

Technická univerzita v Košiciach
Fakulta elektrotechniky a informatiky

QuoVadis Research @ FEI



ročník 6, č.1
Jun 2023

ISSN 2585-9587

QuoVadis Research @ FEI je vedecký časopis, ktorého obsahom sú prezentácie výsledkov vedeckého bádania v oblasti informatických a elektrotechnických vied na Fakulte elektrotechniky a informatiky Technickej univerzity v Košiciach (FEI TUKE). Časopis obsahuje plnotextové vedecké články, ktorých obsahom sú výsledky dizertačných prác doktorandov a ich školiteľov obhájených v aktuálnom období na Fakulte elektrotechniky a informatiky Technickej univerzity v Košiciach. Vedecké články prispievajú k rozšíreniu vedeckých poznatkov v odboroch informatiky a elektrotechniky. Ide o podporu a rozvoj takých odborov ako je výpočtová technika a informatika, počítačové siete, kyberbezpečnosť, mechatronika a automatizácia, kybernetika a umelá inteligencia, multimediálna telekomunikačná technika, aplikovaná informatika, počítačové modelovanie, vybrané oblasti matematiky, hospodárska informatika, elektronika, mikroelektronika a optoelektronika, spracovanie signálov, teoretická elektrotechnika, elektroenergetika a technika vysokých napätí, silnoprúdová elektrotechnika, meracie systémy v elektronike, riadenie elektrotechnickej výroby, materiály a technológie elektroniky, štruktúra a vlastnosti moderných materiálov. Prijaté rukopisy uvádzajú relevantnú medzinárodnú literatúru a popisujú nové objavy, aplikácie nových techník a metód, použitie existujúcej metodológie na nový problém, čím prispievajú k rozširovaniu základných vedeckých poznatkov v danej vednej disciplíne.

Redakčná rada

Šéfredaktor časopisu:

prof. Ing. Liberios Vokorokos, PhD.

Výkonný redaktor:

doc. Ing. Anton Baláž, PhD.

Členovia:

prof. Ing. Jaroslav Porubán, PhD.

odbor: informatika

prof. Ing. Liberios Vokorokos, PhD.

odbor: informatika, kyberbezpečnosť

prof. Ing. Peter Sinčák, CSc.

odbor: kybernetika, umelá inteligencia, inteligentné systémy, aplikovaná informatika

prof. Ing. Ján Paralič, PhD.

odbor: hospodárska informatika, informatika

prof. RNDr. Ján Plavka, CSc.

odbor: aplikovaná matematika, počítačové modelovanie, informatika

prof. Ing. Jozef Juhár, CSc.

odbor: multimediálne telekomunikácie, počítačové siete, informatika

prof. Ing. Michal Kolcun, PhD.

odbor: elektroenergetika a technika vysokých napätí, elektrotechnika

prof. Ing. Dobroslav Kováč, CSc.

odbor: teoretická elektrotechnika, automatizácia, elektrotechnika

prof. Ing. Daniela Perduková, PhD.

odbor: elektrotechnika, mechatronika, riadenie elektrotechnickej výroby

prof. Ing. Alena Pietriková, CSc.

odbor elektrotechnológie a materiály, elektrotechnika

prof. Ing. Ján Šaliga, PhD.

odbor: elektronika, spracovanie signálov, optoelektronika, meranie, elektrotechnika

prof. RNDr. Ján Ziman, CSc.

odbor: štruktúra a vlastnosti moderných materiálov, elektrotechnika

Adresa vydavateľa a redakcie: Redakcia časopisu QuoVadis Research @ FEI, Technická univerzita v Košiciach Fakulta elektrotechniky a informatiky, Letná 9, 042 00 Košice, Slovenská republika, email:quovadis@fei.tuke.sk

Obsah

3D mapovanie dronom vybaveným hĺbkovou kamerou v prostredí bez prístupu GPS navigácie	
<i>Stanislav Alexovič, Milan Lacko</i>	4-11
Výber príznakov založený na riedkej vrstve neurónovej siete	
<i>Peter Bugata, Peter Drotár</i>	12-22
Vybrané aspekty analýzy kvality povrchových úprav dosiek plošných spojov	
<i>Daniel Dzivý, Alena Pietriková</i>	23-31
Capitalizing on Unlabeled Data for Improved Chest X-ray Interpretation: A Self-Supervised Learning Perspective	
<i>Matej Gazda, Jan Plavka</i>	32-37
Advanced Data Processing Methods for Parkinson's Disease Detection from Voice	
<i>Máté Hireš, Peter Drotár</i>	38-44
Využitie algoritmov strojového učenia na detekciu malvéru v operačnom systéme Android	
<i>Michal Hulič, Liberios Vokorokos</i>	45-54
Simulation Model of Small Hydropower Plant	
<i>Richard Olexa, Pavol Fedor</i>	55-68
Detekcia škodlivého softvéru založená na algoritmoch strojového učenia	
<i>Jakub Palša, Liberios Vokorokos</i>	69-79
Aplikácia rozličných foriem strojového učenia pre pohyblivú RAN sieť	
<i>Marek Ružička, Juraj Gazda</i>	80-94
Výskum a diagnostika štandardných izolačných materiálov pre izolátory	
<i>Luboš Šárpataky, Bystrík Dolník</i>	95-115
Výskum alternatívnych kvapalných dielektrík pre vysokonapäťové aplikácie	
<i>Miloš Šárpataky, Juraj Kurimský</i>	116-130

Metódy a prostriedky pre návrh, modelovanie a realizáciu distribuovaných systémov riadenia veľkých fyzikálnych experimentov

Milan Tkáčik, Ján Jadlovský, Slávka Jadlovská *131-140*

3D mapovanie dronom vybaveným hĺbkovou kamerou v prostredí bez prístupu GPS navigácie

Stanislav Alexovič

Katedra elektrotechniky a mechatroniky
Fakulta elektrotechniky a informatiky, TUKE
Košice, Slovensko
stanislav.alexovic@tuke.sk

