
Výsledná priepustnosť v kbps pre simuláciu č. 1

Počet uzlov	AODV	DQL	MC
10	382.13	483.90	456.32
20	362.78	465.94	451.02
30	355.63	450.84	422.44



Obr. 4 Porovnanie priepustnosti v simulácii č. 1

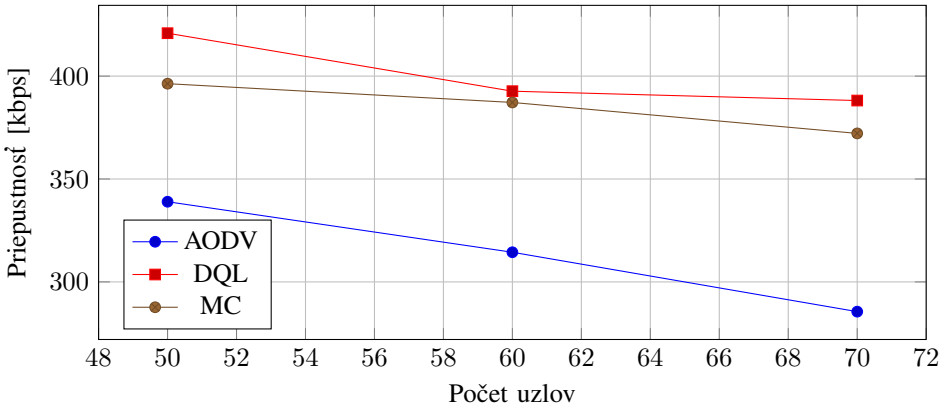
Tabuľka 6
Výsledná priepustnosť v kbps pre simuláciu č. 2

Počet uzlov	AODV	DQL	MC
50	338.96	420.85	396.32
60	314.41	392.67	387.20
70	285.55	388.13	372.16

Priepustnosť (Throughput)

Výsledky priepustnosti sú zhrnuté v tabuľkách 5 a 6, pričom ich porovnanie je graficky znázornené na grafoch 4 a 5.

Podobne ako pri PDR, aj priepustnosť dosahovala najvyššie hodnoty pri použití DQL. Zatiaľ čo priepustnosť AODV výrazne klesala s narastajúcim počtom uzlov (porovnaj napr. hodnoty v tabuľke 6), DQL si zachovával konzistentne vysoký výkon. MC dosahoval lepšie výsledky než AODV, no neprekonal DQL. Táto metrika potvrdzuje, že RL prístupy dokážu optimalizovať prenos dát aj v preťažených topológiách.



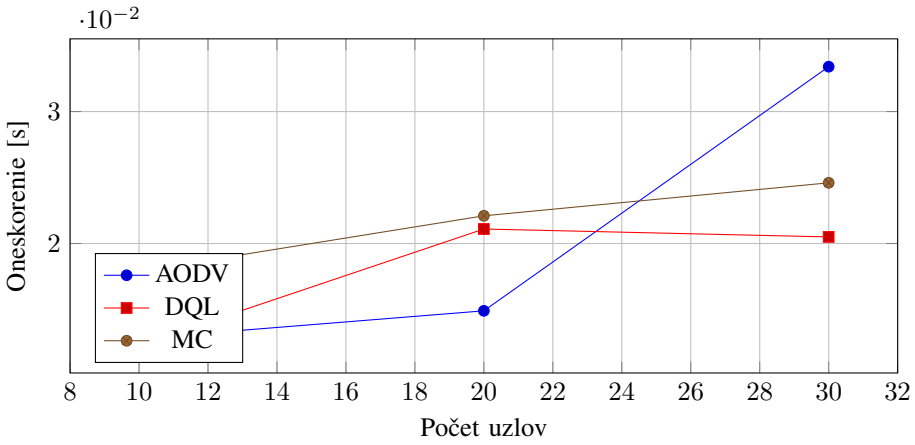
Obr. 5 Porovnanie priepustnosti v simulácii č. 2

Priemerné oneskorenie (Average Delay)

Tabuľky 7 a 8 obsahujú výsledky priemerného oneskorenia pre jednotlivé scenáre, zatiaľ čo grafická vizualizácia je uvedená na grafoch 6 a 7.

Tabuľka 7
Priemerné oneskorenie v sekundách pre simuláciu č. 1

Počet uzlov	AODV	DQL	MC
10	0.0128	0.0123	0.0179
20	0.0149	0.0211	0.0221
30	0.0334	0.0205	0.0246



Obr. 6 Porovnanie oneskorenia v simulácii č. 1

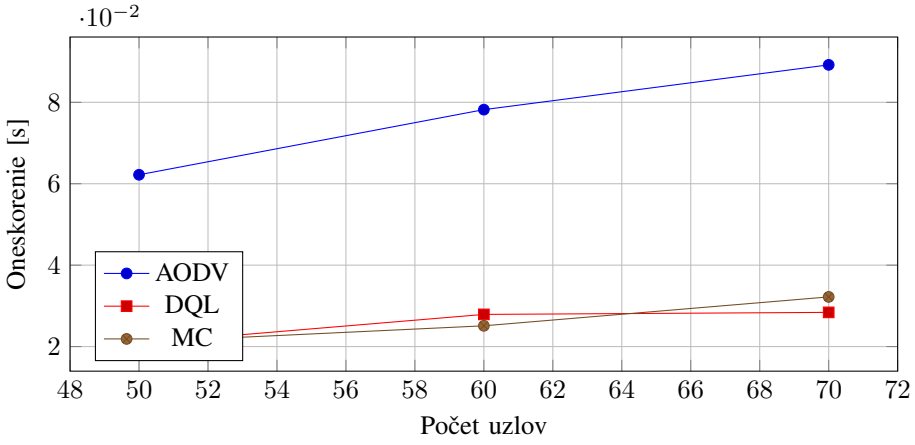
Tabuľka 8
Priemerné oneskorenie v sekundách pre simuláciu č. 2

Počet uzlov	AODV	DQL	MC
50	0.0622	0.0208	0.0211
60	0.0782	0.0279	0.0251
70	0.0892	0.0284	0.0322

Priemerné oneskorenie bolo najnižšie pri použití DQL, najmä pri veľkých topológiách. Naopak, AODV vykazoval dramatický nárast oneskorenia so zvyšujúcim sa počtom uzlov (napr. 0,0892 s pri 70 uzloch v tab. 8). MC bol mierne menej efektívny než DQL, ale stále výrazne lepší než AODV. Tento výsledok poukazuje na schopnosť DQL dynamicky reagovať na meniace sa podmienky v sieti a vyhľadávať efektívne trasy s minimálnou latenciou.

Diskusia

Z výsledkov v tabuľkách 3 až 8 a na grafoch 2 až 7 vyplýva, že algoritmus DQL dosahoval najvyššiu výkonnosť vo všetkých testovaných metrikách. PDR aj priepustnosť boli v prípade



Obr. 7 Porovnanie oneskorenia v simulácii č. 2

DQL konzistentne vyššie než u tradičného AODV a Monte Carlo, pričom rozdiely sa zvyšovali so stúpajúcim počtom uzlov.

Z hľadiska oneskorenia bol DQL taktiež najvýhodnejší – jeho priemerný čas doručenia paketov bol najnižší pri vyššej hustote siete. AODV vykazoval stabilné, no postupne zhoršujúce sa výsledky s rastúcim počtom uzlov, zatiaľ čo MC bol v niektorých prípadoch konkurencieschopný, ale menej adaptívny kvôli epizodickému učeniu.

Tieto výsledky poukazujú na potenciál posilňovacieho učenia v oblasti smerovania v dynamických bezdrôtových sieťach, najmä v scenároch s vysokou mobilitou a zaťažením. Zároveň potvrdzujú pozorovania iných prác, ktoré uvádzajú, že DQL je robustnejší voči zaťaženiu siete a topologickým zmenám [10], [5].

V. Záver

V tomto článku sme porovnali klasický smerovací protokol AODV s dvoma prístupmi založenými na posilňovanom učení – Monte Carlo a Deep Q-Learning – v prostredí mobilnej ad hoc siete simulovanej v NS-3. Výsledky ukazujú, že algoritmus Deep Q-Learning dosahoval konzistentne lepšiu výkonnosť vo všetkých sledovaných metrikách: PDR, priepustnosti aj oneskorení. Získané údaje poukazujú na vysoký potenciál metód umelej inteligencie pri riešení problémov smerovania v dynamických bezdrôtových topológiách. Do budúcnosti plánujeme rozšíriť model o ďalšie parametre siete, ako je spotreba energie či bezpečnostné prvky, a otestovať výkonnosť algoritmov vo väčšom počte reálnych scenárov.

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. Parker. (2024) Ad-hoc networks explained: Understanding their nature and impact. Accessed: Oct. 1, 2024. [Online]. Available: https://agileblue.com/ad-hoc-networks-explained-understanding-their-nature-and-impact
[2] A. Khalid, R. A. Rehman, and B.-S. Kim, "Caching strategies in ndn based wireless ad hoc network: A survey," *ResearchGate*, 2024. [Online]. Available: https://www.researchgate.net/publication/382369505
[3] M. Ghawy and A. Al-quh, "Performance evaluation of aodv protocol," *International Journal of Computer Applications*, 2019.
[4] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
[5] A. El Fergougui *et al.*, "Routing in manets using deep reinforcement learning: A survey," *Computer Communications*, vol. 180, pp. 1–13, 2021.
[6] N. Luong, D. T. Hoang, and D. N. et al., "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.
[7] Y. Yang and H. Liu, "Deep reinforcement learning for routing in dynamic wireless networks," *IEEE Access*, vol. 10, pp. 10 456–10 469, 2022.
[8] V. Mnih, K. Kavukcuoglu, D. Silver *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[9] *NS-3 Manual*, The NS-3 Project, 2024, https://www.nsnam.org/documentation/.
[10] J. Jiang, B. Zhang, Z. Shi, and W. Liu, "A reinforcement learning-based routing protocol for manets," *Computer Networks*, vol. 195, p. 108209, 2021.

Riadenie robotického ramena

[1]*Peter GIROVSKÝ,* [2]*Jaroslava ŽILKOVÁ,* [3]*Tadeáš KMECIK,* [4]*Daniel KOVÁČ*

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

[1]peter.girovsky@tuke.sk, [2]jaroslava.zilkova@tuke.sk, [3]tadeas.kmecik@tuke.sk,
[4]daniel.kovac@student.tuke.sk

***Abstrakt* — V tomto príspevku sa zaoberáme návrhom riadiaceho programu pre 5DOF robotické rameno s použitím servomotorov Dynamixel. V rámci práce bola urobená implementácia algoritmov inverznej kinematiky pre presné ovládanie polohy a integrovanie pákového ovládača. Prostredníctvom programu MATLAB bol vytvorený modulárny systém s možnosťou úloh typu pick-and-place a plánovania trajektórií.**

***Kľúčové slová* — Inverzná kinematika, riadenie robotického ramena, MATLAB, Dynamixel servomotory**

I. Úvod

Ľudská schopnosť manipulovať jednotlivú pozíciu objektov rôzneho tvaru a veľkosti je mnohokrát limitovaná. Často sa človek ocitne v situácii, kde nie je schopný vykonať cielenú úlohu, či už ide o náročné rozmery, hmotnosť daného objektu alebo nepriaznivé prostredie obkolesujúce danú prácu. V takýchto prípadoch robotika dokáže plnohodnotne nahradiť prácu človeka. Robotické rameno v našom prípade má 5 stupňov voľnosti a predstavuje jednoduché a lacné riešenie založené na pohonoch Dynamixel. Na testovanie a samotné ovládanie pohybových charakteristík ramena boli použité priama a inverzná kinematická úloha, ktoré sa počítali pomocou programu MATLAB v ktorom bolo zároveň robené aj ovládanie pomocou pákového ovládača.

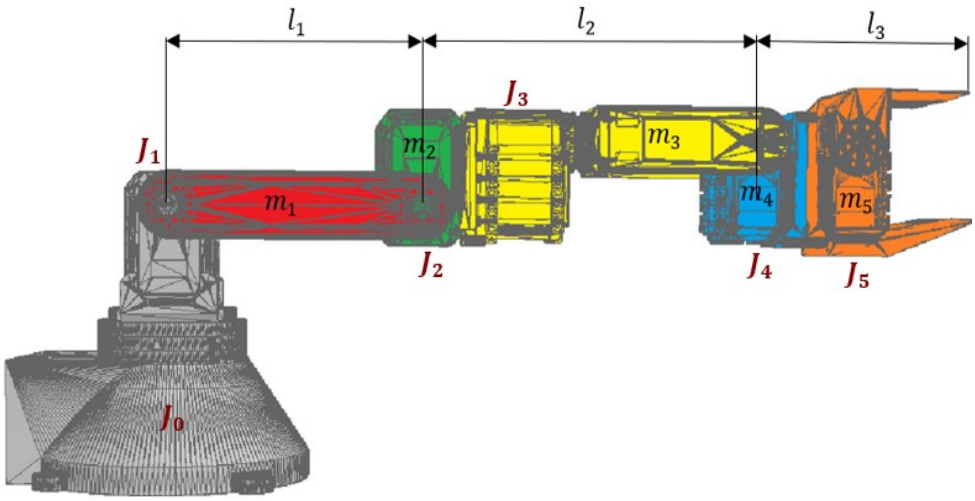
II. Riadenie robotického ramena

V tejto práci sa venujeme riadeniu modulárneho robotického ramena s piatimi stupňami voľnosti. Na zostrojenie robotického ramena, zobrazeného na obr. 1, boli použité servopohony Dynamixel typu AX-12A a MX-28AT. Pre riadenie bola použitá konštrukcia robotického ramena s 5 stupňami voľnosti a kinematickou štruktúrou s reťazcom R-R-R-R-R-R, nakoľko sa v praxi najviac používa. Štruktúra robotického ramena je zobrazená na Obr.1. Uloženie servopohonov Dynamixel, ktoré boli použité ako aktuátory, bude priamo v kĺboch z dôvodu zjednodušenia konštrukcie a modulárnosti. Parametre použitých servopohonov Dynamixel, použitých pre jednotlivé kĺby J1-J5, sú uvedené v Tabuľke 1.

Tabuľka 1
Parametre použitých pohonov Dynamixel

	Servopohony Dynamixel		
	AX-12A	**AX-18A**	**MX-28AT**
Hmotnosť	54,6 g	55,9 g	77 g
Rozsah natočenia	0-300°	0-300°	0-360°
Rozlíšenie	0-1023	0-1023	0-4095
Blokovací moment	1,5 Nm (pri 12 V)	1,8 Nm (pri 12 V)	2,5 Nm (pri 12 V)
Uhlová rýchlosť naprázdno	59 ot/min (pri 12 V)	97 ot/min (pri 12 V)	55 ot/min (pri 12 V)
Prevod	254:1	254:1	193:1

Ovládanie motorov je vyhotovené pomocou MATLAB programu a externého pákového ovládača pre priame ovládanie.



Obr. 1 Robotické rameno.

V Tabuľke 2 sú uvedené parametre robotického ramena s 5DOF aj s koncovým efektorom, kde *m* predstavuje hmotnosť vyznačenej časti a *l* predstavuje dĺžku jednotlivých častí.

Tabuľka 2
Parametre robotického ramena

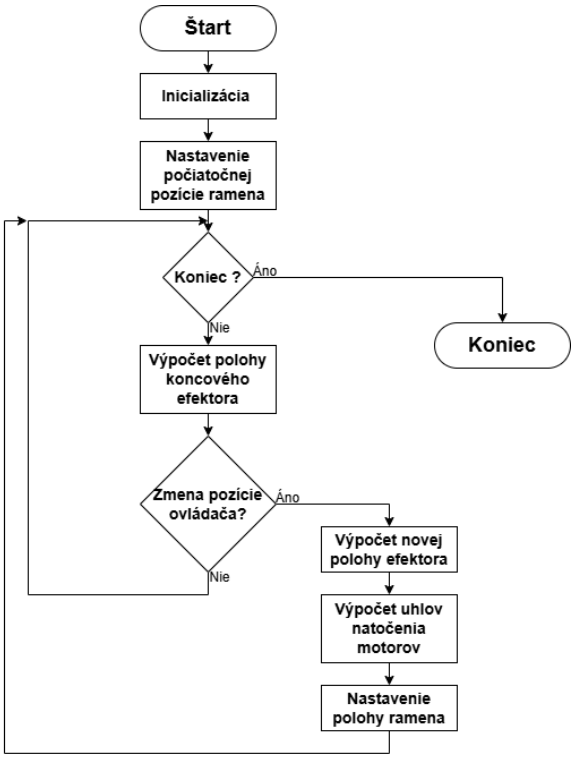
Parameter	**Hodnota**
m_1	31g
m_2	82g
m_3	100g
m_4	59g
m_5	95g
l_1	11cm
l_2	13,7cm
l_3	8,75cm

Samotný riadiaci program robota je rozdelený do 2 programov. Prvý program slúži pre priame ovládanie robotického ramena pomocou pákového ovládača a druhý program pre riadenie robota priamym zadávaným súradníc v osiach x, y, z.

A. Priame ovládanie robotického ramena pomocou pákového ovládača

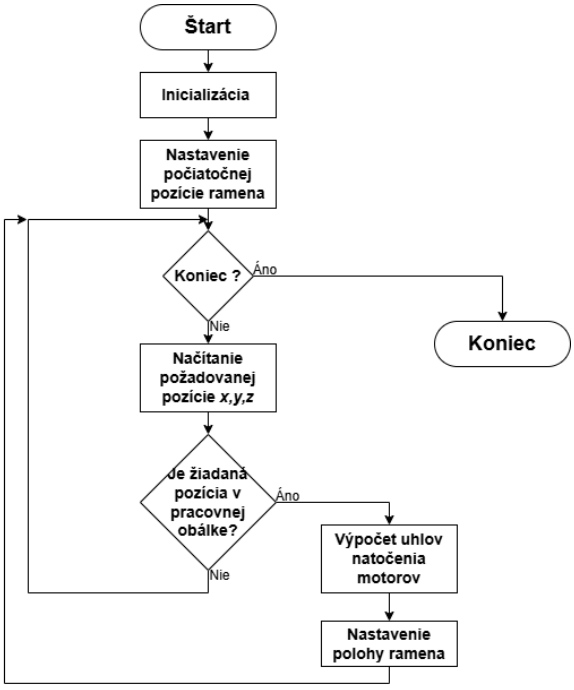
Pákový ovládač pre riadenie pohyby robotického ramena bol pripojený k počítaču cez USB konektor. Pre túto aplikáciu bol použitý pákový ovládač „*phantom hawk flightstick*“ kvôli jeho schopnosti ovládať robotické rameno v osiach x,y,z a zároveň aj využiť rotáciu v osi z. Navyše tento ovládač je vybavený mnohými tlačidlami, čo umožňuje ich využitie pre nastavenie a ovládanie robotického ramena.

Algoritmus priameho ovládania robotického riadenia je zobrazený na Obr.2.



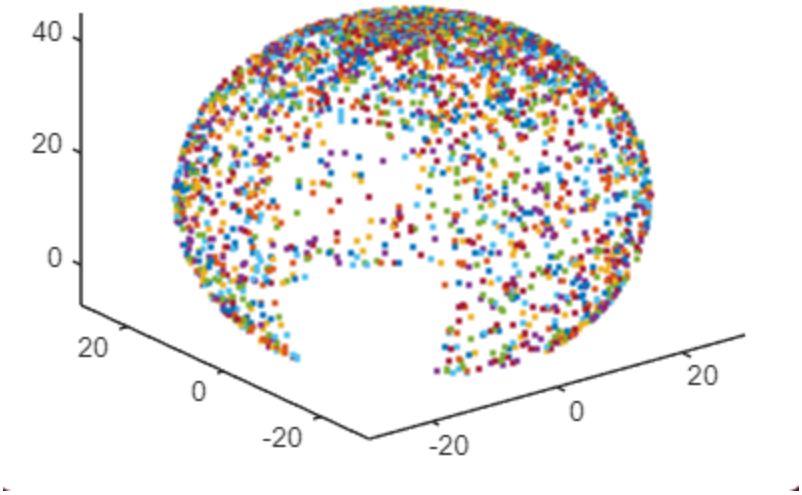
Obr. 2 Algoritmus priameho riadenia.

B. Súradnicové riadenie robotického ramena



Obr. 3 Algoritmus riadenia robotického ramena zadávaním koncových súradníc efektora.

Súradnicové riadenie robotického ramena predstavuje riadiaci program, v ktorom má užívateľ možnosť zvoliť jednotlivé súradnice koncového bodu efektora. Používateľ v tomto prípade zadá *x,y,z* koncové pozície a program následne nastaví robota do pozície, ktorá odpovedá zadanej hodnote koncového bodu. Algoritmus ovládania robotického riadenia pomocou zadávania želaných súradníc je zobrazený na Obr.3. V rámci tohto riadenia sa kontroluje pracovná obálka robotického ramena zobrazená na Obr.4. Pracovná obálka robotického ramena predstavuje pracovný priestor, v ktorom je robotické rameno schopné vykonávať svoj pohyb a jej tvar je závislý najmä od fyzických rozmerov robotického ramena, počtu a druh kĺbov.



Obr. 4 Pracovná obálka robotického ramena.

C. Kinematika robota

V robotike existujú dva hlavné prístupy pri určovaní polohy robotického ramena (kinematiky), na výpočet pozície a orientácie koncového efektora vzhľadom na základňu robota:

- Geometrické metódy - využívajú klasickú euklidovskú geometriu a trigonometrické vzťahy na výpočet polohy a orientácie, využívajú trojuholníky, kružnice a trigonometrické funkcie, zohľadňujú dĺžky článkov a uhly kĺbov.
- Algebraické metódy - využívajú maticovú algebru, najmä homogénne transformačné matice, na popis pohybov a pozícií. Každý článok a kĺb je reprezentovaný transformačnou maticou (napr. Denavit-Hartenbergova reprezentácia) a pozícia efektora sa získa násobením týchto matíc. [1,2,3,4,5]

Porovnanie základných vlastností týchto metód sa nachádza v Tabuľke 3.

Tabuľka 3
Porovnanie vlastností geometrických a algebraických metód

Kritérium	Geometrické metódy	Algebraické metódy
Intuitívnosť	vysoká	Nízka
Presnosť	Stredná	Vysoká
Vhodnosť pre 3D	Obmedzená	Výborná
Výpočtová náročnosť	Nízka	Vyššia
Automatizácia výpočtov	Ťažšia	Jednoduchá
Škálovateľnosť	Slabá	Výborná

Pre riadenie robotického ramena a výpočet kinematiky bola použitá algebraická metóda, pričom pre popis kinematiky bola využitá Denavit-Hartenbergova reprezentácia. Pre každý kĺb robotického ramena bola zostavená transformačná matica podľa všeobecnej rovnice 1. Následne vynásobením jednotlivých transformačných matíc dostaneme celkovú transformačnú maticu koncového efektora. [1,5]

$$T_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \cos\alpha_i & \sin\theta_i \sin\alpha_i & \alpha_i \cos\theta_i \\ \sin\theta_i & \cos\theta_i \cos\alpha_i & -\cos\theta_i \sin\alpha_i & \alpha_i \sin\theta_i \\ 0 & \sin\alpha_1 & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

kde:

θ_i - je uhol rotácie okolo osi z_{i-1}
d_i - je posun pozdĺž osi z_{i-1}

a_i - je dĺžka článku (vzdialenosť medzi osami z_{i-1} a z_i pozdĺž osi x_{i-1})
α_i - je uhol medzi osami z_{i-1} a z_i okolo osi x_{i-1}

Pre overenie správnosti výpočtu pozície robotického ramena pomocou algebraickej metódy je možné za určitých predpokladov použiť aj geometrickú metódu. Samotné robotické rameno sa nachádza v 3D oblasti, ale nakoľko motor nachádzajúci sa v kĺbe *J0* (podľa Obr.1), ovláda natočenie ramena cez uhol $\theta 1$, tak nemá vplyv na dosah samotného ramena. Vynechaním tohto uhla transformujeme robotické rameno z 3D priestoru do 2D, vypočítame pozíciu koncového efektora pomocou geometrickej metódy a samotný uhol $\theta 1$ vieme následne vypočítať zostrojením pravouhlého trojuholníka so stranami odpovedajúcimi súradniciam *x,y*. Teoretické porovnanie výpočtu uhlov natočenia jednotlivých kĺbov pre generované náhodné pozície je uvedené v Tabuľke 4, pričom z uvedených výsledkov je vidno, že takýto zjednodušený prístup pre výpočet uhlov natočenia servomotorov môže fungovať.

Tabuľka 4
Porovnanie prepočítaných uhlov natočenia jednotlivých uhlov servopohonov Dynamixel pomocou geometrickej metódy a algebraickej metódy

Náhodne generované pozície [x, y, z]	**Prepočítané uhly $\theta 1$, $\theta 2$, $\theta 3$, $\theta 4$ pomocou geometrickej metódy**	**Prepočítané uhly $\theta 1$, $\theta 2$, $\theta 3$, $\theta 4$ pomocou algebraickej metódy**
[17, 16, 8]	46.73; 53.70; 160.48; 91.62	46.73; 53.73; 160.45; 91.62
[11, 6, 14]	61.38; 96.71; 138.36; 72.29	61.38; 95.71; 38.93; 72.72
[9, 10, 12]	41.98; 55.33; 91.46; 30.78	41.98; 54.47; 191.04; 32.06
[11, 16, 7]	34.50; 33.58; 85.62; 47.25	34.50; 33.62; 185.65; 47.17
[12, 11, 15]	47.48; 53.30; 198.12; 54.98	47.48; 53.46; 198.24; 54.70

III. ZÁVER

V tomto príspevku bol predstavený návrh a implementácia riadiaceho algoritmu pre 5DOF robotické rameno využívajúce servomotory Dynamixel typu MX-28AT a AX-12A. Navrhnutý riadiaci program slúžil pre priame ovládanie robotického ramena pomocou pákového ovládača a pre súradnicové riadenie robotického ramena. Použitie Denavit-Hartenbergovej metódy nám umožnilo zostaviť rovnice pre priamu a inverznú kinematickú úlohu, pomocou ktorých bolo možné dosiahnuť plynulé ovládanie robotického ramena. Prínosom tejto práce je navrhnutý riadiaci program ktorý, po zmene údajov dĺžky ramien je možné použiť na ovládanie rôznych robotických ramien s piatimi stupňami voľnosti.

POĎAKOVANIE

Tento príspevok vznikol s podporou projektu KEGA 032TUKE-4/2024. Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00018.

LITERATÚRA

[1] John J. Craig. Introduction to Robotics Mechanics and Control. Third Edition Pearson India; 3RD, INTERNATIONAL ECONOMY EDITION. ISBN-10 : 8131718360
[2] Marco Ceccarelli. Fundamentals of Mechanics of Robotic Manipulation (Mechanisms and Machine Science, 112). 2nd ed. 2022 Edition.SPRINGER NATURE, [S.l.], 2022 ISBN-10 3030908461
[3] Ahmed A. Shabana Computational dynamics 2nd ed John Wiley, New York, 2001 ISBN:9780471053262
[4] Hubert Hahn Rigid body dynamics of mechanisms Springer, Berlin, 2002- ISBN: 9783540423737
[5] Mark W. Spong, M. Vidyasagar Robot Modeling and control John Wiley & Sons Inc ISBN: 047161243X

Dopad variability obnoviteľných zdrojov energie na cenotvorbu elektriny: Kvantitatívna korelačná analýza pre slovenský energetický trh

[1]*Marek PAVLÍK,* [2]*Maroš BOSÁK*

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

[1]marek.pavlik@tuke.sk, maros.bosak@student.tuke.sk

***Abstrakt* — V posledných rokoch je možné pozorovať zvýšenú volatilitu ceny elektriny v Európe. Výsledky mnohých štúdii poukazujú na fakt, že obnoviteľné zdroje energie, najmä slnečné a veterné elektrárne spôsobujú nízke, až dokonca záporné ceny elektriny na burze. Tento príspevok sa zameriava na výskum korelácie medzi parametrami (residual load, výroba energie zo slnka, výroba energie z vetra) a cenou elektriny na Slovensku. Korelačná analýza poskytuje pohľad na správanie sa trhu s elektrinou. Štáty v EÚ majú rôzny energetický mix, čo môže zásadne ovplyvniť ceny elektriny v danom štáte. Pre korelačnú analýzu bolo vybrané Slovensko, ktoré má vyvážený energetický mix a ktorý tvoria aj konvenčné zdroje (jadrové a tepelné elektrárne) ale aj obnoviteľné zdroje (vodná, slnečná a veterná energia).**

***Kľúčové slová* — trh s elektrinou, OZE, vplyv OZE na cenu elektriny, korelačná analýza**

I. Vývoj obnoviteľných zdrojov energie v Európe a na Slovensku

V posledných rokoch zaznamenali európske energetické trhy významné zmeny v dynamike cien elektriny. Zatiaľ čo veľkoobchodné ceny boli v minulosti relatívne stabilné a predvídateľné, dnes sa čoraz častejšie stretávame s výraznými výkyvmi, navyše sa objavujú dokonca záporné ceny elektriny. Tento trend volatility ceny elektriny je obzvlášť výrazný po roku 2021, keď trhy zasiahlo niekoľko krízových faktorov: rastúce ceny pohonných hmôt, geopolitické napätie, ako aj rýchly nárast výroby energie z obnoviteľných zdrojov (OZE).

Základ tvorby cien na trhu s elektrinou je založený na mikroekonomických princípoch, najmä na modeloch ponuky a dopytu, kde ceny sú určené prienikom výrobných nákladov a dopytu. Na rozdiel od tradičných komodít nie je možné elektrinu skladovať vo veľkých množstvách, čo robí ceny mimoriadne citlivými na zmeny ponuky a dopytu v reálnom čase. Cena elektriny sa na krátkodobých trhoch určuje najčastejšie formou aukcie, kde sa obchoduje na základe najnižších výrobných nákladov, čo sa označuje ako princíp merit order. Princíp merit order predstavuje základný spôsob, akým sa určuje cena elektriny na liberalizovaných veľkoobchodných trhoch, najmä na krátkodobom trhu s dodávkou v deň vopred. V tomto systéme sa jednotlivé ponuky výrobcov elektrickej energie zoradia podľa výšky ich marginálnych nákladov, teda nákladov potrebných na výrobu jednej MWh elektriny. Výrobcovia s najnižšími nákladmi sú do dodávky zapojení ako prví, zatiaľ čo drahšie zdroje sa pridávajú až vtedy, keď je potrebné pokryť vyšší dopyt. Týmto spôsobom sa vytvára merit order krivka, ktorá predstavuje vzťah medzi cenou a množstvom elektriny dostupnej za daných podmienok. Na základe aktuálneho dopytu sa určí, ktorá elektráreň ako posledná vstúpi do dodávky. Jeho cena sa stáva trhovou cenou pre všetkých akceptovaných výrobcov. Merit order mechanizmus zabezpečuje nielen cenovú efektívnosť, ale aj spoľahlivosť dodávok. Tým, že zvýhodňuje výrobcov s najnižšími marginálnymi nákladmi, podporuje konkurenčné správanie a motivuje výrobcov, aby stanovovali reálne ceny, blízke svojim skutočným nákladom. Súčasne vytvára istotu, že zapojení výrobcovia si pokryjú prevádzkové náklady, čím sa udržiava ich pripravenosť dodávať energiu v čase potreby. [1-5]

V európskom kontexte sa s elektrinou obchoduje prevažne na denných trhoch a vnútrodenných trhoch, kde volatilitu na takýchto trhoch zhoršuje rastúca integrácia variabilných obnoviteľných zdrojov energie, čo vytvára potrebu flexibilných vyrovnávacích mechanizmov za účasti na strane

dopytu. Na trhu Nord Pool bola volatilita cien spojená s hydrologickými podmienkami a obmedzeniami na cezhraničných vedeniach. Na Slovensku je najviac vyrobenej elektrickej energie z jadrových elektrární. Podiel vyrobenej elektrickej energie z jadra je stále nad 50 % z celkovej vyrobenej elektrickej energie. Z pohľadu CO2 je jadrová energia považovaná za "čistú", pretože táto energia je vyrábaná bez emisií. [1-5]

Vodné elektrárne u nás tvoria významnú časť vo výrobe elektrickej energie. Celkový inštalovaný výkon u nás je viac ako 3300 MW. Vodné elektrárne majú približnú produkciu elektrickej energie 5000 GWh, tým sú hneď produkciou za jadrovými elektrárňami.

Slnečné elektrárne majú v našej krajine inštalovaný výkon približne 840 MW. Podľa štatistík je na Slovensku približne 2300 fotovoltických parkov. V tomto celkovom čísle sú už započítané aj malé elektrárne a fotovoltické parky. Z pohľadu budúcnosti je predpoklad, že výstavba fotovoltík či už na rodinných domoch, firmách bude narastať. Výroba energie zo slnka je rovnako ako elektrická energia z jadrových elektrárni považovaná za energiu bez CO2. Uhlie a zemný plyn je v posledných rokoch na Slovensku na zostupe. Kým v minulosti tepelné elektrárne tvorili cca 20 % z podielu vyrobenej elektrickej energie tak v poslednom roku to bolo len približne 5 % ročne, pričom v súčasnosti boli dve tepelné elektrárne (Nováky, Vojany) odstavené. Dôvod poklesu výroby energií z týchto elektrárni je vysoká cena emisných povoleniek, ktorá sa odráža na spaľovaní hlavne uhlia. Veterné elektrárne sú u nás najmenej rozvinutým obnoviteľným zdrojom, pričom na Slovensku tvoria kapacitu približne 3,1 MW, a to je len 1 % z celkového inštalovaného výkonu na Slovensku. Energia vyrobená z týchto elektrárni patrí medzi tie najviac „zelené". Potenciál týchto elektrární je vysoký, keďže netvoria žiadne CO2, nevyrábajú žiadny odpad a nemajú negatívny vplyv na zdravie človeka. [6]

V rokoch 2005 – 2022 sa na Slovensku sa rapídne zvýšil podiel vyrobenej zelenej energie z obnoviteľných zdrojoch zo 6,4 % v roku 2005, na 17,3 % v roku 2020 a 22,9 % v roku 2022. V roku 2024 bol Slovenskej republike dosiahnutý podiel výroby energie z OZE na úrovni 24,2 % z toho: [7]

- Vodné elektrárne 71 %
- Solárne elektrárne 23 %
- Bioenergia vrátane bioplynu 5,99 %
- Veterné elektrárne 0,01 %

Európsky energetický trh prechádza v posledných rokoch zásadnou transformáciou, ktorej hlavnými hnacími silami sú klimatické ciele, tlak na znižovanie emisií a zvyšovanie podielu obnoviteľných zdrojov energie. Tieto ciele sa premietajú do energetických stratégií jednotlivých krajín a ovplyvňujú aj fungovanie trhu s elektrinou. So zvyšujúcim sa podielom veterných a fotovoltických elektrární na výrobe elektriny stúpa potreba analyzovať ich vplyv na cenotvorbu, predovšetkým na krátkodobých trhoch, kde sa cena formuje dynamicky v závislosti od dopytu a ponuky.

Charakteristickou vlastnosťou OZE je ich premenlivosť – výroba z vetra a slnka nie je riadená priamo človekom, ale závisí od klimatických podmienok. To spôsobuje výkyvy vo výrobe, ktoré majú priamy dopad na rovnováhu medzi dopytom a ponukou a tým aj na vývoj ceny elektriny. Na lepšie pochopenie tohto vplyvu sa v energetike využíva pojem residual load – ide o časť dopytu, ktorá nie je pokrytá z OZE a musí byť zabezpečená konvenčnými alebo flexibilnými zdrojmi. Práve vývoj tejto premennej úzko súvisí s výkyvmi cien na trhu.

II. Korelačná analýza – Metodológia výskumu

Pre zistenie vzťahu viacerých parametrov je možné použiť korelačnú analýzu. Korelácia je závislosť medzi dvoma alebo viacerými premennými (napr. ak sa mení jeden parameter mení sa aj druhý a pod.). Hodnota korelačného koeficienta hovorí, aký je vzťah medzi premennými a či je tento vzťah kladný alebo záporný. Pearsonova korelácia meria lineárny vzťah medzi dvoma premennými. Je definovaná vzorcom:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \cdot \sqrt{\sum(Y_i - \bar{Y})^2}} \quad (1)$$

pričom:

- X_i = hodnoty výroby elektriny (napr. jadrová, veterná, výroba z fotovoltiky),
- Y_i = hodnoty ceny elektriny (Day Ahead Auction),
- $\bar{X}$ = priemer výroby elektriny,
- $\bar{Y}$ = priemer ceny elektriny ,
- $\sum$ = suma cez všetky pozorovania (dni).

Poznáme tri druhy korelácie (Obr.1):

1. Kladná korelácia + 1,00 – rast jednej premennej poukazuje na rast druhej premennej (napr. Výroba el. energie a cena rastú spolu)
2. Záporná korelácia - 1,00 – rast jednej premennej poukazuje na klesanie druhej premennej (napr. Pri vyššej výrobe, cena elektriny klesá)
3. Žiadna korelácia – 0 (Výroba nemá vplyv na spotrebu)

Obr. 1 Druhy korelácie - pozitívna korelácia (vľavo), negatívna korelácia (v strede) a bez korelácie (vpravo).[8]

Pre vysvetlenie, ak je situácia, pri ktorom sa skúma závislosť vyrobenej elektrickej energie z OZE a znižovaním emisií CO2 tak platí, čím bude výroba z OZE väčšia, tým budú ostatné elektrárne produkovať menej CO2. Výsledkom je kladná korelácia.

III. Výsledky výskumu – korelačná analýza

V rámci tohto výskumu sa spracovali a porovnávali dáta o výrobe elektriny z OZE a trhových cenách elektriny v časovom rámci niekoľkých rokov, konkrétne 2019 až 2024. Pre určenie mesačných a ročných korelácii boli použité dáta z výroby z fotovoltických a veterných zdrojov energie, celkový dopyt po elektrine a priemerné trhové ceny na dennom trhu (Day Ahead Market). Výpočet residual load je realizovaný ako rozdiel medzi celkovou spotrebou elektriny a výrobou z OZE. Dátové zdroje pochádzajú z verejne dostupných a overených databáz, ako sú energy-charts.

Na vyhodnotenie vzťahu medzi analyzovanými premennými bola použitá korelačná analýza, konkrétne výpočet Pearsonovho korelačného koeficientu (1). Tento štatistický nástroj umožňuje určiť silu a smer závislosti medzi množstvom vyrobenej elektriny z OZE a výslednou cenou elektriny.

Dáta pre Slovensko boli prevzaté z Energy charts s intervalom 1 hod. pre každý deň. Dáta z roku 2019 nie sú zverejnené pre výrobu Wind do mesiaca august. Cieľom bolo porovnať korelácie daných parametrov a zistiť ako sa menia rokmi. Tabuľky 1, 2 a 3 a grafické závislosti (Obr.2) zobrazujú mesačné hodnoty Pearsonovej korelácie medzi troma premennými – výroba z fotovoltiky (Fotovoltika), zvyškové zaťaženie (Residual load) a veterná výroba (Vietor) – a cenou elektriny na Day-Ahead aukcii v období rokov 2019 až 2024. Z výsledkov mesačnej korelácie ale aj ročnej vyplýva, čím je hodnota Residual load vyššia, tým narastá aj cena elektrickej energie, pretože do prevádzky sa zapájajú drahšie zdroje. Čím je hodnota „Fotovoltika“ vyššia, tým je cena elektriny nižšia. Negatívna korelácia znamená, že zvýšená výroba z fotovoltiky znižuje cenu, pretože znižuje potrebu nákladnejších zdrojov. Čím je hodnota „Vietor“ vyššia, tým je cena elektriny nižšia. Negatívna korelácia nám potvrdzuje, že zvýšenie výroby z vetra znižuje cenu elektriny. V prípade dát z roku 2019 pre „Vietor“, v mesiacoch január až august nebola korelácia vypočítaná pretože chýbajú dáta o výrobe energie z vetra.

Tabuľka 1
Hodnoty korelácií pre Slovensko pre roky: 2019, 2020, 2021[9]

	Rok: 2019			Rok: 2020			Rok: 2021		
Mesiac	Fotovoltika	Residual load	Vietor	Solar	Residual load	Wind	Fotovoltika	Residual load	Vietor
1	0,208	0,831	-	-0,164	0,776	-0,273	0,086	0,677	-0,180
2	-0,090	0,655	-	-0,010	0,583	-0,296	-0,243	0,830	-0,185
3	0,133	0,446	-	-0,193	0,797	-0,038	-0,116	0,756	0,091

4	-0,014	0,677	-	0,327	0,583	-0,004	-0,269	0,795	-0,057
5	-0,200	0,704	-	0,218	0,646	-0,070	-0,073	0,684	-0,330
6	0,039	0,811	-	-0,248	0,708	-0,045	0,138	0,625	-0,288
7	0,239	0,769	-	-0,038	0,803	0,046	0,320	0,702	0,242
8	-0,138	0,727	-	-0,135	0,660	0,147	-0,255	0,635	-0,134
9	-0,606	0,741	-0,141	-0,174	0,512	0,388	-0,530	0,598	0,017
10	-0,042	0,653	0,173	-0,362	0,763	-0,003	-0,266	0,332	-0,333
11	0,044	0,728	-0,276	-0,272	0,830	-0,394	-0,275	0,732	0,094
12	0,891	0,904	-0,219	-0,117	0,876	0,045	0,425	0,620	-0,037

Tabuľka 2
Hodnoty korelácií pre Slovensko pre roky: 2022, 2023, 2024[9]

	Rok: 2022			Rok: 2023			Rok: 2024		
Mesiac	Fotovoltika	Residual load	Vietor	Solar	Residual load	Wind	Fotovoltika	Residual load	Vietor
1	0,350	0,803	-0,172	0,320	0,770	-0,443	0,129	0,736	-0,185
2	0,053	0,457	-0,453	0,361	0,812	-0,696	-0,118	0,593	-0,483
3	-0,383	0,600	-0,273	-0,010	0,658	-0,320	0,066	0,788	-0,515
4	0,254	0,582	-0,333	-0,121	0,742	0,047	-0,221	0,824	-0,325
5	-0,331	0,740	-0,272	-0,393	0,663	0,197	-0,436	0,520	-0,244
6	0,183	0,510	-0,016	0,118	0,664	0,107	0,058	0,801	-0,238
7	-0,027	0,593	-0,187	-0,181	0,790	-0,321	-0,061	0,828	-0,319
8	-0,110	0,171	-0,060	0,134	0,618	-0,554	-0,031	0,773	0,073
9	0,417	0,221	-0,408	0,338	0,482	-0,263	0,352	0,133	0,043
10	0,269	0,405	-0,042	0,038	0,775	-0,237	-0,112	0,647	-0,329
11	-0,230	0,761	-0,394	-0,275	0,726	-0,243	-0,256	0,524	0,043
12	-0,286	0,809	-0,362	-0,388	0,903	-0,556	-0,257	0,630	-0,199

Tabuľka 3
Hodnoty ročnej korelácie pre Slovensko pre roky: 2019 až 2024[9]

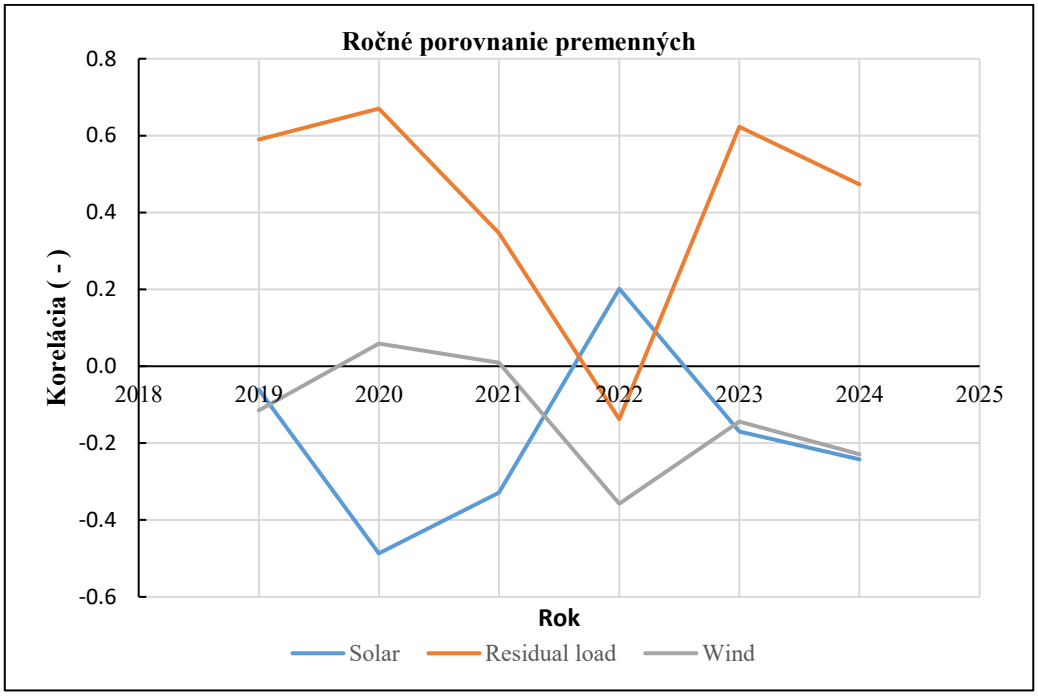
Ročné porovnanie			
Rok	Fotovoltika	Residual load	Vietor
2019	-0,061	0,589	-0,115
2020	-0,487	0,670	0,059
2021	-0,329	0,346	0,009
2022	0,202	-0,138	-0,358
2023	-0,170	0,623	-0,144
2024	-0,242	0,473	-0,229

Residual load, ako premenná vyjadrujúca rozdiel medzi celkovým dopytom a výrobou z obnoviteľných zdrojov, si vo väčšine rokov zachováva pozitívnu koreláciu s trhovou cenou elektriny. Výnimkou je však rok 2022, kedy korelácia prešla do záporného pásma (-0,138) (viď. Tab. 3). Tento výsledok je významný a môže naznačovať neštandardnú situáciu na trhu – napríklad nečakané cenové výkyvy nezávislé od dopytu, zmeny v štruktúre výroby alebo výrazný vplyv externých faktorov (napr. geopolitické udalosti či regulačné zásahy). V ostatných rokoch sa korelácia pohybuje od 0,346 (2021) po 0,670 (2020), čo potvrdzuje, že vyššia hodnota Residual load vo všeobecnosti vedie k rastu cien.