Milan Lacko

Katedra elektrotechniky a mechatroniky
Fakulta elektrotechniky a informatiky, TUKE
Košice, Slovensko
milan.lacko@tuke.sk

Abstrakt— Článok opisuje kvadrokoptéru navrhnutú na 3D rekonštrukciu interiéru pomocou hĺbkovej a sledovacej kamery. Bližšie opisuje zložky softvéru a hardvéru, ktoré sú súčasťou dronu, vrátane sady Holybro X500V2, sledovacej kamery Intel RealSense T265, a hĺbkovej kamery Intel RealSense D435i. Článok sa bližšie zameriava na zostavenie úplného TF stromu, zachytávanie mračna bodov pomocou technológie RtabMap, streamovanie videa a metódami stabilizácie dronu. Ďalej sa článok venuje úspešnému testovaciemu letu v interiéri ako aj schopnosti dronu navigovať sa v interiéroch podmienkach a poskytovať presné údaje na účely následnej 3D rekonštrukcie. Záver článku sa venuje odporúčaniam pre budúci vývoj v autonómnej navigácii a preskúmaní neznámych priestorov. Výsledky potvrdzujú potenciál dronu vytvárať detailné a úplné mapy interiérov ako aj uskutočniteľnosť celého zariadenia.

Kľúčové slova—3D skenovanie, V-SLAM, UAV, dron

Abstract— This paper introduces a quadrotor drone specifically designed for 3D reconstruction of indoor environments, utilizing a depth camera and a tracking camera. The hardware and software components of the drone are detailed, featuring the Holybro X500V2 kit, the Intel RealSense T265 tracking camera, and the Intel RealSense D435i depth camera. The construction of a complete TF tree, point cloud capture with RtabMap, and video streaming are all outlined in the paper. Additionally, the drone's pose stabilization and successful indoor test flight are examined, highlighting its capabilities in navigating through indoor spaces and delivering accurate data for 3D reconstruction. The paper wraps up with suggestions for future advancements, such as autonomous navigation and exploration of unknown areas. The findings showcase the viability and promise of this drone in generating detailed and all-encompassing maps of indoor environments.

I. ÚVOD

V dnešnej technologicky nasýtenej ére sa digitalizácia stala postupne integrovanou súčasťou našich životov. Neustále pokroky v technológiách nám umožnili čoraz vernejšie zachytiť okolitý svet a ponúkli nám nekonečné možnosti jeho objavovania. Doposiaľ tejto sfére dominovalo 2D fotografovanie, ktoré však nedokáže zachytiť celkový obraz reality, pretože zachytením dvojrozmernej podoby sveta prichádzame o celú dimenziu vzácných údajov. Preto sa pozornosť sveta ubera postupne k metódam 3D skenovania, ktoré tieto nedostatky nemajú. To nám poskytuje schopnosť

vytvárať digitálne kópie reality s priestorovými vlastnosťami a nestratiť vzácnu tretiu dimenziu nášho okolia. Preto sa niet čomu čudovať, že 3D skenovanie je čoraz populárnejšie, poskytuje nám totiž nové možnosti ako manipulovať s okolitým prostredím.

Kvalitný výsledok 3D skenovania, s ktorým by sa dalo ďalej pracovať sa nezaobíde bez využitia špecializovaných senzorov. Ako to už býva, modernejšie technológie stavajú na predošlých. Stereovízia ako aj metódy štruktúrovaného svetla v skenovaní využívajú kamery založené na 2D technológiách, princíp Time-of-Flight (ToF), metódy ako LiDAR už vyžadujú využitie špecializovanejších senzorov [1]. Spomínaný princíp ToF využíva kameru na zachytenie informácií o hĺbke. Kamera vysiela svetelný signál, pričom zároveň meria čas, za ktorý sa signál odrazí späť do nej, čím sa vytvára hĺbková mapa objektu alebo prostredia. Táto metóda sa bežne využíva v rozpoznávaní gest a rozšírenej realite, pretože dokáže generovať hĺbkovú mapu v reálnom čase [2]. Druhá spomínaná metóda, LiDAR, funguje na princípe laserov, ktoré merajú vzdialenosť a generujú 3D mračno bodov objektu alebo prostredia a ako taká je táto metóda 3D skenovania využiteľná v rôznych oblastiach. Veľmi často je LiDAR používaný v autách kde sa uplatňuje pri detekcii prekážok alebo aj pri samotnej navigácii v prípade autonómnych vozidiel. Okrem toho sa však používa aj v geológii, kde uľahčuje generovanie detailných 3D modelov terénu [3].

ToF a LiDAR metódy však nie sú zďaleka jediné, ktoré sa využívajú v 3D skenovaní. Niektoré 3D skenery využívajú senzory štruktúrovaného svetla, vyžarujúce svetlo vo forme mriežkového vzoru [4]. Iné využívajú fotogrametriu [5]. Tá v podstate funguje na princípe fotografovania objektu alebo prostredia z rôznych uhlov, ktoré sa následne použijú na tvorbu 3D modelu pomocou softvéru. Samozrejme, každá z týchto technológií má svoje výhody aj nevýhody a konečná voľba senzoru záleží na konkrétnej aplikácii a požiadaviek používateľa.

Precízne metódy 3D skenovania sú čoraz žiadanejšie aj pre dopyt vyvolaný najmä potrebou zaznamenávať čoraz komplexnejšie prostredie s čo možno najvyššou presnosťou. Jedna z perspektívnych oblastí vývoja je, zdá sa, vybavovanie dronov 3D skenovacími senzormi za účelom zaznamenávania údajov v prostredí bez prístupu ku GPS, ako napríklad interiérové podmienky a pod. [6]. Táto technológia má

potenciál spôsobiť revolúciu naprieč odvetvami, od architektúry po inžinierstvo, zábavný priemysel, ale aj vzdelávanie. Používaním dronov na zaznamenávanie 3D údajov z prostredia môžu odborníci získať cenné informácie o podstate fyzického sveta, aby následne mohli robiť lepšie rozhodnutia a vytvárať realistickéjšie 3D simulácie [7].

Rapidný pokrok 3D technológie spôsobuje tlak na vývoj technológie dronov, ktorý zasa tlačí ceny komponentov na výrobu takýchto zariadení nadol. Výsledkom toho vznikajú čoraz cenovo dostupnejšie komponenty a otvára sa celá plejáda možností na využitie týchto riešení v 3D mapovaní. Kvadroptera prezentovaná v tomto článku je výsledkom tohto vývoja udalostí, pretože bola zostrojená práve z takýchto cenovo dostupných materiálov, no napriek tomu dokáže poskytnúť kvalitné 3D mapovanie prostredia.

II. 3D SENZORY

Bez senzorov sa 3D skenovanie nezaobíde, a to aj preto, že ich vlastnosti ovplyvňujú výslednú kvalitu skenovania 3D objektov. V našej práci sme na tento účel použili hĺbkovú kameru Intel RealSense D435i. Bola navrhnutá pre oblasť robotiky a preto mala ideálne vlastnosti pre náš výskum, keďže je relatívne ľahká a malá. Navyše, možno ju využiť nielen v interiéri, ale aj v exteriérových podmienkach. Jej technológia je založená na aktívnej stereovízii, ktorá pozostáva z dvoch infračervených kamier, pričom infračervený projektor vytvára náhodný vzor na skenovanom objekte, ktorý sníma dvojica kamier. Zariadenie má tiež komponent inerciálnej mernej jednotky (IMU). Táto technológia zaznamenáva akceleráciu a rotáciu kamery, čo je ďalej využiteľné v algoritmoch vizuálnej lokalizácie a mapovania (V-SLAM). Napriek tomu, že táto funkcia je už zabudovaná do systému, rozhodli sme sa použiť iný samostatný komponent na meranie jej pohybu.

TABUĽKA I. ŠPECIFIKÁCIA INTEL REALSENSE D435I.

Špecifikácia	Hodnota
Veľkosť zorného poľa (FOV)	87° x 58°
Rozlíšenie hĺbkovej snímky	Max. 1280 x 720
Frekvencia hĺbkovej snímky	Max. 90 fps
RGB senzor FOV (H x V)	69° x 42°
Rozlíšenie RGB snímky	1920 x 1080
Frekvencia RGB snímky	30 fps
Váha	75g
Rozmery	(90 x 25 x 25) mm

Konštrukcia obsahuje sledovaciu kameru od rovnakého výrobcu ako primárna kamera. Konkrétny použitý model sledovacej kamery v našej práci bol Intel RealSense T265. Toto zariadenie využíva V-SLAM technológiu a jeho výstupom je presná pozícia seba samého v skenovanom prostredí. Jeho súčasťou sú dva objektívy fisheye, ktoré spolu majú zorné pole až 163° a úspornú jednotku spracovania obrazu schopnú využívať simultánnu lokalizáciu a mapovanie obrazu. Okrem toho kamera obsahuje aj inertnú mernú jednotku (IMU), ktorá je pre použitie V-SLAM systémov nevyhnutná. Samostatná

kamera nám poskytla presnejšie údaje a odbremenila hlavnú stanicu, ktorá tak nemusela spúšťať výpočtovo náročný algoritmy.

TABUĽKA II. ŠPECIFIKÁCIA INTEL REALSENSE T265.

Špecifikácia	Hodnota
Veľkosť zorného poľa (FOV)	163° x 99°
Rozlíšenie na výstupe	Max. 1280 x 720
Frekvencia výstupu odometrie	Max. 200 fps
Váha	75g
Rozmery	(90 x 25 x 25) mm
Prikon	1,5W

Obe kamery sme zmontovali dohromady, aby sa pohybovali súčasne. Navyše, keďže informácie o polohe získaval dron zo sledovacej kamery a výsledné mračno bodov pochádzalo z údajov hĺbkovej kamery, bolo potrebné vytvoriť transformačný súradnicový systém, aby sme zosúladiť vstupy z oboch zariadení do jednej sústavy. Na tento účel sme použili balík tf z knižnice ROS.

A. Kompletizácia TF stromu

Vo svojej podstate je tf strom riadený acyklický graf (DAG), ktorý predstavuje jednotlivé sústavy súradníc robotického systému a transformácie medzi nimi. Každý jeden uzol stromu je jedna súradnicová sústava, zatiaľ čo podstromy sú znázornením transformácie medzi dvoma súradnicovými sústavami. Jednotlivé uzly v tf strome majú svoje meno a nadradený uzol, ktorý rozhoduje o pozícii každého uzla na sústave súradníc. Okrem toho obsahuje každý uzol aj transformačnú maticu, ktorá upresňuje transformáciu medzi tým-ktorým uzlom a k nemu nadradeným uzlom.

ROS ponúka množstvo knižníc a nástrojov na prácu s tf stromom, nevnímajúc vysielač (tf broadcaster) a prijímač (tf listener) na načítavanie transformácií v reálnom čase a ich vizualizáciu (tf viewer). Na efektívne používanie tf stromu musia mať všetky uzly zadanú hranu k nadradenému uzlu. Hrany sme si zdefinovali pomocou statických transformácií, vytvorených prostredníctvom URDF modelu dronu. Celý dron sa v modeli chová ako pevný objekt a jediná dynamická transformácia sa týka presunu dronu a zmien v jeho pozícii na mape. Je definovaná ako transformácia medzi *odom* (odometria) a *base_link* súradnicou, pričom *base_link* je stredový bod dronu. Ide o jedinú transformáciu, ktorá musí byť vysielať pri každej zmene pozície dronu.