„Fotovoltika“ vykazuje počas sledovaného obdobia prevažne negatívnu koreláciu s cenou (viď. Obr. 2). Výrazný negatívny extrém nastal v roku 2020 (-0,487), čo poukazuje na silný cenotvorný

tlak pri zvýšenej produkcii z fotovoltiky – pravdepodobne v dôsledku prebytku výroby počas slnečných dní a nízkeho dopytu. Naopak, rok 2022 je výnimočný tým, že korelácia sa dostáva do kladného pásma (0,202), čo je v kontexte ostatných rokov odchýlka. Tento pozitívny vzťah môže byť výsledkom nižšej variability vo výrobe z fotovoltiky alebo netypického správania trhu, kde energia z fotovoltiky z rôznych dôvodov prestala tlačiť cenu výrazne nadol.

V prípade vetra (Vietor) je celkový trend menej výrazný, avšak stále zaujímavý. Korelácie sa pohybujú v blízkosti nulovej hodnoty, s mierne negatívnymi až slabými pozitívnymi hodnotami. V roku 2020 sa korelácia vyšplhala až na 0,059, zatiaľ čo v roku 2022 opäť klesla na -0,358 (viď. Obr. 2). To bol zároveň najsilnejší záporný vzťah za celé obdobie. Tento vývoj poukazuje na to, že vplyv veternej energie na cenu sa v čase mení, a to pravdepodobne v závislosti od objemu výroby, sezónnosti, ako aj schopnosti systému flexibilne reagovať na premenlivé podmienky. Treba však poznamenať, že na Slovensku nie je možné predpokladať koreláciu medzi výrobou elektriny z vetra a cenou elektriny pretože podiel výroby z vetra je na Slovensku minimálny.



Obr. 2 Grafická závislosť korelácie pre Slovensko od roku 2019 do roku 2024 s troma premennými: Solar (Fotovoltika), Residual load, Wind (Vietor) voči Day Ahead Auction.[9]

Výsledky potvrdzujú, že korelácia medzi obnoviteľnými zdrojmi a cenou nie je statická a mení sa v závislosti od širších trhových a technických podmienok. Kým residual load je vo väčšine rokov silným pozitívnym ukazovateľom cenového vývoja, fotovoltika a veterná energia vykazujú dynamickejšie správanie s viacerými výkyvmi. Zaujímavosťou sú najmä roky 2020 a 2022, ktoré sa vo všetkých troch premenných líšia od všeobecného trendu.

IV. Záver

Predkladaný príspevok sa venuje problematike vplyvu obnoviteľných zdrojov energie na cenu elektriny na Slovensku. V príspevku bola použitá korelačná analýza pre určenie korelácie medzi residual load, výrobou elektriny zo slnka, veternej energie a cenou elektriny. Korelačná analýza bola aplikovaná na dáta z roku 2019 až 2024. Z výsledkov zobrazených v Tab. 1 až 3 a Obr.2 vyplýva, že pre slovenský trh s elektrinou neexistuje korelácia medzi cenou elektriny a výrobou z veternej energie. Tento výsledok bolo možné predpokladať, nakoľko podiel výroby z veternej energie je na Slovensku veľmi nízky. Podobne je to aj v prípade slnečnej energie, kedy korelácia dosahuje taktiež nízke hodnoty. Výnimkou je korelácia medzi cenou elektriny a residual load. Korelácia medzi týmito dvoma parametrami dosahuje vo väčšine prípadov vysoké hodnoty, nad 0,6. Z výsledkov vyplýva, že čím je vyššia hodnota residual load, tým je vyššia aj cena elektriny. Inak povedané, čím je vyššia hodnota residual load, tým je potrebné zapojiť do výroby elektrickej energie viac konvenčných zdrojov, ktoré majú vyššie prevádzkové náklady, ktoré zvyšujú cenu ponuky na trhu s elektrinou.

POĎAKOVANIE

Táto práca bola podporená Nadáciou ZSE v projekte Dynamika cien elektriny na európskych trhoch: Dátová analýza a ekonomické súvislosti.

REFERENCIE

[1] Sensfuß, F.; Ragwitz, M.; Genoese, M. The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. Energy Policy 2008, 36, 3086–3094, ISSN 0301-4215. https://doi.org/10.1016/j.enpol.2008.03.035.

[2] He, Y. Real Options in the Energy Markets. Ph.D. Thesis, University of Twente, Enschede, The Netherlands, 2007; ISBN 978-90-365-2568-8.

[3] Botterud, A.; Kristiansen, T.; Ilic, M.D. The relationship between spot and futures prices in the Nord Pool electricity market. Energy Econ. 2010, 32, 967–978, ISSN 0140-9883. https://doi.org/10.1016/j.eneco.2009.11.009.

[4] Joskow, P.L. Lessons Learned From Electricity Market Liberalization. Energy J. 2008, 29, 9–42. https://www.jstor.org/stable/27085628.

[5] Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. Int. J. Forecast. 2014, 30, 1030–1081, ISSN 0169-2070. https://doi.org/10.1016/j.ijforecast.2014.08.008..

[6] VÝROBA ELEKTRICKEJ ENERGIE NA SLOVENSKU A JEJ ZDROJE. CenyEnergii.sk, [online]. 1. júna 2021 [cit. 01. 06. 2025]. Dostupné na internete: https://cenyenergii.sk/2021/06/01/vyroba-elektrickej-energie-na-slovensku-a-jej-zdroje/

[7] PODIEL VÝROBY ELEKTRICKEJ ENERGIE Z OBNOVITEĽNÝCH ZDROJOV ENERGIE. Úrad pre reguláciu sieťových odvetví, [online]. 9. februára 2021 [cit. 28. 05. 2025]. Dostupné na internete: https://www.urso.gov.sk/podiel-vyroby-elektrickej-energie-z-obnovitelnych-zdrojov-energie/

[8] GUY-EVANS, Olivia. Correlation in Psychology: Meaning, Types, Examples & Coefficient. Simply Psychology, [online]. 2022 [cit. 19. 5. 2025]. Dostupné na internete: https://www.simplypsychology.org/correlation.html

[9] Bosár, M., Analýza cien elektriny a jej korelácia s výrobou z OZE pre rôzne krajiny, Bakalárska práca

Procedurálny generátor polygonálnych povrchových modelov stromov s využitím L-systémov

[1]*Mykhailo BAKHIR,* [2]*Branislav MADOŠ*

[1,2] Katedra počítačov a informatiky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Letná 9, 040 01 Košice, Slovenská republika

[1]mykhailo.bakhir@student.tuke.sk, [2]branislav.mados@tuke.sk

***Abstrakt* — Jednou z najnáročnejších úloh pri vývoji 3D hier je vytvorenie realistických a vizuálne atraktívnych virtuálnych svetov. Mimoriadne zložitým aspektom je modelovanie prírodného prostredia s veľkým množstvom rastlín, keďže ručné vytváranie každého prvku je mimoriadne časovo náročné a neefektívne. V tomto kontexte sa procedurálne generovanie javí ako aktuálny a perspektívny prístup. V rámci tejto práce boli preskúmané metódy procedurálneho generovania vegetácie s využitím klasického postupu používajúceho bezkontextové gramatiky – L-systémy. Opísaný je proces návrhu vlastných gramatík, ich interpretácie pomocou metódy Turtle interpretera a tvorby textúr. Prínosom tejto práce je návrh generátora a jeho implementácia v programovacom jazyku Python s využitím knižníc na spracovanie 3D geometrie. Navrhované riešenie umožňuje efektívnu automatizáciu tvorby rozmanitej vegetácie pre rozsiahle virtuálne prostredia, ktoré môžu byť integrované do hier využívajúcich aj najmodernejšie herné vývojové prostredia a enginy, ako je Unity alebo Unreal.**

***Kľúčové slová* — 3D hry, modely vegetácie, L-systém, Trimesh Python, Blender, Unreal Engine**

I. Úvod

Procedurálne generovanie obsahu je plne etablované medzi nástrojmi na tvorbu herných svetov. Už v roku 1980 bola táto technológia použitá pri vývoji hry Rogue, ktorá sa stala prvým príkladom jej využitia v hernom priemysle. Postupom času sa táto technológia zdokonaľovala a dnes je ťažké predstaviť si vývoj AAA hry, ktorý by nezahŕňal procedurálne generovanie herného obsahu ako jeden zo zdrojov optimalizácie využitia ľudských, finančných ako aj časových zdrojov.

Z nepreberného množstva známych počítačových hier, v rámci vývoja ktorých bola využitá technológia procedurálneho generovania obsahu ako je prírodné prostredie alebo objekty, možno spomenúť počítačové hry Minecraft, No Man's Sky, Ghost of Tsushima, Skyrim, GTA V alebo Starfield. Dopyt po tejto technológii každoročne rastie, avšak medzi indie vývojármi hier nie je procedurálne vytváranie modelov vegetácie a prírodných prvkov natoľko rozšírené, keďže často poskytuje menšiu mieru kontroly nad výsledkom. Projekty, ktoré sa ju pokúšajú využiť, bývajú vizuálne jednotvárne – časté opakovanie rovnakých textúr a modelov znižuje originálnosť prostredia.

Preto je dôležitým cieľom vytvoriť procedurálny generátor rastlín, ktorý poskytne používateľovi väčšiu kontrolu nad výsledkom a súčasne zabezpečí väčšiu rozmanitosť a originalitu modelov. Motiváciou pre vytvorenie generátora navrhnutého a implementovaného v rámci tohto článku je predpoklad, že v budúcnosti môže byť využitý pri vývoji hier, s možnosťou prispôsobenia podľa potrieb vývojárov.

Procedurálne generovanie rastlín nachádza uplatnenie nielen v hernom priemysle, ale aj v mnohých ďalších oblastiach, kde je potrebná realistická alebo hromadná vizualizácia vegetácie, redukcia manuálnej práce či modelovanie prírodných procesov. Medzi hlavné smery aplikácie patria okrem oblasti tvorby počítačových hier aj filmový a animačný priemysel, architektúra a krajinársky dizajn ale aj botanika, agronómia a ekológia. Tu je možné pomocou procedurálneho generovania modelovať dynamiku procesov rastu a morfogenézy rastlín na základe formálnych modelov, ako sú napríklad L-systémy [1].

Na zhodnotenie výhod procedurálneho vytvárania rastlín je potrebné najprv pochopiť, prečo vývojári začali používať automatizované systémy namiesto ručného modelovania jednotlivých rastlín. Využitie takýchto prístupov má viacero výhod, medzi ktoré patrí úspora času a zdrojov. V priemere sa v hernom priemysle 30% až 40% rozpočtu venuje na vytváranie herného obsahu (generovanie prostredia, objektov, textúr a pod.) [2], pričom procedurálne generovanie môže túto položku výrazne zredukovať. Ďalšou výhodou je variabilita a škálovateľnosť, keď jediný algoritmus môže vytvoriť stovky unikátnych verzií tej istej rastliny s rôznou výškou, tvarom alebo rozvetvením. Pozitívom sú aj menšie nároky na úložisko, keď sa namiesto ukladania veľkého množstva unikátnych 3D modelov ukladajú len pravidlá, na základe ktorých je možné tieto modely generovať. Napriek mnohým výhodám má procedurálne vytváranie rastlín aj viaceré nevýhody v porovnaní s tradičným, manuálnym modelovaním [3], ako je obmedzená kontrola nad výsledkom, opakovateľnosť a stereotypnosť, neprirodzené alebo nerealistické štruktúry, zložitosť nastavenia parametrov a pravidiel alebo obmedzenie možnosti ladenia a opráv.

Častým argumentom proti generovaniu prostredia procedurálnym generátorom je, že vznikne repetitívny, fádny obsah. Preto bol vykonaný výskum [4], ktorý ukázal, že používatelia vnímajú obsah generovaný procedurálnym generátorom a vytvorený dizajnérmi manuálne ako podobne kvalitný, pokiaľ ide o kategórie ako je "zaujímavosť", "zvedavosť na obsah hry" a "chuť vrátiť sa k hre".

Napriek aktívnemu využívaniu procedurálnych generátorov v súčasnom hernom priemysle, najmä na syntézu virtuálnej vegetácie, zostávajú tieto technológie strategicky významné a komerčne cenné. Vývoj vysokokvalitných generátorov si vyžaduje značné zdroje — čas, špecializované odborné znalosti a výpočtovú infraštruktúru. Z tohto dôvodu mnohé spoločnosti uchovávajú svoje algoritmy a nástroje v uzavretom prístupe, pričom ich považujú za intelektuálne vlastníctvo s konkurenčnou výhodou. K dispozícii sú však softvérové riešenia, ktoré podporujú procedurálne generovanie vegetácie, pričom sú k dispozícii tak komerčné riešenia, ako aj riešenia, ktoré sú open source. Patria medzi ne Florasynth Obr.1a [5], SpeedTree Obr. 1b [6] a Plantarium Obr.1c [7]. Vzhľadom na vyššie uvedené sa cieľom tejto práce stalo vytvorenie procedurálneho generátora polygonálnych povrchových modelov stromov, resp. kríkov, ktorý by umožnil generovanie širšej škály rôznych stromov a im podobných kríkov a zabezpečiť tak širšiu dostupnosť takéhoto riešenia.

a) b) c)

Obr. 1 Používateľské rozhrania softvérových riešení procedurálnych generátorov modelov rastlín a) Florasynth, b) SpeedTree a c) Plantarium

II. PROCEDURÁLNY GENERÁTOR

Vytvorenie procedurálneho generátora polygonálnych povrchových modelov stromov si vyžiadalo realizáciu viacerých podúloh, medzi ktoré patrí:

- Vytvorenie textúr pre jednotlivé časti rastliny, ako je textúra kôry, listu, kvetu alebo plodu.
- Vytvorenie bezkontextovej gramatiky, ktorá umožní generovanie derivačného reťazca popisujúceho základnú konštrukciu stromu a vytvorenie samotného generátora derivačných reťazcov.
- Vytvorenie Turtle interpretera, ktorý graficky interpretuje vygenerovaný derivačný reťazec a pretransformuje ho do polygonálneho povrchového modelu stromu.
- Vytvorenie používateľsky prívetivého používateľského rozhrania, ktoré umožní nastavovanie parametrov generovania.

A. Tvorba textúr

Pre generovanie textúry kôry stromu bol využitý algoritmus pre generovanie Voronoiovho šumu. Aby bol vygenerovaný vzor viac pretiahnutý a pripomínal zvislé praskliny na kôre, bola upravená mierka generovanej textúry pozdĺž osi z (z = 0,1), čím sa bunky natiahli v jednom smere. Typ vzdialenosti bol zvolený ako Chebyshev (maximálna vzdialenosť v osiach), čím sa vytvorili hranatejšie a geometricky výraznejšie okraje buniek. Tieto nastavenia spolu umožnili dosiahnuť charakteristickú štruktúru podobnú popraskanej drevenej kôre.

S cieľom zabrániť vzniku pravidelného vzoru a aby sa zabezpečil realistickejší vzhľad materiálu kôry, bol pridaný k textúre slabý vplyv šumu. Tým sa zachoval základný vzor produkovaný algoritmom Voronoi, ale povrchu bola pridaná jemná nerovnosť, ktorá znížila jeho umelosť. Následne bol pridaný aj prvok realistickej zelene, ktorá sa často objavuje na stromovej kôre. Na pridanie zelených prvkov bol použitý ďalší šum. Vďaka tomu sa na povrchu objavili náhodné zelené oblasti, čo zvýšilo dojem prirodzenosti. Je možné vytvárať textúry s rôznym množstvom vegetácie, tak ako ukazuje Obr. 2.

Obr. 2 Rôzne úrovne pokrytia kôry stromu machom, podľa úrovne nastavenia príslušného parametra.

B. Generovanie derivačného reťazca

Na vytvorenie gramatiky je nevyhnutné pochopiť, že mnoho rastlín má symetrickú štruktúru a výrazné opakujúce sa vzory. Phyllotaxia [8] ako oblasť botaniky skúma usporiadanie listov a iných častí rastlín, ktoré je často charakteristické opakujúcimi sa motívmi na rôznych úrovniach. Práve tieto vzory a symetria tvoria základ ich tvaru a štruktúry. Pomocou rekurzie je tak možné efektívne modelovať to, že sa istá logika opakuje viackrát, avšak s rôznymi parametrami. To umožňuje vytvárať komplexné a realistické modely rastlín.

Práve na tento účel je využívaná bezkontextová gramatika, využívajúca prepisovacie pravidlá, ktoré v každej iterácii transformujú generovaný derivačný reťazec, keď jeho symbol prepíšu na sekvenciu symbolov.

Symbol F je napr. možné použiť pre „posun dopredu“. Reťazec symbolov FFF potom znamená posunutie dopredu o tri kroky (takto sú vytvárané kmeň a vetvy). Pomocou symbolu „+“ je možné otočiť sa doprava v ose x, použitie symbolu „-“ predstavuje naopak natočenie doprava v tejto ose. Aby bolo možné vrátiť sa do určitého bodu v reťazca, je tento bod označený pomocou hranatých zátvoriek „[B]“. Symbol „B“ je určený pre vytvorenie vetvy stromu.

Príklad gramatiky umožňujúcej generovanie stromu je uvedený v (1). Predstavuje v textovej forme zapísaný axióm `axiom`, teda reťazec, z ktorého generovanie derivačného reťazca začína a takisto v textovej forme zapísanú množinu prepisovacích pravidiel `rules`, ktoré sú uvedené v tvare A : B. Pri zostavovaní derivačného reťazca sa na jednotlivé symboly, ktoré ho tvoria, aplikujú derivačné pravidlá z množiny `rules` a to tak, že A predstavuje symbol, ktorý bude pri použití daného prepisovacieho pravidla v derivačnom reťazci, v nasledujúcej iterácii jeho zostavovania, nahradený reťazcom symbolov B. Všetky symboly a reťazce symbolov sú v množine `rules` uvedené tak, že sú uzavreté medzi symbolmi `''`. Oddeľovacím znakom jednotlivých pravidiel v množine je znak `;`.

```
axiom = 'FFFL'
rules = {
  'S': 'FFFFFL';
  'L': 'FFF[BCDEL]';
  'B': '[+++FFFf[&&&FFFf]FFFf]';
  'C': '[---FFFf[&&&FFFf]FFFf]';
  'D': '[&&&FFFf[+++FFFf]FFFf]';
  'E': '[^^^FFFf[---FFFf]FFFf]'
}
```
(1)

C. Turtle interpretér

Poslednou dôležitou zložkou systému umožňujúceho generovať polygonálne povrchové modely stromov s využitím L-systémov je grafický interpreter symbolov derivačného reťazca. Príklad takejto grafickej interpretácie možno vidieť na Obr. 3.

Jednotlivé symboly sú potom interpretované napr. takto:

- Symbol **F** – je graficky interpretovaný ako frustum, teda zrezaný ihlan, ktorého dolná aj horná podstava je tvorená pravidelným mnohouholníkom, ktorého počet hrán (segmentov povrchu) je parametrom generovania modelu. Je naň natiahnutá textúra kôry stromu.
- Symbol **f** – je graficky interpretovaný ako list stromu. V najjednoduchšom prípade je interpretovaný ako obdĺžnik, na ktorý je obojstranne natiahnutá textúra listu.

a) b) c)

Obr. 3 Príklady procedurálne generovaného stromu, generovaného podľa zhodnej gramatiky avšak s iným seedom pre vygenerovanie jeho rôznych verzií.

III. Testovanie Dosiahnutého Riešenia

Pre otestovanie dosiahnutého riešenia boli realizované testy generovania geometrie polygonálnych povrchových modelov stromov, pričom bol sledovaný vplyv rôznych natavení parametrov. Test III A bol zameraný na vplyv dĺžky derivačného reťazca, test III B bol zameraný na vplyv počtu segmentov frusta.

A. Testovanie vplyvu dĺžky reťazca na generovanie modelu stromu

Generovanie derivačného reťazca s použitím gramatiky L-systému je iteratívny proces, pričom s každou iteráciou môže dochádzať k signifikantnému nárastu dĺžky tohto derivačného reťazca. Realizovaný test sleduje, ako sa dĺžka tohto reťazca odzrkadľuje na čase vykonávania generovania polygonálneho modelu stromu a ako vplýva na veľkosť súboru, do ktorého je finálna geometria modelu stromu uložená. Dosiahnuté výsledky je možné vidieť v Tab. 1.

Tabuľka 1 Vývoj parametrov generovaného polygonálneho modelu stromu v závislosti od dĺžky derivačného reťazca.

Počet iterácií	Dĺžka reťazca	Čas vykonávania [s]	Veľkosť súboru [KB]
3	111	1,028	2 004
5	388	3,429	7 583
8	756	6,920	15 459
10	1 033	9,600	21 789

B. Testovanie vplyvu počtu segmentov frusta na generovanie modelu stromu

V prípade tejto sady testov bol počet iterácií generovania derivačného reťazca pomocou L-systému pevne nastavený na 6, čo viedlo k derivačnému reťazcu dĺžky 502 znakov. Variabilný bol parameter počet segmentov frusta.

Získané výsledky zhrnuté v Tab. 2 ukazujú, že počet polygónov rastie približne lineárne s počtom segmentov frusta, keďže každý segment prispieva konštantným počtom plôch. Mierne odchýlky sú spôsobené rôznymi faktormi ako delenie segmentov alebo topológia zlučovania polygónových sietí. Pri zdvojnásobení počtu segmentov zo 4 na 8 sa počet polygónov zväčšil o 26 496 (z 64 377 na 90 873), čo je nárast okolo 41%. Ďalšie zdvojnásobenie vedie k ešte výraznejšiemu nárastu – z 90 873 na 143 865 (52 992, resp. 58 %). Veľkosť súboru v KB koreluje s počtom polygónov, no rastie o niečo pomalšie, čo naznačuje, že textová reprezentácia .obj formátu (vrcholy aj steny) absorbuje časť rozšírenia geometrie. Čas vykonania sa mení relatívne málo v porovnaní s počtom polygónov, keď prechod zo 4 na 64 segmentov priniesol nárast času vykonávania z 4,30s na 5,45s, teda len o 27 %.

Na základe vykonaných meraní možno konštatovať, že implementovaný generátor stromov vykazuje stabilné a škálovateľné správanie pri zvyšovaní parametrov ako počet iterácií či segmentov. Použitie fixného seedu zabezpečilo opakovateľnosť testov. Výsledky zároveň ukázali, že aj pri zvýšenej komplexite výstupu zostáva výpočtový čas na prijateľnej úrovni, čo potvrdzuje praktickú využiteľnosť generátora v rôznych aplikáciách – od off-line renderovania až po interaktívne prostredia. Výsledky potvrdzujú, že generátor je nielen výpočtovo efektívny, ale aj flexibilný, čo umožňuje jeho použitie v rôznych kontextoch – od off-line vizualizácií až po real-time generovanie, pričom je možné efektívne využívať aj možnosť úrovní detailov (LOD).

Tabuľka 2 Vývoj parametrov generovaného polygonálneho modelu stromu v závislosti od počtu segmentov, ktoré tvoria jeho jednotlivé frustá.

Počet segmentov	Počet polygónov	Čas vykonania [s]	Veľkosť súboru [KB]
4	64 377	4,30	6 390
8	90 873	4,24	8 198
16	143 865	4,43	11 582
32	249 849	4,76	19 618
64	461 817	5,45	35 312

IV. Záver

Táto práca sa zameriavala na návrh a implementáciu procedurálneho generátora stromov a kríkov. Boli preskúmané spôsoby tvorby gramatiky, jej rozširovania o nové pravidlá a tiež interpretácie výsledných reťazcov pomocou tzv. Turtle Graphics, čo umožňuje vizuálne reprezentovať vývoj rastlinnej štruktúry v 3D priestore. Ďalšou kľúčovou časťou analýzy bolo štúdium knižnice Trimesh, ktorá poskytla potrebné nástroje na tvorbu a manipuláciu s 3D geometriou. Pomocou nej sa podarilo vytvoriť základné stavebné prvky ako kmeň, konáre a listy a tieto následne prepojiť do celistvého modelu stromu.

Výsledkom práce je univerzálne a deterministické riešenie procedurálneho generátora, ktoré umožňuje generovanie rozmanitých trojrozmerných rastlín podľa jednoducho definovateľných gramatických pravidiel. Vďaka použitiu seedu je výstup generátora opakovateľný, čo je dôležité najmä v kontexte simulácií, vizualizácie a testovania. Flexibilita riešenia sa prejavila aj možnosťou generovať rôzne druhy rastlín (napr. kríky s bobuľami) iba zmenou gramatiky, bez potreby zásahu do samotného algoritmu.

Hoci bol systém navrhnutý ako off-line nástroj, jeho potenciál ďalšieho rozvoja je značný. Medzi možné vylepšenia do budúcnosti patrí rozšírenie podpory exportu do moderných 3D formátov ako .gltf, .fbx alebo .usd, ktoré umožňujú prenos animácií, materiálov a pokročilých štruktúr. Ďalšou možnosťou je pridanie podpory pre generovanie v reálnom čase s náhľadom, čo by výrazne zvýšilo interaktivitu a komfort práce s generátorom. Vítané by bolo aj zavedenie samostatného exportu textúr alebo podpora komplexnejších shaderových riešení, ktoré by mohli zlepšiť vizuálnu kvalitu výstupu. Napokon, perspektívnym smerom je aj prepojenie systému so štatistickými alebo biologickými dátami, ktoré by umožnili tvorbu modelov založených na reálnych vzorkách.

Vytvorené riešenie poskytuje funkčný základ pre ďalšiu prácu v oblasti počítačového modelovania rastlín a môže nájsť uplatnenie pri jednoduchších vizualizáciách alebo experimentoch s rastovými modelmi.

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] S. Chuai-Aree, W. JÄGER, H.G. Bock a S. Siripant, "Simulation and Visualization of Plant Growth Using Lindenmayer Systems", In: Modeling, Simulation and Optimization of Complex Processes: Proceedings of the International Conference on High Performance Scientific Computing, March 10–14, 2003, Hanoi, Vietnam. Springer, 2005, ss. 115–126, DOI: https://doi.org/10.1007/3-540-27170-8_9

[2] M. Hendrikx, S. Meijer, J. van Der Velden a A. Iosup, "Procedural Content Generation for Games: A Survey", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP). 2013, roč. 9, ss. 24, DOI: https://doi.org/10.1145/2422956.2422957

[3] K. Baran, C. Krijgsman, S. Lambregts, N. Posner a M. Wisboom, "Procedural automation of general game level generation: the good, the bad and the ugly", Apríl 2020, DOI: https://doi.org/10.13140/RG.2.2.31998.74564

[4] L. Rodrigues, R. Bonidia a J. Brancher, "Procedural versus human level generation: Two sides of the same coin?" International Journal of Human-Computer Studies, 2020, Vol. 141, ISSN 1071-5819, DOI: https://doi.org/10.1016/j.ijhcs.2020.102465

[5] Florasynth Procedural Tree Generator, citované 8.6.2025, dostupné aj z: https://www.florasynth.com/

[6] SpeedTree, citované 8.6.2025, dostupné aj z: https://store.speedtree.com/

[7] Plantarium, citované 8.6.2025, dostupné aj z: https://plant.max-richter.dev/#

[8] S.G. Hotton, "Symmetry of plants", Dizertačná práca, University of California, Santa Cruz, December 1999.

K slovenskému prekladu dotazníkov na meranie používateľskej skúsenosti vo virtuálnej realite

[1]Štefan KOREČKO, [2]Jaroslav MATHES, [3]Branislav SOBOTA

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

[1]stefan.korecko@tuke.sk, [2]jaroslav.mathes@student.tuke.sk [3]branislav.sobota@tuke.sk

Abstrakt **– Dotazníky predstavujú dôležitú súčasť subjektívnych metód hodnotenia používateľskej skúsenosti, zachycujúcich osobný pohľad používateľov na príslušný systém. V priebehu posledných desaťročí bolo navrhnutých množstvo dotazníkov zameraných na rôzne aspekty skúsenosti z ktorých niektoré sa stali v podstate štandardnými. Pri snahe využiť štandardizované dotazníky narážame v podmienkach Slovenskej republiky na problém neexistencie ich ustáleného prekladu. Ako istý impulz pre vypracovanie takéhoto znenia ponúkame v tomto článku slovenský preklad sady štyroch dotazníkov pre hodnotenie virtuálno–realitných systémov, konkrétne System Usability Scale, NASA Task Load Index, Virtual Reality Sickness Questionnaire a Igroup Presence Questionnaire.**

Kľúčové slová **– dotazník, použiteľnosť, používateľská skúsenosť, preklad, virtuálna realita**

I. Úvod

Pod pojmom *používateľská skúsenosť* či *zážitok* (*UX*, User Experience) rozumieme celkovú skúsenosť, ktorú používateľ získa pri interakcii s príslušným produktom, službou alebo systémom. Poskytnúť čo najlepšiu skúsenosť by malo byť cieľom každého takéhoto prostriedku. Pri softvérových produktoch je obzvlášť dôležitá pri systémoch *virtuálnej reality* (*VR*, Virtual Reality), kde je primárnym cieľom poskytnutie čo najautentickejšej a najpohlcujúcejšej formy interakcie používateľa s počítačom vytvoreným *virtuálnym prostredím* (*VP*).

Pri hodnotení systémov zameranom na UX je vhodné kombinovať dva typy metód, subjektívne a objektívne. Tieto dva typy sú chápané v zmysle zaužívaného pohľadu na testovanie UX [1]: objektívne sú založené na meraní pomocou prístrojov, subjektívne, tiež označované ako postojové (attitudinal) [2], na osobnom pohľade účastníkov testovania, vyjadrenom pomocou dotazníkov, rozhovorov a ďalších prístupov.

Tento postup sa dodržiava aj pri vyhodnocovaní VR systémov vytvorených v rámci laboratória inteligentných rozhraní komunikačných a informačných systémov, LIRKIS, na domovskom pracovisku autorov, ako sú systém L-NeRVEn pre neurorehabilitáciu v kolaboratívnej VR pre pacientov po mozgovej príhode [3], [4] či výučbové VR prostredia so scenárom definovaným procesným grafom [5]. Zariadením sprostredkujúcim virtuálne prostredie používateľovi je tu VR prilba (VR headset), zvyčajne z rodiny produktov Meta (Oculus) Quest.

Primárnym zdrojom pre objektívne vyhodnotenie UX na pôde laboratória LIRKIS sú merania miery úspešnosti a času trvania riešenia jednotlivých úloh a videozáznamy, nahrávané počas príslušného testovania z pohľadu prvej a tretej osoby. Záznam z pohľadu prvej osoby je zabezpečený snímaním obrazového výstupu VR prilby, z pohľadu tretej osoby je snímaný samostatnou kamerou, umiestnenou na statíve. Videozáznamy slúžia na zhodnotenie činnosti používateľov vo VP z pohľadu experimentátora, vrátane anomálií v tejto činnosti, a tiež ako sekundárna forma určenia dĺžky plnenia jednotlivých úloh.

Ako subjektívne metódy, zachytávajúce pohľad jednotlivých účastníkov príslušných experimentov, sa primárne využívajú dotazníky a rozhovory. Aby sa zabezpečila čo najväčšia miera kompatibility z inými podobnými meraniami a taktiež miera spoľahlivosti výsledkov, boli pre dotazníkovú časť vybrané štyri značne rozšírené dotazníky, zamerané na rôzne aspekty UX. Prvé dva dotazníky majú široké spektrum využitia a je ich možné použiť aj pre neimerzívnu VR. Sú to *System Usability Scale* [6], [7] pre hodnotenie použiteľnosti systému a *NASA Task Load*

Index [11], kde používateľ vyjadrí do akej miery bolo plnenie úloh v systéme zaťažujúce. Zvyšné dva dotazníky má zmysel použiť iba pri imerzívnej VR (v našom prípade VR prilby). Ide o *Virtual Reality Sickness Questionnaire* [17], zameraný na negatívne aspekty VR zážitku, a *Igroup Presence Questionnaire* [25], zachytávajúci pocit prítomnosti vo virtuálnom prostredí.

Jednou z úloh s ktorou sa bolo potrebné vysporiadať pri aplikácii týchto dotazníkov v Slovenskej republike bolo vypracovanie ich znenia v slovenskom jazyku, keďže ich štandardizované znenie je zvyčajne k dispozícii v angličtine, prípadne niekoľkých ďalších svetových jazykoch. Ako istý krok k štandardizácii týchto dotazníkov v slovenskom jazyku ponúkame v tomto článku nami vypracované slovenské znenie. Slovenské znenie je uvedené v častiach II až V. Záverečná časť VI sa venuje vybraným aspektom prekladu a ďalšiemu postupu.

II. System Usability Scale

System Usability Scale (*SUS*) [6], [7] je veľmi rozšírený štandardný dotazník umožňujúci používateľovi zhodnotiť použiteľnosť systému. Obsahuje 10 položiek (otázok) a odpovede na ne sú vyjadrené na 5-stupňovej Likertovej škále. Z odpovedí sa vypočíta celkové skóre v rozpätí 0 až 100, kde 100 znamená najvyššiu možnú použiteľnosť. SUS dotazník sa bežne používa aj v kontexte VP a VR, napríklad na porovnanie rôznych typov VR prilieb [8], metód interakcie s VP [9] či hodnotenie architektonických simulácií [10].

Anglické znenie	**Preklad**
I think that I would like to use this system frequently.	Myslím si, že by som rád(a) používal(a) tento systém často.
I found the system unnecessarily complex.	Systém mi prišiel zbytočne komplikovaný.
I thought the system was easy to use.	Systém sa mi zdal ľahko použiteľný.
I think that I would need the support of a technical person to be able to use this system.	Myslím si, že by som potreboval(a) technickú podporu aby som vedel(a) používať tento systém.
I found the various functions in this system were well integrated.	Funkcionality systému sú podľa mňa dobre zapracované (integrované).
I thought there was too much inconsistency in this system.	Zdalo sa mi, že v tomto systéme je priveľa nekonzistentností (nezrovnalostí).
I would imagine that most people would learn to use this system very quickly.	Viem si predstaviť, že väčšina ľudí by sa tento systém naučila veľmi rýchlo používať.
I found the system very cumbersome to use.	Tento systém mi príde ťažkopádny na použitie.
I felt very confident using the system.	Pri používaní systému som sa cítil(a) sebaisto.
I needed to learn a lot of things before I could get going with this system.	Než som mohol(la) začať pracovať s týmto systémom, musel(a) som sa naučiť veľa vecí.

Tabuľka 1
Preklad položiek dotazníka System Usability Scale

Pôvodné anglické znenie a preklad jednotlivých položiek dotazníka SUS sú uvedené v Tabuľke 1. Preklad vychádza z už existujúceho slovenského prekladu využívaného vo výučbe[1] na domovskom pracovisku autorov. Odpovede sú preložené zaužívaným spôsobom, “Strongly Disagree” ako “Vôbec nesúhlasím”, “Disagree” ako “Nesúhlasím”, “Not sure” ako “Nie som si istý(á)”, “Agree” ako “Súhlasím” a “Strongly Agree” ako “Úplne súhlasím”.

III. NASA Task Load Index

NASA Task Load Index (*NASA-TLX*) [11] umožňuje posúdiť ako používateľ vníma záťaž spojenú s plnením úloh v príslušnom systéme. Záťaž používateľ hodnotí v zmysle 6 dimenzií, ku každej priradí hodnotenie od 0 do 100 v 5-bodových intervaloch (akoby 21-stupňová Likertová škála). Týmito dimenziami sú záťaž mentálna, fyzická a časová (časová tieseň), ďalej dosiahnutý výkon (úspešnosť), vynaložená námaha a úroveň frustrácie. Príkladmi použitia dotazníka v oblasti VP a VR sú hodnotenie mentálnej záťaže chirurgov pri laparoskopických operáciách [12], štúdia hodnotiaca kognitívnu záťaž pri nakupovaní vo VR [13] či evaluácia VR tréningu zameranom na lekársku prvú pomoc [14]. V poslednom čase sa objavujú rozšírenia NASA-TLX zamerané na VP a VR, napríklad [15], [16], ktoré k pôvodným šiestim pridávajú ďalšie dimenzie. V našom výskume používame pôvodný NASA-TLX, keďže dodatočné dimenzie sú pokryté ostatnými dotazníkmi.

Pôvodné anglické znenie a preklad jednotlivých položiek dotazníka NASA-TLX sú uvedené v Tabuľke 2. Dôvodom použitia množného čísla pre slovo “úloha” je fakt, že v našich testovacích scenároch bolo vždy viacero úloh a vyplnenie dotazníka nasledovalo až po splnení všetkých. Pri odpovediach sú označené iba krajné hodnoty, 0 ako “Veľmi málo” (pôvodne “Very Low”) a 100 ako “Veľmi veľa” (pôvodne “Very High”).

[1] https://kurzy.kpi.fei.tuke.sk/ux/

Anglické znenie	**Preklad**
Mental Demand: How mentally demanding was the task?	Mentálne zaťaženie: Ako mentálne náročné bolo plnenie úloh?
Physical Demand: How physically demanding was the task?	Fyzické zaťaženie: Ako fyzicky náročné bolo plnenie úloh?
Temporal Demand: How hurried or rushed was the pace of the task	Časový nátlak: Ako uponáhľané (pod tlakom) bolo plnenie úloh?
Performance: How successful were you in accomplishing what you were asked to do?	Výkon: Ako úspešní ste boli pri plnení úloh?
Effort: How hard did you have to work to accomplish your level of performance?	Úsilie: Ako ťažko ste museli pracovať aby ste dosiahli to, čo sa vám podarilo?
Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?	Frustrácia: Ako veľmi ste boli neistí, znechutení, podráždení, vystresovaní a otrávení?

Tabuľka 2
Preklad položiek dotazníka NASA Task Load Index

IV. Virtual Reality Sickness Questionnaire

V dotazníku *Virtual Reality Sickness Questionnaire* (*VRSQ*) [17] používateľ hodnotí vnímanie nevoľnosti spôsobenej zážitkom v pohlcujúcej VR. Ide o modifikáciu dotazníka Simulator Sickness Questionnaire (SSQ) [18], z ktorého Kim a kol. [17] vylúčili 5 otázok. VRSQ je určený špeciálne pre posúdenie VP, vnímaných prostredníctvom VR prilieb. Na odpovede slúži 4-stupňová Likertová škála. Aj keď je SSQ stále veľmi rozšíreným aj pre VR [19] a sú identifikované nedostatky VRSQ [20], do našej testovacej sady bolo rozhodnuté zaradiť VRSQ jednak kvôli nižšiemu počtu otázok (9 oproti 14 v SSQ), jednak prihliadnuc na výsledky štúdie [21], z ktorej vyplýva že z psychometrického hľadiska je VRSQ lepšie ako SSQ. Toto rozhodnutie bolo podporené aj pribúdajúcimi prípadmi použitia VRSQ vo vedeckých štúdiách, napríklad pri evaluácii edukačného VP umožňujúceho vtelenie do stromu [22] či porovnaní výučby pomocou pohlcujúcej VR a prezentácie [23]. Z hľadiska našich aplikácií, najmä L-NeRVEn, je relevantné aj použitie vo výskume [24], zameranom na efekty motoricko–kognitívneho tréningu vo VR pre seniorov.

Anglické znenie	**Preklad**
General discomfort	Všeobecný diskomfort
Fatigue	Únava
Eyestrain	Námaha očí
Difficulty focusing	Ťažkosti so zaostrením
Headache	Bolesť hlavy
Fullness of head	Tlak v hlave
Blurred vision	Rozmazané videnie
Dizzy (eyes closed)	Závraty pri zatvorených očiach
Vertigo	Vertigo (točenie hlavy)

Tabuľka 3
Preklad položiek dotazníka Virtual Reality Sickness Questionnaire

Pôvodné anglické znenie a preklad jednotlivých položiek dotazníka VRSQ sú uvedené v Tabuľke 3. Čo sa týka odpovedí, boli jednotlivé stupne Likertovej škály predložené nasledovne: “not at all” ako “žiadny(a)”, “slightly” ako “mierny(a)”, “moderately” ako “stredný(á)” a “very” ako “vážny(a)”.

V. Igroup Presence Questionnaire

Igroup Presence Questionnaire (*IPQ*) [25] umožňuje subjektívne hodnotenie pocitu prítomnosti vo VP. Obsahuje 14 položiek, zameraných na celkový pocit prítomnosti (kategória G) a 3 dimenzie: prítomnosť v priestore (kategória SP), angažovanosť (INV) a reálnosť (REAL). Pre odpovede na jednotlivé položky sa používa 7-stupňová Likertová škála Ako ukazujú aktuálne analýzy [26], [27], IPQ možno na daný účel považovať za spoľahlivý a efektívny prostriedok s dostatočnou citlivosťou.