Pravda je že, pozícia dronu sa získava z *tracking_cam_frame* súradnice a nie z *base_link* súradnice. Práve preto sme museli prepočítavať získanú pozíciu vzhľadom k *odom* súradnici. Transformácia z *base_link* do *tracking_cam_frame* je známa, takže ju vieme použiť na transformáciu údajov do *base_link*.

Na výpočet tejto transformácie bolo potrebné prepočítať údaje o pozícii získané z *tracking_cam_frame* vzhľadom k *odom* súradnici. Museli sme preto vziať do úvahy statickú

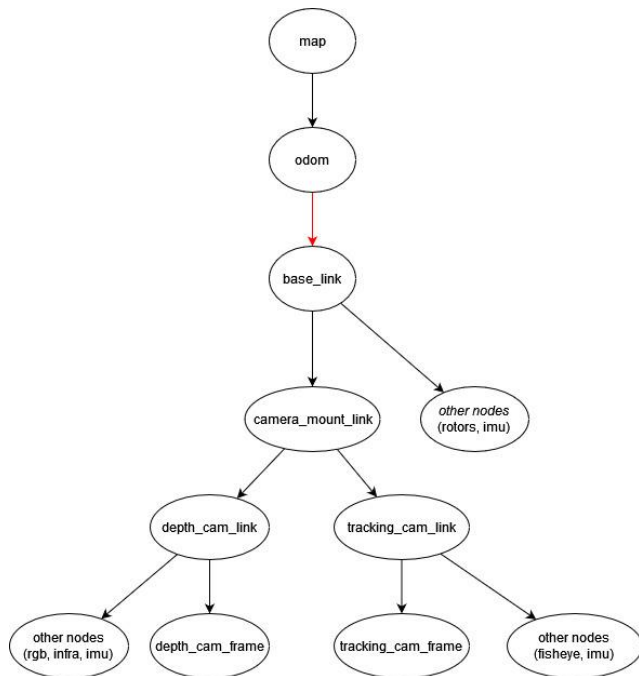
transformáciu medzi *base_link* a *tracking_cam_frame* a zrušiť ju v nameranej polohe.

Nech o_bT definuje homogénnu transformačnú maticu, ktorá popisuje transformáciu medzi *odom* súradnicou *base_link* dronu, tak vzorec bude vyzerat' nasledovne:

$${}^o_bT = {}^b_tT^{-1} \cdot P_t \cdot {}^b_tT \quad (1)$$

pričom b_tT je transformačná matica medzi *base_link* a *tracking_cam_frame*, ${}^b_tT^{-1}$ je inverzná matica z b_tT , a P_t je momentálna poloha, v podobe transformačnej matice vzhľadom na *tracking_cam_frame*.

Uplatnením matice ${}^b_tT^{-1}$ a matice b_tT sme zrušili transformáciu medzi *base_link* a *tracking_cam_frame* a vznikla nám poloha dronu, ako keby pochádzala z *base_link*. Výslednú transformáciu sme teda mohli uplatniť na transformáciu medzi *odom* a *base_link* súradnicami, výsledkom čoho bolo doplnenie chýbajúcej hrany v tf strome.



Obr. 1. Časť tf stromu zobrazuje vzťahy medzi transformačnými maticami.

B. Transformácia bodov z hĺbkovej kamery

Ďalšia transformácia, ktorú sme museli použiť bola transformácia 3D bodov z hĺbkovej kamery. Výsledný hĺbkový obraz sa nachádza v *depth_cam_frame* – vzdialenosť bodov zaznamenávala priamo kamera. Naším cieľom bolo dostať tieto body do *base_link*, ktorý v tejto fáze opisuje polohu dronu podľa údajov o pozícii, ako ju namerala sledovacia kamera.

Na prepočet hĺbkového obrazu sme znova použili balík tf na výpočty a na to, aby sme získali transformačné matice. Bolo nevyhnutné najprv definovať homogénnu transformačnú maticu, určujúcu transformáciu medzi *depth_cam_frame* a *base_link* dronu ako d_bT . Vzorec je nasledovný:

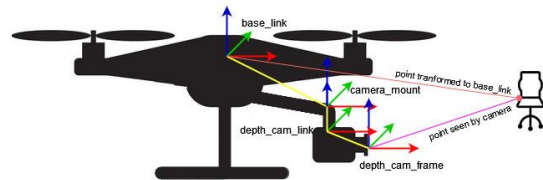
$${}^d_bT = \begin{bmatrix} {}^d_bR & \vec{t} \\ \vec{p} & s \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

pričom d_bR je 3x3 rotačná matica, ktorá popisuje orientáciu kamery vzhľadom na *base_link* súradnicu, 3x1 je vektor posunutia, ktorý opisuje pôvodnú pozíciu kamery vzhľadom k pôvodnej súradnici *base_link*, \vec{p} je 1x3 vektor perspektívy, ktorý sme nepoužili a preto je jeho hodnota $\vec{0}$; s je škálovací faktor a má vždy hodnotu 1.

Na to, aby sme 3D bod P_d získaný z hĺbkovej kamery presunuli do *base_link* súradnice môžeme uplatniť transformačnú maticu nasledovne:

$$P_b = {}^d_bT \cdot P_d \quad (3)$$

pričom P_b je transformovaný bod v *base_link* súradnici. Tým, že spojíme sledovaciu aj hĺbkovú kameru do spoločného súradnicového systému nám umožní získať presnejšie a úplnejšie prostredie okolo dronu. Zarovnaním oboch kamier do súradnice *base_link* sme získali konzistentný obraz sveta, ktorý je nevyhnutný na lokalizáciu a mapovanie.



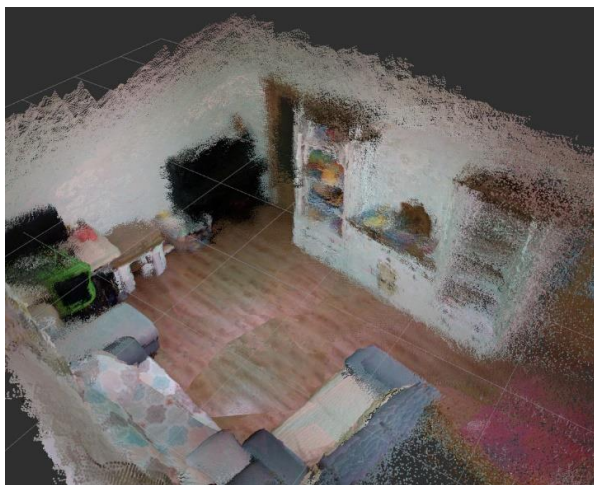
Obr. 2. Transformácia polohy videného objektu do súradnic *base_link*.

C. 3D rekonštrukcia z hĺbkového obrazu

Proces vytvárania 3D mapy vyžaduje používanie hĺbkových snímok získaných z hĺbkovej kamery. Hĺbkový snímok je vo svojej podstate dvojrozmerná matica, v ktorej každý pixel zodpovedá vzdialenosti medzi objektom a kamerou. Na to, aby sa vytvorila 3D mapa z hĺbkového snímku sa musia údaje o hĺbke konvertovať do mračna bodov. To sa vykonáva pomocou softvérových balíkov RTAB-Map. Tie dokážu prečítať hĺbkový obraz a vyrátať 3D súradnice každého bodu na scéne [8]. Okrem toho tieto balíky ponúkajú aj lokalizáciu, nielen mapovanie. Po tom, čo sa vytvorí mračno bodov, softvér pomocou neho vygeneruje 3D rekonštrukciu prostredia. Sledovať pohyby hĺbkovej kamery a zrovnať ho s každým mračnom bodov nám umožnila odometria zo sledovacej kamery. Pomocou odometrie dokáže softvér presne odhadnúť polohu a orientáciu hĺbkovej kamery pre každý zaznamenaný hĺbkový snímok. V ďalšom kroku softvér skombinuje viacero mračien do jednej 3D mapy.

Prostredníctvom údajov z odometrie sledovacej kamery náš dron dokázal odhadovať svoju pozíciu a orientáciu na mape. Vďaka tomu zvládne samostatne sa navigovať v prostredí a vykonávať úlohy podľa svojej momentálnej pozície.

RTAB-Map dokáže okrem vyššie spomenutých funkcií aj detegovať uzatvorenie slučiek, čo pomáha opravovať chyby, ktoré sa mohli časom nahromadiť na mape. Tým, že RTAB-Map rozpoznáva oblasti mapy, v ktorých sa dron nachádzal predtým môže následne upraviť a spresniť odhad polohy dronu a vylepšiť tak presnosť mapy.



Obr. 3. Mračno bodov miestnosti vytvorené pomocou RTAB-Map.



Obr. 4. Plne zostavený dron bez ochrannej kliečky

III. DRON

V našej práci sme použili dron skonštruovaný zo sady Holybro X500V2 kit. Táto sada obsahuje kostru z uhlíkového vlákna, 4 BLDC motory spolu s elektronickými regulátormi otáčok (ESC). Dron bol vybavený aj riadiacou jednotkou Pixhawk 6x, ktorá je vybavená open-source firmvérom PX4. Pixhawk 6x sme si zvolili pre jeho spoľahlivosť a všestrannosť, dá sa totiž integrovať s rôznymi senzormi a komponentmi. Dron sme ďalej vybavili aj telemetrickým rádiom s frekvenciou 433Mhz SIK za účelom komunikácie medzi pozemnou stanicou a dronom.

Pre potreby diaľkového ovládania sme dron vybavili aj 7-kanálovým prijímačom Futaba R617FS. Čo sa týka batérií, ako zdroj sme použili akumulátory 4S 4500mAh 14,8V LiPo, ktoré sme napojili na ostatné komponenty pomocou modulu napájania

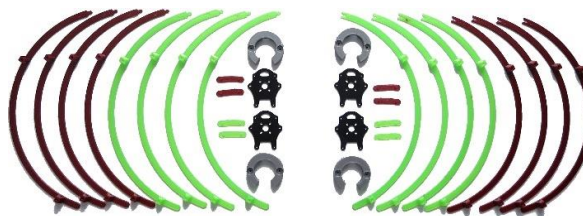
PM03D. Modul napájania monitoroval napätie a prúd a chránil systém aj pred prípadným prepätím.

TABULKA III. ŠPECIFIKÁCIA KOMPONENTOV DRONU.

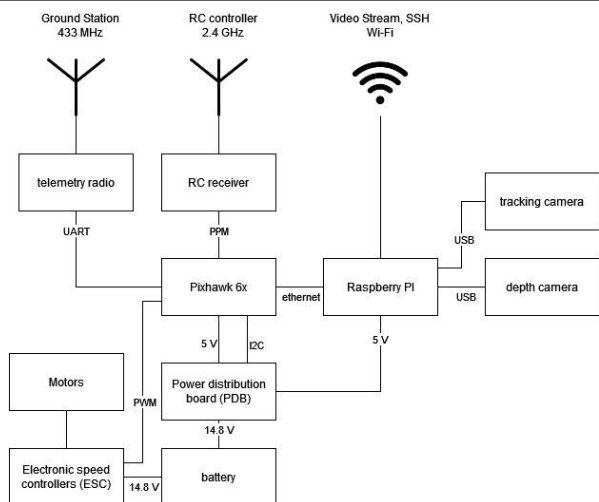
Špecifikácia	Hodnota
Kostra	Holybro X500V2
Motory	Holybro 2216 KV920
ESC	BLHeli S ESC 20A
Vrtule	1045"
Bateria	4S 4500MAh 14,8V LiPo
Telemetrické rádio	SIK Telemetry Radio V3 433MHz
Dialkový ovladač	1,5W
Ovladač letu	Pixhawk 6X
Palubný počítač	Raspberry Pi 4B 8GB

Navigáciu v lete zabezpečoval palubný počítač Raspberry Pi 4B 8GB. Palubný počítač bol pripojený k riadiacej jednotke prostredníctvom ethernetového kábla a bol v ňom spustený robotický operačný systém (ROS2). Vďaka nemu mohol systém komunikovať a zároveň vyhodnocovať výstup z hĺbkovej a sledovacej kamery. Palubný počítač poskytoval aj WiFi pripojenie na komunikáciu s pozemnou stanicou.