Dotazník IPQ bol pôvodne pripravený v nemčine, preto v Tabuľke 4 okrem slovenského a anglického znenia uvádzame aj nemecké. Keďže ide v našej sade o najkomplikovanejší dotazník s rôznymi kategóriami položiek a typmi odpovedí, uvádza Tabuľka 4 aj konkrétne označenie položiek dotazníka podľa príslušnej kategórie (stĺpec “Pol.”) a typy odpovedí (stĺpec “Odp.”). Pri samotnom preklade bolo nemecké znenie smerodajné najmä pri položkách *REAL1* a *REAL3*, ktorých anglické znenie je úplne rovnaké.

Na rozdiel od predchádzajúcich troch dotazníkov, ktoré mali odpoveď pre každú položku rovnako slovne označenú, v prípade IPQ existuje až 7 rôznych typov odpovedí. Preto ich pôvodné a slovenské znenie uvádzame v samostatnej Tabuľke 5. A práve v prípade týchto odpovedí sme

Pol.	Nemecké znenie	Anglické znenie	Preklad	Odp.
G1	In der computererzeugten Welt hatte ich den Eindruck, dort gewesen zu sein...	In the computer generated world I had a sense of ”being there”	V počítačom generovanom svete som mal(a) pocit, ako keby som ”bol(a) tam”.	O1
SP1	Ich hatte das Gefühl, daß die virtuelle Umgebung hinter mir weitergeht.	Somehow I felt that the virtual world surrounded me.	Mal(a) som pocit, že ma virtuálny svet obklopuje.	O2
SP2	Ich hatte das Gefühl, nur Bilder zu sehen.	I felt like I was just perceiving pictures.	Mal(a) som pocit, ze vnímam (vidim) iba obrázky.	O2
SP3	Ich hatte nicht das Gefühl, in dem virtuellen Raum zu sein.	I did not feel present in the virtual space.	Necítil(a) som sa vo virtuálnom svete prítomný(á).	O3
SP4	Ich hatte das Gefühl, in dem virtuellen Raum zu handeln statt etwas von außen zu bedienen.	I had a sense of acting in the virtual space, rather than operating something from outside.	Mal(a) som pocit, že naozaj konám vo virtuálnom svete a nie, že ho zvonku ovládam.	O2
SP5	Ich fühlte mich im virtuellen Raum anwesend.	I felt present in the virtual space.	Cítil(a) som sa vo virtuálnom svete prítomný(á).	O2
INV1	Wie bewußt war Ihnen die reale Welt, während Sie sich durch die virtuelle Welt bewegten (z.B. Geräusche, Raumtemperatur, andere Personen etc.)?	How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?	Do akej miery ste si boli vedomí reálneho sveta (teplota, zvuky, iní ľudia,...) počas vašej aktivity vo virtuálnom svete?	O4
INV2	Meine reale Umgebung war mir nicht mehr bewußt.	I was not aware of my real environment.	Neuvedomoval(a) som si okolitý reálny svet.	O2
INV3	Ich achtete noch auf die reale Umgebung.	I still paid attention to the real environment.	Stále som venoval pozornosť reálnemu prostrediu.	O2
INV4	Meine Aufmerksamkeit war von der virtuellen Welt völlig in Bann gezogen.	I was completely captivated by the virtual world.	Virtuálny svet úplne upútal moju pozornosť.	O2
REAL1	Wie real erschien Ihnen die virtuelle Umgebung?	How real did the virtual world seem to you?	Ako velmi reálny vám pripadal virtuálny svet?	O5
REAL2	Wie sehr glich Ihr Erleben der virtuellen Umgebung dem Erleben einer realen Umgebung?	How much did your experience in the virtual environment seem consistent with your real world experience ?	Ako veľmi sa zážitok vo virtuálnom svete zhodoval so zážitkom v reálnom svete?	O6
REAL3	Wie real erschien Ihnen die virtuelle Welt?	How real did the virtual world seem to you?	Nakoľko reálny sa vám zdal virtuálny svet?	O7
REAL4	Die virtuelle Welt erschien mir wirklicher als die reale Welt.	The virtual world seemed more realistic than the real world.	Virtualny svet sa mi zdal skutočnejší ako ten reálny.	O2

Tabuľka 4
Preklad položiek dotazníka Igroup Presence Questionnaire

Odp.	Nemecké znenie	Anglické znenie	Preklad
O1	überhaupt nicht – sehr stark	not at all – very much	Vôbec nesúhlasím – Úplne súhlasím
O2	trifft gar nicht zu – trifft völlig zu	fully disagree – fully agree	Vôbec nesúhlasím – Úplne súhlasím
O3	hatte nicht das Gefühl – hatte das Gefühl	did not feel – felt present	Vôbec nesúhlasím – Úplne súhlasím
O4	extrem bewußt – mittelmäßig bewußt – unbewußt	extremely aware – moderately aware – not aware at all	Úplne vedomý(á) – Priemerne vedomý(á) – Nevedomý(á)
O5	vollkommen real – weder noch – gar nicht real	completely real – not real at all	Úplne reálny – Ani ani – Vôbec nie reálny
O6	überhaupt nicht – etwas – vollständig	not consistent – moderately consistent – very consistent	Vôbec sa nezhodoval – Trocha sa zhodoval – Úplne sa zhodoval
O7	wie eine vorgestellte Welt – nicht zu unterscheiden von der realen Welt	about as real as an imagined world – indistinguishable from the real world	Iba natoľko ako vymyslený svet – Nevedel(a) som ho odlíšiť od reálneho sveta

Tabuľka 5
Preklad odpovedí na položky dotazníka Igroup Presence Questionnaire

si dovolili azda najväčšiu odchýlku od pôvodného znenia, konkrétne pre odpovede O1 až O3, ktoré prekladáme rovnako. Dôvodom je, že tieto odpovede majú vzhľadom na znenie príslušných položiek rovnaký význam a ich znenie je rovnaké ako pri odpovediach dotazníka SUS. Bolo však potrebné upraviť aj spôsob vyhodnotenia, konkrétne pri položke SP3, kde je teraz interpretácia jednotlivých stupňov Likertovej škály opačná voči pôvodnému zneniu.

VI. Záver

V tomto článku sme predstavili slovenský preklad sady štyroch široko používaných dotazníkov zameraných na rôzne aspekty používateľskej skúsenosti (používateľského zážitku), určenej pre

hodnotenie virtuálno–realitných aplikácií. Predložené znenie sa snaží v maximálnej miere zachovať pôvodný význam položiek jednotlivých dotazníkov, no nie je ich "otrockým" prekladom. V prípade SUS dotazníka je to zreteľné pri preklade položky 5, kde je pre slovo "integrateď" v slovenskom znení okrem priameho prekladu ("integrované") uvedené aj slovo ("zapracované"). Taktiež v položke 6 sa pre "inconsistency" používa aj preklad "nezrovnalosť". V prvotnej slovenskej verzii boli prítomné len priame preklady, no najmä starší účastníci mali problém správne porozumieť ich významu v kontexte plnených úloh. Uvádzanie alternatívnych formulácií pre lepšie pochopenie je prítomné aj pri dotazníkoch NASA-TLX (3. položka) a VRSQ (9. položka). V prípade dotazníka IPQ sa pri väčšine položiek vychádzalo z anglickej verzie, znalosť nemeckej verzie však bola dôležitá pri asistencii účastníkom pri jeho vypĺňaní. Tu si asistenciu vyžadovala najmä položka SP4, kde nebolo všetkým účastníkom zrejmé, že pod slovným spojením "zvonku ovládam" sa myslí niečo ako ovládanie myšou a klávesnicou na klasickom počítači. Ďalšia úprava, zjednotenie 3 rôznych typov odpovedí dotazníka IPQ do jednej a to rovnakej ako v SUS dotazníku, mala za cieľ znížiť zaťaženie účastníkov vyhodnotení bez vplyvu na význam položiek. Ďalším krokom k štandardizácii slovenského znenia týchto dotazníkov by mala byť širšia diskusia ich potencionálnych používateľov. Samozrejme, v konečnom dôsledku by nemalo chýbať ani experimentálne overenie či je zachovaná konzistentnosť s ich pôvodným znením.

Poďakovanie

Tento článok vznikol vďaka podpore projektu APVV-21-0105 "Dôveryhodná interakcia človek–robot a terapeut–pacient vo virtuálnej realite"

Literatúra

[1] K. Tcha-Tokey, E. Loup-Escande, O. Christmann, and S. Richir, "A questionnaire to measure the user experience in immersive virtual environments," in *Proceedings of the 2016 Virtual Reality International Conference*, ser. VRIC '16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: https://doi.org/10.1145/2927929.2927955

[2] M. García, S. Cano, and F. Moreira, "User experience (ux) evaluation in virtual reality (vr)," in *Information Systems and Technologies*, A. Rocha, H. Adeli, G. Dzemyda, F. Moreira, and V. Colla, Eds. Cham: Springer Nature Switzerland, 2024, pp. 207–215.

[3] v. Korečko, P. Nehila, and B. Sobota, "Game engine based application for neurorehabilitation in collaborative virtual reality," in *2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2023, pp. 313–317.

[4] Š. Korečko, B. Sobota, P. Nehila, and D. Chmura, "Architecture, enhancements and perspective of the l-nerven virtual environment for neurorehabilitation," in *2024 IEEE 17th International Scientific Conference on Informatics (Informatics)*, 2024, pp. 152–156.

[5] Š. Korečko, B. Sobota, M. Mattová, and A. Kašela, "Petri-net-driven educational scenarios in web-based extended reality," in *2022 IEEE 16th International Scientific Conference on Informatics (Informatics)*, 2022, pp. 163–167.

[6] J. Brooke, "SUS: a 'quick and dirty' usability scale," in *Usability evaluation in industry*, P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester, Eds. Taylor and Francis Ltd, 1996, ch. 21, pp. 189–194.

[7] ——, "Sus: a retrospective," *Journal of usability studies*, vol. 8, no. 2, pp. 29–40, 2013.

[8] R. Webster and J. F. D. Jr., "System usability scale (sus): Oculus rift® dk2 and samsung gear vr®," in *2017 ASEE Annual Conference & Exposition*. Columbus, Ohio: ASEE Conferences, June 2017, https://peer.asee.org/28899.

[9] J.-N. Voigt-Antons, T. Kojic, D. Ali, and S. Möller, "Influence of hand tracking as a way of interaction in virtual reality on user experience," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–4.

[10] A. Globa, M. Yildirim, K. Lyu, and O. Gocer, "Evaluation of interactive vr environments for architecture," in *HUMAN-CENTRIC - Proceedings of the 28th CAADRIA Conference, Ahmedabad, 18-24 March 2023*, 01 2023, pp. 351–360.

[11] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006. [Online]. Available: https://doi.org/10.1177/154193120605000909

[12] B. Zheng, X. Jiang, G. Tien, A. Meneghetti, O. N. M. Panton, and M. S. Atkins, "Workload assessment of surgeons: correlation between nasa tlx and blinks," *Surgical Endoscopy*, vol. 26, no. 10, pp. 2746–2750, 2012.

[13] J. A. Holdnack and P. F. Brennan, "Usability and effectiveness of immersive virtual grocery shopping for assessing cognitive fatigue in healthy controls: Protocol for a randomized controlled trial," *JMIR Research Protocols*, vol. 10, no. 8, p. e28073, 2021. [Online]. Available: https://www.researchprotocols.org/2021/8/e28073/

[14] M. H. Knudsen, N. Breindahl, T.-S. Dalsgaard, D. Isbye, A. G. Mølbak, G. Tiwald, M. B. S. Svendsen, L. Konge, J. Bergström, and T. Todsen, "Using virtual reality head-mounted displays to assess skills in emergency medicine: Validity study," *Journal of Medical Internet Research*, vol. 25, p. e45210, 2023. [Online]. Available: https://www.jmir.org/2023/1/e45210

[15] D. Harris, M. Wilson, and S. Vine, "Development and validation of a simulation workload measure: the simulation task load index (sim-tlx)," *Virtual Reality*, vol. 24, no. 4, pp. 557–566, Dec 2020. [Online]. Available: https://doi.org/10.1007/s10055-019-00422-9

[16] A. Vidal-Balea, P. Fraga-Lamas, and T. M. Fernández-Caramés, "Advancing nasa-tlx: Automatic user interaction analysis for workload evaluation in xr scenarios," in *2024 IEEE Gaming, Entertainment, and Media Conference (GEM)*, 2024, pp. 1–6.

[17] H. K. Kim, J. Park, Y. Choi, and M. Choe, "Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment," *Applied Ergonomics*, vol. 69, pp. 66–73, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000368701730282X

[18] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. L. and, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993. [Online]. Available: https://doi.org/10.1207/s15327108ijap0303_3

[19] N. Tian, P. Lopes, and R. Boulic, "A review of cybersickness in head-mounted displays: raising attention to individual susceptibility," *Virtual Reality*, vol. 26, no. 4, pp. 1409–1441, Dec 2022. [Online]. Available: https://doi.org/10.1007/s10055-022-00638-2

[20] J. Josupeit, "Cybersickness as the virtual reality sickness questionnaire (vrsq) measures it!? –an environment-specific revision of the vrsq," *Frontiers in Virtual Reality*, vol. 4, 2023. [Online]. Available: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2023.1291078

[21] V. Sevinc and M. I. Berkman, "Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments," *Applied Ergonomics*, vol. 82, p. 102958, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003687019301759

[22] P. Spangenberger, S.-C. Freytag, and S. M. Geiger, "Embodying nature in immersive virtual reality: Are multisensory stimuli vital to affect nature connectedness and pro-environmental behaviour?" *Computers & Education*, vol. 212, p. 104964, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131523002415

[23] A. P. Lawson and R. E. Mayer, "Individual differences in executive function affect learning with immersive virtual reality," *Journal of Computer Assisted Learning*, vol. 40, no. 3, pp. 1068–1082, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12925

[24] R. Y. C. Kwan, J. Liu, O. S. K. Sin, K. N. K. Fong, J. Qin, J. C. Y. Wong, and C. Lai, "Effects of virtual reality motor-cognitive training for older people with cognitive frailty: Multicentered randomized controlled trial," *J Med Internet Res*, vol. 26, p. e57809, Sep 2024. [Online]. Available: https://www.jmir.org/2024/1/e57809

[25] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 266–281, 06 2001. [Online]. Available: https://doi.org/10.1162/105474601300343603

[26] M. Melo, G. Gonçalves, j. Vasconcelos-Raposo, and M. Bessa, "How much presence is enough? qualitative scales for interpreting the igroup presence questionnaire score," *IEEE Access*, vol. 11, pp. 24 675–24 685, 2023.

[27] T. Q. Tran, T. Langlotz, J. Young, T. W. Schubert, and H. Regenbrecht, "Classifying presence scores: Insights and analysis from two decades of the igroup presence questionnaire (ipq)," *ACM Trans. Comput.-Hum. Interact.*, vol. 31, no. 5, Nov. 2024. [Online]. Available: https://doi.org/10.1145/3689046

Adaptive Multi-Physics vSLAM for Industrial AR: Prototype Implementation and Early Validation

[1]*Ivana NOVÁKOVÁ,* [2]*Miroslav MICHALKO*

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

[1]ivana.novakova2@tuke.sk, [2]miroslav.michalko@tuke.sk

***Abstract* — Based on our previous theoretical frameworks, this article presents the first functional prototype of AMP-vSLAM—an adaptive multi-physical visual SLAM pipeline tailored for demanding industrial conditions by utilizing augmented or mixed reality as a working tool in the industry. By combining visual, IMU, and environmental sensors through Bayesian weighting and temperature-sensitive calibration, the system maintains the map's integrity even under substantial changes in lighting, mechanical shocks, and thermal drift. Experiments in a robotic welding cell demonstrate up to 48% lower absolute trajectory error and 2,1 times higher retention rate of elements compared to ORB-SLAM3, while maintaining real-time overlay on the MS HoloLens 2 device for mixed reality.**

***Keywords* — AMP-vSLAM, AR Algorithm, Augmented Reality, vSLAM, hololens 2, industry**

I. INTRODUCTION

Augmented Reality (AR) is a technology with high potential for application in industry. [1]Although it is not a state-of-the-art technology, it still faces significant challenges, especially in a dynamic environment. Implementation involves overlaying graphic 3D content onto devices, materials, and workflows where AR shortens training cycles, reduces human errors, and supports flawless maintenance. However, successful deployment depends on the accuracy of registration to centimeters or even millimeters, as well as a good understanding of the dynamics of the scene. The visualized system of simultaneous localization and mapping (vSLAM) has become the de facto standard for localization for head-mounted devices; however, its performance sharply deteriorates when factory conditions move away from the controlled lighting and vibration environments for which most algorithms were designed.[2]

Recent surveys report 25-60% tracking failures when intense flickering lighting, thermal drifting, or mechanical shaking are present, highlighting the need for extensions that are aware of physical laws.

Several research streams have begun to address individual stressors: adaptive exposure control for low lighting in vSLAM [3], vibration-resistant inertial tubes [4], and temperature-compensated camera models [5]. Nevertheless, these efforts remain fragmented; none integrates the full spectrum of current physical phenomena that characterize the challenging conditions of production lines, such as sparks from arc welding, vibrations caused by conveyors, dust, and other solid particles in the air, or rapidly heated air. Furthermore, classical sensor fusion schemes assign static weights to data from cameras and IMUs, ignoring the fact that the reliability of each modality fluctuates depending on environmental conditions. [6].

Our previous work introduced a conceptual framework for an *Adaptive Multi-Physics vSLAM* (AMP-vSLAM) approach that augments every stage of the SLAM loop—feature detection, state estimation, and loop closure—with real-time feedback from light, temperature, vibration, and acoustic sensors. This paper advances that concept in three ways: prototype realization, Quantitative Validation**,** and Operational Insights.

II. AMP-vSLAM ARCHITECTURE

Based on previous theoretical work, this article presents the first functional prototype of AMP-vSLAM—adaptive multiphysics visual SLAM developed for augmented reality applications primarily for HMD in a dynamic industrial environment. AMP-vSLAM is built from five tightly coupled layers. This section explains each layer's role, data flow, and algorithmic details. The proposed AMP-vSLAM system features a self-learning module that continually adjusts the weighting of each sensor according to its current data quality. A Bayesian reliability analysis constantly evaluates visual, inertial, thermal, and acoustic streams—noisy sensors are automatically down-weighted, while more stable ones receive greater influence. All measurements and contextual information are archived; machine-learning models mine this history to identify the most effective sensor weights under similar conditions and instantly apply these predicted coefficients to the next round of data fusion.

A. Sensing layer

It is necessary to integrate multiple data types to develop an adaptive and robust algorithm, such as RGB-D cameras, stereo sensors, IMU, temperature sensors, vibration sensors, acoustic sensors, and GPS with local reference points to form a complete multi-sensor input..All sensors publish ROS 2 topics time-stamped in hardware and synchronized with global clocks with a deviation of ≤ 50 µs (PTP + HW tags). Their data must be precisely synchronized in time to avoid accumulating positioning errors when using multiple sensors. Keeping the deviation below 50 µs is accurate enough for utilizing augmented reality applications in an industrial environment so that graphical content, movement, and physical data can be reliably integrated.

Table 1.

Sensing layer

Sensor	Key specs	Physical phenomenon	Samp. rate
RGB-D stereo (Intel D455)	1280 × 720 @ 90 fps, 95 mm baseline	Scene depth	90 Hz
IMU (Bosch BMI270)	2000 dps, ±16 g	Vibration, motion	400 Hz
Thermocouples (Type K)	−40 … +125 °C, ±0.25 °C	Thermal drift	10 Hz
MEMS vibrometer	0–2 kHz, 0–2 g	Shocks, resonances	400 Hz
Lux-meter	0–2000 lx, 12-bit	Illumination	25 Hz

B. Pre-processing & Temporal Alignment

Timestamp regression – every message is mapped to *t_{global}* via linear drift correction.
Kalman-IMU denoiser – two-stage filter: (i) low-pass (120 Hz cutoff) for grinder vibrations; (ii) iterative EKF estimating
x = [p, v, q, b_a, b_g].
Rolling-shutter rectification – each pixel (u,v) is corrected:

$$\acute{u} = u - \frac{\Delta T_{row}}{T_{frame}}\ \omega\, x\, u \qquad (i)$$

where *Δt _{row}* is row exposure and *ω* is the IMU angular rate.
In a rolling-shutter sensor, each row starts integrating slightly later than the previous one.If the camera rotates during that period, every row "sees" the scene from a slightly different pose, producing geometric distortions (bent lines, wobbles).

C. Adaptive Multi-Sensor Fusion

Bayesian reliability index: Every 100 ms a score is computed for sensor i*:*

$$\omega_T = exp\left(\frac{r_i}{2\sigma_i^2}\right) \qquad (ii)$$

with *r_i* the factor-graph residual and *σ_i* the variance of the last 50 samples.

Weighted EKF-SLAM

In a conventional extended Kalman-Filter SLAM (EKF-SLAM) the state-transition equation is:

$$xk = Fx_{k-1} + Bu_k + w_k$$

We insert a diagonal weight block in F:

$$F_{weight} = diag\ (w_{vis}, F_{vis}, w_{imu}, F_{imu}, w_{temp}, F\theta,..$$

F_{vis}, F_{imu} F_{θ} are the motion-model sub-matrices contributed by the visual stream, the inertial sensors, and the temperature channel, respectively

For example, when welding arcs illuminate the camera, its residuals increase. In that case, the Bayesian module sets $w_{vis} \rightarrow 0$, and the visual part of the motion model is effectively silenced. Conversely, in the case of a physical phenomenon such as increased vibrations, the IMU residuals increase and w decreases, allowing vision to prevail in the prediction step. Since the weights scale like a state transition matrix, so do the associated covariance blocks. The overall update and prediction cycle remains consistent, meaning that the output information from low-quality sensors will not be rejected; they only receive a proportionally smaller influence until their residuals stabilize.

D. *Temperature-Dependent Camera Re-calibration*

Camera intrinsics $K = [fx, fy, cx, cy]$ update online:

$$fx(T) = fx0(1 + \alpha f(T - T0)), cx(T) = cx0 + \alpha c(T - T0).$$

E. *Vibration & Shock Mitigation*

Raw accelerations undergo FFT; if RMS > 0.5 g at dominant frequency *f_v*, a notch IIR filter (bandwidth 1.5 × *f_v*) is inserted. Simultaneously, *w_{imu}* is lowered, letting stereo close loops more aggressively.

F. CNN-Based Dynamic Scene Filter

The dynamic scene filter is based on MobileNet V3 and enhanced by a two-level Feature Pyramid Network (FPN). Thanks to optimization, it is a lightweight convolutional neural network (CNN) that can recognize objects over a wide range, from a few millimeters, such as dust particles, to life-size equipment in a factory. Each frame taken from the camera is reduced to a resolution of 320 × 180 pixels, allowing for quick inference while maintaining sufficient spatial details for accurate motion tracking and reducing memory bandwidth usage. Deployed using NVIDIA TensorRT with FP16 precision, the network processes each frame in approximately 4.2 milliseconds. This minimal load ensures that it meets the real-time requirement of 60 Hz for augmented reality applications.

G. *Local Map & Optimisation*

The mapping backbone is deliberately split into two complementary tiers so that pure geometry can be updated at video rate on the GPU while slower-changing physical fields evolve in the background. A sliding window of 120 keyframes resides entirely on the GPU; older frames are compressed and pushed to host RAM. The geometric layer streams a decimated surfel mesh at 10 Hz to HoloLens 2.

On request, the physics layer overlays false-color volumes (e.g., heat map) for maintenance diagnostics, sampled via GPU ray-marching at < 3 ms. Separating geometry from environment metadata lets the tracker stay in real-time, even in harsh factories. At the same time, loop-closure decisions are “physics-aware,” reducing false positives by 37 % and lowering drift by 22 % in long (> 300 m) sequences.

Spatial Rendering & AR Interface
Remote streaming pipeline**.** The system employs Microsoft's Remote Rendering SDK to offload heavy geometry processing to an edge server with an RTX A4000. Scene meshes are decimated to ≈1.2 M triangles, Draco-compressed, and packetized over a Wi-Fi 6E 6 GHz link using QUIC/UDP with forward-error correction. End-to-end pose-to-photon latency stays below 35 ms, while network jitter is under 8 ms during peak traffic bursts, ensuring a stable hologram.

Self-Learning Module & Historical Database
The self-learning module is achieved through continuous performance recording, mining contextual patterns, and shifting updated priors into HMD. Adaptation is thus accomplished without human intervention - which is key for industrial factories where environmental conditions evolve more dynamically and manual tuning is more frequent.

III. Experimental Methodology

Measurements serve for realistic comparisons of the proposed AMP-vSLAM algorithm with currently state-of-the-art visual-inertial SLAM systems.
For testing, an industrial environment with welding machines was chosen, where during production, the air temperature varies from 25 °C (at rest) to 55 °C (continuous welding). Accelerometers placed on the floor report vibrations of RMS 0.4–0.9 g in the range of 60–140 Hz, and lighting fluctuates between 200 lx and 1,450 lx during welding, when the intensity of illumination increases due to the arc sparking from the welder. The HMD chosen was the Microsoft Hololens 2 with AMP-vSLAM. During testing, the operator had to follow three pre-defined trajectories. At the end of the path, they stopped near the arc for safety to avoid exposure to heat and glare. The route took approximately 3 minutes, and the total distance of the path was around 110 m.

Algorithm differences were assessed with paired two-tailed t-tests ($\alpha = 0.05$). Latency and feature retention were averaged over 80 % of frames to exclude start-up transients.

Table 2.

Metric	*Definition*	Sampling	Purpose
Absolute Trajectory Error (ATE)	RMS distance between estimated and ground-truth positions after SE(3) alignment	full trajectory	overall drift
Relative Pose Error (RPE)	RMS error of frame-to-frame motion over 1 s intervals	sliding window	local consistency
Feature-retention %	inliers ÷ total matched features after RANSAC	per keyframe	robustness of tracking
Frame latency	capture-to-render time measured with a photodiode on HoloLens display	every 30th frame	AR responsiveness

Table 3.

Condition (60 s run)	*ORB-SLAM3 ATE [cm]*	AMP-vSLAM ATE [cm]	Improvement
Stable 400 lx, 28 °C	*1.9 ±0.3*	1.1 ±0.2	0,42
Flicker 1 kHz, 800–1500 lx	6.7 ±1.1	3.4 ±0.6	0,49
Vibration 0.8 g RMS	5.2 ±0.8	2.7 ±0.4	0,48
Heat-soak 55 °C	4.6 ±0.7	2.5 ±0.5	0,46

IV. Conclusion

This study focused on the initial testing of the proposed AMP-vSLAM algorithm designed for dynamic industrial environments with many variable factors, where classical SLAM exhibits significant errors. By developing a more robust algorithm and adding physical sensors, adaptive filters, and learning weights, we demonstrated benefits for further development. Modularity allows for the disconnection or addition of sensors without changing the core, enabling the system to scale from prototype manufacturing plants, smaller assembly lines, to large operations.

References

[1] Jeevan S. Devagiri, Sidike Paheding, Quamar Niyaz, Xiaoli Yang, Samantha Smith,Augmented Reality and Artificial Intelligence in industry: Trends, tools, and future challenges,Expert Systems with Applications,Volume 207,2022,118002,ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2022.118002.

[2] S. -Y. Yu and H. Zhu, "Low-Illumination Image Enhancement Algorithm Based on a Physical Lighting Model," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 28-37, Jan. 2019, doi: 10.1109/TCSVT.2017.2763180.W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[3] Wang, Kunfeng & Zhao, Kaichun & Lu, Wenshuai & Wan, Zhenghua & Bai, Yang & You, Zheng. (2024). Temperature Drift and Compensation in Stereo Visual Odometry. IEEE Sensors Journal. PP. 1-1. 10.1109/JSEN.2024.3477530. B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.

[4] Huang, X.; Chen, X.; Zhang, N.; He, H.; Feng, S. ADM-SLAM: Accurate and Fast Dynamic Visual SLAM with Adaptive Feature Point Extraction, Deeplabv3pro, and Multi-View Geometry. *Sensors* **2024**, *24*, 3578. https://doi.org/10.3390/s241135781J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.

[5] Cai, Y.; Ou, Y.; Qin, T. Improving SLAM Techniques with Integrated Multi-Sensor Fusion for 3D Reconstruction. *Sensors* **2024**, *24*, 2033. https://doi.org/10.3390/s24072033

[6] Huadong Wu, M. Siegel and S. Ablay, "Sensor fusion using Dempster-Shafer theory II: static weighting and Kalman filter-like dynamic weighting," *Proceedings of the 20th IEEE Instrumentation Technology Conference (Cat. No.03CH37412)*, Vail, CO, USA, 2003, pp. 907-912 vol.2, doi: 10.1109/IMTC.2003.1207885.

Meranie a modelovanie torzného kmitania v točivých zariadeniach poháňaných spínaným reluktančným motorom

Peter BOBER

Katedra elektrotechniky a mechatroniky, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovensko

peter.bober@tuke.sk

***Abstrakt* — Mechanické kmitanie sa prirodzene vyskytuje v mnohých systémoch. V torzných zariadeniach poháňaných spínaným reluktančným motorom (SRM) môže byť veľkosť torzného kmitania vyššia vzhľadom na prirodzene pulzujúci moment SRM. V článku je uvedené meranie kmitania momentu na meracom stojane so SRM a modelovanie torznej sústavy. Výsledky simulačných experimentov dobre odrážajú javy sledované na meracom stojane, čo je základ pre ďalšie skúmanie možností potláčania kmitania.**

***Keywords* — modelovanie, rezonancia, spínaný reluktančný motor, torzné kmitanie, točivé zariadenie**

I. Úvod

Mechanické kmitanie sa prirodzene vyskytuje v mnohých systémoch. Je prejavom cyklickej premeny energie z jednej formy na inú a naspäť do pôvodnej formy. Časť energie sa však pri každom cykle premení do takej formy, že opúšťa systém. Preto je kmitanie tlmené a bez dodávania energie z vonkajšieho prostredia sa po čase zastaví. Frekvencia premeny energie závisí od štruktúry a zloženia systému a každý systém má jednu alebo viac prirodzených frekvencií, na ktorých kmitá. Energia sa do systému alebo von zo systému prenáša v každom cykle v dávkach. Ak frekvencia dodávania energie zodpovedá prirodzenej frekvencii a prírastok je väčší ako úbytok, potom sa energia v systéme hromadí a amplitúda kmitov narastá. Tento jav nazývame rezonancia. Energia sa môže pri rezonancii nahromadiť v takom množstve, že spôsobí deštrukciu systému.

V točivých sústavách majú jednotlivé časti zotrvačnú hmotnosť a pružnosť a dochádza k premene kinetickej energie na deformačnú a naopak. Prejaví sa to vo forme priečneho, axiálneho a torzného kmitania [1]. Vplyvom trenia, ložísk alebo samotným motorom vzniká v točivej sústave pulzácia momentu spôsobujúca torzné kmitanie. Ak sa frekvencia momentových pulzov priblíži k prirodzenej frekvencii sústavy dochádza k rezonancii. Mechanické vibrácie na rezonančných frekvenciách môžu významne ovplyvniť výkon, bezpečnosť a životnosť točivých sústav.

Na zmiernenie vibrácií v točivých aj netočivých sústavách boli vyvinuté rôzne metódy a zariadenia. Podľa [2] ich môžeme rozdeliť na:

- pasívne tlmenie,
- aktívne tlmenie,
- poloaktívne tlmenie.

Pasívne tlmenie využíva rôzne tlmiace materiály a konštrukcie na odoberanie energie z vibrujúceho systému. Ďalšou možnosťou je zmena prirodzených rezonančných frekvencií pomocou ladeného hmotnostného tlmiča. Aktívne tlmenie využíva snímače a aktuátory na identifikáciu vibrácií a na aktívne pôsobenie proti kmitaniu. Deje sa to buď samostatným aktuátorom alebo zmenou parametrov regulátorov v spätnoväzobnej slučke hlavného aktuátora. Poloaktívne tlmenie je kombináciou pasívneho a aktívneho tlmenia. Mení niektoré

charakteristiky systému, napríklad tuhosť, a tým posúva frekvenciu prirodzenej rezonancie mimo pracovnú oblasť [3].

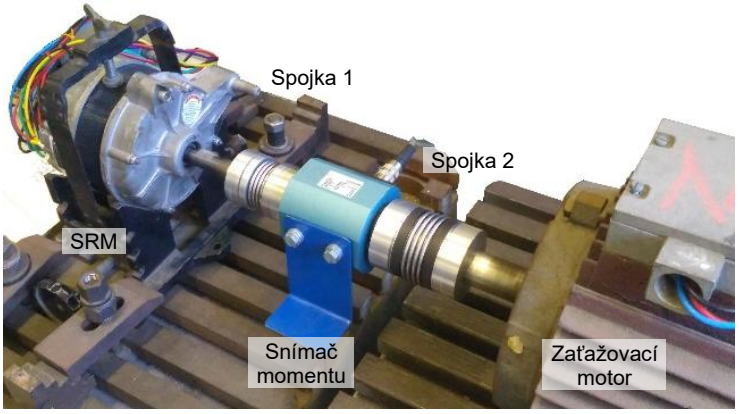
Na analýzu a návrh rôznych metód tlmenia sú potrebné modely. Pojednanie o komplexnosti kmitania v rotačných zariadeniach a rôznych prístupoch k modelovanie je uvedené v [1].

V tomto článku sú výsledky merania torzného kmitania na meracom stojane so SRM, ktoré dokumentujú aktuálnu situáciu. Na účely analýzy torzného kmitania a návrh metód na jeho potlačenie je v ďalšej časti uvedený jednoduchý model torznej sústavy a jeho vlastnosti sú dokumentované výsledkami simulačných experimentov.

II. Meranie torzného kmitania na meracom stojane

A. Meracie pracovisko

Meracie pracovisko sa skladná zo spínaného reluktančného motora, záťaže predstavovanej indukčným motorom (IM), výkonového meniča s riadiacim mikropočítačom a z osobného počítača na riadenie experimentov a záznam údajov. Medzi SRM a IM je snímač momentu pripojený dvoma spojkami (obr. 1). Označenie a vlastnosti jednotlivých komponentov sú v tab. 1.



Obr. 1 Merací stojan so spínaným reluktančným motorom.

Tabuľka 1
Zariadenia meracieho pracoviska

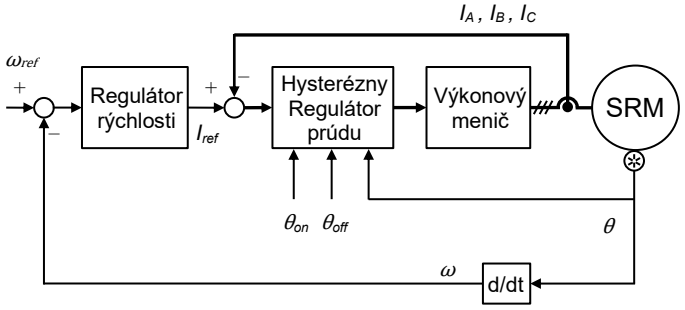
Zariadenie	Opis
SRM	Trojfázový 12/8 pólový spínaný reluktančný motor určený pre pohony s premenlivou rýchlosťou, výkon 200 W, napätie 120 V, prúd 2,5 A, maximálny prúd 6 A.
Indukčný motor	Štandardný indukčný motor, 2p = 6, výkon 3 kW, 3 x 380V
Snímač momentu	Snímač krútiaceho momentu KISTLER 4520A s vyhodnocovacím modulom CoMo 4700BP0UA, vzorkovanie 10 kHz, nastaviteľný filter, analógový výstup, USB výstup s periódou merania od 5 ms do 10 s.
Výkonový menič	Trojfázový asymetrický H-mostík, prototyp, jednosmerné napätie 120 V
Riadiaci procesor meniča	Vývojová doska s mikrokontrolérom STM32F303RET6, vzorkovanie 50 μs
Snímač	Snímač absolútnej polohy 360°, rozlíšenie 12 bitov (0,0878°)
Spojka 1: BK2 30/69	Priemer 55 mm, dĺžka 69 mm, moment zotrv. $0{,}12.10^{-3}$ kg.m^2, tuhosť 39.10^3 Nm/rad, hmotnosť 0,26 kg
Spojka 2: BK2 60/83	Priemer 66 mm, dĺžka 83 mm, moment zotrv. $0{,}32.10^{-3}$ kg.m^2, tuhosť 76.10^3 Nm/rad, hmotnosť 0,48 kg

B. Spínaný reluktančný motor

Spínané reluktančné motory majú jednoduchšiu konštrukciu ako asynchrónne motory a dokážu pracovať pri vyšších teplotách. Pri správnom riadení majú vyššiu účinnosť a náklady na výrobu sú nižšie [4]. Na druhej strane vyžadujú zložitejšie riadenie, majú nižší účinník a vyšší pulzujúci moment vyplývajúci z princípu činnosti. Každé vinutie viacfázového statora vytvára moment len pre obmedzený uhol rotácie v ktorom sa mení reluktancia a tým pádom aj indukčnosť. Trvalý moment sa vytvorí postupným spínaním prúdu vo fázach statora, pričom pri prechode z jednej fáze na druhú dochádza zvyčajne k poklesu momentu. Tým sa vytvorí kmitanie. SRM sa používajú v priemyselných pohonoch a elektrických vozidlách.

C. *Riadenie spínaného reluktančného motora*

Spínaný reluktančný motor potrebuje k svojej činnosti výkonový menič a riadiaci obvod, ktorý vytvára prúdové impulzy vo vinutiach jednotlivých fáz statora podľa uhla natočenia rotora. Uhol a rýchlosť otáčania sa získavajú zo snímača polohy. Výkonový menič je trojfázový asymetrický H mostík, ktorý umožní vytvárať tri nezávisle unipolárne prúdy I_a, I_b a I_c pre tri fázy statora. Na obr. 2 je štruktúra riadiaceho obvodu SRM s hysteréznym regulátorom prúdov. PI regulátor rýchlosti vytvára referenčnú hodnotu prúdu tečúceho motorom I_{ref}. Hysterézny regulátor prúdu zapína prúdový impulz pri uhle rotora θ_{on}, vypína pri uhle θ_{off} a medzi týmito dvoma uhlami udržiava veľkosť prúdu na hodnote I_{ref}. Prúdové impulzy sú pre každú fázu posunuté. Riadiaci algoritmus je naprogramovaný v jazyku C. Perióda riadiacej slučky je 50 μs.



Obr. 2 Bloková schéma riadenia SRM s regulátorom rýchlosti a hysteréznym regulátorom prúdu.

D. *Výsledky merania prúdu a torzného momentu*

Na meracom pracovisku bola vykonaná séria experimentov pri ktorých sa zaznamenal priebeh prúdu jednej fázy statora a torzný moment zo snímača. Na obr. 3 sú priebehy v ustálenom stave pri rýchlostiach 300, 640, 900 a 1200 ot/min. Pri rýchlosti 600 ot/min sa objavilo rezonančné kmitanie a preto bolo meranie uskutočnené pri 640 ot/min. Zaťažovací moment bol udržiavaný na strednej hodnote 1 Nm brzdením indukčným motorom, cez ktorý prechádzal jednosmerný prúd.

Na obrázku je možné vidieť typické priebehy prúdového impulzu, kde je hodnota prúdu regulovaná hysteréznym regulátorom. Tvar nábehu, dobehu a regulácie je ovplyvňovaný zmenou indukčnosti počas otáčania rotora. Torzný moment má strednú hodnotu 1 Nm a pri všetkých rýchlostiach obsahuje výraznú kmitavú zložku na rezonančnej frekvencii rotačnej sústavy 245 Hz, čo zodpovedá 612 ot/min pri 24 prúdových impulzoch na jednu mechanickú otáčku. Snímačom umiestneným podľa obr. 1 nie je možné merať kmitanie momentu motora ale iba kmitanie celej sústavy.

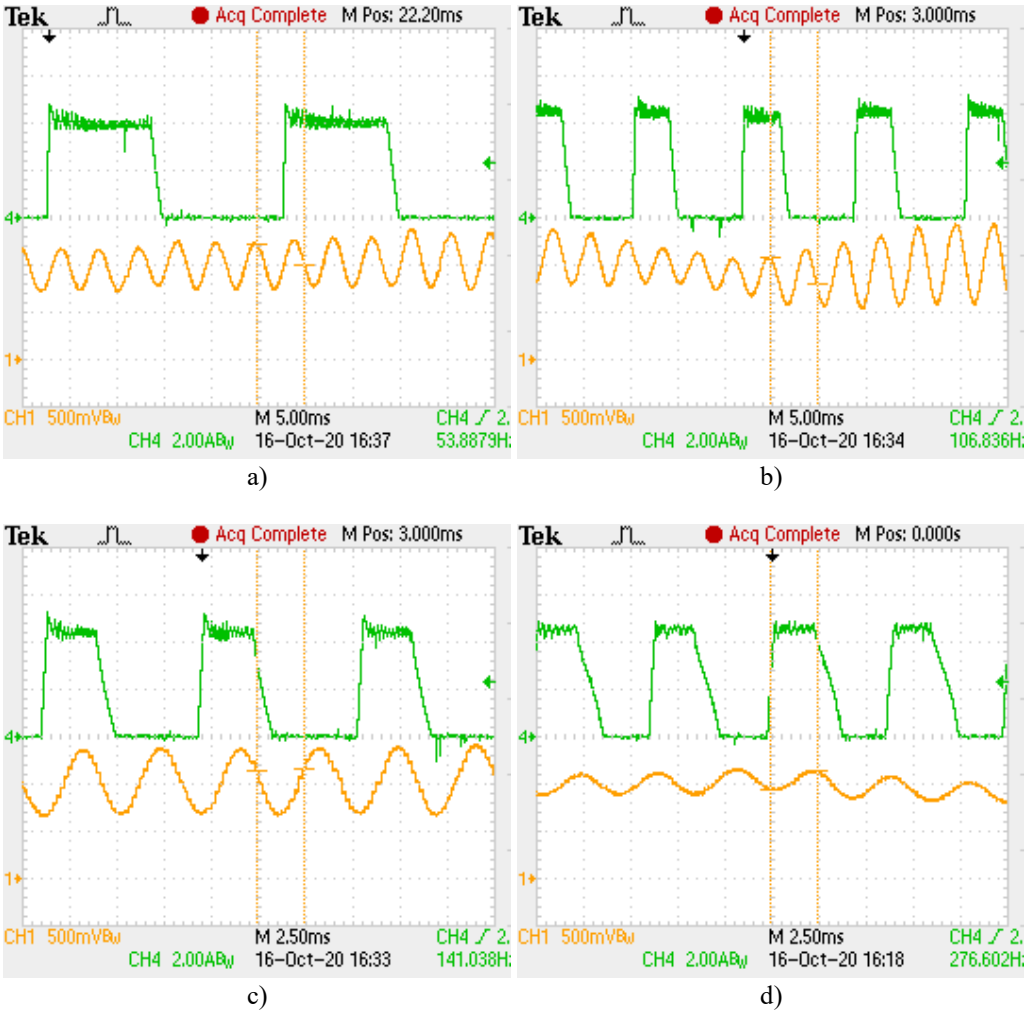
III. MODELOVANIE TORZNÉHO KMITANIA

Torzná sústava na obr. 1 sa skladá z viacerých zotrvačných hmotností a pružných členov. Aj samotný hriadeľ SRM je pomerne dlhý a má nezanedbateľnú pružnosť. Pre prvé priblíženie boli pri modelovaní zohľadné viaceré zjednodušenia:

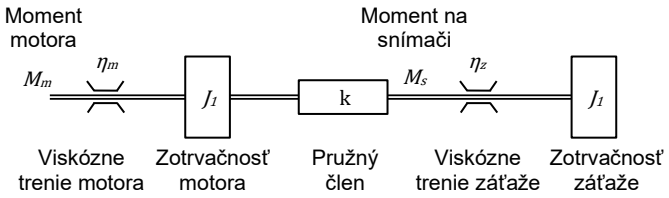
- model bude uvažovať len torzné kmitanie,
- model neuvažuje torzné kmitanie zapríčinené ložiskami a nesúosovosťou hriadeľov,
- trenie bude len viskózne,
- v modeli bude len jeden pružný člen.