Okrem hlavných komponentov dronu sme na jeho kostru namontovali aj držiak na obe kamery a hĺbkovú kameru umiestnili na prednú časť držiaka, smerom dopredu. Kamera slúžila na zaznamenávanie informácií o hĺbke prostredia využívaného na 3D mapovanie v neskorších fázach spracovania získaných údajov. Sledovaciu kameru sme namontovali na spodok tohto držiaka a nasmerovali ju nadol. Využitá bola na vizuálnu odometriu a jej funkciou bolo odhadovať polohu a orientáciu dronu na základe viditeľných znakov prostredia. Sledovacia kamera spolupracovala s firmvérom PX4 a RTAB-Map softvérom a poskytovala tak presné lokalizačné a mapovacie schopnosti. Vďaka kombinácii uvedených komponentov zvládol dron samostatne lietať, navigovať sa, mapovať prostredie a odosielať nazbierané údaje do pozemnej stanice na ďalšie spracovanie a analýzu.

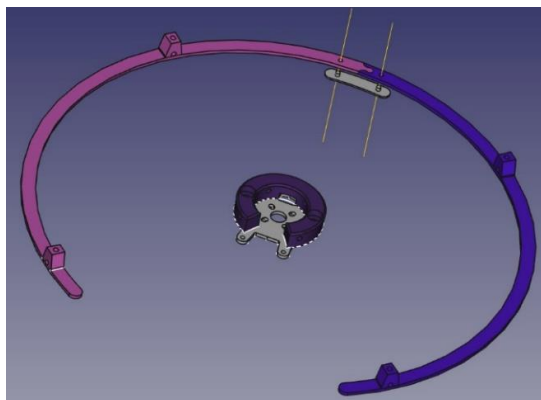


Obr. 5. Vytlačené časti ochrannej kliečky na 3D tlačiarňami.



Obr. 6. Bloková schéma všetkých komponentov použitých na stavbu dronu.

Keďže bolo pre nás dôležité ochrániť nielen prostredie, ale aj zraniteľné komponenty dronu počas letu, navrhli sme ochrannú klietku a pripevnili ju na dron. Navrhnutá bola v bezplatnom softvéri freeCAD a následne sme ju vytlačili na 3D tlačiarňami z odolného materiálu. Jednotlivé časti klievky sme spájali tyčami z uhlíkového vlákna, aby sme dosiahli čo najľahšiu no stále odolnú konštrukciu. Keďže nám však pôvodná základná doska motora neumožňovala bezpečne upevniť ochrannú klietku na dron, vytlačili sme si na 3D tlačiarňami aj novú základnú dosku motora. Ochranná kliečka totiž bola kľúčová na fungovanie dronu v interiéri. Akákoľvek kolízia dronu so stenami alebo nábytkom by poškodila jeho vrtule čo by viedlo k pádu dronu a pravdepodobnému poškodeniu aj iných dôležitých komponentov. Ochranná kliečka teda umožnila dronu pohybovať sa v interiéri bez veľkého rizika poškodenia jeho súčastí alebo prostredia.



Obr. 7. Dizajn ochrannej klievky v softvéri FreeCAD.

A. Streamovanie videa

Ovládanie dronu sa nezaobíde bez funkcie streamovania video záznamu, keďže ide o kľúčový aspekt pilotáže. V našej práci sme používali stream s tromi zložkami informácií: prvou bolo video z RGB senzora, druhú zložku tvorilo video z hĺbkového senzora (hĺbková mapa) a tretia boli metadáta s

časovou pečiatkou o momentálnej pozícii vo forme súradníc X, Y, Z. Ideálnym riešením by bolo aj streamovanie vygenerované 3D mapy ale to nebolo možné kvôli rýchlosti bezdrôtového pripojenia, a to najmä na miestach so slabším signálom.

Na vytvorenie streamu sme použili Gstreamer s funkciou RSTP servera, ktorý bežal na samostatnom vlákne v ROS uzle, zodpovednom za spracúvanie dát zo senzoru. Snímky čítal priamo senzor pomocou librealsense SDK a následne ich konvertoval pomocou OpenCV. OpenCV softvér spájal obe snímky dohromady a pridával k nim metadáta. Výsledné OpenCV snímka s metadátami putovala do RSTP servera a ten streamoval video užívateľovi. Stream bolo možné sledovať na pozemnej stanici, a to buď priamo, alebo cez akýkoľvek videoprehrávač schopný sledovať streamované video, ako napríklad VLC.



Obr. 8. Streamovanie videa. Hore: obrázok z RGB senzora; Dole: obrázok z hĺbkového senzora

Oneskorenie vytvoreného streamu sa nám podarilo dostať pod hodnotu 300ms. Čo bolo pre bežného pozorovateľa nepostrehnuteľné. Oneskorenie videa sme merali tak, že kamera snímala obrazovku, na ktorej sa zobrazoval stream z kamery. Toto vytvorilo nekonečný zrkadlový obraz kde sme mohli vidieť čas jednotlivých snímok. Porovnaním týchto časov sme získali oneskorenie videa.

Pre kódovanie videa do H264 sme využili hardvérovú akceleráciu, ktorú Raspberry Pi poskytoval. Toto umožnilo znížiť záťaž palubného počítača a vďaka tomu streamovanie zaberalo iba 7 % jadra hlavného procesora, čo je dôkazom efektivity celého systému.



Obr. 9. Meranie latencie streamovania.

B. Ovládanie a navigácia

Dron sme ovládali poväčšine manuálne. Treba však povedať, že dron mohol lietať aj samostatne, keďže sme mu nahrali vopred pripravenú 2D mapu. Podarilo sa nám to kombináciou funkcie lokalizácie RTAB-Map a balíka ROS MoveBase so schopnosťou vytvárania trajektórie. MoveBase je veľmi nápomocný nástroj, ktorý ponúka možnosti algoritmov pre plánovanie trajektórií. Vďaka týmto funkciám sa môže dron s ľahkosťou navigovať na predpripravenú cieľovú pozíciu na mape. Ešte pred samostatným letom sme teda vytvorili mapu, aby MoveBase mal k dispozícii nevyhnutné informácie o prostredí. Kľúčovú úlohu v zisťovaní polohy dronu tu zohrávala sledovacia kamera. Používali sme ju spolu s plánovacími algoritmi, aby sme vygenerovali bezpečnú a efektívnu trajektóriu do cieľa. Po vyrataní trajektórie sa údaje posielali do riadiacej jednotky dronu, ktorá ich ďalej použila pre navigáciu do cieľového bodu. Samostatný a bezpečný let dronu aj v komplexnom interiérovom prostredí sa nám podarilo docieľiť prostredníctvom kombinácie funkcií RTAB-Map a MoveBase.

IV. STABILIZÁCIA DRONU

Aj keď dron lietal najmä v interiéri, spozorovali sme, že externé faktory, ako vírenie vzduchu spôsobené vrtuľami dronu dokážu vo výraznej miere ovplyvniť jeho let. 3D mapovanie vyžaduje ustálenú pozíciu a akákoľvek nestabilita dronu môžu negatívne ovplyvniť výslednú kvalitu mračna bodov. Stabilizácia letu dronu sa dá dosiahnuť využitím sledovacej kamery a PID regulátora v riadiacej jednotke Pixhawk. PID regulátor je vo svojej podstate súčet proporčnej, integračnej a derivatívnej zložky.

$$u(t) = P(t) + I(t) + D(t) \quad (4)$$

Pričom $u(t)$ je korekčný signál poslaný do motorov dronu v reálnom čase t .

Každá zložka môže byť definovaná ako

$$P(t) = K_p e(t) \quad (5)$$

Pričom $P(t)$ je výsledok proporčnej zložky v čase t , K_p je proporčný zisk a $e(t)$ je chyba v čase t .

$$I(t) = K_i \int_0^t e(t) dt \quad (6)$$

Pričom $I(t)$ je výsledok integračnej zložky v čase t , K_i je integračný zisk, $e(t)$ je chyba v čase t a integrál je od 0 do t .

$$D(t) = K_d \frac{de(t)}{dt} \quad (7)$$

Pričom $D(t)$ je výsledok derivatívnej zložky v čase t , K_d je derivatívny zisk a $\frac{de(t)}{dt}$ je pomer zmeny chyby v čase t .

Výsledný vzorec preto vyzerá nasledovne:

$$u(t) = K_p e(t) + K_i \int_0^t e(t) dt + K_d \frac{de(t)}{dt} \quad (8)$$

Na to, aby sme vedeli definovať stabilizáciu polohy v 3D potrebujeme definovať vektor požadovanej polohy $\vec{p}_{des} = [x_{des} \ y_{des} \ z_{des}]$ a vektor momentálnej pozície $\vec{p}_{cur} = [x_{cur} \ y_{cur} \ z_{cur}]$. Momentálnu pozíciu poskytuje sledovacia kamera dronu. Táto poloha sa porovnáva s požadovanou polohou a systém vytvorí vektor chyby $\vec{e} = \vec{p}_{des} - \vec{p}_{cur}$. Chybu následne použije PID regulátor na výpočet výstupného vektora, ktorý sa použije na úpravu rýchlosti dronu.

Na výpočet nasledujúcej korekčnej hodnoty musíme vziať do úvahy momentálnu chybu a čas, ktorý uplynul od poslednej korekcie. Vznikne nám tak nasledovná rovnica:

$$\vec{u} = K_p \vec{e} + K_i I + K_d D \quad (9)$$

pričom I a D sú integračná a diferencovaná chybová premenná.

Integrovaná chybová premenná je výsledkom súčtu všetkých predošlých chýb v čase:

$$I = I + \vec{e} dt \quad (10)$$

pričom dt je čas, ktorý uplynul od posledného výpočtu. Takto sa zbavíme chyby, keď je systém v pokoji.

Diferencovaná chybová premenná je vypočítaná ako pomer zmeny v chybe:

$$D = \frac{(\vec{e} - \vec{e}_{prev})}{dt} \quad (11)$$

pričom \vec{e}_{prev} je chyba z predošlého výpočtu. To nám umožní vziať do úvahy náhle zmeny v chybovej premennej a vyhnúť sa prekročeniu hodnôt.

Napokon je tu výstupný vektor \vec{u} , ktorý použijeme na úpravu rýchlosti dronu v súradniciach x , y a z .

$$\vec{u} = \vec{v} = [v_x \quad v_y \quad v_z] \quad (12)$$

Pričom v_x , v_y a v_z sú rýchlosti v smeroch x , y a z . Poloha dronu sa následne aktualizuje na základe rýchlostí v krátkom časovom intervale a PID regulátor sa prepočíta v novom chybovým vektorom. Neustálym opakovaním tohto procesu dron dokáže udržiavať svoju polohu a upraviť ju podľa zmien v prostredí alebo na základe vonkajších faktorov ako sú vietor alebo vírenie vzduchu.



Obr. 10. Dron počas letu v interiéri.