Štruktúra výsledného modelu s dvoma zotrvačnými hmotnosťami a jedným pružným členom je na obr. 4. Parametre modelu v tab. 2 boli odhadnuté na základe známych parametrov z tab. 1. Teoretická rezonančná frekvencia modelu je 241,66 Hz, čo je pomerne dobrá zhoda s nameranou frekvenciou 245 Hz.

Model bol zostavený v Matlab-Simulink. Na modelovanie SRM bol použitý blok *Switched Reluctance Machine* z knižnice *Simscape/Electrical/Electromechanical/Reluctance&Stepper* s parametrami podľa merania v [5].



Obr. 3 Prúd jednou fázou a moment zo snímača pre rýchlosti a) 300 ot/min, b) 640 ot/min, c) 900 ot/min, d) 1200 ot/min. Snímač momentu má konštantu 1 Nm/V, frekvencia kmitania momentu je 245 Hz.



Obr. 4 Štruktúra torznej sústavy s dvoma zotrvačnými hmotami a pružným členom.

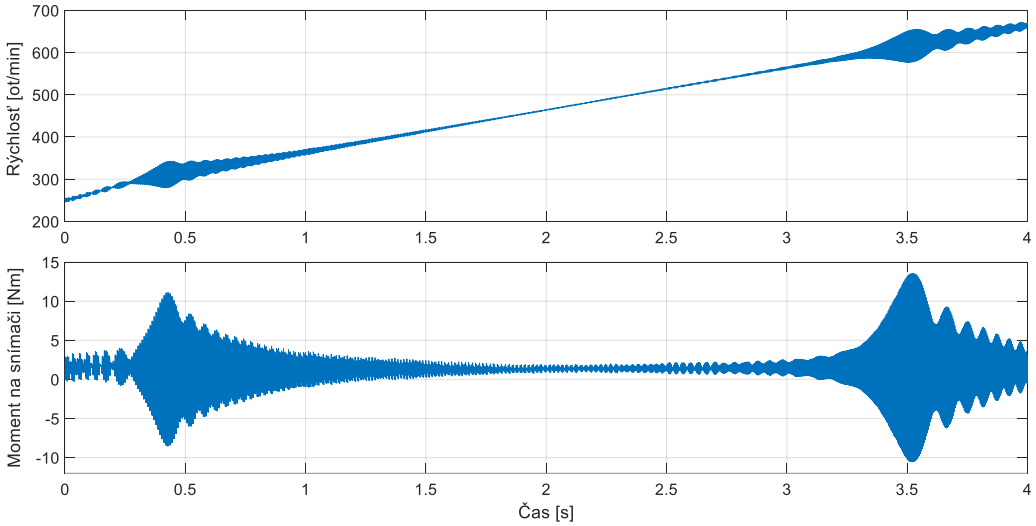
Tabuľka 2
Parametre a symboly modelu torznej sústavy

Symbol	Hodnota	Opis
η_m	0,0002 Nm.s.rad^{-1}	Koeficient viskózneho trenia motora.
η_z	0,01 Nm.s.rad^{-1}	Koeficient viskózneho trenia záťaže.
k	4500 Nm.rad^{-1}	Tuhosť všetkých pružných častí.
J_1	0,002 kg.m^2	Moment zotrvačnosti motora, spojky motora.
J_2	0,081 kg.m^2	Moment zotrvačnosti záťaže, spojky záťaže.
f_{rez}	241.66 Hz	Teoretická vlastná frekvencia torznej sústavy.
M_m	-	Reluktančný moment motora.
M_z	-	Moment na záťaži snímaný snímačom.

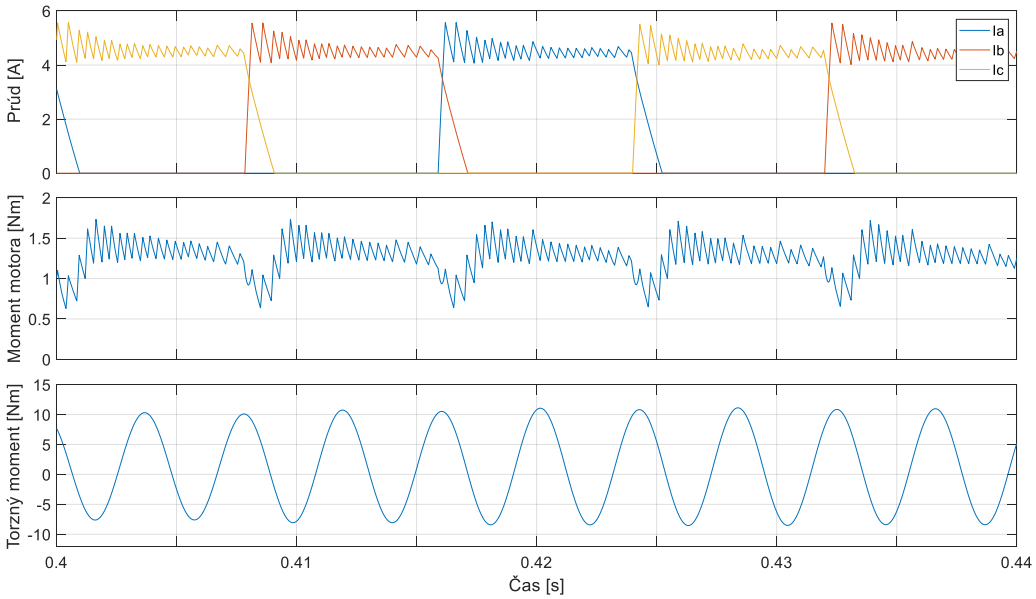
IV. VÝSLEDKY SIMULAČNÝCH EXPERIMENTOV

Na obr. 5 je priebeh rýchlosti a momentu na snímači pri rozbehu sústavy po rampe žiadanej rýchlosti. Ak sú impulzy motora celočíselným podielom alebo násobkom prirodzenej frekvencie, objaví sa rezonancia a amplitúda kmitania momentu narastie. Pri postupnom zvyšovaní rýchlosti motora sa v priebehu momentu objavujú rázy spôsobené skladaním prirodzenej frekvencie sústavy a premenlivej frekvencie kmitania momentu motora.

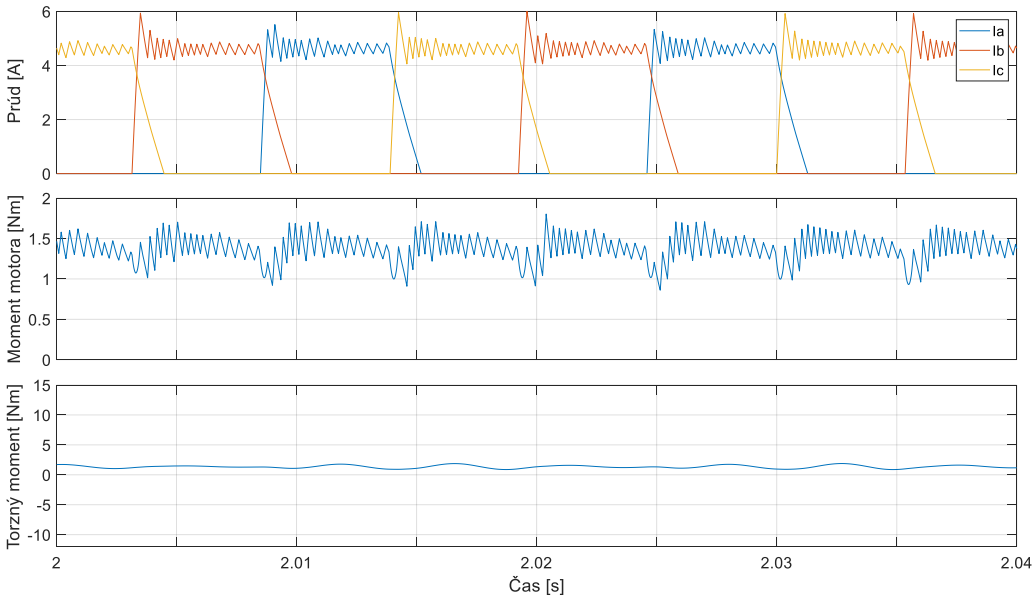
Na obr. 6, obr. 7 a obr. 8 sú zobrazené priebehy fázových prúdov, reluktančného momentu motora a momentu zo snímača pri rýchlostiach v oblasti prvej rezonancie, medzi rezonanciami a v oblasti druhej rezonancie. K prvej rezonancii dochádza približne pri frekvencii pulzov momentu motora 118 Hz, čo je polovica prirodzenej frekvencie sústavy (obr. 6). V oblasti medzi rezonanciami je v signáli zo snímača momentu stále prítomné kmitanie, hoci s malou amplitúdou. Ak sa frekvencia kmitania momentu zhoduje s prirodzenou frekvenciou, amplitúda kmitov je najvyššia (obr. 7).



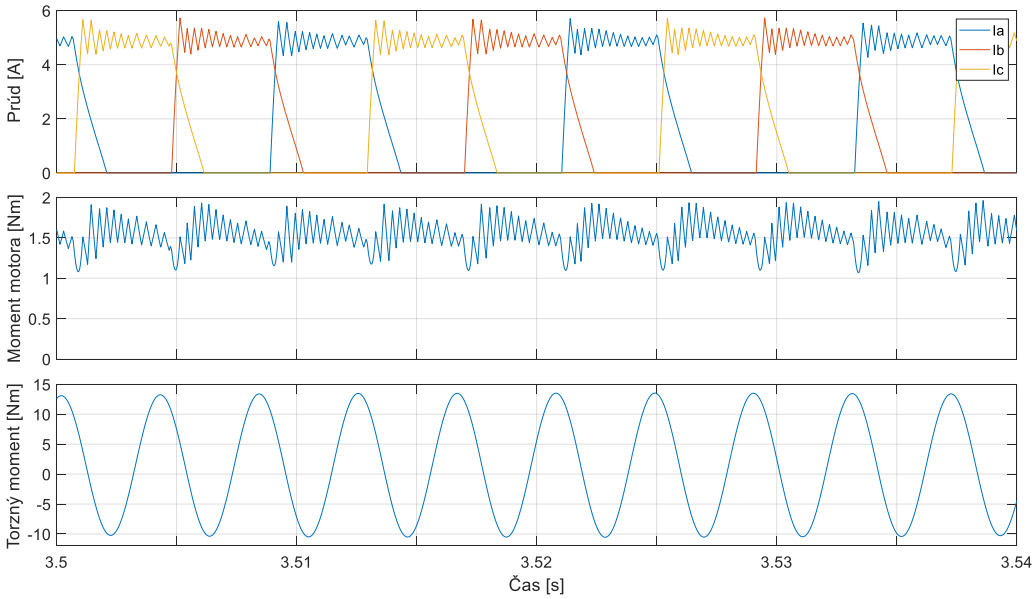
Obr. 5 Simulácia uhlovej rýchlosti a momentu na hriadeli záťaže pri rozbehu.



Obr. 6 Prúd, moment motora a moment na záťaži, rýchlosť 295 ot/min, frekvencia pulzov momentu 118 Hz.



Obr. 7 Prúd, moment motora a moment na záťaži, rýchlosť 455 ot/min, frekvencia pulzov momentu 174 Hz.



Obr. 8 Prúd, moment motora a moment na záťaži, rýchlosť 605 ot/min, frekvencia pulzov momentu 242 Hz.

V. ZÁVER

Torzné kmitanie v točivých zariadeniach sa vyskytuje prirodzene. Jeho amplitúda sa zvyšuje prítomnosťou kmitania momentu od motora alebo od ďalších častí zariadenia, napríklad od ložísk. Ďalej na veľkosť kmitania má zásadný vplyv blízkosť frekvencie vonkajších momentových kmitov k prirodzenej frekvencii sústavy.

Výsledky simulácie torzného kmitania v sústave poháňanej spínaným reluktančným motorom, uvedené v tomto článku, veľmi dobre zodpovedajú pozorovanému javu pri prevádzke motora na meracom stojane napriek tomu, že model je veľmi zjednodušený. Dobrá zhoda modelu s reálnym javom umožní analýzu možností zníženia amplitúdy kmitania. Spínaný reluktančný motor prirodzene vytvára kmitavý moment, čo sa pravdepodobne bude dať pri vhodnom spôsobe riadenia využiť na aktívne potláčanie torzných vibrácií v točivých sústavách.

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.

REFERENCES

[1] P. A. Patil and C. Teodoriu, "A comparative review of modelling and controlling torsional vibrations and experimentation using laboratory setups," *Journal of Petroleum Science and Engineering*, vol. 112, pp. 227-238, 2013. doi: 10.1016/j.petrol.2013.11.008.

[2] S. Pasupuleti, "Vibration Control and Noise Reduction in Mechanical Systems: Techniques and Applications," *Int. J. Core Eng. Manage*., vol. 6, no. 8, 2020.

[3] S. M. Savaresi, C. Poussot-Vassal, C. Spelta, O. Sename, and L. Dugard, "Semi-Active Suspension Technologies and Models, " in *Semi-Active Suspension Control Design for Vehicles*, S. M. Savaresi, C. Poussot-Vassal, C. Spelta, O. Sename, and L. Dugard, Eds. Butterworth-Heinemann, 2010, pp. 15–39. DOI: 10.1016/B978-0-08-096678-6.00002-X

[4] Z. Yang, F. Shang, I. P. Brown and M. Krishnamurthy, "Comparative Study of Interior Permanent Magnet, Induction, and Switched Reluctance Motor Drives for EV and HEV Applications," in *IEEE Transactions on Transportation Electrification*, vol. 1, no. 3, pp. 245-254, Oct. 2015, doi: 10.1109/TTE.2015.2470092.

[5] P. Bober and Ž. Ferková, "Meranie indukčnosti fáz spínaného reluktančného motora," in *Electrical Engineering and Informatics 15: Proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košic*e, Košice, Slovensko, Technická univerzita v Košiciach, 2024, s. 445-448. [Online]. Dostupné na: https://eei.fei.tuke.sk/wp-content/uploads/2024/07/EEI_15.pdf.

Využitie elektronického vodoznaku a siete ANFIS pri ochrane medicínskych dokumentov

Ján VAŠČÁK

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

jan.vascak@tuke.sk

Abstrakt **– V medicínskej oblasti je integrita a dôveryhodnosť medicínskych dokumentov mimoriadne dôležitá, vzhľadom na to, aký vplyv môžu mať na zdravie pacienta. Ideálne by tieto údaje mali byť odolné voči zmenám, ktoré môžu nastať počas prenosu a zároveň nepostrehnuteľné, aby ich nemohli zmeniť potenciálni útočníci. Jedným zo spôsobov, ako to dosiahnuť, je použitie elektronického vodoznaku. Pre potreby nastavenia parametrov, pomocou ktorých sa takýto vodoznak generuje, je použitá neurónová sieť typu ANFIS. Na vybraných medicínskych dátach boli vykonané experimenty, ktoré dosiahli sľubné výsledky.**

Keywords **– ANFIS, elektronický vodoznak, medicínske dokumenty**

I. Úvod

Integrita a dôveryhodnosť medicínskych obrázkov a dokumentov je vzhľadom na liečebnú starostlivosť pacientov mimoriadne dôležitá. Keďže bolo zaznamenané množstvo snáh o ich pozmenenie a tieto snahy majú rastúci trend s potenciálne vážnymi následkami [1], tak by tieto údaje mali byť odolné voči zmenám, ktoré by sa snažili zaviesť potenciálni útočníci. Jedným zo spôsobov, ako to dosiahnuť, je ochrana pomocou tzv. elektronického vodoznaku, ktorý je odolný voči rôznym kybernetickým útokom a zachováva originalitu daného záznamu [2], pričom robustnosť a percepčná neviditeľnosť sú v tomto obzvlášť dôležité a mali by byť zachované počas celého procesu tvorby a vloženia vodoznaku do dokumentu. Z tohto dôvodu sa z daného dokumentu extrahujú tzv. oblasti záujmu (OZ) [3], ktoré je potrebné obzvlášť zabezpečiť voči kybernetickým údajom, nakoľko obsahujú obzvlášť dôležité časti informácií. V ďalšom sa budeme zaoberať problematikou medicínskych dokumentov ako napr. röntgenové snímky, ktorých časti môžu obsahovať závažné informácie o zdraví pacienta a obzvlášť tieto časti, t.j. teda OZ, podliehajú skúmaniu odborníkov.

Obzvlášť dôležitý je spôsob tvorby vodoznaku. Existuje niekoľko základných prístupov využívajúcich celú radu metód rozkladu ako napr. singulárna dekompozícia hodnôt (SDH) [4], waveletová transformácia a Schurova dekompozícia [5], či Arnoldova transformácia [6].

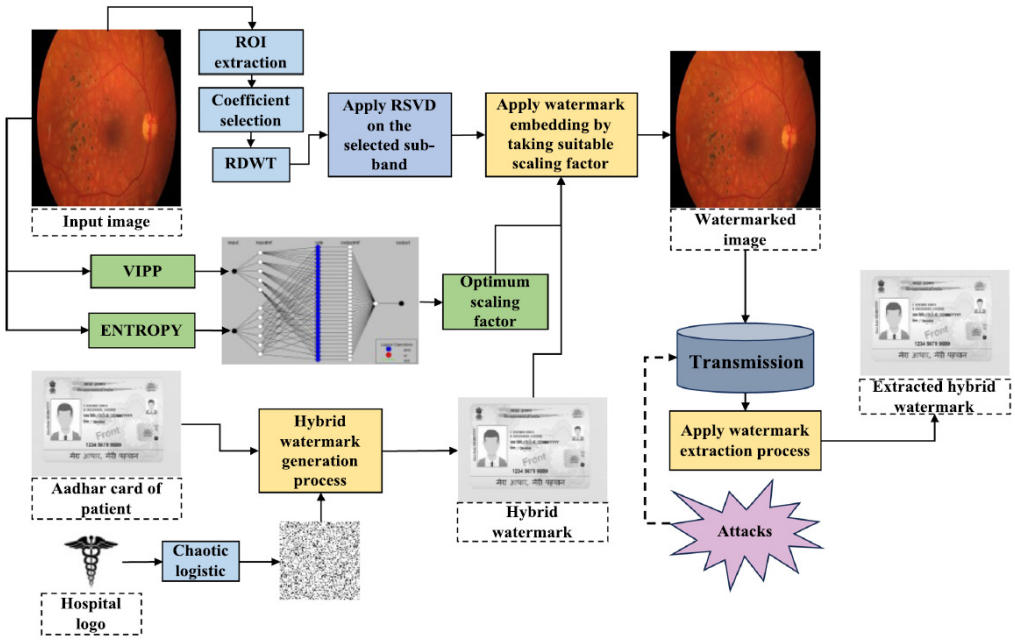
Vodoznak je možné využiť nielen na detekciu toho, či došlo k neautorizovanej úprave daného dokumentu, ale aj k určeniu miesta, kde bol dokument pozmenený a dokonca k jeho pôvodnej rekonštrukcii. Za týmto účelom sa využívajú metódy ako tzv. krehké bloky [7], SDH, či celočíselná vlnová transformácia chaotickej mapy [8]. Reverzibilný vodoznak [9] s využitím rozdielového rozšírenia a prahovania histogramu pixelového obrazu je možné využiť aj na rekonštrukciu poškodeného dokumentu a predstavuje technológiu bezpečného prenosu dokumentov, ktorá je odolná voči väčšine rušivých prostriedkov, ako sú napr. rôzne typy šumov. Jedná sa však o výpočtovo veľmi náročnú metódu, preto sa hľadajú alternatívne prístupy, ktoré sú schopné zabezpečiť rekonštrukciu daného dokumentu. Na tomto mieste stoja za zmienku dva prístupy. Po prvé sa jedná o kombináciu náhodnej dekompozície singulárnej hodnoty (NDSH), Hessenbergovej dekompozície, redundantnej diskrétnej vlnovej transformácie (RDVT) a diskrétnej kosínusovej transformácie (DKT) [10]. Druhý prístup je založený na kombinácii Henonovej mapy pre zabezpečenie bezpečného prenosu, fuzzy logiky pre určenie faktora mierky a konvexnej neurónovej siete pre zvýšenie robustnosti celkového riešenia [11]. Faktor mierky práve určuje, do akej miery bude vodoznak viditeľný a jeho nastavenie nie je triviálne. Vystupuje tu celá škála neurčitostí a z tohto dôvodu sa na jej určenie využíva fuzzy logika. Keďže problémom môže byť vytvorenie bázy znalostí, na základe ktorej fuzzy inferenčný systém pracuje, je v tomto článku spomenutá

adaptácia bázy znalostí pomocou systému ANFIS [12]. V ďalšom sa práve budeme zaoberať využitím tejto adaptívnej neurónovej siete.

II. Spôsob využitia elektronického vodoznaku

Práca s digitálnym vodoznakom sa v princípe vykonáva v nasledujúcich fázach: tvorba vodoznaku, vkladanie vodoznaku, kontrola neautorizovanej úpravy a obnova pôvodného obsahu dokumentu. Na Obr. 1 je zobrazený pôvodný postup vkladania vodoznaku do dokumentu podľa [1].

V tejto metóde je pôvodný obrázok najprv rozdelený na jednotlivé OZ pomocou algoritmu založeného na metóde najstabilnejších extremálnych oblastí (NEO). Následne sa vytvorí vodoznak, v našom prípade spojením karty poistenca a loga poisťovne pomocou algoritmov RDVT a SDH. Vodoznak je vytvorený kombináciou dominantných komponentov, ktoré sú výsledkom SDH. Následne sa vypočíta tzv. variácia obrázka podľa pixelov (VOPP) a entropia hostiteľského obrázka. Tieto hodnoty sa použijú ako vstupy do systému ANFIS (Adaptive Neuro Fuzzy Inference System), aby sa určil optimálny faktor mierky aplikovaný na vodoznaku počas procesu vkladania. Nakoniec sa vodoznak vloží do OZ hostiteľského obrázka tiež použitím vyššie spomenutých metód RDVT, SDH a DKT.

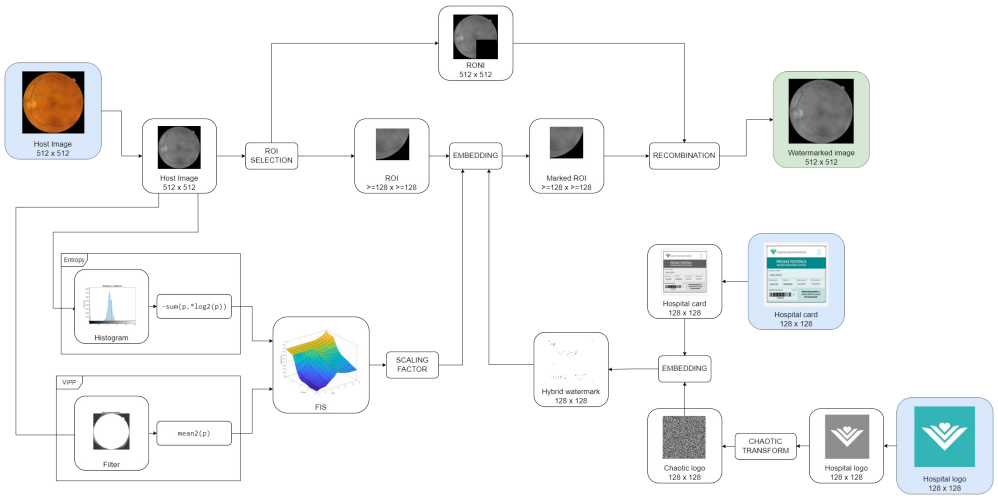


Obr. 1 Postup vkladania vodoznaku do dokumentu podľa [1].

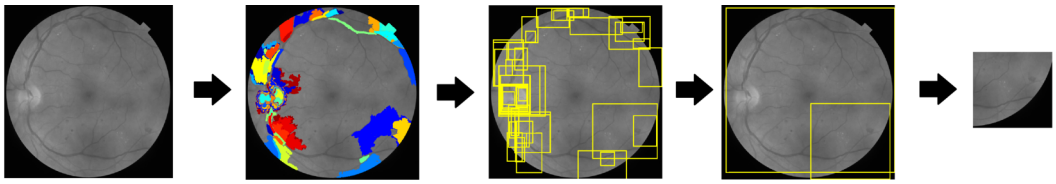
Vzhľadom na zvýšené výpočtové nároky pôvodného návrhu, kde sú vkladanie a extrakcia vodoznaku riešené v rámci jednej funkcie, bol zvolený praktickejší spôsob tak, že vkladanie a extrakcia sú oddelené, čo lepšie odráža spôsob, akým by bol tento proces realizovaný v praxi, viď Obr. 2.

Algoritmus založený na NEO extrahuje OZ s premenlivou veľkosťou z hostiteľského obrázka. Najprv sa detegujú NEO oblasti, následne sa okolo nich vykreslia ohraničenia (bounding boxy). Tieto sa orezávajú a spájajú dohromady a nakoniec sa odstránia tie, ktoré sú príliš malé. Ako OZ sa vyberie najmenšia oblasť zo zostávajúcich, viď Obr. 3. Napriek svojmu názvu nemusí OZ obsahovať žiadne zaujímavé alebo významné časti obrázka. Dôležité je, že proces jeho výberu je deterministický vzhľadom na hostiteľský obrázok.

V ďalšom si ozrejmíme niektoré základné pojmy. Entropia je miera náhodnosti v obrázku. Najprv sa obrázok prevedie na histogram a následne sa ako výsledok vypočíta entropia tohto histogramu. VOPP je miera variability pixelov v obrázku, kde sa najprv obrázok spracuje filtrom, ktorý spája susedné pixely s podobnými hodnotami a následne sa výsledok vypočíta ako priemer filtrovaného obrázka. Na určenie faktoru mierky vodoznaku sa použil systém ANFIS, ktorý ho vypočítavá na základe vyššie uvedených veličín entropie a VOPP. Faktor mierky určuje jas vodoznaku, kde nižšie hodnoty zachovávajú lepšiu neviditeľnosť, zatiaľ čo vyššie hodnoty lepšie zachovávajú integritu v prípade manipulácie s obrázkom, preto je dôležité nájsť vhodnú rovnováhu na základe hostiteľského obrázka.



Obr. 2 Upravený proces vkladania vodoznaku.



Obr. 3 Proces extrakcie OZ.

III. Evaluačné kritériá

Pre vyhodnotenie experimentov boli zvolené nasledovné evaluačné kritériá:

a) NCC: Normalized cross correlation, normalizovaná verzia cross correlation, meria lineárnu asociáciu medzi dvomi premennými (obrázkami), minimum = -1, žiadna korelácia, maximum = 1, úplná korelácia, čím väčšia, tým lepšie:

$$NCC(x,y) = \frac{\sum((x - x_m) \cdot (y - y_m))}{\sqrt{\sum(x - x_m)^2 \cdot \sum(y - y_m)^2}}. \tag{1}$$

x = Hodnoty pixelov 1. obrázka, x_m = Priemer hodnôt pixelov 1. obrázka
y = Hodnoty pixelov 2. obrázka, y_m = Priemer hodnôt pixelov 2. obrázka

b) BER: Bit error rate, počet bitov obrázka, ktoré sú odlišné, deleno celkovým počtom bitov. Minimum = 0, úplne identický obrázok, maximum = 1, úplne odlišný obrázok, čím menší, tým lepšie:

$$BER(x,y) = \frac{\text{Počet bitov, ktoré sa nezhodujú medzi x a y}}{\text{Celkový počet bitov}}. \tag{2}$$

c) PSNR: Peak signal-to-noise ratio, metrika ktorá slúži na zistenie kvality rekonštrukcie alebo kompresie obrázka. Kvantifikuje integritu medzi originálom a rekonštrukciou. Udáva sa v decibeloch. Minimum = 0 db, úplne odlišný obrázok, maximum = teoreticky nekonečno, úplne identický obrázok, čím väčšia hodnota, tým lepšie:

$$PSNR(x,y) = 10\log_{10}(\frac{M^2}{MSE(x,y)}). \tag{3}$$

M = Maximálna možná hodnota pixelu, pre gray scale obrázky = 255.

d) SSIM: Structural Similarity Index Measure, metrika pre vyhodnocovanie kvality obrázka. Na rozdiel od iných metrík je viac citlivá na zmeny viditeľné ľudskému oku. Minimum = -1, úplne odlišný obrázok, max = 1, úplne identický obrázok, čím väčšia hodnota, tým lepšie:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \tag{4}$$

$$c_1 = (k_1 L)^2. \tag{5}$$

$$c_2 = (k_2 L)^2. \tag{6}$$

μ_x = vzorkový priemer 1. obrázka, μ_y = vzorkový priemer druhého obrázka
σ_x = vzorkový rozptyl 1. obrázka, σ_y = vzorkový rozptyl 2. obrázka
σ_{xy} = vzorková kovariancia medzi 1. a 2. obrázkom
k_1 = konštanta 0.01, k_2 = konštanta 0.03
L = dynamický rozsah pixelových hodnôt, pre gray scale = $2^8 - 1 = 255$

IV. Experimenty a ich vyhodnotenie

Na vyhodnotenie efektivity implementácie bolo použitých 6 obrázkov z verejných databánk, 3 z databanky RFMiD (Retinal Fundus Multi-disease image Dataset) [13] a 3 z databanky TCIA (The Cancer Imaging Archive) [14].

Obr. 4 Obrázky sietnic z RFMiD

Obr. 5 Obrázky ultrazvukov pečene z TCIA

Nižšie sú v tabuľkách uvedené súhrnné výsledky podľa použitých evaluačných kritérií.

Najprv bola analyzovaná podobnosť pôvodného obrázka s obrázkom, do ktorého bol vložený vodoznak. Získané hodnoty v Tabuľke 1 naznačujú, že aj vizuálne bude ťažké vodoznak identifikovať, takže návrh vodoznaku a jeho implementácia do dokumentu (tuná obrázka) boli úspešné.

Tabuľka 1
Pôvodný obrázok verzus obrázok s vodoznakom – metriky

Sample	NCC	BER	PSNR	SSIM
Eye_1	1	0.0060043	71.7777	0.99998
Eye_2	1	0.010365	70.3296	0.99995
Eye_3	1	0.004612	73.412	0.99999
$Liver_1$	1	0.0064468	72.7353	0.99999
$Liver_2$	1	0.0083313	71.9779	0.99999
$Liver_3$	1	0.00090408	81.7854	1

Následne bol obrázok označený vodoznakom, vykonala sa na ňom transformácia a z obrázka sa extrahoval vodoznak. Takto bolo možné porovnať integritu preneseného vodoznaku s pôvodným. Pre experimenty boli vybrané nasledovné štyri typy transformácií v podobe šumov a filtrov.

1) Šum typu Salt & pepper (SP) – prejavuje sa vo forme náhodných svetlých a tmavých bodov (pixelov) rozptýlených po obrázku. Tento šum vzniká často v dôsledku chýb pri prenose alebo poruchách snímania obrazu, výsledky viď Tabuľka 2.
2) Gaußovský šum (GŠ) – šum s normálnym (Gaußovským) rozdelením pravdepodobnosti. Najčastejšie sa vyskytuje pri spracovaní obrazu alebo zvuku ako dôsledok náhodných variácií v senzore alebo elektronike. Tento šum pridáva jemné náhodné odchýlky k pixelovým hodnotám, výsledky viď Tabuľka 3.
3) Pohybové rozmazanie (PR) – vizuálny efekt, ktorý vzniká, keď sa objekt alebo kamera pohybujú počas expozície. Výsledkom je rozmazanie objektov v smere pohybu, čo spôsobuje stratu detailov. V spracovaní obrazu sa často simuluje pomocou konvolučných filtrov, výsledky viď Tabuľka 4.
4) Gaußov dolnopriepustný filter (GDF) – typ lineárneho filtra, ktorý prepúšťa nízkofrekvenčné zložky obrazu a potláča vysokofrekvenčný šum a detaily. Je založený na Gaußovskej funkcii, ktorá určuje váhy okolitých pixelov pri vyhladzovaní obrazu, výsledky viď Tabuľka 5.

Tabuľka 2
NCC a BER pre šum typu SP

SP	Eye_1	Eye_2	Eye_3	$Liver_1$	$Liver_2$	$Liver_3$
NCC	0.9977	0.9974	0.9927	0.9967	0.9961	0.9969
BER	0.54379	0.5527	0.69022	0.5936	0.62277	0.5127

Tabuľka 3
NCC a BER pre GŠ

GŠ	Eye_1	Eye_2	Eye_3	$Liver_1$	$Liver_2$	$Liver_3$
NCC	0.9676	0.9363	0.9802	0.9727	0.9705	0.9546
BER	0.087033	0.086206	0.079361	0.086453	0.088092	0.079831

Tabuľka 4
NCC a BER pre PR

PR	Eye_1	Eye_2	Eye_3	$Liver_1$	$Liver_2$	$Liver_3$
NCC	0.9938	0.99882	0.99704	0.9959	0.99528	0.99658
BER	0.31595	0.35783	0.42953	0.45152	0.45607	0.44971

Tabuľka 5
NCC a BER pre GDF

GDF	Eye_1	Eye_2	Eye_3	$Liver_1$	$Liver_2$	$Liver_3$
NCC	0.99976	0.99992	0.99992	0.99958	0.9994	0.99981
BER	0.096484	0.064514	0.064514	0.11401	0.12419	0.093921

V. Záver

Výsledky v obidvoch častiach vyhodnocovania boli veľmi dobré a porovnateľné voči iným, avšak komplikovanejším metódam, čo ukazuje, že systém ANFIS sa dokáže prispôsobiť konkrétnym typom dokumentov. V budúcnosti by bolo zaujímavé generovať farebný vodoznak, keďže medicínske dokumenty sa v stále väčšej miere vyskytujú aj vo farebnej verzii. Istý potenciál má aj možnosť použiť väčšie dátové súbory pre trénovanie systému ANFIS, alebo skúsiť iný adaptačný mechanizmus.

Poďakovanie

Financované EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 17R05-04-V01-00004, Kompetenčné a vzdelávacie centrum kybernetickej a informačnej bezpečnosti TUKE.

LITERATÚRA

[1] V. K. S. Divyanshu Awasthi, Priyank Khare, "ANFISmark: ANFIS-based secure watermarking approach for telemedicine applications," *Springer Neural Computing and Applications*, vol. 37, pp. 8677–8698, 2025.
[2] L. K. Hui, S. Liew, and J. M. Zain, "Region of interest-based tamper detection and lossless recovery watermarking scheme (ROI-DR) on ultrasound medical images," *Journal of Digital Imaging*, vol. 30, no. 3, pp. 328–349, 2017.
[3] P. Singh, K. J. Devi, H. K. Thakkar, and K. Kotecha, "Region-based hybrid medical image watermarking scheme for robust and secured transmission in IoMT," *IEEE Access*, vol. 10, pp. 8974–8993, 2022.
[4] R. Pandey, A. K. Singh, B. Kumar, and A. Mohan, "Iris based secure NROI multiple eye image watermarking for teleophthalmology," *Multimedia Tools and Applications*, vol. 75, no. 22, pp. 14 381–14 397, 2016.
[5] A. Tiwari and V. K. Srivastava, "Novel schemes for the improvement of lifting wavelet transform-based image watermarking using Schur decomposition," *The Journal of Supercomputing*, vol. 79, no. 12, pp. 13 142–13 179, 2023.
[6] R. Keshavarzian and A. Aghagolzadeh, "ROI based robust and secure image watermarking using DWT and Arnold map," *AEU – International Journal of Electronics and Communications*, vol. 70, no. 3, pp. 278–288, 2016.
[7] R. Eswaraiah and E. Sreenivasa Reddy, "Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest," *IET Image Processing*, vol. 9, no. 8, pp. 615–625, 2015.
[8] D. Ravichandran, P. Praveenkumar, S. Rajagopalan, J. B. B. Rayappan, and R. Amirtharajan, "ROI-based medical image watermarking for accurate tamper detection, localisation and recovery," *Medical & Biological Engineering & Computing*, vol. 59, no. 6, pp. 1355–1372, 2021.
[9] A. F. Qasim, R. Aspin, F. Meziane, and P. Hogg, "ROI-based reversible watermarking scheme for ensuring the integrity and authenticity of DICOM MR images," *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 16 433–16 463, 2019.
[10] D. Awasthi, P. Khare, and V. K. Srivastava, "PBNHWA: NIfTI image watermarking with aid of PSO and BO in wavelet domain with its authentication for telemedicine applications," *Multimedia Tools and Applications*, pp. 1–33, 2024.
[11] D. Awasthi, P. Khare, and V. Kumar Srivastava, "RFDB: Robust watermarking scheme with Fuzzy-DnCNN using blockchain technique for identity verification," *Expert Systems with Applications*, vol. 255, p. 124554, 2024.
[12] A. Abraham, *Adaptation of Fuzzy Inference System Using Neural Learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 53–83.
[13] Kaggle, "Retinal disease classification," [online] https://www.kaggle.com/datasets/andrewmvd/retinal-disease-classification/data.
[14] TCIA, "The cancer imaging archive," [online] https://www.cancerimagingarchive.net/.

Zvyšovanie odolnosti smerovania v MANET - IoT

[1] *Kevin SENDROVIČ, Ján PAPAJ*

[1]Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach, Slovenská republika

[1]kevin.sendrovic@student.tuke.sk, jan.papaj@tuke.sk

Abstrakt **– Mobilné ad-hoc siete (MANET) predstavujú kľúčovú infraštruktúru pre podporu komunikácie v dynamických a decentralizovaných prostrediach, obzvlášť pri integrácii s IoT zariadeniami (MANET-IoT). Vzhľadom na ich vysokú mobilitu, obmedzené zdroje a nestabilnú topológiu čelia smerovacie protokoly výzvam, ako sú časté prerušenia spojení, výpadky uzlov či strata konektivity. Táto práca sa zameriava na analýzu správania vybraných proaktívnych smerovacích protokolov (Babel, OLSRv1, OLSRv2) v extrémnych podmienkach s vysokým percentom zlyhaní uzlov a náhodnou mobilitou. Výsledky simulácií ukazujú výrazné rozdiely vo výkonnosti protokolov, pričom protokoly Babel a OLSRv2 preukázali vyššiu mieru robustnosti a prispôsobivosti. Zistenia poukazujú na význam správneho výberu smerovacieho protokolu v závislosti od prevádzkových podmienok a poskytujú praktické odporúčania pre návrh spoľahlivých MANET-IoT sietí.**

Kľúčové slová **– MANET-IoT, smerovanie, odolnosť, simulátor MeshLab**

I. Úvod

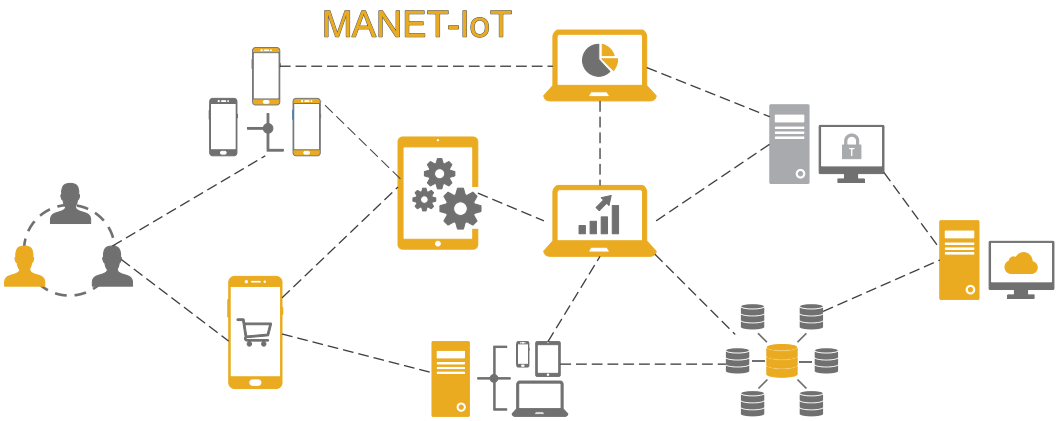
Mobilné ad hoc siete (MANET – Mobile Ad Hoc Networks) predstavujú osobitý typ bezdrôtovej siete, ktorá funguje bez potreby centrálnej infraštruktúry. Vznikajú dynamicky z množiny autonómnych zariadení schopných vzájomnej bezdrôtovej komunikácie. Medzi tieto zariadenia patria najmä smartfóny, senzory, roboty alebo drony, ktoré si medzi sebou samostatne odovzdávajú dáta a spoločne vytvárajú dočasnú, neustále sa meniacu sieť. Každý uzol v sieti zároveň plní úlohu smerovača, čo znamená, že sa podieľa nielen na prenose vlastných dát, ale aj na smerovaní dátových paketov iných uzlov [1], [2].

Charakteristickou črtou MANET sietí je vysoká dynamika zmeny topológie. Topológia siete sa mení v reálnom čase v dôsledku pohybu zariadení, ich dostupnosti a prevádzkového stavu. Táto vlastnosť ich predurčuje na použitie v prostrediach, kde nie je dostupná stabilná infraštruktúra alebo kde je potrebná rýchla a flexibilná sieťová konektivita. Vojenské jednotky využívajú MANET siete na koordináciu taktických operácií v teréne bez rizika centrálneho zlyhania alebo kompromitácie spojenia [1], [3], [4]. Rovnako v prípade prírodných katastrof môžu záchranné tímy vytvoriť dočasnú komunikačnú infraštruktúru z dostupných mobilných zariadení. V kontexte inteligentných miest sa MANET siete uplatňujú napríklad pri koordinácii autonómnych vozidiel alebo v systémoch environmentálneho monitoringu.

Napriek mnohým výhodám čelia MANET siete aj viacerým výzvam. Vysoká miera mobility, časté zmeny topológie a obmedzené zdroje zariadení (výpočtový výkon, kapacita batérie, šírka pásma) výrazne ovplyvňujú stabilitu a spoľahlivosť komunikácie. Navyše absencia centrálnej autority otvára dvere bezpečnostným hrozbám, ako sú neautorizované prístupy, falošné uzly alebo šírenie zavádzajúcich informácií. Zariadenia v MANET sieti musia preto neustále prispôsobovať svoje smerovanie aktuálnemu stavu siete, aby nedochádzalo k stratám dát alebo výpadkom spojenia. V prípade nízkej efektivity smerovania môže dôjsť aj k plytvaniu energetickými zdrojmi, čo má negatívny dopad na životnosť a udržateľnosť siete [5].

Efektívne smerovanie je preto kľúčovým faktorom pre fungovanie MANET-IoT sietí (Obr. 1). Vzhľadom na rozmanité aplikačné scenáre je potrebné vyhodnocovať rôzne smerovacie protokoly z pohľadu ich spoľahlivosti, priepustnosti, latencie, energetickej efektivity a schopnosti adaptácie. Medzi najčastejšie používané patria protokoly AODV (Ad hoc On-Demand Distance Vector) a OLSR (Optimized Link State Routing), ktoré reprezentujú dva rozdielne prístupy – reaktívny a proaktívny [6].

OLSRv2, ako vylepšená verzia OLSR, prináša modulárnejšiu architektúru, podporu IPv6 a lepšie prispôsobenie požiadavkám moderných mobilných sietí. Zachováva si výhodu nízkej latencie v



Obr. 1 Princíp činnosti MANET-IoT sieť.

stabilnejších častiach siete, no v prostredí s rýchlou a neustálou zmenou spojení sa môže potýkať s oneskorením aktualizácií topológie, čo znižuje efektivitu smerovania.

Každý z týchto protokolov má svoje výhody aj obmedzenia, ktoré sa prejavujú v závislosti od charakteru nasadenej siete. AODV je vhodný pre dynamické a menej predvídateľné siete, kde šetrí šírku pásma, no za cenu vyššej latencie. OLSR/OLSRv2 vynikajú v prostredí s vysokou hustotou a stabilnejšou štruktúrou [4].

Vzhľadom na vyššie uvedené skutočnosti je vhodné tieto protokoly podrobiť analýze a experimentálnemu porovnaniu v realistických podmienkach, ktoré reflektujú dynamiku reálnych mestských prostredí. Cieľom tejto štúdie je zhodnotiť výkonnosť vybraných smerovacích protokolov z pohľadu kľúčových metrík siete – spoľahlivosť, priepustnosť, latencia a odolnosť voči dynamickej zmene topológie – a navrhnúť ich vhodnosť pre rôzne typy nasadenia v oblasti IoT a smart cities [7], [5].

A. MANET-IoT a smerovacie algorimy

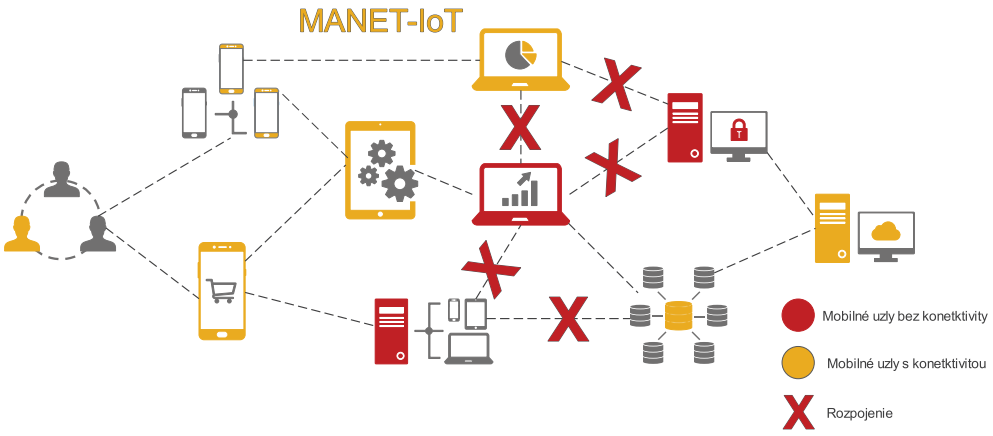
MANET-IoT siete predstavujú technologický paradox – na jednej strane musia byť dostatočne odolné voči neustále sa meniacim podmienkam v dynamickej topológii siete, zatiaľ čo na druhej strane musia rešpektovať prísne energetické a výpočtové obmedzenia zariadení v prostredí internetu vecí (IoT). V prostredí s vysokou mierou mobility, výpadkami signálu, rušením alebo energetickými obmedzeniami (napr. vybitie batérií senzorov) dochádza k náhlym výpadkom uzlov, čo si vyžaduje okamžitú adaptáciu sieťovej štruktúry. Zároveň však tieto siete operujú na zariadeniach, ktoré často nedisponujú dostatočným výpočtovým výkonom alebo energetickými rezervami na vykonávanie komplexných algoritmov či časté aktualizácie topológie.

Výzvy však nespočívajú len v samotnej dynamike, ale aj v rozmanitosti aplikačných scenárov. V kontexte inteligentných miest sa vyžaduje takmer okamžitá reakcia na zmenu polohy zariadení, napríklad pri riadení dopravy v reálnom čase. V precíznom poľnohospodárstve musia senzory optimalizovať spotrebu energie, keďže ich fyzická údržba v teréne je často neefektívna alebo nemožná. V krízových situáciách, ako je kolaps infraštruktúry, zohráva MANET kľúčovú úlohu pri zabezpečení ad hoc komunikácie medzi záchrannými jednotkami.

Súčasné smerovacie protokoly sa s týmito požiadavkami vyrovnávajú rôznorodo. Proaktívne riešenia, ako napríklad OLSRv2, udržiavajú aktuálnu topológiu prostredníctvom pravidelnej výmeny riadiacich správ, čo zabezpečuje rýchlu dostupnosť trás, ale za cenu vyššej záťaže a spotreby energie. Naopak, reaktívne alebo distribuované prístupy, ako B.A.T.M.A.N.-adv, smerujú pakety skôr lokálne – na najlepšieho suseda – čím znižujú výpočtovú náročnosť, no môžu spôsobiť oneskorenie pri vyhľadávaní ciest v prípade náhlej zmeny siete. Zatiaľ neexistuje univerzálne riešenie, ktoré by dokázalo efektívne pokryť všetky požiadavky vyplývajúce z heterogénnych podmienok prevádzky MANET-IoT sietí.

Táto štúdia sa preto zameriava na komparatívnu analýzu vybraných protokolov – Babel, OLSRv1 a OLSRv2 – ktoré reprezentujú odlišné prístupy k smerovaniu v dynamických sieťach. Cieľom experimentálnych simulácií je zhodnotiť ich výkonnosť a stabilitu v prostredí s vysokou mierou mobility, výpadkov a šumov.

Analýza napokon smeruje k identifikácii protokolu, ktorý v jednotlivých podmienkach dosahuje najlepšiu rovnováhu medzi spoľahlivosťou, výkonom a energetickou efektivitou. Táto rovnováha je kľúčová bez ohľadu na to, či ide o zber dát z autonómnych traktorov na poliach, komunikáciu medzi dronmi počas pátracích operácií, alebo koordináciu zložiek integrovaného záchranného systému v krízovej situácii.



Obr. 2 MANET-IoT siet a rozpojenia komunikačných liniek

B. Odolnosť smerovania v MANET-IoT sieťach

Odolnosť siete MANET-IoT predstavuje schopnosť odolávať narušeniam siete a výpadkom zariadení v sieti, a naďalej poskytovať služby v komunikácii medzi jednotlivými zariadeniami IoT [2]. Kombinácia samoorganizujúcich sa MANET a IoT technológií prináša nové výzvy škálovateľnosti aj spoľahlivosti sietí, čo podrobne rozoberajú Gatete a Harubwira vo svojej štúdii [3]. Štruktúra odolnosti, definujúca redundanciu, robustnosť, vynaliezavosť a rýchlosť obnovy komunikácie, často čerpá z normy ISO, ale prehľady uvádzajú rozsiahle porovnanie existujúcich prístupov k ich implementácii v smart prostredí IoT.

V MANET-IoT hlavnou príčinou narušenia odolnosti je mobilita uzlov, zlyhania spojení a výpadky uzlov [1] Odolná sieť by mala udržiavať spojenie a výkonnosť medzi IoT zariadeniami aj napriek týmto nepriaznivým výzvam. Kľúčové aspekty odolnosti zahŕňajú redundanciu, robustnosť (Obr. 2), vynaliezavosť a rýchlosť obnovy. Menej husté siete môžu mať menej redundantných trás, a preto sú náchylnejšie na rozdelenie pri zlyhaní uzlov, zatiaľ čo hustejšie siete môžu trpieť vyššou mierou rušenia. Začiatočné umiestnenie uzlov v MESH sieti má teda výrazný vplyv na základnú odolnosť celej IoT siete.

Mobilita uzlov alebo zariadení spôsobuje časté zmeny v topológii siete, čo vedie k prerušovaniu spojení medzi zariadeniami. Vyššia mobilita vo všeobecnosti vedie k zníženiu stability siete. Smerovacie protokoly musia byť schopné odolať týmto dynamickým a náhodným zmenám, aby dokázali udržať stabilné spojenia. Rôzne modely mobility (napr. Random Waypoint) umožňujú simulovať rôzne pohybové vzorce, ktoré majú rôzny dopad na výkonnosť smerovania [3], [4]. Správny výber modelu mobility je preto dôležitý pre realistickú simuláciu a hodnotenie protokolov pre MESH topológie.

Zlyhania určitých uzlov môžu viesť k strate spojení a prerušeniu komunikácie, čo si vyžaduje presmerovanie dátových tokov [5]. Vplyv takýchto zlyhaní závisí od topologickej dôležitosti uzla a dostupnosti alternatívnych ciest (Obr. 2). Viacnásobné alebo trvalé zlyhania môžu spôsobiť tzv. rozdelenie siete na izolované segmenty, čím sa znemožní vzájomná komunikácia [7]. Odolné protokoly by preto mali minimalizovať pravdepodobnosť tohto stavu.

II. Simulácia výpadkov uzlov a mobility v MANET-IoT sieti

V rámci simulácie sme sa zamerali na hodnotenie vplyvu zlyhaní uzlov na výkonnostné parametre siete, konkrétne na pomer doručených paketov, priepustnosť a latenciu. Výpadky uzlov, ktoré narúšajú topológiu siete, môžu výrazne degradovať kvalitu komunikácie. Strata kľúčových trás vedie k zvýšenej záťaži na zostávajúcich uzloch a spôsobuje vznik úzkych miest, čo následne vedie k stratám paketov a zvýšenej latencii (Tab. 1).

Na účely testovania sme implementovali rôzne scenáre s náhodným odpojovaním časti uzlov a zároveň simulovali mobilitu v simulačnom prostredí simulátora MeshLab [8], [9], [10]. Cieľom bolo otestovať, ako si rôzne smerovacie protokoly poradia s výpadkami až 30, 50 a 75 % náhodne vybraných uzlov. Takéto extrémne podmienky reflektujú situácie v krízových nasadeniach alebo v prostredí s vysokou dynamikou, typickou pre MANET-IoT siete.

Odolné smerovacie protokoly by mali byť schopné zachovať základnú konektivitu a prijateľnú úroveň výkonnosti aj pri výpadkoch, pričom dôležitým aspektom je schopnosť rýchlej rekonfigurácie trás a detekcie zlyhaní. Napríklad reálne nasadenia v priemyselných sieťach, potvrdzujú, že

použitie redundantných rádiových kanálov výrazne znižuje dopad porúch na celkovú spoľahlivosť siete [3],[4].

Naše simulácie hodnotili, do akej miery sú testované protokoly schopné zabezpečiť stabilnú komunikáciu aj v prípade náhlych a rozsiahlych porúch. Výsledky ukázali rozdiely v rýchlosti obnovy trás, stabilite doručovania a schopnosti protokolov adaptovať sa na meniace sa podmienky v topológii siete.

A. Výpadok 30% spojení v procese smerovania v MANET-IoT

V rámci simulácie bolo v každej iterácii náhodne odpojených až 30 % uzlov z, čo umožnilo otestovať schopnosť proaktívnych smerovacích protokolov adaptovať sa na čiastočne poškodenú topológiu siete a obnoviť spojenia medzi uzlami. Okrem toho boli všetky uzly nasadené v pohybe s náhodnou rýchlosťou v intervale 10–200 metrov a náhodným smerom pohybu.

Výsledky simulácie potvrdzujú, že aj pri miernej úrovni výpadkov môže kombinácia mobility uzlov a ich náhodného odpojovania výrazne ovplyvniť stabilitu spojení a celkovú výkonnosť siete.

Analýza výsledkov ukazuje, že pri 30 % výpadku dosahovala priemerná úspešnosť doručenia paketov hodnoty približne 46 % pre protokol Babel, 41 % pre OLSRv2 a len 20 % pre OLSRv1. Protokol Babel tak v danom scenári dosiahol najlepšie výsledky, aj keď s vyššou variabilitou medzi iteráciami. OLSRv2 vykazoval porovnateľnú, no stabilnejšiu výkonnosť, zatiaľ čo OLSRv1 zaostával s trvalo nízkymi hodnotami úspešnosti doručenia paketov.

Počas väčšiny iterácií sa úspešnosť doručenia paketov pri protokoloch Babel a OLSRv2 pohybovala v rozsahu 30–60 %, pričom v niektorých prípadoch protokol Babel prekročil 70 %. Tento fakt naznačuje ich schopnosť zachovať konektivitu siete aj pri zvýšenej dynamike a čiastočných výpadkoch uzlov. Naopak, OLSRv1 vykazoval úspešnosť doručenia prevažne v rozsahu 10–25 %, čo signalizuje jeho obmedzenú schopnosť adaptácie na meniace sa sieťové podmienky.

Výkonnosť protokolu OLSRv2 bola výrazne podporená mechanizmom Fish-Eye Routing a optimalizovaným formátom správ, ktoré umožňujú efektívnejšie a častejšie šírenie topologických informácií v okolí uzla bez výrazného zaťaženia siete. Protokol Babel efektívne využíval časté HELLO správy, čo mu umožňovalo flexibilne reagovať na zmeny sieťovej topológie. Naopak, OLSRv1 postrádal tieto pokročilé mechanizmy, čo negatívne vplývalo na jeho schopnosť obnoviť trasy po topologických zmenách.

Pri 30 % výpadku uzlov dochádzalo často k rozdeleniu siete na viacero izolovaných častí, pričom niektoré uzly ostávali úplne odrezané od zvyšku siete. Tieto výsledky zdôrazňujú citlivosť proaktívnych protokolov na dynamiku siete, nakoľko kontinuálne udržiavanie celkovej topológie spôsobuje, že náhle pohyby a výpadky môžu narušiť konzistenciu a presnosť smerovacích tabuliek pred ich aktualizáciou. Výrazné rozdiely v úspešnosti doručenia paketov sú dôsledkom súčasného výskytu mobility a výpadkov, ktoré pravidelne narúšajú sieťovú štruktúru, avšak občas umožňujú vznik stabilnejších prepojení.

B. Výpadok 50% spojení v procese smerovania v MANET-Iot

V rámci simulácie bola testovaná odolnosť vybraných smerovacích protokolov v extrémnych podmienkach, keď dochádzalo k náhodnému výpadku až 50 % uzlov a súčasne k náhodnému pohybu uzlov v rozsahu 10 až 200 metrov počas 50 iterácií. Takto navrhnuté scenáre reflektujú vysoko dynamické a nestabilné prostredia typické pre MANET-IoT siete. Výsledky ukazujú, že napriek výrazným stratám konektivity dokázali niektoré protokoly udržať funkčné spojenie a dosiahnuť pomerne priaznivú mieru doručenia paketov.

Najlepšie výkonnostné charakteristiky preukázal protokol Babel, ktorý dosiahol priemernú úspešnosť doručenia okolo 50 %. V mnohých iteráciách si udržiaval hodnoty medzi 30–70 %, pričom niekoľkokrát prekročil hranicu 80 %. Táto výkonnosť poukazuje na jeho schopnosť efektívne reagovať na zmeny topológie a prispôsobiť sa výpadkom uzlov, čo je dôsledkom jeho reakčného mechanizmu na báze smerovacích metrík a adaptívneho aktualizačného procesu.

Druhým najvýkonnejším bol OLSRv2, ktorý dosiahol priemerne 41 % úspešnosť doručenia. Hoci v niektorých iteráciách prekonal Babel, celkovo sa ukázal ako menej konzistentný. Rozsah jeho výkonu sa pohyboval od približne 20 % do vyše 80 %, čo naznačuje určitú adaptabilitu – podporenú najmä technikou Fish-Eye smerovania a efektívnejšou propagáciou informácií o topológii v porovnaní so svojím predchodcom.

Naopak, OLSRv1 vykazoval najnižšiu stabilitu aj úspešnosť, pričom priemerná hodnota doručenia bola len okolo 23 %. Väčšina iterácií sa pohybovala v pásme 15–25 %, pričom výnimočne sa úspešnosť zvýšila nad 30 %. Slabý výkon tohto protokolu naznačuje jeho obmedzenú schopnosť prispôsobiť sa rýchlo sa meniacim podmienkam siete a nízku efektivitu v rekonštrukcii ciest po výpadkoch uzlov.

Parameter	Hodnota
Simulátor	meshnet-lab
Počet uzlov	100
Smerovacie protokoly	Babel, OLSRv1, OLSRv2
Trvanie simulácie	100 s
Počet iterácií	50
Fyzikálna vrstva	IEEE 802.11n (2.4 GHz)

Tabuľka 1
Simulačné parametre v prostredí simulátora meshnet-lab [9]

Získané výsledky potvrdzujú, že moderné a flexibilnejšie proaktívne protokoly, ako Babel a OLSRv2, dokážu zabezpečiť relatívne spoľahlivú komunikáciu aj v prípade výrazných výpadkov a mobility uzlov. Protokol OLSRv1 sa naopak javí ako menej vhodný pre vysoko dynamické MANET-IoT prostredia, kde je nevyhnutná rýchla adaptácia a robustnosť voči stratám konektivity.

C. Výpadok 75% spojení V procese smerovania v MANET-Iot

V poslednom simulačnom scenári sme testovali správanie smerovacích protokolov v mimoriadne náročných podmienkach, kde v každej iterácii došlo k náhodnému odpojeniu až 75% uzlov v sieti. Súčasne boli všetky uzly vystavené náhodnému pohybu v rozsahu 10 až 200 metrov. Tieto faktory zásadne narušili sieťovú topológiu a predstavovali extrémnu výzvu pre udržiavanie spojenia a smerovanie paketov.

Výsledky preukázali značné kolísanie výkonnosti všetkých testovaných protokolov, čo bolo vzhľadom na rozsah výpadkov očakávané. Protokol Babel sa aj v týchto podmienkach ukázal ako najodolnejší, pričom dosiahol priemernú úspešnosť doručenia na úrovni približne 43%. V niektorých iteráciách sa jeho výkon priblížil, alebo dokonca prekročil hranicu 70%, čo poukazuje na jeho schopnosť adaptácie a efektívnej obnovy trás aj pri masívnom výpadku siete.

OLSRv2 dosiahol priemernú úspešnosť približne 38%. Hoci v niektorých prípadoch prekonal Babel, jeho výkonnosť bola menej konzistentná. Výsledky naznačujú, že implementované optimalizácie ako napríklad Fish-Eye Routing síce zvyšujú jeho schopnosť prispôsobiť sa dynamickej topológii, no nedokážu vždy zabezpečiť rýchlu reakciu na drastické zmeny v štruktúre siete.

Naopak, protokol OLSRv1 potvrdil svoje obmedzenia. S priemernou úspešnosťou doručenia okolo 23% výrazne zaostal za oboma novšími protokolmi. Výkonnosť sa pohybovala prevažne v rozmedzí 20–25%, pričom len zriedkavo dosahovala vyššie hodnoty. Slabá adaptabilita a absencia moderných smerovacích mechanizmov sa v tejto simulácii prejavili obzvlášť výrazne.

Záverom možno konštatovať, že aj pri výpadku väčšiny siete môžu pokročilejšie proaktívne protokoly ako Babel a OLSRv2 zabezpečiť určitý stupeň konektivity. Ich schopnosť rýchlo obnovovať trasy a flexibilne reagovať na zmeny v topológii im umožňuje efektívnejšie zvládať extrémne dynamické prostredia. Naproti tomu tradičné riešenia ako OLSRv1 sa v takýchto scenároch ukazujú ako nevhodné. Výsledky zároveň poukazujú na prirodzené limity proaktívnych prístupov – ak topológia siete kolabuje rýchlejšie, než sa stihnú aktualizovať smerovacie tabuľky, dochádza k výraznému zníženiu efektivity smerovania.

III. Záver

Na základe vykonaných simulácií možno konštatovať, že výkonnosť smerovacích protokolov v MANET-IoT sieťach je výrazne determinovaná dynamikou siete, mierou výpadkov uzlov a variabilitou topológie. Výsledky jednoznačne ukazujú, že žiadny z testovaných protokolov neposkytuje univerzálne optimálne riešenie pre všetky scenáre.

V rámci proaktívneho prístupu preukázal protokol Babel najvyššiu odolnosť a schopnosť udržať stabilné smerovanie aj pri masívnych výpadkoch a vysokej mobilite uzlov. OLSRv2 dosahoval uspokojivú úroveň prispôsobivosti, avšak jeho výkon bol menej konzistentný. Naproti tomu OLSRv1 sa ukázal ako najmenej vhodný pre vysoko dynamické prostredia, keďže jeho výkonnosť výrazne degradovala pri zvýšenej sieťovej nestabilite.

Zistenia naznačujú, že výber smerovacieho protokolu by mal byť starostlivo prispôsobený predpokladaným prevádzkovým podmienkam siete (Tab. 2). V stabilných scenároch sa rozdiely medzi protokolmi minimalizujú, no v prostrediach s vysokou mierou mobility a častými zlyhaniami uzlov sa stáva správna voľba protokolu kritickým faktorom pre zachovanie spoľahlivosti a výkonnosti komunikácie. Výsledky tak poskytujú relevantné odporúčania pre návrh robustných a adaptívnych MANET-IoT sietí.

Zároveň sa ukazuje, že aj odolné proaktívne protokoly, ako Babel a OLSRv2, sú obmedzené inherentnými nevýhodami svojho mechanizmu – najmä požiadavkou na pravidelnú aktualizáciu

Mobilita	**Protokol**	**Hlavná výhoda**
Vysoká mobilita	Babel(Proaktívny)	Rýchla detekcia zmien Nízka strata paketov
Rozsiahle topológie	OLSRv1 (Proaktívny)	Záložné trasy a tunelovanie pre efektívne reagiovanie na výpadky v sieti.
Stabilita siete	OLSRv2 (Proaktívny)	Spoľahlivosť a podpora IPv6 pre napríklad smart city

Tabuľka 2
Odporúčania na základe simulačných výsledkov získaných v prostredí simulátora meshnet-lab

topológie. V prostrediach, kde sa sieť mení rýchlejšie, než sa stihnú distribuovať nové smerovacie informácie, dochádza k využívaniu zastaraných trás, čo znižuje efektivitu smerovania.

Z tohto dôvodu je do budúcnosti vhodné uvažovať o hybridných smerovacích riešeniach, ktoré kombinujú výhody proaktívnych a reaktívnych prístupov, a súčasne optimalizovať periodicitu a rozsah aktualizačných mechanizmov s cieľom zvýšiť robustnosť siete v reálnych a nepredvídateľných podmienkach.

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] V. K. Quy, N. T. Ban, D. Van Anh, N. M. Quy, and D. C. Nguyen, "An adaptive gateway selection mechanism for manet-iot applications in 5g networks," *IEEE Sensors Journal*, vol. 23, no. 19, pp. 23 704–23 712, 2023.

[2] Y. Liu, L. Chen, H. Xu, D. Yang, T. Cao, and Y. Liu, "Joint neighbor discovery and topology exchange with directional antennas in multi-hop manets," *IEEE Access*, vol. 13, pp. 95 218–95 226, 2025.

[3] D. Xue and W. Huang, "Smart agriculture wireless sensor routing protocol and node location algorithm based on internet of things technology," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 24 967–24 973, 2021.

[4] M. J. AL-Mashhadani and K. Karoui, "Rule-based and ai-based routing protocols for mobile ad hoc networks: A comparative review," in *2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA)*, 2025, pp. 1–12.

[5] N. Schweitzer, A. Stulman, R. D. Margalit, and A. Shabtai, "Contradiction based gray-hole attack minimization for ad-hoc networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2174–2183, 2017.

[6] K. Kiran, N. P. Kaushik, S. Sharath, P. D. Shenoy, K. R. Venugopal, and V. T. Prabhu, "Experimental evaluation of batman and batman-adv routing protocols in a mobile testbed," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, 2018, pp. 1538–1543.

[7] 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.

[8] M. Kadadha, H. Abualola, H. Otrok, R. Mizouni, S. Singh, F. Betene, and J.-P. Giacalone, "A cluster-based quality-of-service optimized link state routing protocol for mesh networks," in *2022 International Wireless Communications and Mobile Computing (IWCMC)*, 2022, pp. 336–341.

[9] P. Cignoni, A. Muntoni, G. Ranzuglia, and M. Callieri, "Meshlab."

[10] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Eurographics Italian Chapter Conference*, V. Scarano, R. D. Chiara, and U. Erra, Eds. The Eurographics Association, 2008.

Distributed edge architecture for control of selected laboratory models

Ján SPIŠÁK, Peter PAPCUN

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

jan.spisak@student.tuke.sk, peter.papcun@tuke.sk

Abstract **– The goal of this contribution is to enable the control and monitoring of selected laboratory models using the Internet of Things concept using an edge device based on the REST API. The laboratory models with which we will work in this contribution are located in the Smart Industry – Industry 4.0 laboratory and the Think Lab: IoT & Vision & Robotics laboratory of the Technical University of Košice. After implementing this REST API, we will conduct experiments that focus on comparing latency using different network transmission media: Ethernet, Wi-Fi, and 5G.**

Keywords **– IoT, Edge architecture, Client-Server, Flask, PLC, ESP32, Raspberry Pi 5, Arduino**

I. Introduction

With the advent of the fourth industrial revolution, concepts such as digitalization and the Internet of Things (IoT) have become widespread, changing the way various industries operate. IoT enables the interconnection of devices, automation and real-time monitoring, which increases efficiency and enables remote control, especially in industry and laboratories. An important element of this technology is the edge architecture, which reduces latency and improves system reliability. This allows for faster response and the creation of real-time visualizations of processes.

The reliable functioning of such an architecture also depends on the network elements used. Ethernet is commonly used, but in cases where this is not possible, wireless solutions such as Wi-Fi or 5G are applied.

This contribution focuses on the design and implementation of a distributed system based on edge architecture and REST API, designed to control selected laboratory models. It includes an analysis of IoT architectures and network solutions, system design, its practical implementation and testing with various types of transmission media, as well as an evaluation of the results.

II. Analytical Overview of the Contribution

The IoT is a network of devices, such as sensors and actuators, that are connected to the Internet and communicate with each other. IoT brings benefits in everyday life, but also challenges in the area of security and data protection.

An extension of this concept is the Internet of Everything (IoE), which connects things, people, data and processes. The goal is efficient communication and automation, but this also places demands on security and interoperability.

These technologies are the basis of the fourth industrial revolution and use different types of data processing and distributed computing: cloud computing ([1], [2]), fog computing ([1], [2]), edge computing ([3], [4]).

A. Communication models

1) Client-server model: A device (client) sends requests to a server, which processes and responds. This model is common in IoT, for example when transmitting data from sensors.

2) Peer-to-Peer model (P2P): In P2P models, devices communicate directly with each other without a central server. As [5] states, there are different types of P2P networks – pure, hybrid, structured, or unstructured – that differ in the degree of decentralization and the way data is indexed.

B. TCP/IP Model

The TCP/IP model (Fig. 1), which forms the basis of the Internet, consists of four layers: application, transport, network, and access. Each layer provides specific network functions. The equivalent is the seven-layer OSI model, which provides a more detailed breakdown of these functions.

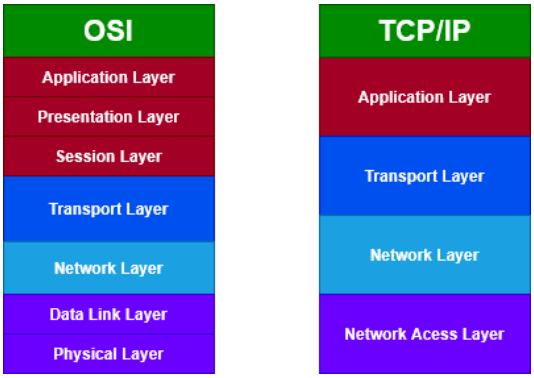


Fig. 1 Comparison of TCP/IP and OSI models

C. REST API and endpoints

REST (Representational State Transfer) is an architectural style for designing web services [6]. REST APIs enable communication between a client and a server using HTTP methods such as *GET*, *POST*, *PUT*, or *DELETE*. REST API is simple, easily scalable, and is mainly used in combination with JSON as a data exchange format. In the field of cybersecurity, it is essential to ensure authentication, authorization, encrypted communication, and protection against attacks such as SQL injection, man-in-the-middle, and brute-force when designing and using REST APIs.

D. Computer networks

Computer networks are systems that allow devices such as computers, mobile phones, Internet of Things devices, and others to communicate with each other to share data, resources, and services. These networks can be of various types, such as local area network (LAN), metropolitan area network (MAN), wide area network (WAN), personal area network (PAN), body area network (BAN), or global area network (Internet). The basic components of computer networks are devices such as routers, switches, access points, and firewalls that provide data routing, security, and connectivity between servers and their clients. Computer networks use the following transmission media [7]:

- wired (twisted pair, coaxial cable, fiber optic cable),
- wireless (radio waves, microwave connections, infrared light).

III. Design of an edge architecture based on REST API

This chapter is dedicated to the design of a REST API-based communication architecture for monitoring and controlling selected laboratory models. The web application will contain four specific interfaces for these models:

- programmable logic controller (PLC) model Case [8],
- PLC model Traversa [9],
- IoT model Tube (ESP32) [10],
- IoT model House (Arduino + Raspberry Pi 5) [11].

It will be possible to use the application to obtain data, monitor the activity of the models, control active members, and display the development of trends on the graph. Since the models are in two laboratories, each will have its own web application created in Flask and deployed on the Raspberry Pi 5 edge server. Fig. 2 presents the entire architecture. The use of two edge servers will provide response, load balancing and possible further extensions.

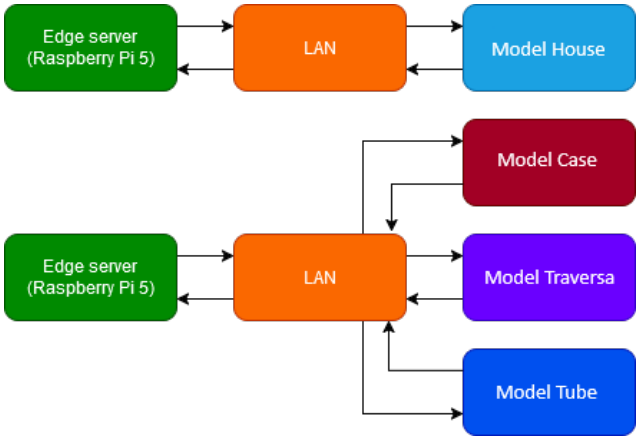


Fig. 2 Architecture of connecting laboratory models with edge servers

A. Design of communication architecture for PLC devices

PLC devices differ from classic IoT devices in that they do not communicate actively with the web application themselves. To obtain data, the server must initiate communication, which is not common in the classic client-server model. To communicate with the PLC in the Python environment, the pylogix [12] library is used, which can obtain a list of data from the PLC after entering the IP address.

In order for the server to be able to obtain this data regularly, there are two solutions:

- A helper script running in a cycle on the edge device (e.g. Raspberry Pi 5), which will ensure continuous data transfer to the web application.
- JavaScript in the GUI, which obtains data at regular intervals during an active user connection.

Alternatively, it is possible to create an independent script or device that would function as an ”intelligent agent”. This would ensure continuous data acquisition and processing, thus replacing the user and supporting process automation. Both approaches have advantages and disadvantages, but both allow sending current data to the PLC. This method of communication will be used by the Case and Traversa models.

B. Design of communication architecture for REST devices via JSON POST

This type of communication is most often used to exchange data between a web application and an IoT device. The IoT device regularly sends data from sensors and actuator status in JSON format. The server processes this data, stores it, and responds back in JSON format as well. This exchange cycle continues as long as the device is connected. This approach is used by the laboratory IoT model House.

C. Design of communication architecture for REST devices via parametric GET

Communication via parametric GET is similar to the previous method and is often used due to its simplicity and speed. The IoT device sends a GET request with the specified parameters in the URL and the server responds with an HTML document containing instructions for the device.

However, if the application requires standard HTML5 with a full structure, it is more difficult to obtain the necessary data from the document (it must first be processed – parsed). If an application like Flask allows this, it is more advantageous to send only clean data in a simple format (e.g. as CSV), which will significantly simplify processing on the IoT device side.

This form of communication will be used by the IoT model Tube. As with PLC devices, it is also advisable to add a separate interface to the web application that will allow other devices to directly communicate with the models using GET and POST requests.

D. Design of screens for control and monitoring interfaces

Before implementing the web application, we designed a basic user interface (GUI) design using Figma, which serves as a visual basis for creating individual screens. Overview of the designed screens:

- Home screen: Displays brief information about the site and provides links to other parts of the application.

- GUI screen: Used to monitor and control laboratory models through an intuitive interface with forms and data visualization.
- Knowledge article screen: Provides the user with basic information, an image of the model, and instructions for using the GET and POST methods.

IV. Implementation of edge devices based on REST API

In this chapter, we will gradually review the communication architecture for each laboratory model. We will also describe the implementation of individual web control interfaces within the REST API. In total, 6 control and monitoring interfaces were created during the implementation: *Case GUI, Traversa GUI, Universal PLC GUI, Universal REST GUI, House GUI, Tube GUI.*

A. Implementation of the interface for Case GUI

1) Implementation of communication architecture: Initially, we needed to retrieve data from the Case PLC model via the pylogix library to get a list of available tags. After retrieving these tags, we could retrieve the values and identify which sensor or actuator is connected to each pin. Since not all tags were used, we retrieved the list of available tags only once, which accelerated the communication between the PLC and the Flask server. The JavaScript on the web page updates the data as long as the user has the main page of the Case PLC web interface open.

2) Web interface implementation: The main page (GUI) contains a navigation panel on the left and an interactive control interface on the right Fig. 3. The design was created to faithfully copy the PLC model Case, which provides users with the same overview of the model's status as the HMI interface. The user can monitor sensors and send instructions such as turning on lights, setting a value on the 8-segment display, or turning on the backlight of the buttons. The user can interact with the 8-segment display in two ways: by clicking on the segments or by setting a specific numerical value via the input field. The navigation panel offers basic functions:

- Home: return to the application's home page,
- Data API: display of current data in JSON format,
- About: help on functionalities and correct use of GET and POST methods,
- Logs: records of interactions between the user and the model.

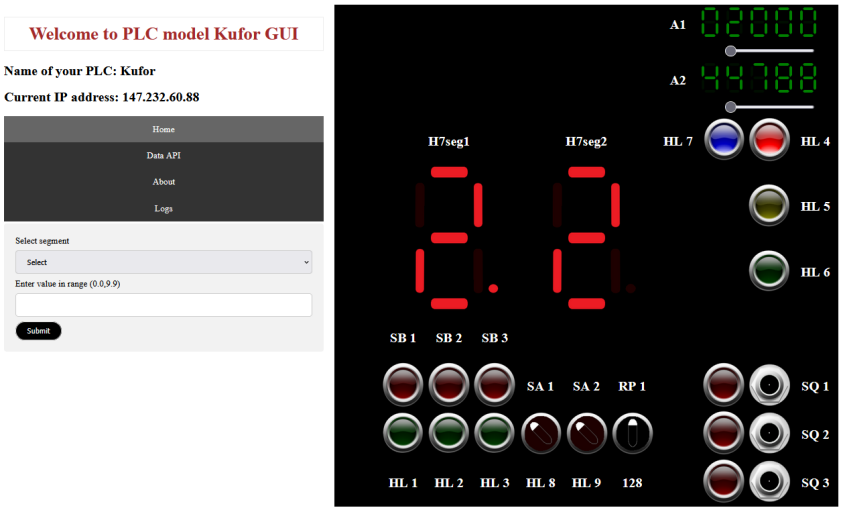


Fig. 3 Web interface GUI for PLC model Case

3) Communication interface for third-party applications: The web application allows third parties to retrieve current data from the Case model via GET method and write new values via POST with JSON data. POST functionality is limited to setting one tag per request.

To avoid multiple requests to the PLC device, we have implemented a mechanism that allows only one user or application to update data. If the user uses the GUI page, he is responsible for updating the data. If a third-party application queries the data, he only gets a local copy. After the user's connection is terminated, the Data API takes over the data update, unless the third-party application creates a new request. After the user reconnects, he gets priority in updating the data.

Two other interfaces: one specific for the Traversa PLC model and the other universal for any PLC device from the Rockwell Automation manufacturer use the same library to exchange data. Let's focus therefore on another interface that uses the POST method with JSON content.

B. Implementation of the interface for house GUI

1) Implementation of communication architecture: The IoT model House uses the Arduino Due controller, so it was necessary to implement an edge device (Raspberry Pi 5) that communicates with Arduino via a serial UART connection. At the beginning, we were given the code running on Arduino to see what messages it could send. This model had already been implemented in the past, but only the controller and the model with the corresponding sensors and active members remained.

Based on this information, we created a client application that provides communication between Raspberry Pi and Arduino via UART. Commands with the "?" characters are used to obtain sensor data, while commands with the "=" characters allow writing instruction data for actuators. After stopping data, the application sends a request via the JSON POST method to a web application, which processes the request and returns a response with the current state of the variables. This process runs in an infinite loop.

2) Web interface implementation: The main page contains a navigation panel on the left and two control panels on the right of the Fig. 4. The first panel, *Lighting*, allows you to control the lighting and blinds in the 4 rooms of the House model. The user can manually turn on/off the lights, adjust the lighting intensity and control the automatic light regulation according to the light intensity requirement (Expected Luminosity). This parameter is sent to the Arduino, which performs the regulation based on the feedback from the sensors.

The second panel, "Heating", allows you to control the Peltier unit and the temperature in the rooms. There are manual and automatic modes, where the automatic mode sets the desired temperature and the Arduino regulates the temperature based on the temperature sensor. When the automatic mode is turned on, the previous manual settings are reset. If the user wants to manually control the components again, they must turn off the automatic mode.

The navigation panel contains links to: *Home, Graphs, Data API, About, Logs*.

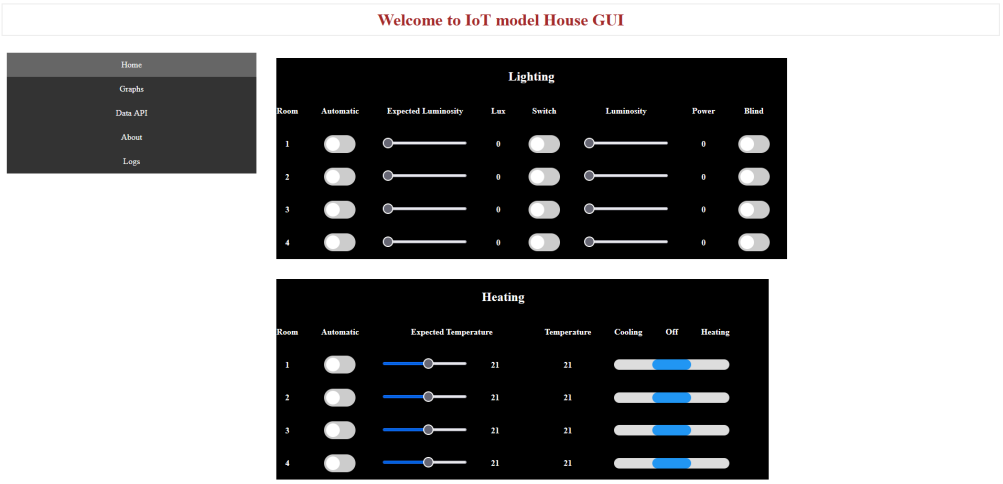


Fig. 4 Web interface GUI for PLC model House

3) Communication interface for third-party applications: Third-party applications can retrieve current data from the IoT model House and send new instructions via JSON POST. Parameters that can be set include: *turning lights on/off; setting light intensity; turning automatic mode for lights on/off; setting Lux value for automatic mode; controlling blinds; turning peltier unit on/off; setting temperature for automatic temperature control*.

These parameters can be set for each of the 4 rooms, and the JSON message must include the room number. Each communication can contain only one parameter. A similar communication is also used by the universal interface for any REST device. Therefore, we will focus on the last type of communication we used, namely parametric GET.

C. Implementation of the interface for tube GUI

1) Implementation of communication architecture: The data exchange between the ESP32 and the web interface takes place in an infinite loop, where the ESP32 regularly sends data about the height of the ball in the tube and the motor power (PWM) via GET request. The web application processes this data and returns an HTML page with the current values. Originally, only one value of the ball height and motor power was sent, but due to problems with the fluctuating latency of

the Wi-Fi signal, the Tube model was generated faster than the communication unit could process and send it.

To solve this problem, we modified the reading from the serial connection so that all records are read with each reading and then edited into CSV format (for the ball height and motor power). This data is then regularly sent via GET request, which guarantees the supply of current data in real time and eliminates the problem of a weak Wi-Fi signal.

2) Web interface implementation: The web interface contains a navigation panel on the left and two forms on the right of the Fig. 5. The first form is used to set the desired ball height and its rotation angle. The second form allows you to set these parameters using the position [x, y], where the center (point [0, 0]) is in the middle of the tube. There is also an ON/OFF button to turn the ball control on and off.

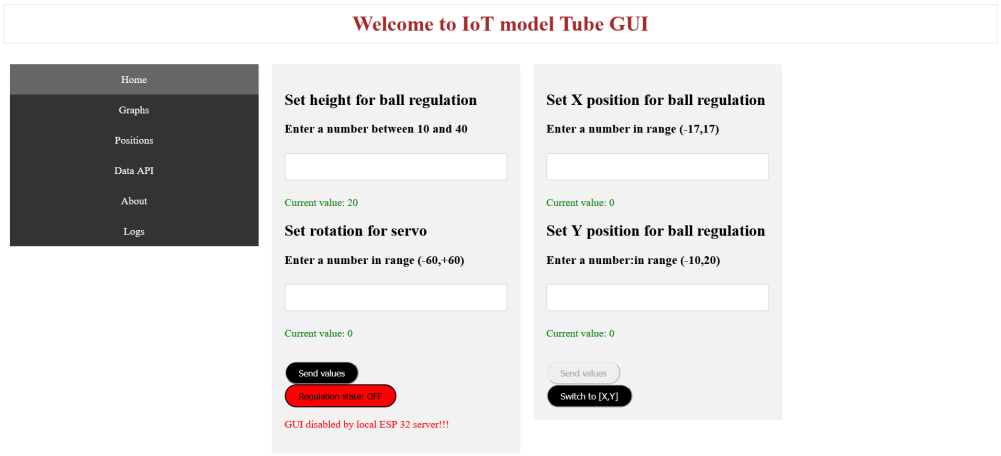


Fig. 5 Web interface GUI for PLC model Tube

In addition to the basic links, the navigation panel also has two specific to the Tube model:

- Graphs – Displays graphs of current trends in the ball control in the tube. The range of the monitored period can be set at the bottom of the page.
- Positions – Displays all valid positions for setting the ball height and the tube rotation angle.

3) Communication interface for third-party applications: In addition to retrieving current data about the Tube model, third-party applications can also send new control instructions via JSON POST. Possible settings include:

- enable/disable control,
- set the height of the ball and the angle of rotation of the tube,
- set the height and angle using the [x, y] position.

V. Evaluation and Latency Experiments

The goal of evaluating our solution is to compare the responses (latency) between 5G networks and commonly used networks such as Wi-Fi or Ethernet.

In the experiment, two Raspberry Pi 5 units, located on different floors of the building, will be used to test the network response when communicating between the web application and PLC devices over different types of connections (Ethernet, Wi-Fi, 5G), while comparing latency in normal operation and outside of it.

Basic points of the experiment:

- Devices: One Raspberry Pi 5 unit serves as an edge server in the V113-114 laboratory, the other as a client in the V011 laboratory.
- Tested requests: The client will send three types of GET requests - to the home page and data from two PLC models (/, /caseDataAPI, /traversaDataAPI).
- Goal of the experiment: Compare the response speed (latency) for different types of network connections - Ethernet, Wi-Fi, 5G.
- Limitations: Parameters other than download or upload speed are not taken into account, only response time.
- Ethernet: Stable connection via gigabit cable (Cat5e), using a switch in the laboratory.
- Wi-Fi: Network (802.11ac, 5.32 GHz), accessible to TUKE students and staff.
- 5G network: TUKE's private 5G network, connected via Waveshare 5G modem with QUECTEL RM520N-GL chip (Sub-6 GHz, 2.4 Gbps down / 900 Mbps up).

- Expectations: Wireless networks (Wi-Fi, 5G) will be slower due to additional routing compared to a direct Ethernet connection.

A. Measuring the response between the web application and the PLC models

Initially, the response measurements were performed between the web application and the PLC models Case and Traversa, using a Python script also used in real data collection. Five measurements were performed for each model, each with 100 requests.

Subsequently, the response between the Raspberry Pi 5 client device and the web server was measured in several scenarios:

- outside active network use (from 16:40 to 17:30)
- during network utilization (from 10:30 to 11:20)

Testing was performed over three transmission media: Ethernet, Wi-Fi, and 5G. A total of 90 measurements were performed (30 for each network type, 5 measurements for each of the 3 tested web endpoints).

All results are summarized in 6 tables, which contains:

- statistics for each network type and network status (busy / unbusy)
- response when loading the initial web page (Table 1, Table 4)
- response when loading data from the Case (Table 2, Table 5) and Traversa models (Table 3, Table 6)

Table 1
Web application home page response measurement statistics during active network usage

	Ethernet	Wi-Fi	5G
Minimal	2 ms	8 ms	30 ms
Maximal	15 ms	567 ms	64 ms
Average	5ms	57 ms	35 ms
Deviation	3 ms	83 ms	2 ms
Median	4 ms	24 ms	35 ms
P90	11 ms	152 ms	38 ms
P95	12 ms	235 ms	38 ms
P99	12 ms	416 ms	39 ms

Table 2
/caseDataAPI endpoint response measurement statistics during active network usage

	Ethernet	Wi-Fi	5G
Minimal	399 ms	401 ms	396 ms
Maximal	461 ms	6472 ms	474 ms
Average	412 ms	473 ms	429 ms
Deviation	11 ms	396 ms	9 ms
Median	408 ms	420 ms	428 ms
P90	431 ms	506 ms	438 ms
P95	437 ms	597 ms	443 ms
P99	452 ms	963 ms	454 ms

Table 3
/traversaDataAPI endpoint response measurement statistics during active network usage

	Ethernet	Wi-Fi	5G
Minimal	709 ms	758 ms	783 ms
Maximal	819 ms	3799 ms	859 ms
Average	751 ms	835 ms	798 ms
Deviation	18 ms	212 ms	12 ms
Median	753 ms	780 ms	793 ms
P90	775 ms	949 ms	818 ms
P95	781 ms	1065 ms	820 ms
P99	791 ms	1582 ms	834 ms

Table 4
Web application home page response measurement statistics outside of active network usage

	Ethernet	Wi-Fi	5G
Minimal	2 ms	6 ms	24 ms
Maximal	4 ms	85 ms	32 ms
Average	3 ms	11 ms	30 ms
Deviation	0 ms	6 ms	1 ms
Median	3 ms	10 ms	30 ms
P90	3 ms	13 ms	30 ms
P95	3 ms	14 ms	31 ms
P99	3 ms	17 ms	31 ms

Table 5
/caseDataAPI endpoint response measurement statistics outside active network usage

	Ethernet	Wi-Fi	5G
Minimal	374 ms	380 ms	401 ms
Maximal	472 ms	452 ms	482 ms
Average	389 ms	396 ms	418 ms
Deviation	16 ms	12 ms	15 ms
Median	382 ms	391 ms	411 ms
P90	417 ms	417 ms	443 ms
P95	421 ms	423 ms	448 ms
P99	433 ms	434 ms	455 ms

Table 6
Endpoint response measurement statistics /traversaDataAPI outside active network usage

	Ethernet	Wi-Fi	5G
Minimal	620 ms	742 ms	690 ms
Maximal	703 ms	1829 ms	805 ms
Average	649 ms	772 ms	726 ms
Deviation	17 ms	58 ms	20 ms
Median	649 ms	767 ms	723 ms
P90	669 ms	798 ms	761 ms
P95	675 ms	802 ms	767 ms
P99	692 ms	815 ms	780 ms

B. Findings

Ethernet was the most stable with the lowest latency regardless of network load.

Wi-Fi had a better average latency than 5G when the network was not busy (by 19 ms), but it deteriorated significantly during busy times (up to 46 ms slowdown, maximum response of 567 ms).

The 5G network was more stable and faster than Wi-Fi during the day, especially in real-world conditions, reaching a maximum response of only 64 ms.

These measurements confirm that while Wi-Fi can perform well under ideal (laboratory) conditions, 5G provides a more reliable response in real-world use. However, the results only take into account latency – one of several indicators of network quality.

C. Web application evaluation

The Flask web application includes 6 control and monitoring interfaces for interacting with laboratory models in real time. Four of them are model-specific, two are universal. A communication interface has been added to support third-party applications. The application is automatically launched using Gunicorn on Raspberry Pi 5 edge servers.

Each interface includes a help file with instructions for use, as well as support for third-party applications. Interaction with the interfaces is recorded, allowing for later retrieval of the last user. Some interfaces also include a graphical display of trends in data obtained from the models. The application allows you to add additional interfaces and models for future bachelor's or master's theses.

VI. Conclusion and future works

The aim of the contribution was to create a web application for controlling and monitoring laboratory models via edge servers (Raspberry Pi 5). These models are located in the Smart Industry and Think Lab laboratories at the Technical University of Košice. It was necessary to analyze the technical possibilities, design a REST API using Flask, and provide a communication interface for external applications (GET, POST). All steps were successfully implemented, but both the application and the models still have room for improvement.

The application was designed to be independent of external devices, which allows its deployment on different servers. For greater reliability, it would be appropriate to expand it with a database for storing important data (logs, configurations), which are currently stored only in temporary JSON files. With a higher system load, it will be necessary to ensure file management when multiple workers work in parallel.

Latency was tested via Ethernet, Wi-Fi, and 5G networks. Ethernet remains the fastest and most stable, while Wi-Fi comes close under ideal conditions but suffers from fluctuations. Although 5G showed the highest latency, it was surprising with its stability. Further testing is recommended, for example, with a mobile robots in the TUKE area.

Acknowledgment

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 17R05-04-V01-00004, Competence and Training Center for Cyber and Information Security TUKE.

References

[1] E. Kajati, P. Papcun, C. Liu, R. Y. Zhong, J. Koziorek, and I. Zolotova, "Cloud based cyber-physical systems: Network evaluation study," *Advanced Engineering Informatics*, vol. 42, p. 100988, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474034619305610

[2] M. ABDELSHKOUR, "IoT, from Cloud to Fog Computing," cisco, Accessed: 2025-01-12. [Online]. Available: https://blogs.cisco.com/perspectives/iot-from-cloud-to-fog-computing

[3] P. Papcun, E. Kajati, D. Cupkova, J. Mocnej, M. Miskuf, and I. Zolotova, "Edge-enabled iot gateway criteria selection and evaluation," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 13, p. e5219, 2020.

[4] J. Mocnej, M. Miškuf, P. Papcun, and I. Zolotová, "Impact of edge computing paradigm on energy consumption in iot," *IFAC-PapersOnLine*, vol. 51, no. 6, pp. 162–167, 2018, 15th IFAC Conference on Programmable Devices and Embedded Systems PDeS 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896318308917

[5] "Peer-to-Peer (P2P) Architecture," geeksforgeeks, Accessed: 2025-01-13. [Online]. Available: https://www.geeksforgeeks.org/peer-to-peer-p2p-architecture/

[6] A. Brecko, F. Burda, P. Papcun, and E. Kajati, "Applicability of opc ua and rest in edge computing," in *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2022, pp. 000 255–000 260.

[7] "Network Media Types," mossé Cyber Security Institute, Accessed: 2024-12-27. [Online]. Available: https://library.mosse-institute.com/articles/2023/09/network-media-types/network-media-types.html

[8] B. ŠIMKO, "Návrh a realizácia simulačných a komunikačných aplikácií pre laboratórny model mobilný kufor," *Technická univerzita v Košiciach, Bachelor thesis*, 2016.

[9] E. KAJÁTI, "Návrh a realizácia riadenia laboratórneho modelu traverza pomocou cloudových technológií," *Technická univerzita v Košiciach, Diploma thesis*, 2016.

[10] J. SPIŠÁK, "Riadenie a komunikačná architektúra laboratórneho iiot modelu trubica," *Technická univerzita v Košiciach, Bachelor thesis*, 2023.

[11] M. KUC, "Domácnosť internetu vecí," *Technická univerzita v Košiciach, Bachelor thesis*, 2017.

[12] "Network Media Types," dmroeder, Accessed: 2025-01-28. [Online]. Available: https://github.com/dmroeder/pylogix

Security of Native, Docker, and Kubernetes Deployments: A Comparative Analysis for Edge and IoT Applications

[1]Ľubomír URBLÍK, [1]Michal ILEK, [1]Erik KAJÁTI, [1]Iveta ZOLOTOVÁ

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

[1]lubomir.urblik@tuke.sk, [1]michal.ilek@student.tuke.sk, [1]erik.kajati@tuke.sk, [1]iveta.zolotova@tuke.sk

Abstract **– The proliferation of edge computing and IoT has fundamentally transformed the landscape of application deployment, demanding new approaches to security, scalability, and operational efficiency. This article provides a critical comparative analysis of security in native, Docker, and Kubernetes-based deployments, using the context of a real-world IoT educational panel application. Drawing on recent academic literature (2022–2025), we evaluate isolation, attack surface, vulnerability management, and operational risks. The findings show that while containerization and orchestration offer substantial security improvements over native deployments, they introduce new complexities and require careful configuration, especially in resource-constrained edge environments.**

Keywords **– cybersecurity, Docker, edge computing, Kubernetes**

I. Introduction

The proliferation of edge computing and IoT has fundamentally transformed the landscape of application deployment, demanding new approaches to security, scalability, and operational efficiency. Traditional native deployments, once the default for embedded and edge systems, are increasingly being replaced or augmented by containerization technologies such as Docker and orchestration platforms like Kubernetes. This shift is motivated by the need for portability, consistency, and ease of management across heterogeneous hardware and distributed environments [1], [2], [3].

However, the adoption of these technologies in edge and IoT scenarios raises important questions about their security implications. Edge devices often operate in untrusted environments, are physically accessible to attackers, and may have limited computational resources for defense mechanisms [4], [5]. This article examines the security characteristics of native, Docker, and Kubernetes deployments.

II. Deployment Models and Security Fundamentals

A. Native Deployments

Native deployment refers to installing and running application components directly on the host operating system, without any additional abstraction or isolation layers. This approach, common in traditional embedded and IoT systems, offers minimal overhead and direct access to hardware resources. However, it also exposes the system to significant security risks. All processes share the same kernel, and a vulnerability in one component can compromise the entire system [4], [6]. Manual configuration of access controls, firewalls, and updates is required, which is error-prone and difficult to automate across a fleet of heterogeneous devices.

Native deployments on edge devices are particularly susceptible to privilege escalation attacks and lateral movement by adversaries. Since all applications and services share the same operating system kernel and user space, a single compromised process can potentially access sensitive information or disrupt other processes running on the device. This risk is amplified in IoT and edge scenarios, where physical access to devices is often possible and attackers may exploit

unpatched vulnerabilities or weak configurations [4]. Furthermore, the lack of automated update mechanisms in many native deployments increases the likelihood that known vulnerabilities will remain unpatched for extended periods.

B. Containerization with Docker

Docker provides OS-level virtualization by encapsulating applications and their dependencies into containers, which are isolated from each other using Linux namespaces and cgroups. This model reduces the attack surface by limiting the scope of compromise to individual containers, provided that containers do not run with elevated privileges [3], [7]. However, all containers share the host kernel, and kernel vulnerabilities can still lead to full system compromise. The security of Docker deployments also depends on the provenance and hygiene of container images, as malicious or outdated images can introduce vulnerabilities [8].

Recent studies have shown that container escape vulnerabilities, while rare, are a significant concern. For example, a misconfigured container with the `--privileged` flag or mounted host directories can provide an adversary with a pathway to the underlying host system [7]. Image provenance and regular vulnerability scanning are now considered best practices for Docker deployments, as the use of untrusted or outdated images can introduce critical vulnerabilities into the environment [8]. Automated tools such as Trivy and Clair are increasingly used to scan images for known CVEs prior to deployment [3].

C. Orchestration with Kubernetes

Kubernetes extends containerization by providing automated deployment, scaling, and management of containerized applications. It introduces additional security mechanisms such as Role-Based Access Control (RBAC), network policies, and secret management [6], [9]. However, Kubernetes clusters are complex to configure securely, especially in edge environments where resources are limited and operational expertise may be lacking. Misconfigurations, such as overly permissive RBAC policies or exposed APIs, are common sources of vulnerabilities [10].

Kubernetes' security model is built around the principle of least privilege, but in practice, many clusters are deployed with default or overly broad permissions for convenience. This can lead to privilege escalation, lateral movement, and data exfiltration if an attacker gains access to a compromised pod or service account [10]. Network policies are essential for restricting inter-pod communication, but are not enabled by default in many distributions. Secret management is another area of concern: while Kubernetes provides encrypted storage for secrets in etcd, improper handling or exposure through environment variables can still occur [6].

III. Security Analysis: Native vs Docker vs Kubernetes

A. Isolation and Containment

Isolation is a cornerstone of deployment security. In native deployments, process isolation relies solely on the operating system, and privilege separation is weak. A successful exploit in one application can easily escalate to system-wide compromise [4]. Docker containers improve isolation by leveraging kernel features, but the shared kernel remains a critical point of failure. Recent research shows that container escape vulnerabilities continue to be discovered and exploited in the wild [7], [3]. Kubernetes further enhances isolation by segmenting workloads into pods and namespaces, and by supporting network segmentation through policies. However, the effectiveness of these mechanisms depends on correct configuration and up-to-date software [6]. It is worth noting that the granularity of isolation in Kubernetes allows for more sophisticated segmentation of workloads, especially when combined with network policies and pod security standards. This enables organizations to implement microsegmentation, reducing the blast radius of any single compromise. However, the increased complexity of configuration and management can introduce new risks if not properly handled [10].

B. Attack Surface and Vulnerability Management

The attack surface in native deployments is typically large, as all system services and libraries are exposed. Patch management is manual and inconsistent, increasing the risk of unpatched vulnerabilities [4]. Docker reduces the attack surface by packaging only required dependencies, but the use of third-party images can reintroduce risk if images are not regularly updated or scanned for vulnerabilities [8]. Kubernetes introduces additional components (API server, etcd, controllers) that must be secured. The complexity of the control plane and the number of moving parts can lead to overlooked vulnerabilities, particularly in multi-tenant or federated edge deployments [10], [6].

The use of minimal base images, such as Alpine or Distroless, is now a recommended best practice for reducing the attack surface in containerized deployments [8]. Automated vulnerability scanning and continuous monitoring are essential for maintaining a secure environment. In Kubernetes, policy-driven vulnerability management is possible through tools such as OPA/Gatekeeper and Kyverno, which enforce compliance with security standards at the cluster level.

C. Operational Security and Resource Constraints

Edge environments impose unique operational constraints. Devices often lack hardware security modules, have limited CPU and memory, and may be deployed in physically untrusted locations [1], [5]. Native deployments offer minimal overhead but little in the way of automated security controls. Docker is lightweight and can be adapted to resource-constrained devices, but security features such as image scanning and runtime monitoring may incur performance penalties [7]. Kubernetes, while offering the richest set of security controls, has significant resource requirements, even with lightweight distributions such as K3s or KubeEdge [11], [6]. The trade-off between security and operational feasibility is a key consideration for edge deployments.

Physical security is a frequently overlooked aspect of edge deployments. Devices deployed in the field may be physically accessible to attackers, who can attempt to extract data or tamper with system integrity. Recent research recommends the use of hardware-backed root of trust, secure boot, and encrypted storage to mitigate these risks [4], [12]. In practice, however, such features are often unavailable or disabled due to cost or complexity.

D. Case Study Context: IoT Educational Panel

The IoT application described in Ilek's thesis provides a practical context for comparing these deployment models [13]. The application consists of a microcontroller-based sensor panel, a Python Flask server, a React frontend, and a Redis database. The system was deployed natively, in Docker containers, and on a local Kubernetes cluster. The evaluation included not only performance and manageability but also an assessment of the security posture in each scenario, as summarized in Table 1.

Table 1
Security Comparison of Deployment Models

Criteria	**Native**	**Docker**	**Kubernetes**
Isolation	Minimal	Moderate (container)	Strong (pod, network)
Attack Surface	Large	Reduced	Controlled, but complex
Vulnerability Management	Manual	Image-based, semi-automated	Policy-driven, automated
Resource Overhead	Low	Low–moderate	Moderate–high
Operational Complexity	Low	Moderate	High
Edge Suitability	High	High	Moderate (K3s, KubeEdge)

IV. Edge-Specific Security Challenges and Future Trends

Edge environments amplify security challenges due to hardware heterogeneity, physical accessibility, network fragmentation, and resource scarcity. Containerization mitigates these through immutable images, microsegmentation, and declarative security policies. Kubernetes enforces state during network partitions, while Docker's lightweight nature suits resource-constrained devices. However, inconsistent security capabilities across ARM/x86 architectures complicate policy enforcement, and intermittent connectivity hinders centralized monitoring. Autonomous threat detection and hardware-rooted trust mechanisms (e.g., TPMs) are critical for tamper-prone edge locations [4], [5].

Recent trends in edge security include the adoption of zero-trust architectures, confidential computing, and service mesh technologies. Zero-trust principles, such as mutual TLS (mTLS) for all service-to-service communication and strict identity management, are increasingly recommended for both containerized and orchestrated environments [12]. Confidential computing, which provides encrypted memory and secure enclaves, is emerging as a solution for protecting sensitive computations even in the event of host compromise. Service meshes such as Istio and Linkerd are being adapted for edge use cases, providing fine-grained control over traffic, security policies, and observability [12].

Lightweight Kubernetes distributions, such as K3s and KubeEdge, are specifically designed to address the resource constraints of edge environments while maintaining essential security features [11]. These platforms strip down the control plane and reduce memory and CPU requirements, making them suitable for deployment on single-board computers and other constrained devices. However, the responsibility for secure configuration and ongoing management remains with the operator, and the risk of misconfiguration is significant.

V. Discussion: Best Practices and Recommendations

Recent literature emphasizes the need for defense-in-depth strategies in containerized and orchestrated environments. For Docker, this includes using minimal base images, regularly scanning images for vulnerabilities, running containers with least privilege, and isolating sensitive workloads [8], [7], [3]. Kubernetes deployments should enforce RBAC, use network policies to restrict pod communication, and secure the control plane with strong authentication and encryption [6], [10], [9]. In edge scenarios, lightweight Kubernetes distributions (e.g., K3s) and automated configuration management tools are recommended to balance security and resource efficiency [11].

Despite these advances, challenges remain. The shared kernel in Docker remains a fundamental risk, and Kubernetes' complexity can lead to misconfiguration. Edge deployments are particularly vulnerable to physical attacks and require additional measures such as secure boot, device attestation, and encrypted storage [4], [5], [12]. Zero-trust architectures and confidential computing are emerging as promising approaches for future edge-native security [12].

VI. Conclusion

The comparative analysis demonstrates that containerization and orchestration technologies, when properly configured and maintained, offer significant security improvements over native deployments in edge and IoT contexts. However, these benefits are not automatic and require a disciplined approach to configuration, vulnerability management, and operational monitoring. Native deployments, while simple and resource-efficient, lack effective isolation and automated security controls. Docker provides a practical balance of security and performance for most edge scenarios, provided that image hygiene and kernel security are maintained. Kubernetes offers the most comprehensive security features but at the cost of increased complexity and resource consumption. The choice of deployment model should be guided by the specific security requirements, operational constraints, and threat landscape of the target environment.

Acknowledgment

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 17R05-04-V01-00004, Competence and Training Center for Cyber and Information Security TUKE.

References

[1] I. Čilić, P. Krivić, I. Podnar Žarko, and M. Kušek, "Performance evaluation of container orchestration tools in edge computing environments," *Sensors*, vol. 23, no. 8, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/8/4008

[2] B. R. Cherukuri, "Edge computing vs. cloud computing: A comparative analysis for real-time ai applications," *Int. J. Multidiscip. Res*, vol. 6, pp. 1–17, 2024.

[3] S. Gayathri, P. Nivash, and T. Manikumar, "Docker for iot edge computing," *International Journal of Creative Research Thoughts*, vol. 12, no. 4, pp. 918–927, 2024.

[4] A. Ometov, O. L. Molua, M. Komarov, and J. Nurmi, "A survey of security in cloud, edge, and fog computing," *Sensors*, vol. 22, no. 3, p. 927, 2022.

[5] L. Urblik, E. Kajati, P. Papcun, and I. Zolotová, "Containerization in edge intelligence: a review," *Electronics*, vol. 13, no. 7, p. 1335, 2024.

[6] D. Yakubov and D. Hästbacka, "Comparative analysis of lightweight kubernetes distributions for edge computing: Security, resilience and maintainability," in *Service-Oriented and Cloud Computing*, C. Pahl, A. Janes, T. Cerny, V. Lenarduzzi, and M. Esposito, Eds. Cham: Springer Nature Switzerland, 2025, p. 96–104.

[7] R. Gupta and K. Nahrstedt, "Performance characterization of containers in edge computing," 2025. [Online]. Available: https://arxiv.org/abs/2505.02082

[8] V. Ajith, T. Cyriac, C. Chavda, A. T. Kiyani, V. Chennareddy, and K. Ali, "Analyzing docker vulnerabilities through static and dynamic methods and enhancing iot security with aws iot core, cloudwatch, and guardduty," *IoT*, vol. 5, no. 3, p. 592–607, 2024.

[9] S. Koksal, F. O. Catak, and Y. Dalveren, "Flexible and lightweight mitigation framework for distributed denial-of-service attacks in container-based edge networks using kubernetes," *IEEE Access*, vol. 12, pp. 172 980–172 991, 2024.

[10] R. Hat, "Kubernetes adoption, security, and market trends report 2024," 2024. [Online]. Available: https://www.redhat.com/en/resources/kubernetes-adoption-security-market-trends-overview

[11] K. Lee, Y. Choi, and B. Tak, "Performance analysis of microvms and containers for edge computing: A focus on file and network i/o," *SSRN*, 2025. [Online]. Available: https://www.ssrn.com/abstract=5193593

[12] J. Thijsman, M. Sebrechts, F. D. Turck, and B. Volckaert, "Trusting the cloud-native edge: Remotely attested kubernetes workers," 2024. [Online]. Available: https://arxiv.org/abs/2405.10131

[13] M. Ilek, "Iot aplikácia - lokálne a cloudové riešenie s kontajnerizáciou," Ph.D. dissertation, Technical University of Košice, 2023.

Design and implementation of an automated timetable generation system using genetic and memetic algorithms

[1]Marek SABOL, [2]Michaela BAČÍKOVÁ

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

[1]marek.sabol.2@student.tuke.sk, [2]micheala.bacikova@tuke.sk

Abstract **– This paper presents the design and implementation of an automated timetable generation system for the Department of Computers and Informatics at Technical University of Košice (TUKE). The system integrates modern web technologies and evolutionary algorithms to efficiently generate and manage timetables. The system is built using a back-end developed in Python, a user interface created with React, and MySQL as the database. For generation of the timetable itself, we used genetic and memetic algorithms. The system was tested with real academic data and demonstrated high usability and accuracy in timetable generation.**

Keywords **– genetic algorithm, memetic algorithm, timetable generation, timetable**

I. Introduction

The manual creation of a university schedule is often a time-consuming task and may lead to errors, especially when dealing with complex constraints such as room availability, teacher preferences, and course requirements. As universities grow, there is a clear need for automated solutions that can handle these complex timetables more efficiently and accurately.

This paper focuses on the design and implementation of an automated timetable generation system developed specifically for the Department of Computers and Informatics at the Technical University of Košice. The system uses evolutionary algorithms, specifically genetic and memetic algorithms, to solve the problem of creating timetables. In the next sections, we describe the background of the algorithms used, how they were applied, the structure of the system, how it was implemented, and how well it worked in tests using real data.

II. Applied Algorithms

A. Genetic algorithm

According to Assi et al. [1], genetic algorithms are optimization techniques inspired by the process of natural selection. In our system, potential timetable solutions are represented as individuals within a population. Each individual encodes a complete timetable configuration.

The algorithm starts by generating an initial population of randomly created solutions. Each of the individuals is then evaluated on the basis of how well it is built. This evaluation is scored based on predefined constraints: room capacities, teacher's availability, room availability, not having lecture for same year of study at the same time as the laboratory exercise, soft constraints such as teacher's preferred days and hours of lessons, preferred rooms. The best individuals are then selected to produce the next generation using functionality crossover. This mixes segments of two solutions and includes mutation which brings a little random changes in order to keep diversity.

Schmitt's theory of mutation [2] says that if we use, for example, "00010010", mutation can happen on eight position, which changes the original to "00010011". This can happen on any bit position. The matrix m models the flipping of one bit of the genome (Fig. 1.

$$m^{(1)} = \mathbf{f} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbf{M}_2.$$

Fig. 1 The matrix describing mutation

B. Genetic Algorithm Parameterization

Proper parameterization is critical to the success and efficiency of a genetic algorithm. The most influential parameters include:

- Population size,
- Number of generations,
- Crossover and mutation probabilities.

Population Size: The optimal population size in evolutionary algorithms is largely dependent on the specific problem. According to Patil and Pawal [3], a population too small may result in insufficient genetic diversity, which can reduce the algorithm's ability to explore the solution space, increasing the risk of getting trapped in local optima. However, a population that is too large increases the computational load, which in turn may slow down the overall execution of the algorithm without significantly improving the quality of the results.

Number of Generations: Boyabatli and Sabuncuoglu [4] explain the effect of the maximum number of generations on the number of cycles that the algorithm continues to evolve the population. If the number of generations is too low, the algorithm might terminate before reaching a high-quality solution. On the other hand, too many generations may unnecessarily extend runtime without yielding significant improvements. The optimal number of generations depends on the complexity of the problem and the configuration of other parameters. In general, more complex problems with many constraints require a higher number of generations to reach convergence.

Crossover and Mutation Probabilities: Crossover and mutation are fundamental genetic operators that influence population diversity according to Hassanat et al. [5]. The commonly used settings for these parameters include a high crossover probability (up to 90%) and a relatively low mutation probability (around 3%). These values can vary depending on the problem being solved and the algorithmic context. In some implementations, dynamic adaptation strategies are employed, where the probabilities are adjusted throughout the algorithm's run to improve performance. In certain scenarios, the crossover probability may even be set to 100%, emphasizing exploration through recombination.

C. Memetic algorithm

The memetic algorithm builds on the genetic algorithm by integrating local search techniques. Rahman et al. [6] explain that, while the genetic algorithm explores globally, the memetic algorithm also refines individual solutions locally, in order to achieve faster and more accurate convergence. However, it's speed depends on the local search function. Evaluation functionality is used as well; however, takes more time in larger datasets, making the memetic algorithm slower than the regular genetic algorithm.

III. User interface design

To ensure clarity and user-friendliness, we carefully selected colors on the basis of their psychological impact and readability. The background was chosen in neutral gray, which promotes calmness and visual balance. Action elements are colored blue, while delete buttons are highlighted in red to emphasize their critical function. The colors were selected according to the theory of Ferris and Zhang [7].

The layouts and initial designs of the user interface were created in Figma, focusing on consistency between the management of teachers, rooms, and courses. The goal was to make user navigation easier and to make the system more learnable and effective.To ensure a proper generation of schedules, it is necessary to maintain structured data for teachers, rooms, and courses. Each of these has a dedicated page within the application for clear organization and easy access. To enhance usability, the interface includes instant filtering features. As the user types into a filter field, the results are dynamically narrowed down in real time. We chose colors for the units of the timetable as close as possible to the current scheduling system in the department, MAIS[1]. As can be seen in Fig. 2, lectures we chose yellow, and for laboratory exercises we chose

[1]https://mais.tuke.sk/

orange to be easily identified. MAIS is usable as a schedule filter and display and can serve as an inspiration in this case. However, it does not enable the schedule *generation* and *management* of the units of the schedule, the most innovative features of our solution. The Figma design of the tool can be seen in Fig. 3.

Fig. 2 Showcase of cards used for timetable units (classes are on the left, lectures on the right)

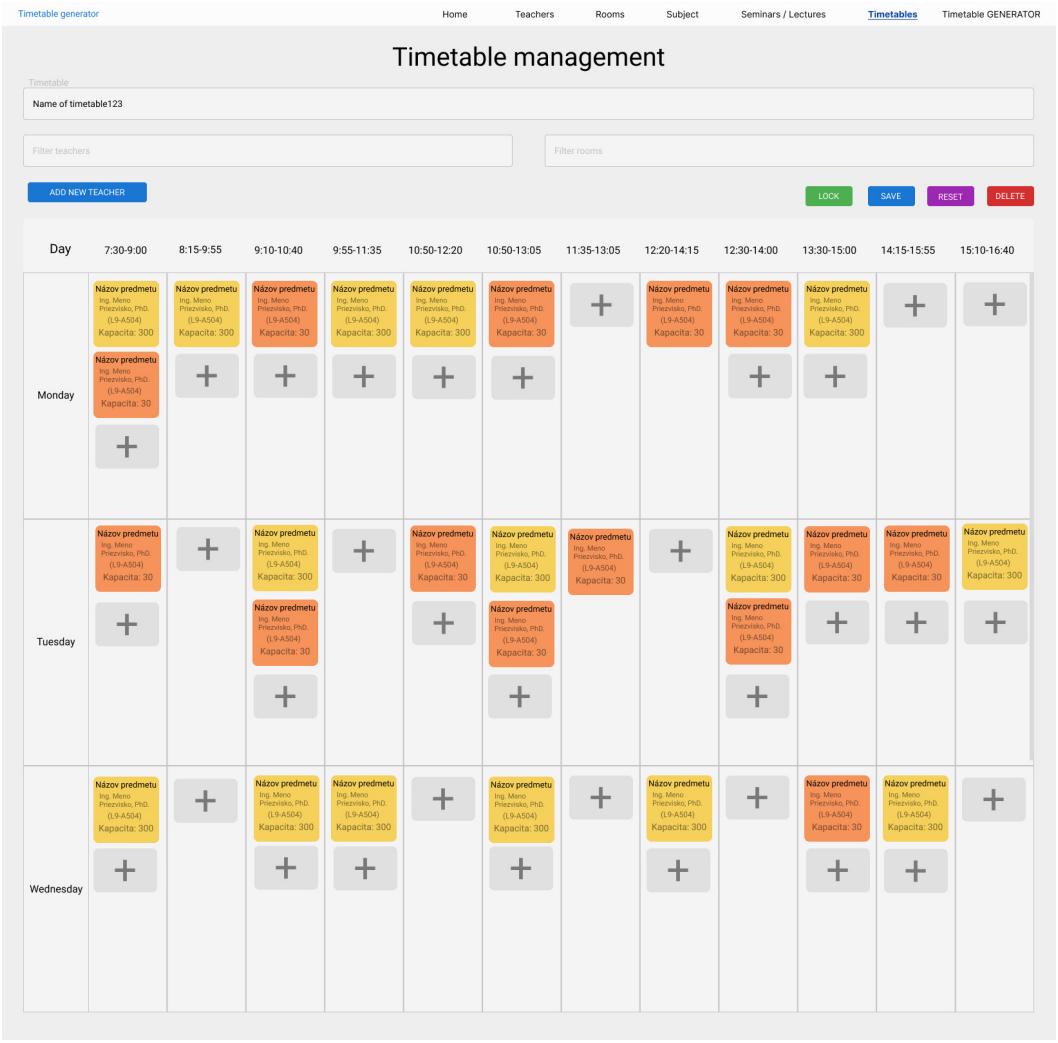


Fig. 3 Design proposal for Timetable management. Columns display time during the day and rows list days of the week.

IV. Architecture and Implementation

The system is designed with a clear separation of concerns to ensure maintainability and scalability. It consists of three core components:

- *Front-end (React)*: Responsible for displaying data and enabling user interaction. It includes views for managing teachers, rooms, courses, and the schedule itself.
- *Back-end (Python)*: Implements the application logic and handles communication between the front-end and the database via a RESTful API. Built using the Flask framework, the back-end processes requests for timetable management and generation using genetic and memetic algorithms.

- *Database Layer (MySQL)*: Stores structured data such as teacher information, room types, courses, and generated schedules.

On the back-end side, classes were used for data management, while every class holds a group of data: *Course, Teacher, Room, TimeBlock* and *TimetableSlot*. A function used for evaluation needs specific rules to calculate the fitness score for each individual. The rules were set as follows:

- Violation of soft constraints related to preferred teaching days: **-20**
- Violation of soft constraints related to preferred teaching hours: **-10**
- Teacher timetable conflict – a teacher assigned to two lessons at the same time: **-30**
- Room timetable conflict – two lessons set in the same room simultaneously: **-50**
- Overlap between a lecture and a laboratory exercises for the same year group (e.g., a lecture and an exercise for 3rd year students at the same time): **-50**
- Exceeding the max. number of hours assigned to a teacher based on their workload: **-(2x number of excess hours $\times$ 2)**
- Failure to utilize a computer-equipped room when required: **-10**
- Failure to use a preferred room: **-20**

The *genetic algorithm* will be illustrated by a simplified pseudoalgorithm in Lst. 1

Listing 1 Simplified Pseudoalgorithm of the genetic algorithm

```
initialize population
evaluate fittness
select parents
crossover
mutate
run evolution loop
return the best schedule
```

The added functionality is in the *memetic algorithm*, implemented on top of the genetic algorithm code, with several enhancements. We introduced the following additional functions:

- exploration function,
- local search,
- dynamic mutation.

The *exploration function* introduces new solutions to the population during generations. Its goal is to expand the search space and prevent premature convergence. The *local search* aims to find the best possible solution within the current generation. The *dynamic mutation* adjusts the mutation rate based on the current state of the algorithm. It helps prevent stagnation and ensures efficient exploration of all existing possibilities.

The user interface was implemented according to its designs, updated after a short qualitative user testing. The most important updates were in the schedule management. When the user opens the schedule management, the last generated schedule is automatically opened. It is also possible to choose the timetable by its name that can be set during the generation process. We implemented 3 filters for teachers, rooms, and courses. These filters do not need any submit button, the data is filtered automatically after each letter. The schedule interface enables the user to move the units by using drag&drop. Units can be placed only in a free time slot without causing a conflict with another unit.

We implemented an additional feature enabling to lock a unit illustrated in Fig. 4. When a unit is locked, it is not possible to move it, and no other unit can be dropped into its place. This feature enables the users to fix some units of the schedule and be able to safely update the rest of the schedule. When the unit is locked, it is marked with a green color. When another unit is being dragged, the slots are highlighted with red or yellow border according to their availability, to better inform the user. Locked units are marked green in the figure and their border is red to prevent the placement of the other unit in the same place.

V. Testing and evaluation

To test both algorithms, we have set the data as close to reality as possible. We have run 100 tests for both algorithms. Both algorithms were 100% successful in creating schedules without any errors. The algorithms were able to create enough timetable units for large capacity courses and also to work with teachers' preferences. The *genetic algorithm* was able to find a solution in average 67.31 seconds, and it took 42.44 generations in average. *Memetic algorithm* was a

Fig. 4 A showcase of locked timetable units

little slower: 111 seconds and 36.55 generations in average. The memetic algorithm took longer because of the complexity of our evaluation function that is called multiple times in this type of algorithm.

We chose two most average generations from both algorithms and tried to compare their divergention. The results are illustrated by Fig. 5 and 6. The red line in the graphs represents the average fittness score and the blue line represents the best fittness score.

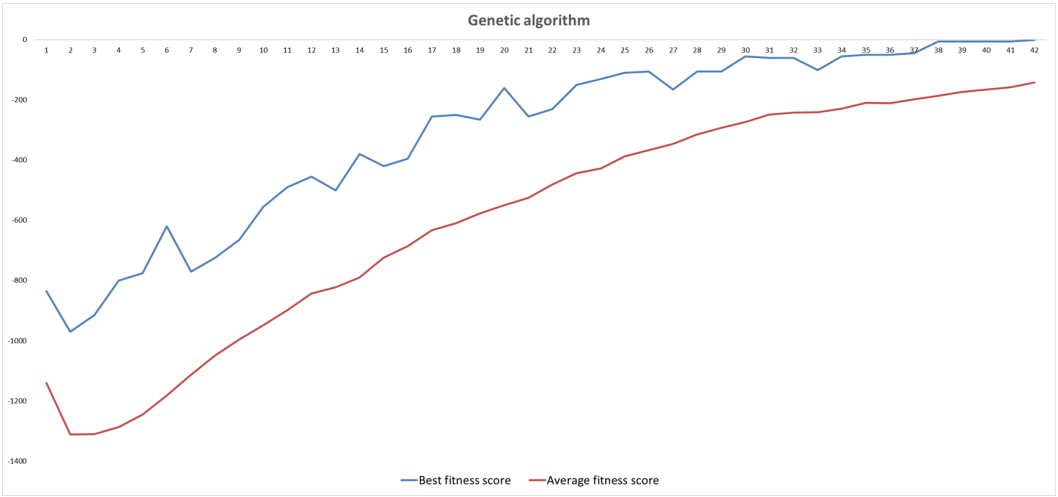


Fig. 5 Graph generation by genetic algorithm

The user experience was evaluated by qualitative user testing and a structured interview. System of Usability Scale (SUS) [8] with standard 10 questions has been used to formally evaluate the usability of the interface. The test participant was an end user, i.e. an employee of Department of Computers and Informatics at TUKE responsible for creating schedules. A short scenario has been defined for the employee to test the system in order to make the process easier.

In the structured interview, we asked her about her current process of creating a schedule and about her opinion on the tool. Her current work consists of creating tables in *.doc for each room. For any update, she needs to copy&paste the units inside the complicated document. Such a process is highly error-prone. She described the new tool as very effective and easy to use. She thought it would be a significant help in her work. The calculated SUS score was 85%, which we can say is very positive.

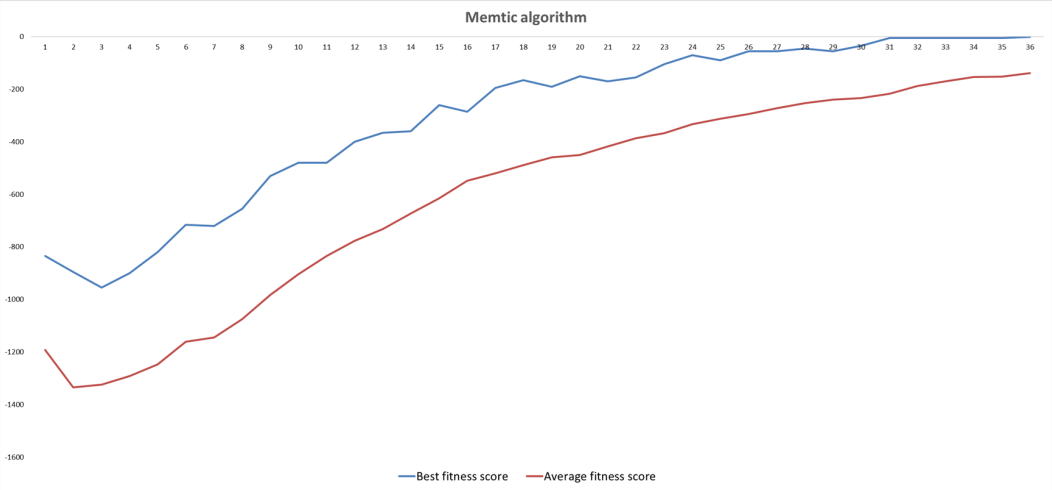


Fig. 6 Graph generation by memetic algorithm

VI. Conclusion

In this paper, we presented the design and implementation of an automated university schedule generation system. The tool was tailored for the Department of Computers and Informatics at TUKE but it can be considered a universal tool for any university as well.

Our system architecture composed of a React front-end, Python back-end, and a MySQL database ensures scalability, maintainability, and ease of use. Through a careful user interface design and real-time interaction features, such as filtering and drag&drop capabilities, the system significantly simplifies the creation and management of schedules for university administrators. Testing with real academic data confirmed the effectiveness of both implemented algorithms. Both approaches, genetic and memetic, produced valid and accurate timetables within acceptable timeframes. The user testing confirmed that the front-end is usable and user-friendly.

Acknowledgement

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-23-0408.

References

[1] M. Assi, B. Halawi, and R. A. Haraty, "Genetic algorithm analysis using the graph coloring method for solving the university timetable problem," *Procedia Computer Science*, vol. 126, pp. 899–906, 2018.
[2] L. M. Schmitt, "Theory of genetic algorithms," *Theoretical Computer Science*, vol. 259, no. 1-2, pp. 1–61, 2001.
[3] V. Patil and D. Pawar, "The optimal crossover or mutation rates in genetic algorithm: A review," *International Journal of Applied Engineering and Technology*, vol. 5, no. 3, pp. 38–41, 2015.
[4] O. Boyabatli and I. Sabuncuoglu, "Parameter selection in genetic algorithms," *Journal of Systemics, Cybernetics and Informatics*, vol. 4, no. 2, p. 78, 2004.
[5] A. Hassanat, K. Almohammadi, E. Alkafaween, E. Abunawas, A. Hammouri, and V. S. Prasath, "Choosing mutation and crossover ratios for genetic algorithms—a review with a new dynamic approach," *Information*, vol. 10, no. 12, p. 390, 2019.
[6] H. F. Rahman, R. K. Chakrabortty, and M. J. Ryan, "Memetic algorithm for solving resource constrained project scheduling problems," *Automation in Construction*, vol. 111, p. 103052, 2020.
[7] K. Ferris and S. Zhang, "A framework for selecting and optimizing color scheme in web design," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2016, pp. 532–541.
[8] J. Brooke, "Sus: A "quick and dirty" usability scale," in *UsabilityEvaluationin Industry*. Taylor and Francis, 1996, pp. 189–194.

Japanese Pronunciation Component for Self-Learners

[1]*Raman LISTAREVICH,* [2]*Michaela BAČÍKOVÁ*

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

[1]raman.listarevich@student.tuke.sk, [2]michaela.bacikova@tuke.sk

Abstract **– This paper presents the design, implementation and experimental evaluation of a web-based component for automatic Japanese pronunciation assessment, specifically tailored to self-learners. The goal is to enable users to receive immediate, intelligible feedback on their pronunciation without the need for a language instructor. Following an analysis of existing solutions (including commercial APIs like Azure Speech) that revealed insufficient accuracy and lack of detailed feedback on specific pronunciation errors, we developed a custom multi-level evaluation system. The system integrates speech segmentation at the word and phoneme levels, visual analysis of pronunciation via approximated decibel curves, and correlation-based scoring using Dynamic Time Warping (DTW) alignment and Pearson's correlation coefficient. We hypothesized that this multi-level, feedback-rich approach would more effectively improve learners' pronunciation than conventional one-dimensional scoring methods. The system was tested in a series of user experiments, which confirmed its ability to accurately identify and localize pronunciation errors while significantly supporting the learning process. A comparative study showed that our approach yields greater pronunciation improvement than baseline solutions, and a user experience evaluation demonstrated that the tool meets the practical needs of novice learners. These findings validate the effectiveness of the proposed system and indicate its promise as a foundation for next-generation pronunciation training tools, potentially extendable to other languages.**

Keywords **– pronunciation assessment, Japanese, feedback, phoneme analysis, Dynamic Time Warping (DTW), User Experience (UX), machine learning, speech-to-text**

I. Introduction

Pronunciation is a critical component of language proficiency, yet self-learners often struggle to improve their pronunciation without expert feedback. In the context of learning Japanese—a language with subtleties in phonemic length, pitch accent, and rhythm—mispronunciation can hinder communication [1]. Traditionally, learners rely on native speakers or instructors for correction, but such guidance may not be readily available in self-study settings. This gap has motivated the development of computer-assisted pronunciation training (CAPT) tools that can provide automated feedback. However, existing solutions have notable limitations. Mainstream speech evaluation APIs and commercial pronunciation apps often provide only an overall pronunciation score or limited feedback, lacking the precision to pinpoint specific pronunciation errors. Their accuracy in evaluating non-native speech may be insufficient for nuanced guidance, and they seldom offer insight into where and what to improve in an utterance. Consequently, learners might receive a generic score without understanding whether their issue lies in a particular sound, timing, or intonation pattern.

To address these limitations, we propose a multi-level pronunciation assessment system designed for Japanese self-learners. The core idea is to combine fine-grained phonetic analysis with intuitive feedback visualizations. Instead of treating an utterance as a monolithic whole, our system evaluates pronunciation at multiple levels - the phrase, each word in the phrase, and each phoneme in the words. At each level, it analyzes different aspects of pronunciation quality and provides specific, actionable feedback. For example, the system can highlight if the error is in the duration of a vowel in a particular word or a pitch drop at the end of the phrase. We achieve this by integrating several techniques: speech segmentation to align the learners' speech with a reference, calculation of timing deviations for fluency, and acoustic similarity measures using DTW and Pearson correlation to compare the learners' audio with the reference model.

By visualizing the results, the component gives learners a clear picture of their pronunciation performance.

Our central hypothesis is that this rich feedback approach will lead to more effective learning - in other words, a custom multi-level pronunciation assessment system can significantly improve pronunciation outcomes for self-learners of Japanese, compared to one-dimensional scoring methods. In this paper, we detail the system's methodology and an experimental evaluation of its effectiveness. We conducted user studies to validate whether the system's feedback indeed helps learners improve and to compare its performance against alternative solutions (including Azure's API and SpeechSuper). We also assessed the user experience to ensure the tool is accessible and engaging to its target users. The remainder of the paper is organized as follows: The Methodology section describes the system design and experimental setup. The Results section presents quantitative findings from two experiments and qualitative observations. In the Discussion, we interpret the results, examine the hypothesis in light of the evidence, and outline the system's contributions and practical implications. Finally, the Conclusion summarizes the work and suggests future directions.

II. Methodology

Our pronunciation assessment system was developed as a web-based interactive tool designed to provide real-time, multi-level feedback to learners of Japanese. Users are presented with a Japanese phrase or word and are prompted to record their own pronunciation. A native-speaker reference recording is provided as the target. The system then compares the user's utterance against this reference using a series of segmented analyses and acoustic similarity measures.

A. Segmentation and Alignment

Both the reference and user utterances are segmented for analysis. The system first aligns the input at the word level, then further segments each word into phonemes using the known transcription and a pronunciation lexicon for Japanese. This allows for direct comparison of corresponding segments between the user and the reference at multiple levels of granularity.

B. Pronunciation Metrics

Three core metrics are computed to assess pronunciation:

- **Tempo (Timing Accuracy)**: This metric measures the relative duration of spoken segments. For each word and phoneme, the system calculates the deviation in duration from the reference. A normalized score from 0 to 1 is assigned, where 1.0 indicates perfect temporal alignment. This helps identify whether the learner is speaking too quickly, too slowly, or elongating specific sounds.
- **Fluency (Rhythmic Stability)**: Fluency is defined as the internal consistency of phoneme timing within a word. It is computed as the average relative timing deviation across all phonemes in a word compared to the reference. A high fluency score suggests smooth, uninterrupted pronunciation, while lower scores typically indicate hesitations, irregular pacing, or unnatural pauses.
- **Correlation (Acoustic Similarity)**: To assess acoustic similarity independent of timing and volume, we use a Dynamic Time Warping (DTW)-assisted Pearson correlation. After non-linear alignment via DTW, we compute the correlation coefficient between the aligned acoustic features (e.g., loudness envelopes, spectral curves) of the user and reference. This metric, reported as a percentage, reflects how closely the learners' audio matches the reference in quality, regardless of pace.

C. Multi-Level Feedback Architecture

These metrics are integrated into a hierarchical feedback system that operates at three levels:

- **Phrase-Level Feedback**: The system provides an overall pronunciation score for the entire phrase, alongside aggregate tempo, fluency, and correlation scores.
- **Word-Level Feedback**: Each word in the phrase is scored individually. Words with significantly lower scores are highlighted, allowing learners to identify specific problematic regions.
- **Phoneme-Level Feedback**: Within each word, the system analyzes individual phonemes, identifying the most divergent segments and explaining the nature of the deviation (e.g., overly long vowel or unclear consonant). The feedback includes visual comparisons of loudness curves between user and reference for each phoneme, with deviation points marked.

D. User Interface and Visualization

The user interface was designed with a strong emphasis on clarity, interactivity, and immediate interpretability of feedback. It is implemented as a web application using the `React` framework and supports real-time communication with the evaluation backend through REST APIs.

Feedback is structured across three hierarchical levels: phrase, word, and phoneme. After each pronunciation attempt, the system displays a transcript of the target phrase, where each word is color-coded based on its pronunciation accuracy. The color scheme (e.g., green for correct, orange for moderate deviation, red for significant error) enables users to quickly identify problematic regions without reading numerical data.

At the word level, users can expand a detailed view that shows individual phoneme evaluations. Each phoneme is also color-coded and supplemented with numeric scores for tempo, fluency, and correlation. This allows learners to pinpoint whether issues stem from articulation timing, rhythm irregularities, or acoustic mismatch with the reference.

The visualizations are rendered using the `Recharts` library and dynamically update in response to user interactions. Hover actions reveal exact timing and correlation values, while clicking on a word or phoneme replays the corresponding audio segment. This bidirectional link between visualization and playback supports a tight feedback loop, helping users correlate what they hear with what they see.

Additional interface components include:

- A pronunciation score summary bar for the whole phrase, displayed prominently after each attempt;
- Graphical representations of DTW path alignment and correlation distribution across the utterance;
- Tooltip-based explanations for metrics and color codes to support novice users;
- A mode-switching panel that allows users to toggle between high-level and detailed views without losing context.

E. Implementation Details

The pronunciation evaluation system was implemented as a modular web-based application built on a microservice architecture. The backend and frontend are connected via a RESTful API, enabling flexible integration, scalability, and maintainability.

Server-side: The server logic is implemented in `.NET` and interacts with both external services and internal components. Audio input is recorded in the browser and sent to the server, where it is processed in near real-time. Speech transcription is performed using IBM Speech-to-Text (STT) via the official IBM Cloud SDK. The STT module returns both the recognized text and word-level timestamps.

Based on this transcription, the server identifies high-level pronunciation errors:

- *Insertions* – extra words not present in the reference,
- *Omissions* – missing words,
- *Substitutions* – incorrect words.

The system computes initial metrics such as completeness (word match rate) and tempo (relative duration deviation). Each word is then isolated as a standalone `.wav` file, preserving high-quality audio for further analysis.

Phoneme-level segmentation is performed using the `Julius Speech Segmentation Kit`, triggered via command-line Perl scripts. Input text is automatically converted into hiragana to comply with the format required by Julius. The segmentation output is a `.lab` file containing phoneme timestamps and labels, which forms the basis for low-level acoustic analysis.

The use of Dynamic Time Warping (DTW) to align time-series data and capture acoustic variation has been previously validated in speech processing as competitive with human-level perception under certain conditions [2]. To support intuitive feedback, amplitude contours are approximated with Chebyshev polynomials and stored as aligned DTW paths. All results, including metrics and visualization data, are returned as a structured `.json` file. The server exposes dedicated REST endpoints for each evaluation stage, ensuring compatibility with existing systems and supporting future extensibility.

Client-side: The frontend is built using the `React` framework. It includes an interactive feedback component that guides the user through pronunciation exercises. Key features include:

- Phrase and word selection or random generation from a predefined library;
- Audio playback of native reference and user recording;
- Dynamic switching between phrase, word, and phoneme-level analysis;
- Visualization of decibel curves, correlation scores, and error localization using the `Recharts` library;

- Integration with a dictionary and help modules developed in the initial prototype.

The user interface emphasizes clarity and responsiveness, enabling learners to see, understand, and correct their pronunciation mistakes in an interactive manner. All visual and numerical feedback is synchronized across hierarchical levels, helping users internalize pronunciation patterns and improve over repeated interactions. The importance of perceptual feedback and user engagement in pronunciation tools has been echoed in recent studies evaluating AI-driven speech interfaces for learners [3].

Unlike conventional deep learning-based approaches that require extensive labeled datasets and training cycles [4], this system operates without reliance on large-scale pre-trained models. Instead, it enables dynamic evaluation of arbitrary vocabulary by directly comparing user speech to reference recordings, allowing for domain-agnostic extensibility and lightweight deployment.

III. Experimental Setup

We conducted two primary experiments to evaluate the effectiveness of our pronunciation feedback system and compare it to baseline approaches. All participants were adult Slovak native speakers with no prior knowledge of Japanese, representing typical beginner self-learners.

- **Experiment 1: Comparative Learning Efficacy Study**
  This controlled study assessed pronunciation improvement using three different systems. Nine university students (ages 20-23, technical background, no prior Japanese or phonetics training) were randomly assigned to three groups:
  - Group A used a prototype based on Azure Speech API (representing commercial-style feedback),
  - Group B used SpeechSuper (a deep-learning-based pronunciation scoring demo),
  - Group C used our developed multi-level feedback system.

  Each participant completed three training sessions (10-15 minutes each) using their assigned system. Session content consisted of randomized beginner-level Japanese words and phrases (e.g., greetings, common terms). Participants received only the feedback provided by their assigned system (e.g., numeric score or detailed multi-level feedback) and could repeat items freely within the session.
  We recorded two primary metrics: the average pronunciation score per session (normalized to a percentage scale across systems) and the relative improvement from the first to the final session. Additionally, we measured system reliability by tracking the proportion of failed pronunciation evaluations (e.g., unrecognized input or missing score).
- **Experiment 2: Intensive Session Learning Curve Study**
  This experiment analyzed short-term learning behavior during a single extended session. Four new participants completed a one-time training with our system, practicing a fixed set of 65 beginner-level Japanese items delivered sequentially via a web interface.
  A minimum pronunciation score of 65% was required to pass each item. Participants could retry an item up to five times; after five failed attempts, the item was skipped automatically. This threshold-based design encouraged immediate correction and adaptation.
  Each session lasted approximately 40-50 minutes. Collected metrics included average pronunciation scores (perceptual tempo, fluency, phoneme correlation), average time per item, number of attempts, and number of skipped items—analyzed separately for the first and second halves of the session. We also recorded observational notes on participant engagement and behavioral changes across the session.
- **User Experience Assessment**
  To evaluate usability, we conducted a separate study using the System Usability Scale (SUS). Twelve additional participants (not involved in previous experiments) tested the system by analyzing a Japanese phrase without prior instruction, then completed the SUS questionnaire. Participants also answered open-ended questions on aspects they found confusing or helpful. This qualitative feedback was used to identify interface issues (e.g., difficulty interpreting phoneme-level visualizations) and assess the perceived usefulness of the feedback mechanisms.

A. Results of Experiment 1 - Comparative Study

The multi-session comparative experiment evaluated pronunciation improvement and system effectiveness across three feedback systems: Azure Speech API, SpeechSuper, and our custom multi-level pronunciation evaluation system. Each system was tested across three sessions by distinct participant groups.

Despite starting with the **lowest initial average score** (69.9), our system enabled the **highest relative improvement** across sessions, reaching 81.6 in Session 3 – a gain of **+15.8%**. In contrast,

Azure API users improved from 76.1 to 81.9 (+7.6%), and SpeechSuper users from 82.5 to 85.8 (+4.0%). These results, summarized in Table 1, confirm that the detailed, structured feedback provided by our system leads to more substantial learning gains, even when starting from a lower performance baseline.

Error rates (failed recognition attempts) also varied. SpeechSuper had the lowest average error rate (approx. 13%), followed by Azure API (reduced from 28% to 23%), and our system (from 30% to 27%). Although our approach initially showed slightly higher error rates, it remained stable across sessions, indicating acceptable robustness for real-world use.

These outcomes suggest that learners using our system not only improved faster but also benefited from more pedagogically effective guidance. The structured, multi-level feedback appears to better support learners in identifying and correcting specific pronunciation issues than the simpler, aggregated scoring provided by baseline systems.

Table 1
Experiment 1 Results - Evaluation of Learning Rate and Effectiveness

System / Session	Average Score	Relative Gain	Recognition Errors (%)
Azure Speech API			
Session 1	76.1	—	28%
Session 2	78.0	+2.5%	26%
Session 3	81.9	+7.6%	23%
SpeechSuper			
Session 1	82.5	—	12%
Session 2	83.4	+1.1%	14%
Session 3	85.8	+4.0%	13%
Our Algorithm			
Session 1	69.9	—	30%
Session 2	74.2	+6.1%	29%
Session 3	81.6	+15.8%	27%

B. Results of Experiment 2 - Intensive Session

In the intensive single-session training using our system, participants demonstrated measurable within-session improvements, confirming the effectiveness of real-time, multi-level feedback.

For analysis, the session was split into two halves: the first 32 phrases and the remaining 33. During the first half, participants were still adapting to the feedback mechanism. The average tempo score was 83, fluency 71, and correlation 84. On average, participants skipped 4.25 items due to the five-attempt rule and required 1.8 attempts to complete an item successfully. This phase took approximately 25 minutes. These metrics are detailed in Table 2.

In the second half (phrases 33–65), performance improved across all recorded metrics. Tempo rose to 89, fluency to 72, and correlation to 85. Skipped items dropped to 2.4, and average attempts decreased to 1.6 per item. Participants completed this part in about 21 minutes, despite a higher number of items. A summary is presented in Table 3.

These improvements indicate growing familiarity with both the pronunciation challenges and the feedback interface. The reduction in skipped items and attempts per item demonstrates improved efficiency and confidence. Learners were able to correct timing and articulation issues more accurately on subsequent tries, suggesting rapid skill acquisition.

Qualitative observations supported this trend. Users appeared more comfortable interacting with phoneme-level feedback in the second half and often responded with targeted corrections in rhythm and duration without needing external guidance.

In summary, the experiment confirmed that our system facilitates not only progressive improvement across sessions, but also immediate learning gains within a single focused session. These results validate its applicability for both ongoing study and high-intensity pronunciation practice.

Table 2
Results of Experiment with Repetition of Basic Word and Phrase Set
(First 32 Phrases)

Indicator	Value
Average time	25 minutes
Average number of skipped phrases	4.25
Average tempo score	83
Average fluency score	71
Average correlation score	84
Average number of attempts per success	1.8

Table 3
Results of Experiment with Repetition of Basic Word and Phrase Set
(Phrases 33–65)

Indicator	Value
Average time	21 minutes
Average number of skipped phrases	2.4
Average tempo score	89
Average fluency score	72
Average correlation score	85
Average number of attempts per success	1.6

C. Results of experiments with User Experience and Feedback

The usability test and participant feedback provided insight into how learners perceived the system. The overall System Usability Scale (SUS) score was 72.1, which is above the commonly accepted 68-point threshold for "acceptable" usability and within the average range for consumer software. According to standard SUS interpretations, a score in the low 70s suggests the system is fairly easy to use and without major usability flaws (users were generally able to complete tasks and understand the interface). Indeed, most participants navigated through the multi-level feedback (phrase/word/phoneme views) without needing assistance, indicating that the interface design was intuitive in terms of basic navigation. They especially appreciated the clear separation of feedback levels and the ability to drill down into more detail on demand. In open-ended feedback, a majority of users reported that the visual feedback and structured scores were helpful and motivating. For example, several users noted that seeing color-coded highlights on the specific parts of words where they went wrong made it easier to focus their practice. One user mentioned that "the ability to switch between the phrase overview and individual word analysis was extremely useful", as it gave a big picture and a detailed view as needed - this aligns with other comments, where the multi-level navigation was seen as a major advantage of our system. The correlation score in particular was cited as a favorite feature at the word level; users interpreted it as a clear indicator of how close they sounded to the native pronunciation of that word. The numeric scoring with explanatory labels (e.g. fluency, tempo) provided a sense of tangible goals to work toward, which users found more informative than a single opaque score.

IV. Discussion

The results support our hypothesis: a custom multi-level pronunciation assessment system significantly improves Japanese pronunciation for self-learners. Users of our system showed greater improvement than those using Azure's API or SpeechSuper, likely due to our system's detailed and specific feedback. Instead of a single score, it highlights issues like vowel length, timing, and fluency, allowing learners to focus on concrete problems. This led to more effective self-correction, unlike the generic scores from baseline systems. Compared to deep-learning systems like SpeechSuper, our explainable metrics (DTW, correlation) gave clearer feedback. Black-box models may assess well but don't teach. Our rule-based system, though simpler, aligns with aspects of speech learners can understand and act on.

Limitations include small sample size and short training periods. A longer-term study could verify sustained improvement. We also lacked human rater validation, though our metrics are standard proxies. Future improvements include faster DTW processing, better phoneme alignment, and selective use of ML for classifying common errors — provided interpretability is maintained. The system is adaptable to other languages. The multi-level feedback framework is language-agnostic; only the phoneme set and error types change. Applied to tonal or phonetically different languages, it may require new visualizations but can follow the same principles.

V. Acknowledgement

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-23-0408.

References

[1] H. Kubozono, Ed., *Handbook of Japanese Phonetics and Phonology*. Berlin, München, Boston: De Gruyter Mouton, 2015. [Online]. Available: https://doi.org/10.1515/9781614511984

[2] N. Minematsu, Y. Fujisawa, and S. Nakagawa, "Performance comparison among hmm, dtw, and human abilities in terms of identifying stress patterns of word utterances," in *6th International Conference on Spoken Language Processing (ICSLP 2000)*, 2000, pp. vol. 1, 617–620.

[3] E. A.-R. Amin, "Efl students' perception of using ai text-to-speech apps in learning pronunciation," *Migration Letters*, vol. 21, no. 3, p. 887–903, Jan. 2024. [Online]. Available: https://migrationletters.com/index.php/ml/article/view/6905

[4] M. Haque, A. Verma, J. S. R. Alex, and N. Venkatesan, *Experimental Evaluation of CNN Architecture for Speech Recognition*, 01 2020, pp. 507–514.

Effective Collection, Structuring and Analysis of Medical Data

[1]Michal POPOVIČ, [2]Michaela BAČÍKOVÁ

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

[1]michal.popovic@student.tuke.sk, [2]michaela.bacikova@tuke.sk

Abstract **– This paper presents the design and implementation of a system for the effective collection, storage and visualization of medical data, specifically blood oxygenation and heart rate values. The solution combines the MAX30100 and MAX30102 sensors with an ESP8266 microcontroller that provides wireless data transmission to a web server built on NestJS and PostgreSQL technologies. The goal was to create a functional data chain with potential for use in machine learning. In the practical part and with the help of experiments, the obtained data were compared with the outputs of a certified oximeter. The MAX30102 sensor showed high instability, while the MAX30100 provided more consistent results and was able to capture physiological trends over time. Tests were also performed to verify the reliability of the sensors under various conditions such as finger movement, sensor movement, change in pressure, and others.**

Keywords **– blood oxygen saturation, heart rate, medical data, medical sensors, programming**

I. Introduction

Acquisition and analysis of medical data is an increasingly important part of modern healthcare. Vavrinsky et al. [1] explain their use for diagnosis and treatment on a very common holter sensors that represent a perfect example. Digital technologies and wearable devices, such as smart watches and fitness bracelets, are revolutionizing health monitoring. These devices can measure a wide range of body data including heart rate, physical activity levels, quality of sleep, and other health indicators. Despite these technological advances, there is a significant gap in the ability of these devices to effectively structure and analyze the data collected, limiting their usefulness in clinical practice.

The main problem is that existing wearable devices, although capable of collecting large amounts of data, cannot effectively structure and analyze these data in a way that makes them easily accessible and interpretable for both clinicians and patients. As a consequence, quality and comprehensive information that would enable accurate monitoring and diagnostic of the health status of patients is not available to physicians.

The aim of this paper is to develop a prototype system for collecting, analyzing and storing medical sensor data, which will consist of hardware and software and will allow easy access, visualization, and management of the data via a web application. The system is intended to streamline patient-physician communication, improve the quality of health monitoring, and promote better information in clinical practice.

II. Analysis and Solution Proposal

The analysis and design of the solution represented a key phase of the system development, during which the requirements, technical constraints and possibilities of integration of individual components were identified. The goal was to design an efficient and reliable architecture that would ensure seamless collection, processing and visualization of medical data. Therefore, this section addressed the selection of appropriate hardware and software technologies, the design of the data flow, the method of communication between devices, and the definition of a database structure for data storage and management.

A. *Medical data security*

Medical data security encompasses four key phases in its lifecycle: collection, transformation, analysis and knowledge creation. As the authors Abouelmehdi, Beni-Hessane et al. [2] inform, it is already important to ensure the trustworthiness of sources, protection from unauthorized access, and anonymization of data at the collection stage. The best practices of the privacy and security of big medical data in healthcare were described by Shah and Callahan [3] and Boulemehdi et al. [4].

In this work, we have followed the above principles in designing the solution, taking care to secure the data at every stage of its processing. In transformation, filtering and classification prepare the data for further processing without compromising its integrity. Once stored in a secure repository, this is followed by analysis, where careful anonymization is necessary to avoid the disclosure of personal data. In the last stage, knowledge creation, it is necessary to ensure the protection of sensitive information that can be used by healthcare professionals in decision-making.

B. *Types of sensors used in wearables*

In the field of wearable devices, a wide range of sensors are used to monitor various physiological parameters. These include for example: **mechanical**, **electrical**, **optical** or **chemical**.

These sensors provide the wearable devices with a number of functions that are necessary for the collection of individual data.

Mechanical sensors are divided into piezoresistive, capacitive, iontronic and piezoelectric. Heikenfeld et al. [5] state that piezoresistive ones respond to mechanical deformation, allowing tracking of movements or pulse. Capacitive sensors detect pressure changes through the capacitive response. Iontronic sensors work with changes in the electrical properties of the skin and are characterized by high sensitivity. Piezoelectric sensors generate an electrical charge in response to mechanical stress, which is useful for monitoring vibration or heart rate.

Electrical sensors measure changes in the electrical resistivity or conductivity of the skin. They are used in recording biopotentials such as ECG, EEG or EMG, and require good electrical contact with the skin for a good signal.

Optical sensors use light sources and detectors to sense signals from the body based on the absorption and scattering of light in tissues. They are key for measuring heart rate, blood oxygenation or blood flow, often based on photoplethysmography [1].

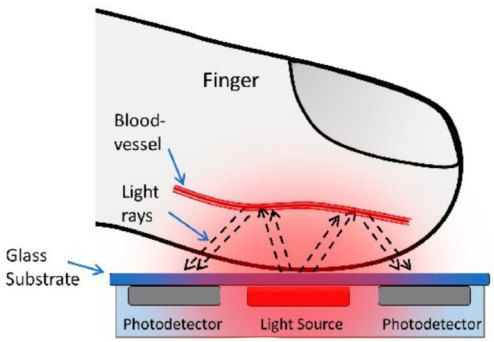


Fig. 1 Optical sensor data collection principle [6]

.

In (Fig. 1) a demonstration of the functionality of the optical sensor is shown. Elsamnah et al. [6] present a method for measuring pulse frequency using the photoplethysmographic principle.

Chemical sensors analyse body fluids such as sweat, saliva or tears in order to detect metabolites, electrolytes and other chemically active substances that provide important information about the physiological state.

C. *Comparison of control platforms*

In this work, we have analyzed three popular control platforms suitable for medical data collection projects: *Arduino, Raspberry Pi and ESP8266.* Each of these platforms has its own advantages and limitations that need to be considered when designing a wearable device.

[1] A non-invasive optical method used to measure blood volume changes in peripheral parts of the body, typically the fingers, wrists or ears.

Arduino is a simple, energy-efficient platform with the possibility of expansion with Wi-Fi modules, suitable for basic monitoring tasks. *Raspberry Pi* offers higher performance and advanced encryption, but its larger size makes it less suitable for wearable devices. The *ESP8266* is a compact microcontroller with integrated Wi-Fi and encryption support, ideal for small wireless applications.

We considered four main criteria when selecting a suitable platform: connectivity, security level, compactness and ease of implementation. A comparison of these features is shown in Table 1.

Table 1
Comparison of Arduino, Raspberry Pi, and ESP8266

Feature	Arduino	Raspberry Pi	ESP8266
Connectivity	Add-on modules	Wi-Fi and Ethernet	Wi-Fi
Security	Limited	Robust encryption	SSL/TLS
Compactness	Requires modules	Larger	Very compact
Ease of Implementation	Simple	More complex, Raspbian OS	Similar to Arduino

A comparison of these units showed that each has its strengths and weaknesses strengths in terms of connectivity, data security, compactness and ease of implementation. From the analysis performed, it is concluded that the ESP8266 control unit is the most suitable platform to solve the medical data acquisition problem. Thanks to its capabilities and compact size, it perfectly meets the requirements of a wearable device communicating wirelessly with the corresponding server.

D. MAX30100 and MAX30102 sensors comparison

The MAX30102 and MAX30100 sensors are integrated biosensor modules designed to measure pulse oximetry and heart rate. Both devices use a similar principle based on infrared and red LEDs, but differ in selected technical parameters that affect their use in specific applications. Table 2 provides a comparison of the main features of both sensors[2].

Table 2
Comparison of Specifications: MAX30102 vs. MAX30100

Specification	MAX30102	MAX30100
Supply Voltage	1.8V to 3.3V	1.8V to 3.3V
Standby Power Consumption	0.7µA	1.0µA
Active Mode Power Consumption	600µA	650µA
Interface	I2C-compatible	I2C-compatible
Operating Temperature Range	-40°C to +85°C	-40°C to +85°C
Sampling Frequency	50Hz to 3200Hz	50Hz to 1000Hz
RED LED Wavelength	660nm	660nm
IR LED Wavelength	880nm	880nm

E. System solution design

This section describes the design of a system solution for efficient medical data acquisition. The solution consists of three main entities.

The first entity (patient) involves data acquisition using MAX30102 or MAX30100 sensors connected to the ESP8266 controller. Heart rate and blood oxygenation data are then wirelessly sent to a local server.

The second entity (system) represents the local server that receives, processes and stores the data in the database. Communication with the server is done via TCP/IP protocol and a simple HTTP-based API.

The third entity (physician) uses a web application to visualize, analyze and manage the data, thus providing efficient access to health information.

III. Implementation

The *NestJS* framework was used to implement the server side, which allows for a modular and scalable architecture. *PostgreSQL* was chosen as the database system, while communication with the database was implemented through *Prisma ORM*, which ensured simple and type-safe data operations. The database part of the server application was containerized using *Docker*, allowing for easy deployment and isolation of the database.

[2]Based on official datasheets.

The web application was created in *AngularJS* and is used to visualize the measured values. For this purpose, charts created using the *Chart.js* library and the *PrimeNG* component library were used, providing a responsive and user-friendly interface.

Communication between the *ESP8266* microcontroller and the server, as well as between the server and the web application, was handled using *API* connections using the HTTP protocol.

IV. Experiments and evaluation

This section focuses on evaluating the accuracy of the *MAX30100* and *MAX30102* sensors in measuring blood oxygenation (SpO2) and heart rate. The aim was to verify how reliably these sensors record medical data under different measurement conditions. Measurements were taken that recorded values at short and long term intervals. Also, ancillary influences on the measurement were evaluated, such as:

- finger movement, sensor movement, skin conditions and others

Measured sensor data were compared with reference values obtained from a certified medical pulse oximeter. Several methods were used to evaluate the accuracy and agreement between measurements, such as calculating the mean difference between values or **Pearson correlation**.

A. Comparison with certified medical oximeter

The key was to compare the accuracy of the measurements obtained with the sensors MAX30100 and
MAX30102 with data from a certified medical oximeter. Testing included both **short-term** and **long-term** measurements, looking at the accuracy of of the values obtained at different time intervals.

1) Max30100 sensor: Short-term measurements: Each measurement took approximately 3 minutes, during which the sensor data was recorded at three points (start, middle and end of the measurement) in parallel, a value was recorded using a medical oximeter. A total of 30 measurements were performed in different physiological and body states.

The measurement results show that the MAX30100 sensor systematically recorded slightly lower blood oxygen saturation (SpO2) values compared to the medical oximeter. The mean values measured by the MAX30100 sensor ranged from $96\%-97$ %, while the medical oximeter recorded values predominantly of 99 %, occasionally 98 %.

A similar trend was observed for heart rate, where the MAX30100 showed values approximately 4 beats per minute lower than the medical oximeter. Table 3 shows fragments of the measurements.

Table 3
MAX30100 and certified oximeter measurement data

Č. m	TF MAX30100	TF Oximeter	SpO2 MAX30100	SpO2 Oximeter
1	77.2	80	96.0	99
5	97.5	101	95.2	99
10	95.3	98	97.0	99
15	97.3	102	96.4	99
20	103.1	107	95.6	99
25	103.7	109	97.5	99
30	107.8	110	96.4	99

To verify the consistency of the measurements, the **Pearson correlation** [7] was also calculated between the values obtained by the MAX30100 sensor and the certified medical oximeter.

$$r = \frac{\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum \left(X_i - \overline{X}\right)^2 \sum \left(Y_i - \overline{Y}\right)^2}}$$

Where:

- X_i, Y_i - values of variables X a Y,
- $\overline{X}, \overline{Y}$ - average values X a Y,

The correlation coefficient for heart rate *r = 0.99* indicates a strong positive correlation, which is also reflected in the plot (Figure 2).

The correlation coefficient for SpO2 *r = -0.13*, on the other hand, indicates that the measured values were not dependent on each other. This phenomenon is supported by the fact that the

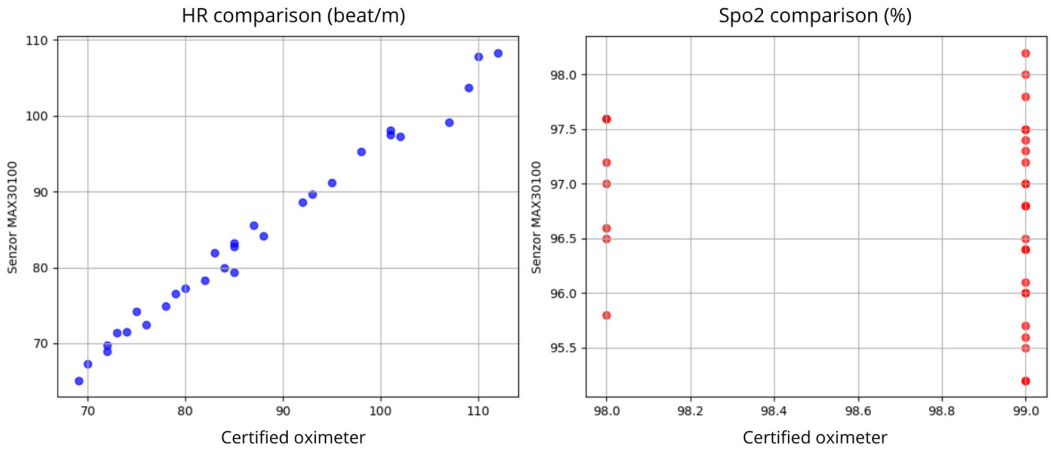


Fig. 2 Measurements comparison MAX30100/Certifikovaný Oximeter

values measured by the certified oximeter were predominantly 99%, while the values measured by the sensor fluctuated slightly around the values 96-97%.

These results show that although the MAX30100 sensor exhibits slight deviations from the certified device, its outputs are consistent and allow reliable monitoring of physiological parameters.

2) Max30100 sensor: long-term measurements: Measurements were taken after more strenuous physical activity, with the aim of tracking the decline in heart rate over time as the heart rate gradually returns to a resting state. After physical exertion, the heart rate was was higher at the start of the measurement and gradually decreased, reflecting the physiological process of calming the body after increased physical exertion.

The results are visualized in the graph 3, which shows the progression of the measured heart rate values from both devices.

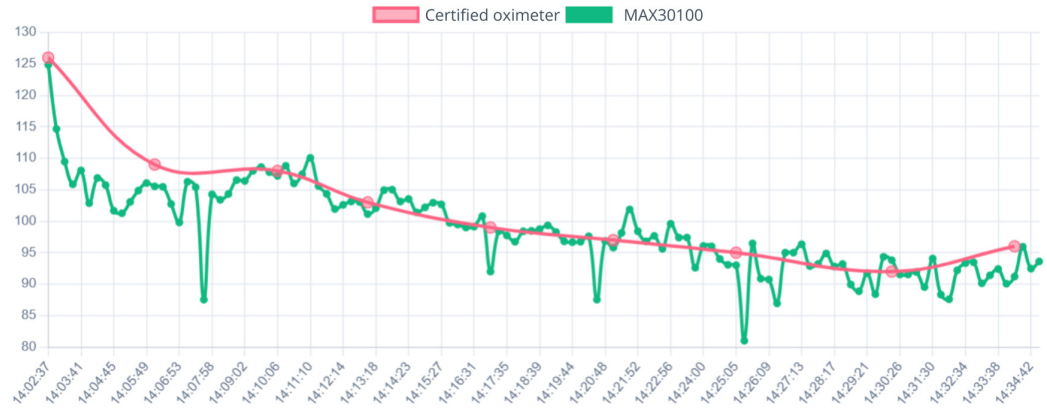


Fig. 3 Heart rate long-term measurement comparison

The measurement results show that the readings from the MAX30100 sensor reflected the values from the certified oximeter. As with the short-term measurements, the sensor systematically slightly underestimated the results compared to the medical oximeter.

The graph shows the short-term fluctuations in the readings from the sensor, which are due to sensitivity to movement or change in finger pressure. However, these fluctuations do not affect the long-term trend, and the average values remain comparable to the certified oximeter.

3) Sensor MAX30102: During testing of the **MAX30100** and **MAX30102** sensors, the accuracy and stability of the measurements were verified. However, the MAX30102 sensor exhibited significant errors - saturation values were unrealistic and heart rate held steady at an unrealistic $214u/m$. These results indicate unreliable signal processing and an inability of the sensor to correctly detect cardiac activity.

B. Stability and accuracy of sensor measurements

Testing of sensors for measuring physiological parameters has shown that the accuracy and stability of the measured values are strongly influenced by factors such as the pressure of the

attached finger, its position, movement during measurement and the condition of the skin. Even small fluctuations in finger pressure or orientation caused sudden variations in the data, with a return to stable values requiring the acquisition of multiple new data samples. The graph (fig. 4) shows the effect of finger movements on heart rate and blood oxygenation measurements, where significant fluctuations appeared after each movement, which later stabilized.

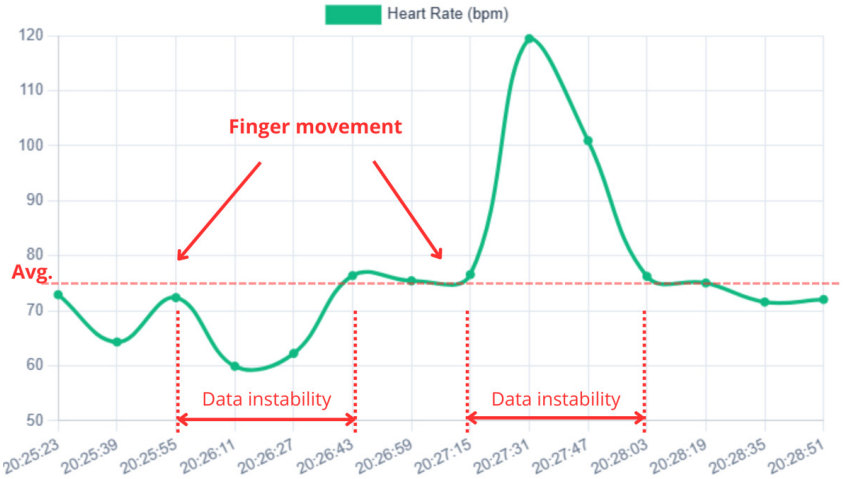


Fig. 4 Signal instability caused by finger movement (HR)

The signal quality was also reduced by biological factors such as humidity, sweat or skin pollution, which change the optical properties of the surface. Equally problematic was the high sensitivity of the sensors to movement and shocks; even slight movement of the sensor caused degradation of the data. These findings highlight the need to ensure stable and controlled conditions during measurement: uniform pressure, minimal movement, dry and clean skin surface.

V. Conclusion

Practical testing revealed significant differences between the sensors. The MAX30102 proved unsuitable for medical use due to inaccurate and unstable data. MAX30100 on the other hand provided consistently reliable heart rate and blood oxygenation measurements. While short-term MAX30100 readings slightly underestimated values compared to a certified oximeter, the sensor accurately captured trends and dynamics, as evidenced by high correlation. Long-term measurements further confirmed its ability to track physiological changes over time, mirroring the readings of the reference device.

Tests also highlighted the MAX30100's sensitivity to external factors, such as movement, position changes, and skin conditions (e.g. moisture, dirt). These elements can negatively impact measurement quality and must be considered when designing wearable devices.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-23-0408. The authors thank Siemens co. for professional domain and technical support during our work.

References

[1] J. D. M. W. A. Z. T. S. H. VAVRINSKY, Erik; SUBJAK, "Application of modern multisensor holter in diagnosis and treatment." *Sensors*, vol. 20, no. 9, p. 22, 2020.
[2] K. Abouelmehdi, A. Beni-Hessane, and H. Khaloufi, "Big healthcare data: preserving security and privacy," *Journal of big data*, vol. 5, no. 1, pp. 1–18, 2018.
[3] N. H. Shah and A. Callahan, "Machine learning in healthcare," in *Key Advances in Clinical Informatics*. Academic Press, 2017, pp. 279–291.
[4] A. K. H. BOUELMEHDI, Karim; BENI-HESSANE, "Big healthcare data: preserving security and privacy," *Journal of big data*, vol. 5, no. 1, p. 18, 2018.
[5] J. Heikenfeld, A. Jajack, J. Rogers, P. Gutruf, L. Tian, T. Pan, R. Li, M. Khine, J. Kim, and J. Wang, "Wearable sensors: modalities, challenges, and prospects," *Lab on a Chip*, vol. 18, no. 2, pp. 217–248, 2018.
[6] F. Elsamnah, A. Bilgaiyan, M. Affiq, C.-H. Shim, H. Ishidai, and R. Hattori, "Reflectance-based organic pulse meter sensor for wireless monitoring of photoplethysmogram signal," *Biosensors*, vol. 9, no. 3, p. 87, 2019.
[7] statstutor.ac.uk, "Statistic support for students at www.statstutor.ac.uk, statistical education through problem solving, pearson's correlation coefficient," last edited: 2004-01-19 by E. Crawford, Accessed: 2025-04-14. [Online]. Available: https://www.statstutor.ac.uk/resources/uploaded/pearsons.pdf

Web Touch Component for Self-Learners of Japanese Hiragana

[1]*Marek BARNOVSKÝ,* [2]*Michaela BAČÍKOVÁ*

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

[1]marek.barnovsky@student.tuke.sk, [2]michaela.bacikova@tuke.sk

Abstract **– This paper presents the design and implementation of a web-based handwriting learning component for self-learners. The tool was developed as an interactive touch interface with adaptability to various script types. Japanese hiragana was chosen as the model script due to its lower familiarity and suitability for testing. As part of the work, an experiment was conducted comparing learning on paper with learning in a digital environment. The results showed higher success rates, fewer attempts, and shorter testing times in the digital group. It was also confirmed that skills acquired through drawing on a touchscreen can be transferred to traditional handwriting. The web application was developed using the React library and is publicly available online.**

Keywords **– handwriting learning, Hiragana, self-learning, handwriting comparison algorithm**

I. Introduction

The acquisition of handwriting skills is a complex process involving the integration of motor control, visual perception, and cognitive memory. Although it may appear simple, handwriting requires the development of fine motor patterns and a clear understanding of stroke sequences. According to a study by Nonaka [1], even early learners of hiragana must overcome significant motor and perceptual challenges to master basic writing skills.

To address these challenges, this study presents a modular web-based handwriting learning component, designed primarily for self-directed learners. The system provides adaptive feedback, animated stroke guidance, and a progression model based on user performance. While the tool currently focuses on hiragana—a phonetic script unfamiliar to most non-native learners—it is built to be extensible to other writing systems. As highlighted by Rose [2], the unique structure of the Japanese writing system makes it particularly suitable for testing strategies in character acquisition.

The component leverages touchscreen technology to capture and evaluate user-drawn strokes in real time. Stroke accuracy is assessed using a custom algorithm that combines spatial and temporal metrics. This interactive approach facilitates deeper engagement and retention. In line with findings by Yamamoto [3], interactive mobile systems have been shown to enhance learner motivation and autonomy in digital environments.

The main goal of this work is to develop an accessible handwriting learning system that offers personalized support, encourages consistent practice, and effectively bridges the gap between digital training and traditional handwriting. In particular, the study aims to verify the hypothesis that characters learned through interactive digital training can be accurately reproduced by users using traditional pen-and-paper writing. This addresses the crucial question of whether motor skills acquired in a virtual environment can successfully transfer to real-world contexts.

II. Methodology

The system is built around a custom *stroke comparison algorithm* designed to evaluate the accuracy of user-drawn characters against predefined reference templates. The algorithm provides immediate visual feedback and drives an adaptive learning mechanism that adjusts difficulty based on user performance.

A. *Stroke Comparison Process*

The comparison process integrates both *geometric* and *temporal (trajectory-based)* analysis to assess stroke similarity in terms of shape and order.

- In the *temporal phase*, each stroke is *resampled* into a sequence of evenly spaced points, preserving the stroke's original direction and segmentation.
- In the *geometric phase*, *pairwise Euclidean distances* are computed between corresponding points of the user stroke and the reference stroke.

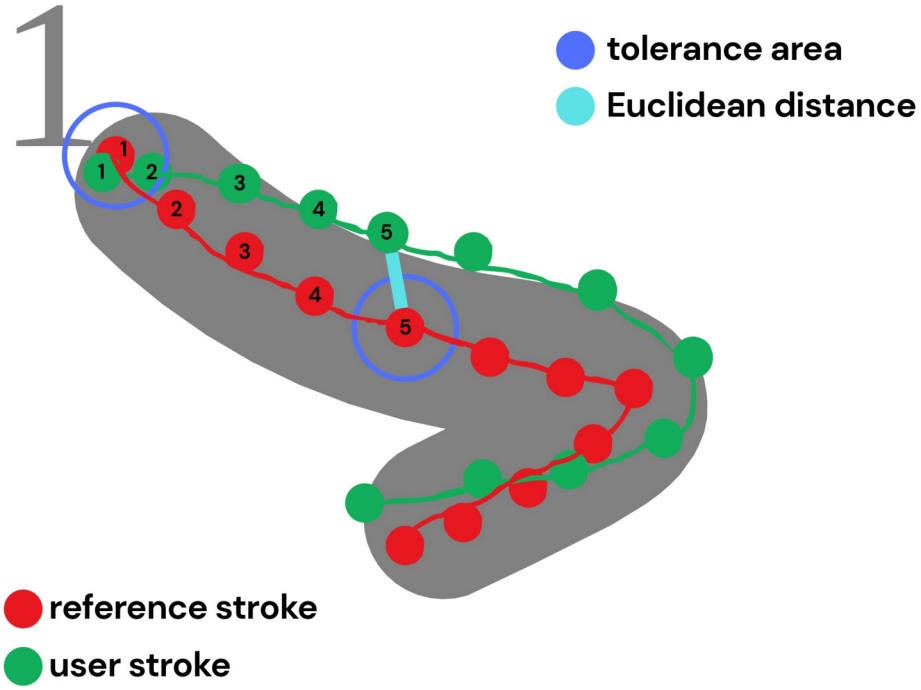


Fig. 1 Visualization of the stroke comparison process.

A key concept in the evaluation is the *tolerance area*, which defines an acceptable margin of deviation. If the Euclidean distance between two corresponding points is within this predefined *threshold*, the point is accepted as a perfect match and does not contribute to the penalty. If the distance exceeds the threshold, the distance is included in the accumulated penalty.

The final *similarity score* is expressed as a percentage and is computed as follows:

$$D = \sum_{i=1}^{n} \delta\left(\|u_i - r_i\|\right) \tag{1}$$

$$\delta(d) = \begin{cases} 0, & \text{if } d < T \\ d, & \text{if } d \geq T \end{cases} \tag{2}$$

$$D_{\max} = n \cdot C \tag{3}$$

$$\text{Similarity} = \left(1 - \frac{D}{D_{\max}}\right) \cdot 100 \tag{4}$$

Where:

- u_i and r_i are the i-th user and reference points, respectively.
- T is the *tolerance threshold*.
- C is a constant representing the *maximum allowed deviation*.
- n is the total number of compared points.
- D is the accumulated penalty for points outside the tolerance area.

Preprocessing and Adaptivity

To ensure robust and consistent comparison, the reference templates are preprocessed using *adaptive subdivision of Bézier curves* [4], resulting in a uniform distribution of sampling points even in complex shapes. Both user and reference strokes are then normalized to contain the same number of points.

The algorithm is governed by three key parameters: the *tolerance radius*, the *distance threshold*, and the *maximum deviation constant*. These values are dynamically adjusted based on the user's current learning level and performance. Repeated errors can trigger a reduction in difficulty, reinforcing learning through repetition—a core idea of *adaptive learning models*.

To identify optimal settings for stroke evaluation, a dedicated experimental interface was developed, allowing interactive testing of algorithm parameters. More than 100 experiments were conducted with various parameter combinations. Based on qualitative and quantitative results, five configurations were selected and verified through user testing.

The selected configurations were designed to balance sensitivity and user experience. Excessively strict settings were avoided to prevent user frustration, while overly lenient configurations were refined to maintain evaluation integrity. Among the tested parameters, the `tolerance` value used for adaptive Bézier subdivision was kept constant across all modes. A value of 0.03 was chosen as an optimal compromise between comparison precision and performance efficiency.

The five final configurations—ranging from lenient to strict—are shown in Table 1.

Table 1
Stroke comparison algorithm settings for predefined modes

Mode	maxPossibleDistanceConstant	distanceThreshold
EASY	35	8
E-M	35	6
MEDIUM	30	5
M-H	30	4
HARD	25	3

The *MEDIUM* configuration was selected as the default. Other modes allow for increasing or decreasing strictness based on user skill level.

B. Touch-Based Web Component

The interactive learning system was implemented as a modular web application using the *React* framework and designed as a *Progressive Web App (PWA)*. This architecture ensures cross-platform compatibility, offline functionality, and the option to install the app directly on mobile devices.

At the core of the interface lies a *Canvas component*, which captures user-drawn strokes via touch or mouse input and overlays them onto a reference character displayed in the background. The canvas supports undo functionality, automatic scaling, and dynamic visual cues such as stroke endpoints and direction markers, which are progressively removed in later stages using a *fading method* [5] to encourage memory-based reproduction.

The system uses a single *SVG template* per character, representing stroke paths and order in a resolution-independent format. Rather than storing multiple versions, the application dynamically modifies these templates to suit each learning level — hiding or simplifying strokes, animating stroke order, or showing only key points. This flexible SVG handling streamlines asset management while enabling consistent visual guidance across phases.

Each character is composed of individual strokes, and user input is evaluated in real time. Correct strokes are highlighted in green, incorrect ones in red, with optional animated replays offered to support correction. In testing phases, all assistance is disabled to simulate traditional assessments.

Beyond stroke rendering, the application includes components for navigating between learning and testing modes, selecting characters, and viewing performance summaries. An internal state management system monitors user progress and adaptively adjusts difficulty: repeated mistakes trigger increased guidance, while consistent accuracy allows users to advance.

Together, the real-time canvas interaction, dynamic SVG rendering, and adaptive feedback create an effective and responsive environment for acquiring handwriting skills digitally.

III. Experimental Design

To evaluate the effectiveness and usability of the proposed handwriting learning system, two complementary approaches were used: a controlled learning experiment and a standardized System Usability Scale (SUS) survey.

Both experiments were conducted using *Google Forms*, which allowed participants to submit structured responses efficiently and consistently. The collected data were then exported and analyzed in *Microsoft Excel* using basic statistical methods, including averages, medians, and standard deviations. This workflow ensured accurate data handling while supporting transparency and reproducibility in the evaluation process.

Participant Groups and Learning Setup

The experiment was conducted with two groups, each consisting of ten participants. The first group (*Group A*) used the interactive, web-based handwriting component on touchscreen devices. The second group (*Group B*) followed a traditional paper-based approach using worksheets specifically created for this study.

These worksheets were carefully designed to replicate the progression logic of the digital system. In particular, they implemented the *fading method* [5], which gradually reduces visual support (e.g., stroke paths and direction indicators) across successive writing attempts. This ensured that both groups experienced a comparable pedagogical structure, despite differences in interaction modality.

Participants were tasked with learning to write two Japanese hiragana characters—*(A)* and *(O)*—which were selected due to their different visual structures and similar stroke counts, making them suitable for evaluating learning effectiveness across varied character types.

Experimental Phases

The study consisted of three structured phases:

a) Learning Phase: Group A participants used the digital interface, which provided dynamic feedback, animated stroke guidance, and real-time evaluation. Group B completed the same tasks on paper, progressing through manually prepared worksheets simulating the digital fading-based structure.

b) Testing Phase: All participants, regardless of their learning method, were asked to reproduce both characters from memory using the digital system. This uniform testing condition allowed direct comparison of performance under identical evaluation criteria.

c) Transfer Phase: Only Group A participants (digital learners) were additionally asked to reproduce the characters on paper. This step was intended to test the study's central hypothesis: whether motor skills and stroke sequences acquired via a touchscreen interface can successfully transfer to traditional handwriting with pen and paper.

Anticipated Challenges

Before the experiment, several potential confounding variables were considered. These included:

- differences in prior exposure to Japanese characters,
- variations in fine motor skill development,
- individual motivation levels,
- and usability challenges linked to unfamiliarity with the interface.

To mitigate these risks, instructions were standardized, all participants were given equal time for learning, and usability was separately assessed using the SUS questionnaire.

IV. Evaluation and Results

To evaluate perceived usability, the System Usability Scale was administered to an initial group of 30 participants aged between 20 and 40. Two responses were excluded due to uniform maximum scores combined with vague open-ended feedback, suggesting inattentive participation. The final dataset thus included 28 valid responses.

The average SUS score was *85.80*, with a median of 87.5 and a standard deviation of 8.77. Scores ranged from 67.5 to 97.5. These results place the system well within the "excellent" usability range. Participants described the interface as intuitive, responsive, and effective for independent learning. Such results support the system's potential use in autonomous educational contexts.

Collected Metrics

During testing, several key metrics were collected:

- *Similarity score* — automatically calculated as a percentage based on geometric stroke matching.
- *Time spent* — measuring efficiency in both learning and testing phases.
- *Number of attempts* — indicating how many repetitions were needed to achieve correct reproduction.

Comparative Results

The digital group outperformed the paper group across most metrics. Participants learning through the interactive system reached higher accuracy with fewer attempts and in shorter time. Figure 2 shows a boxplot comparison of success rates between the two groups.

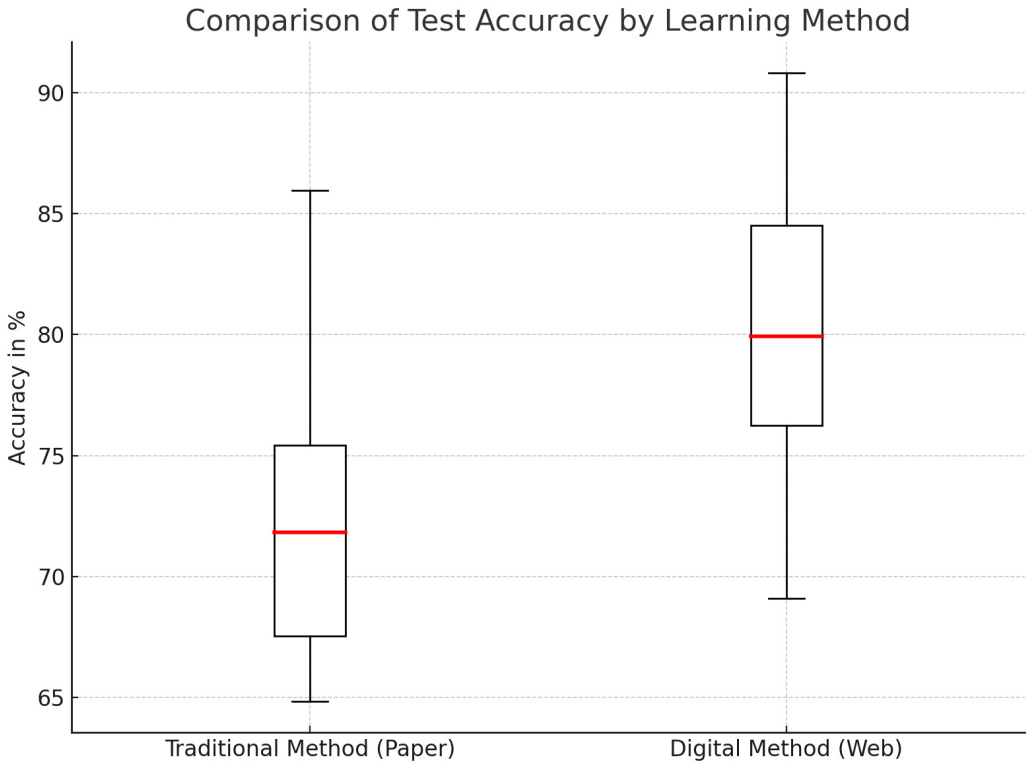


Fig. 2 Comparison of character reproduction success rate between traditional (paper) and digital (web) learning groups.

The digital group achieved a median success rate of *80%*, with most participants scoring above 75%. The paper group showed a lower median of *72%* and greater performance variance. These results suggest that the digital interface promoted not only more accurate but also more consistent learning outcomes.

Hypothesis Confirmation and Additional Observations

The study confirmed the central hypothesis: *users trained via the digital component were able to accurately reproduce the characters on paper*. All Group A participants successfully wrote both characters by hand, validating the transferability of acquired motor skills from digital to traditional writing environments.

While the overall results were consistent, several additional observations provide valuable insights into the usability and pedagogical effectiveness of the system. Notably, participants from Group B—who had trained solely with paper—were exposed to the digital handwriting component for the first time only during the testing phase. This lack of prior interaction may have temporarily affected their performance, particularly in terms of navigating the canvas interface or understanding the layout of the drawing space. Despite this, they still achieved relatively strong results, indicating that the system is sufficiently intuitive even for first-time users with no prior digital handwriting experience.

A particularly illustrative case involved a participant who originally trained on character *(A)* using only the paper-based method. Due to a misunderstanding of the testing instructions, this participant initially skipped the test phase and instead decided to repeat the learning process for both characters using the web-based system. This offered a unique opportunity to compare the same learner's experience across both methods.

In the paper-based learning, the participant spent 5 minutes and 10 seconds mastering character A. However, during digital learning, it took them 10 minutes and 5 seconds to reach the same outcome. This extended duration was a direct result of the system's real-time feedback repeatedly flagging an incorrectly drawn stroke—an issue that had gone unnoticed in the paper version. This highlights a key advantage of the digital system: its ability to *immediately identify and correct mistakes*, thereby preventing incorrect motor patterns from being reinforced.

Interestingly, the opposite trend was observed for character *(O)*. In this case, the participant spent 6 minutes and 1 second during the paper-based training, but only 1 minute and 44 seconds when learning digitally. While previous exposure likely played a role in accelerating performance, the result also demonstrates the system's *adaptive capacity*: it does not hinder progress when the user performs correctly, allowing smooth advancement without unnecessary repetition. This dual behavior—targeted intervention when needed, and non-intrusive progression otherwise—underscores the pedagogical advantage of an interactive, feedback-driven environment.

Taken together, these observations reinforce the value of digital handwriting tools not only in improving accuracy and retention, but also in offering structured, individualized guidance that paper-based methods cannot provide.

V. Conclusion

This article presented a modular, web-based handwriting learning component optimized for self-directed acquisition of Japanese hiragana. Developed as a Progressive Web App using React, the system integrates a real-time stroke comparison algorithm, adaptive difficulty settings, and dynamic SVG rendering within a touchscreen-friendly interface.

The component was evaluated through a controlled experiment comparing it to traditional paper-based learning. Results demonstrated that users of the digital system achieved higher accuracy, required fewer attempts, and completed tasks in less time. The study also confirmed that skills acquired in the digital environment successfully transferred to handwriting on paper, validating the system's pedagogical effectiveness.

Notable user cases further highlighted the advantages of real-time feedback and adaptive guidance, emphasizing the tool's potential in correcting errors and supporting autonomous progression. Future work will focus on expanding script support, integrating learning analytics, and refining personalization features to enhance long-term learning outcomes.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-23-0408.

The authors thank all participants who contributed to the experimental evaluation of the system. Their participation was essential in validating the results and improving the design of the learning component. We also acknowledge the support and feedback provided during the development phase, which helped refine both the user interface and the evaluation methodology.

References

[1] T. Nonaka, "Cultural entrainment of motor skill development: Learning to write hiragana in japanese primary school," *Developmental Psychobiology*, vol. 59, no. 6, pp. 749–766, 2017.

[2] H. Rose, *The Japanese writing system: Challenges, strategies and self-regulation for learning kanji*. Multilingual Matters, 2017, vol. 116.

[3] N. Yamamoto, "An interactive learning system using smartphone: Improving students' learning motivation and self-learning," in *2014 Ninth International Conference on Broadband and Wireless Computing, Communication and Applications*, 2014, pp. 428–431.

[4] M. Demchenko, "Adaptive subdivision of bezier curves," https://agg.sourceforge.net/antigrain.com/research/adaptive_bezier/, accessed: 2025-03-29.

[5] H. Heuer and J. Lüttgen, "Motor learning with fading and growing haptic guidance," *Experimental brain research*, vol. 232, no. 7, pp. 2229–2242, 2014.

Web Touch Component for Self-Learners of Japanese Katakana

[1]*Andrea KANÓCZOVÁ,* [2]*Michaela BAČÍKOVÁ*

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

[1]andrea.kanoczova@student.tuke.sk, [2]michaela.bacikova@tuke.sk

Abstract **– Learning Japanese Katakana presents several challenges, including its limited appearance in everyday life in countries outside of Asia. The visual similarity of Katakana characters and the need for regular practice present a challenge. In this paper, we describe a novel web touch component for self-learners that enables the users to learn the characters by drilling each character on a canvas. The component adapts to the user's current knowledge by gradually removing visual aids from the canvas, making the component adaptable to the user. When designing the component, we focused on usability and overall user experience.**

Keywords **– Application for self-learners, Katakana, touch user interface, writing assessment**

I. Introduction

The Japanese alphabets present challenges for learners whose native language uses a phonemic alphabet, particularly within the Indo-European language group [1]. Writing difficulties can be so significant that they lead to demotivation or even abandonment of language learning [2], specifically in case of self-learners. Reading is essential for learning, and the "real Japanese" always comes after learning the two syllabic alphabets (Hiragana and Katakana, both 46 characters) and one large semantic alphabet (Kanji, 2,136 characters for daily life).

One challenging factor is that, despite being essential for Japanese literacy [3], Katakana is relatively less present in academic settings compared to Hiragana and Kanji. Learners also struggle with the visual similarity of certain Katakana characters, making them difficult to remember and distinguish [4], emphasizing the need for regular practice.

It is a known fact that *drilling characters helps us to learn any alphabet effectively*. It is possible on paper; however, it presents a challenge for the self-learners, since without a native speaker's feedback, it is hard to learn the proper writing. The touch component presented in this work supports frequent engagement with Katakana, with the aim of improving writing skills and supporting self-learning via drilling. The advantage, as opposed to standard paper and pen, is the automatic recognition of correct and incorrect strokes and feedback on correct or incorrect writing. Moreover, it is adaptive, removing visual aids gradually according to the user's progress.

II. Related Work

Writing was chosen as the primary method on the basis of research showing its strong link to memory retention. Longcamp et al. [5] found that handwriting activates sensori-motor brain areas, creating durable cerebral representations of letters that aid in recognition and long-term recall. Similarly, James et al. [6] showed that handwriting training, whether tracing or freehand, significantly improves letter learning in preliterate children. Participants in their study showed higher brain activity than those using keyboards. This aligns with the "Theory of Embodied Cognition" by Kiefer and Trumpp [7], which suggests that active sensory-motor engagement, like handwriting, enhances and accelerates learning more effectively than passive methods such as typing or rote memorization.

III. Design and Implementation

To successfully implement writing as a teaching method, it was necessary to verify the accuracy of handwriting. An offline handwriting recognition approach based on geometric similarity was adopted, using the Hausdorff Distance (HD) metric, calculated as follows:

$$h(A, B) = d_B(A) + d_A(B) \tag{1}$$

In the formula stated above, the HD is calculated by first finding the shortest distance from each point $a \in A$ (user stroke coordinates) to all points $b \in B$ (template stroke). The greatest of these minimum distances (i.e., the supremum of the infima) is then taken as the final result [8].

The core user scenario involves handwriting practice. A character is first displayed, its coordinates are extracted, and visual style modifications provide writing cues. The user input for each character stroke is then captured and analyzed for geometric similarity to the template. If the final stroke meets the similarity threshold, the lesson is marked as complete.

IV. The First Iteration Prototype

To support handwriting input, a touch component for free-hand entry was implemented as part of a native mobile application, enabling intuitive haptic interaction and convenient access. The application supports all the interactions needed for tracing-based learning, although its graphic design remains minimal. It includes a navigation panel, a tab panel to switch between related features, and a central writing workspace. The *home screen* offers access to training exercises grouped into lessons. The *exercise screen* includes a touch-enabling canvas, a Latin phonetic transcription of the character, and an audio playback button for pronunciation. To guide users, the correct stroke order and shape are visually indicated: blue lines mark the current stroke, while completed ones appear in black. Once a stroke is drawn correctly, the previously inactive *Continue* button turns blue, signaling progress.

Japanese characters are transcribed using the *Hepburn system* [9], which aligns closely with Slovak-to-English pronunciation, making it intuitive for users.

A. Learning Methods Comparison Experiment

The prototype was tested in an experiment that compared character writing training using the touch component with a traditional pen-and-paper method (PaP). Two equally sized groups (*"app" group* and *"PaP" group*) were given the same task: train a set of 16 Katakana vowels (*a, e, i, o, u*) for a fixed time at set intervals. The goal was to evaluate the effectiveness of learning and user experience. The participants were six persons having no previous knowledge of the Japanese language. Drawing on similar experiments by Bonneton et al. [10] and Blumenthal [11], the following hypotheses were formulated:

- There will be no significant differences between the groups in their ability to reproduce the characters, as touch writing is expected to yield high learning success.
- The digital component will lead to more consistent training.
- Users will perceive the app as more comfortable for practice.
- The immediate feedback provided by the app may be distracting and require adjustment.

After the training period, participants completed a final test, recalling and writing the characters into a guiding grid on a testing list based on a phonetic prompt.

When comparing the groups, the app users had fewer issues with the guiding grid, despite its slightly different size from the training version. However, they all reported not paying much attention to it during training and focusing only on the template, which may explain the reduced precision in character reproduction.

Both groups found the number of trained characters appropriate, with only one person preferring less and one preferring more characters, but suggested adding more varied exercises. The PaP group requested more challenging tasks to boost engagement, while the app group felt that repeated tracing became monotonous.

While PaP participants recalled the characters, they struggled with proportions and orientation without a reference. The app group, on the other hand, struggled and once even failed to recall characters and none attempted to write from memory, citing low confidence. Two also reported difficulty in concentrating during practice.

These findings suggest that increasing the difficulty of tracing tasks and adding exercises that reinforce Katakana-Latin associations could improve engagement and retention.

The negative reaction to character mixing in the final test suggests that grouping similar characters during practice may be more effective. However, this observation relates more to teaching methodology than to the touch component itself.

Participants often formed associations between Katakana and Latin letters (4/6), which is not always helpful, as multiple Katakana characters can resemble the same Latin letter, or none at all. A more effective strategy may be to focus on the differences and similarities between the Katakana characters themselves.

Only one participant likened characters to objects (e.g. a spider, a needle), but did not use them for memorization. Another participant mentioned that they might have used object associations with a different character set.

The initial hypotheses were only partially supported. Both groups found their learning methods comfortable. The app group appreciated the flexibility of mobile practice, but their results suggest a lower focus during learning. Adherence to the schedule was not an issue and no one reported the application's feedback as distracting.

B. User Testing

At the end of this prototype iteration, the six participants involved in the learning methods experiment evaluated the user interface in a user testing. In this iteration, the user interface supported the following features: free-hand writing, accuracy evaluation using the HD method, selecting activities via the home screen, and user registration and login.

The first identified issue was the lack of support for dark mode. This missing feature negatively impacted the usability rating to a significant extent. A minor issue noted was the low contrast between the icon buttons representing completed and uncompleted activities.

Four participants reported difficulties with writing, including difficulty maintaining alignment with the template guidelines during freehand input, and an overly strict evaluation of writing accuracy. All participants were confused about how many exercises they were supposed to complete during an activity. Most also found the home screen unclear without verbal instructions, indicating a need to include additional information on both.

Additional feedback included increasing the size of the writing area for characters and the *Continue* button, clearly distinguishing the design of the login confirmation button from the one that switches to the registration screen, and allowing users to view their password while typing it.

One participant in the PaP group noted that the component was harder to use because they could not see multiple characters at once and, therefore, could not compare them. This observation, along with the idea that noticing differences between characters might aid in learning, led to the addition of cards showing similar characters in the second iteration.

V. The Final Application

The UI modifications in the second iteration of the component were based on feedback from the evaluation of the first iteration.

In login and registration screens, the text readability issue was resolved in dark mode. The confirmation button and the registration link were visually differentiated, and an option to view and confirm the password was added. The contrast between completed and uncompleted activities has been increased on the home screen (see Fig. 1). Subheadings with a short description of the content have been added to the chapter cards.

A summary screen was added at the end of each lesson (Fig. 3), showing the success rate of completed exercises and the time it took to finish them. The graphic content and title of the screen changed based on the success rate.

Features were added to automatically play and replay the pronunciation of a character on the writing exercise screens (Fig. 2). Additionally, cards were introduced to show characters visually similar to the current one. Other minor modifications included increasing stroke thickness, enlarging the *Continue* button, and the Latin letter. Dark mode support has been added for all screens (Fig. 2).

A. Adjustments to the Recognition

Although majority of users did not request a change in evaluation sensitivity, based on observation of interactions, the threshold was adjusted slightly from 20 to 25. This was to reflect the expectation of greater tolerance when producing first strokes.

B. Adaptation to Users' Skills

The component has been expanded to include two additional *levels of writing difficulty* and two new exercises were added: character identification and sound identification. The difficulties were adjusted by the users' skills as follows.

The default difficulty level is 1, where visual aids are given for all character strokes. After writing the character correctly multiple times, the user continues on the *second level* where one random stroke of the character becomes invisible. If the user makes a mistake, the difficulty resets to the first level.

Fig. 1 Home screen.

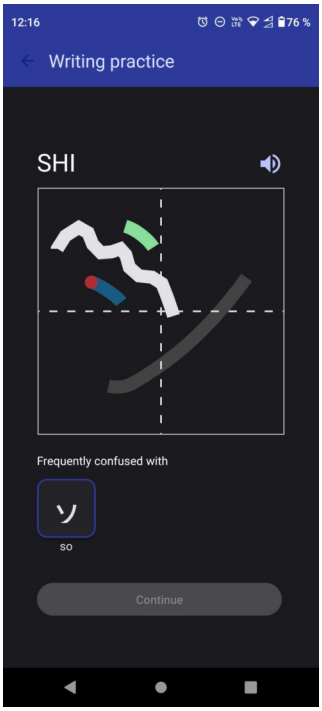


Fig. 2 Drawing screen.

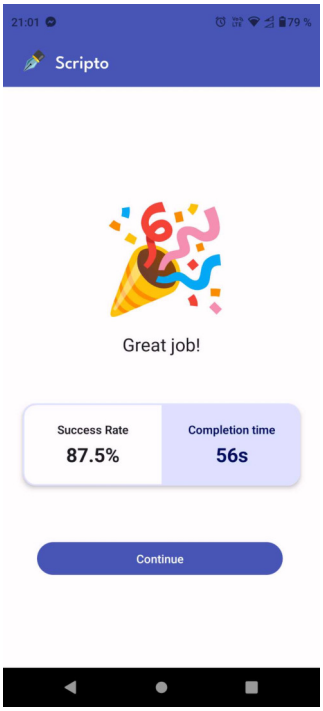


Fig. 3 Results screen.

After writing the character with missing strokes correctly multiple times, the user goes to the *third level*. All strokes are hidden, so the user writes the character completely independently. When the user makes mistakes in this level, they degrade to the second level.

As we can see, when the user makes mistakes in the particular level, the component degrades the difficulty to help the user remember better. After degrading, the level is increased after the next successful character. When making too many mistakes in the particular level (the number was determined to 5), the user is degraded another level. This way the learner gradually learns the characters until they can write them without any visual aids.

The travel up and down between difficulties has been designed in order to support gradual learning as well. When the user learns one character, it is easy to repeat it again. With every new character, the remembering the correct writing of any character becomes harder. This way we not only support correct writing of one character, but we help to remember the whole alphabet correctly. The idea is the user will train until they make no mistakes in all characters.

C. Methods of Comparing Character Strokes

In addition to the original approach using HD, two new methods, Fréchet Distance (FD) [12] and Shape Context (SC) [13], have been implemented as alternative ways of comparing font similarity.

FD is calculated using the following equation:

$$D(f,g) := \inf_{\alpha,\beta} max_{t\in[0,1]}\, d(f(\alpha(t)), g(\beta(t))) \tag{2}$$

Its result is the minimum possible maximum distance between corresponding points along two curves, considering all continuous and monotonic parameterizations of the curves. After implementing FD, it was found that the results were nearly identical to those produced by the HD. However, FD offers an additional advantage of evaluating stroke direction, which led to HD being deprioritized in the subsequent evaluation.

The SC method represents a shape by computing histograms of the relative positions of all other points with respect to each reference point on the shape. Shape similarity is then evaluated by matching their point-wise histograms and determining the cost of the best-fitting transformation (typically affine) that aligns the shapes. The histograms describing the shape via relative positions of coordinates can be expressed as follows [14]:

$$h_i^k = \#\{q \neq p_i : (q - p_i) \in \text{bin}(k)\}. \tag{3}$$

Key limitations of SC included its insensitivity to orientation and stroke length, as well as the lack of stroke direction assessment. The issue of significantly differing stroke lengths was addressed by introducing a tolerance threshold. If the difference between the user's stroke and the template did not exceed 25, the stroke was accepted.

Initial attempts at using angle deviation penalties to resolve the stroke orientation issue were ineffective. Ultimately, checking the distance between the first control points of the user's stroke and the template proved to be effective. If this distance exceeded 25, the stroke was rejected. This approach also reliably enforces correct stroke direction. Differences in the number of control points in strokes written with different speed were mitigated through interpolation and resampling, but increasing the sampling frequency was also considered.

VI. Second Evaluation

The second round of user evaluation involved independently completing activities and filling out a related questionnaire. *11 users* participated this time, none of whom had prior knowledge of Japanese language or script. In this section we will summarize the results from the evaluation.

A. Exercise Difficulty

Almost 40% of participants found the character identification exercise challenging. Their responses suggest that some began lessons with unfamiliar characters rather than starting with the testing lesson, which made the exercise more difficult. This is supported by the fact that, otherwise, 9 out of 11 users rated the exercise type as *easy to use*. Most users found the three answer choices for identification exercises sufficient. One user suggested gradually increasing the number of options to match lesson difficulty.

Writing-related exercises received feedback centered on usage difficulty or the UI shared across all three difficulty levels.

- The first two writing exercise types were rated as easy to use by 9 users.
- For the third type, all participants agreed it was easy to use, even though it was considered the most challenging to complete.
- Tracing and partial tracing exercises received similar difficulty ratings, with tracing being only slightly easier.

In the tracing exercise, 6 users immediately understood what to do, 4 hesitated at first but figured it out, and one was confused.

In the free writing exercise, *10 users recalled the character strokes from previous exercises*. However, 4 believed they would likely write with some inaccuracies.

B. Pronunciation Feature

The sound identification exercise was rated as very easy, with its purpose clear to all users. All but one participant agreed that a separate pronunciation exercise was useful; one found pronunciation within other exercise types sufficient. The option to replay the character sound was used by 7 participants. Replaying pronunciation was not something most users would use regularly, even though they did use it during evaluation.

C. Suggestions for Improvement

In user suggestions, the most frequently mentioned improvement was to *reduce the strictness* of the similarity evaluation in writing exercises, as this was a common source of difficulty. In summary, the suggestions for improvement included:

- reducing evaluation strictness,
- not locking lessons (only marking them as completed),
- adding gamification elements.
- more detailed feedback (e.g., number of correct strokes),
- display of time in minutes,
- summary of mistakes,

Stroke starting dots were sufficient for 8 users, 3 would prefer *more guidance* (e.g., arrows, numbers). In the free writing exercise, 2 participants would prefer no hints at all. The increase in exercise difficulty was seen as appropriate by all, though 3 users would add additional aids.

D. Other Results and Suggestions

The login and registration screens were rated as very clear, understandable, and easy to use by a strong majority (9). Only 2 users rated the difficulty higher than the minimum, scoring 2/5 and 3/5. Suggestions for improvement included implementing persistent login across sessions, stricter password criteria, and two-factor authentication. 9 participants understood the purpose of the home screen, and 8 found it clear. However, two users were unsure how to begin learning the characters, and three had difficulty understanding the scope of learning represented by the chapter cards.

In open-ended responses, users suggested adding a *tutorial* or on-boarding screens. One participant was confused about what was being practiced within a lesson card, and another was misled by the similar numbering of chapters and lessons.

Cards with similar characters were considered helpful by 8 out of 11 participants. The progress indicator was understood by all participants and seen as suitable. The results screen was considered useful by all, though by some only partially. 10 rated its design positively, one neutrally. For the text and image on the results screen, 8 participants found them adequate for their performance. 7 found them motivating, and 5 found them fun.

Theme switching posed no issues for any participant. 7 out of 11 preferred the dark theme, mainly for eye comfort and visibility, but also for aesthetics and environmental reasons. Two users preferred warm-colored themes (cream, beige). All participants were satisfied with the font and its visibility.

E. System Usability Scale Evaluation

The final iteration was evaluated using the System Usabilty Scale metric, where the system scored 83% out of 100%, indicating very good usability.

VII. Comparison of Evaluation Methods FD and SC

The evaluation also included a set of questions comparing participants' perceived similarity of their writing to the reference characters with automated evaluation using the FD and SC methods. Participants were divided into two groups (5:6), each testing a different evaluation method.

User satisfaction was higher with the SC method (72%) than with FD method (56.6%). *Difficulties* with evaluation were reported by 77% of participants in the FD group and 58% in the SC group. All responses indicating a mismatch between evaluation and perceived accuracy came from the FD group, where only one participant considered the evaluation accurate. *In the SC group, all participants felt the evaluation was accurate*, although more of them experienced (partial or full) discomfort. Two participants in the FD group reported being demotivated by discrepancies in evaluation, none in the SC group did.

The results suggest a greater suitability of the SC method, but due to the small sample size, this should be seen as a trend rather than a definitive conclusion.

VIII. Conclusion

The goal of this work was to develop an interactive touch component for self-learners of the Japanese Katakana alphabet. The component is part of an application that supports learning through writing and recognition, with the flexibility to accommodate various teaching methodologies. Comparison of teaching methods showed that the group using the component had lower writing success, but both groups found the learning experience comfortable and adhered equally well to the schedule. Participants appreciated the flexibility of practice in the component.

In the second iteration, new activity types were added, along with evaluation method settings, full dark mode support, a redesigned home screen, and a results screen. Between 72–100% of users rated the exercises as easy to use. Features such as theme switching, progress indicators, character group cards, results screen, and pronunciation recognition practice were almost universally positively received.

Some issues arose with how users interpreted interface cards on the home screen, suggesting a need for short chapter descriptions. There were methodological shortcomings in the home screen content, which contributed to users encountering unfamiliar characters. The current lesson button should also be visually better distinguished from chapter icons. On the results screen, several users suggested that stroke mistake count, rather than just letter mistake count, should be included in the evaluation metric. Participants also suggested adding an introductory guide to explain the types of exercises and the layout of the home screen.

The most negatively rated aspect was the strictness of the evaluation method. 63% of participants expressed dissatisfaction, with some feeling demotivated, and they preferred metrics that provide

clearer feedback on errors. To better define evaluation thresholds, it may be helpful to quantify similarity based on user perception, for example, by creating a dataset of stroke patterns commonly perceived as mistakes. However, this would require more in-depth research into handwriting.

Currently, the component acts as a tool for building a course from modular components, but does not implement any specific methodology. Adding one would be a valuable extension beyond the current scope of the work. The component is already close to supporting additional writing systems. Arabic and Indian scripts are suggested as suitable extensions, as fewer tools currently exist for them compared to Japanese, Cyrillic, or Thai scripts.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-23-0408.

The authors thank all the participants involved in the experiments and surveys, whose insights provided valuable contributions to this work.

References

[1] N. Yiitolu and M. Reichelt, *L2 writing beyond English*. Multilingual Matters, 2019.
[2] F. Kato, "Efficacy of intervention strategies in learning success rates," *Foreign Language Annals*, vol. 35, no. 1, pp. 61–72, 2002.
[3] B. S. Handini, "Teacher's difficulties in teaching japanese writing system (sakubun)," *Humanitatis: Journal of Language and Literature*, vol. 4, no. 2, pp. 22–27, 2018.
[4] O. Udeshika, "A study on use of katakana letters by learners of japanese as a foreign language," 2016.
[5] M. Longcamp, J.-L. Anton, M. Roth, and J.-L. Velay, "Visual presentation of single letters activates a premotor area involved in writing," *Neuroimage*, vol. 19, no. 4, pp. 1492–1500, 2003.
[6] K. H. James and L. Engelhardt, "The effects of handwriting experience on functional brain development in pre-literate children," *Trends in Neuroscience and Education*, vol. 1, no. 1, pp. 32–42, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211949312000038
[7] M. Kiefer and N. M. Trumpp, "Embodiment theory and education: The foundations of cognition in perception and action," *Trends in Neuroscience and Education*, vol. 1, no. 1, pp. 15–20, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221194931200004X
[8] A. A. Taha and A. Hanbury, "An efficient algorithm for calculating the exact hausdorff distance," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2153–2163, 2015.
[9] M. Kai, *JAPANESE CHARACTERS*. Cambridge Scholars Publishing, 2013, p. 2–3.
[10] N. Bonneton-Botté, S. Fleury, N. Girard, M. Le Magadou, A. Cherbonnier, M. Renault, E. Anquetil, and E. Jamet, "Can tablet apps support the learning of handwriting? an investigation of learning outcomes in kindergarten classroom," *Computers & Education*, vol. 151, p. 103831, 2020.
[11] S. Blumenthal and Y. Blumenthal, "Tablet or paper and pen? examining mode effects on german elementary school students' computation skills with curriculum-based measurements." *International Journal of Educational Methodology*, vol. 6, no. 4, pp. 669–680, 2020.
[12] K. Buchin, M. Buchin, and Y. Wang, "Exact algorithms for partial curve matching via the fréchet distance," in *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2009, pp. 645–654.
[13] G. Mori, S. Belongie, and J. Malik, "Efficient shape matching using shape contexts," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, pp. 1832–7, 12 2005.
[14] D. Qian, T. Chen, and H. Qiao, "Background of shape contexts for point matching," *Pattern Recognition Letters*, vol. 84, pp. 114–119, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865516301787

Author's index

A
Andrišov Dávid 35

B
Babitskyi Dmytro 554
Bačík Ján 62
Bačíková Michaela 624, 630, 636, 642, 648
Badár Jozef 560
Bakhir Mykhailo 577
Barnovský Marek 642
Basarik Tomáš 27, 246, 445
Berecký Kamil 218
Bober Peter 593
Bosák Maroš 571
Bubeňková Lenka 172, 181
Buček Leo 239
Buček Tomáš 400, 408, 545

Č
Čigáš Tomáš 194
Čorňák Daniel 476

D
Drutarovský Miloš 218, 250, 257, 271, 392

Ď
Ďurovský František 318, 428

E
Eliaš Matej 400

F
Fabišík Adam 285
Feciľak Peter 279, 285, 291
Fedor Pavol 148

G
Garnek Lukáš 440
Genči Ján 400, 408, 545
Gibová Zuzana 299
Girovský Peter 27, 94, 102, 225, 355, 536, 550, 566
Gordan Daniel 27, 246, 445
Gučik Christián 107

H
Haluška Renát 465, 471
Haneszova Hajnalka 8
Herasymchuk Nazarii 115
Hirko Matej 482
Hliva Samuel 344
Horváth Marek 311, 351, 423
Hric Ladislav 75
Hric Matej 428
Hryhoriev Pavlo 489
Hryhoriev Yurii 496
Hudák Roman 434

Ch
Chodarev Sergej 344

I
Ilek Michal 620
Ivanoc Dominik 159

J
Jakubčo Valér 465
Jakubec Viliam 519
Jenča Viliam 43
Jurík Patrik 372
Jusko Miroslav 21, 80, 87

K
Kačmarik Branislav 50
Kainz Ondrej 107, 115, 122, 130, 165, 264
Kajáti Erik 531, 620
Kalejová Veronika 264
Kanálik Tomáš 142
Kanóczová Andrea 648
Kantor Dušan 455
Kaňuch Ján 94, 225, 536, 550
Kardoš Slavomír 554
Katonová Erika Abigail 279, 285, 291
Kelemen Peter Daniel 189
Khainas Andrii 137
Kijac Jozef 372
Király Jozef 189
Kmecik Tadeáš 27, 246, 445, 550, 566
Kondáš Nikolas 225
Korečko Štefan 152, 582
Kosiuk Veronika 531
Kovalčín Jozef 271
Kováč Daniel 566
Kováč Ondrej 137, 142
Kováčová Antónia 137, 142
Kövesi Mikuláš 417
Král Richard 207
Kupcová Eva 372, 379, 385, 525

L
Labanc Daniel León 56
Lacko Milan 307, 445, 519
Lapanik Tsimafei 385
Lenger Tomáš 80, 87, 202
Leonenko Anna 360
Liashenko Oleksandr 449
Listarevich Raman 630
Liščinská Zuzana 35, 43, 50, 56, 326, 333
Livovský Ľubomír 339
Lukács Peter 21, 80, 87, 202
Lukáč Martin 379

M
Mackanič Pavol 550
Maček Dávid 560
Madoš Branislav 159, 194, 239, 449, 577
Magyar Ján 434
Mahdal Filip 80
Majirská Karin 366
Marcin Daniel 307
Matejová Miroslava 360, 366
Mathes Jaroslav 582
Medveď Dušan 476, 482, 489, 496, 501, 507, 513
Michalek Eduard 351
Michaliková Simona 311
Michalko Miroslav 165, 172, 181, 264, 588
Murin Miroslav 107, 115, 122, 130, 165, 172, 181, 264

N
Nagy Christopher 461
Nemergut Peter 67, 102
Nguyen Martin 213
Nováková Ivana 107, 115, 122, 130, 588

O
Olearník Tibor 392
Onderko Matej 536
Oravec Jakub 35, 43, 50, 56, 326, 333
Ovseník Ľuboš 35, 43, 50, 56, 326, 333

P
Pacanovský Martin 545
Pajkoš Jaroslav 250
Papaj Ján 417, 455, 560, 606
Papcun Peter 612
Paralič Ján 360, 366
Pavlík Marek 571
Pavlus Maroš 471
Pástor Marek 246
Penička Matúš 172
Perduková Daniela 75, 148
Petija Rastislav 279, 285, 291
Pietriková Alena 67
Pleva Matúš 379, 385, 465, 471, 525
Podracký Erik 318
Polačková Karolína 257
Popovič Michal 636
Proner Lukáš 423

R
Redaj Denis 152
Rešovský Norbert 94
Rzuhovský Filip 122

S
Sabol Marek 624
Savenko Karina 501
Sendrovič Kevin 606
Sobota Branislav 152, 582
Sokolová Zuzana 372
Sopkovič Kristián 379, 385, 525
Spišák Ján 612
Stankovych Alina 408
Strela Gabriel 279

Szaffko Alex 291
Szomosi Vladimír 507

Š

Ševec Jerguš 87
Škulec Jakub 130

T

Tomko Samuel 355

U

Urblík Ľubomír 531, 620

V

Vargoško Tomáš 165
Vaščák Ján 600
Vaško Matúš 181
Vehec Igor 440
Vince Tibor 207
Vozdrahanov Artem 513

Z

Zajac Radovan 15
Zbojovský Ján 461
Zolotová Iveta 620

Ž

Žilková Jaroslava 355, 566

Electrical Engineering and Informatics 16

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, 658 pages

Date of publication: July 2025
Edition: 1^{st}
Pages: 658
Available on: eei.fei.tuke.sk

ISBN 978-80-553-4854-4