Obr. 11. Mračno bodov interiéru vytvorený lietajúcim

V. VÝSLEDKY

Počas testovacieho letu sa dron dokázal úspešne navigovať v dome a zachytil úplné 3D mračno bodov prostredia pomocou sledovacej a hĺbkovej kamery. Dron vedel preletieť vďaka stabilizácii v lete aj úzke priestory ako dvere bez toho, aby nabúral do prekážok. Ak sa aj nejaké drobné kolízie so stenami a dverami vyskytli, vďaka ochrannej klietke mohol dron pokračovať v mapovaní priestoru. Streamované video z hĺbkovej kamery umožňovalo monitorovať let dronu v reálnom čase a dron vďaka tomu počas svojho letu neodbočil z kurzu. Výsledné 3D mračno bodov sme neskôr stiahli z dronu a výsledky ukázali, že systém dokázal zmapovať interiérové prostredie. V budúcom výskume by sme sa mohli zamerať na

preskúvanie neznámych prostredí pomocou nášho systému. Okrem toho by sme mohli vylepšiť streamovacie schopnosti a zbieranie údajov.

VI. ZÁVER

Tento článok opisuje kvadroptéru navrhnutú na vytvorenie 3D modelu interiéru pomocou kombinácie hĺbkovej kamery na zaznamenávanie informácií o hĺbke a sledovacej kamery, ktorá je zodpovedná za zaznamenávanie polohy v priestore. V článku sa kladie osobitný dôraz na hardvérové a softvérové komponenty použité na postavenie dronu. Komponenty boli síce ľahko dostupné, ich zapojenie sa ukázalo ako pomerne náročné. Funkčné zapojenie komponentov a chod dronu vyžadovalo inžiniersku odbornosť a zručnosti s vývojom softvéru.

Počas výskumu sme sa museli popasovať aj s bugom v USB porte Raspberry Pi. Tento bug spôsobil, že sledovacia kamera prestala v istom bode posielať snímky zo zaznamenávania polohy. Keď sme sa pokúšali problém odstrániť u výrobcu, zistili sme, že ten ukončil podporu a museli sme preto vyriešiť tento problém vlastnými silami. Vyvinuli sme na tento účel vlastný kód, ktorý komunikoval priamo s kamerou tak, že RealSense-ROS wrapper sme obišli. Náš vlastný kód fungoval hladšie a efektívnejšie čítal údaje zo sledovacej kamery a údajov o polohe. Vďaka nemu sme sa vyhli aj vnútornému pretečeniu zásobníka.

Vytvorené mračno bodov bolo pomerne presnou reprezentáciou interiéru a obsahovalo ucelenú a detailnú mapu skenovaného prostredia. Vo výslednom letovom test dron lietal aj v úzkom priestore bez straty svojej polohy a zbieral kvalitné dáta pre 3D rekonštrukciu.

V ďalšom výskume by sme sa mohli zamerať na zlepšenie autonómie dronu, poprípade doceliť, aby dron sa vedel navigovať aj v komplexnom prostredí bez ľudského zásahu. Náš výskum preukázal, že kvadroptéry sú využiteľné v oblasti 3D skenovania interiéru a majú potenciál v budúcnosti posunúť túto oblasť vedy.

ACKNOWLEDGMENT

Táto práca bola podporená v rámci projektu APVV-18-0436 Agentúrou na podporu výskumu a vývoja.

REFERENCES

- [1] C. Lee, H. Song, B. Choi, Y. S. Ho, "3D scene capturing using stereoscopic cameras and a time-of-flight camera", in IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1370-1376, August 2011, doi: 10.1109/TCE.2011.6018896.
- [2] O. Choi, S. Le, "Wide range stereo time-of-flight camera", 2012 19th IEEE International Conference on Image Processing, 2012, pp. 557-560, doi: 10.1109/ICIP.2012.6466920.
- [3] D. Bastos, P. P. Monteiro, A. S. R. Oliveira, M. V. Drummond, "An Overview of Li-DAR Requirements and Techniques for Autonomous Driving", 2021 Telecoms Conference (ConfTELE), 2021, pp. 1-6, doi: 10.1109/ConfTELE50222.2021.9435580.
- [4] M. Atif, S. Lee, "Adaptive Pattern Resolution for Structured Light 3D Camera System" 2018 IEEE SENSORS, 2018, pp. 1-4, doi: 10.1109/ICSENS.2018.8589640.

-
- [5] E. P. Baltsavias, "A comparison between photogrammetry and laser scanning" 1999 ISPRS Journal of photogrammetry and Remote Sensing, 54(2-3), 83-94, doi: 10.1016/S0924-2716(99)00014-3.
 - [6] I. Motawa, A. Kardakou, "Unmanned aerial vehicles (UAVs) for inspection in construction and building industry" 2018 Paper presented at the 16th International Operation & Maintenance Conference, Cairo, Egypt.
 - [7] H. Song, W. S. Yoo, W. Zatar, "Interactive Bridge Inspection Research using Drone" 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), Los Alamitos, CA, USA, 2022, pp. 1002-1005, doi: 10.1109/COMPSAC54236.2022.00155.
 - [8] M. Labbé, F. Michaud, "RTAB-Map as an Open-Source Lidar and Visual SLAM Library for Large-Scale and Long-Term Online Operation" in Journal of Field Robotics, vol. 36, no. 2, pp. 416–446, 2019. (Wiley)

Výber príznakov založený na riedkej vrstve neurónovej siete

Peter Bugata

Katedra počítačov a informatiky
Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach
Košice, Slovensko
peter.bugata@student.tuke.sk

Peter Drotár

Katedra počítačov a informatiky
Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach
Košice, Slovensko
peter.drotar@tuke.sk

Abstract—Data dimensionality reduction using feature selection is an important part of data preprocessing in machine learning. It leads to improving the performance of prediction models, saving the computing resources and shortening the calculation duration. Neural networks have experienced tremendous success in solving many nonlinear learning problems, and research shows that they can be the basis for the development of feature selection methods. In this work, we present a new feature selection method based on the use of a sparse layer of a feedforward neural network. The method introduces two sets of conditions, the satisfaction of which leads to a sparse feature-selection layer. The proposed method is evaluated on synthetic and real-world data in terms of success in identifying relevant features and impact on prediction accuracy. Experimental results show that the method achieves comparable or better performance than popular traditional feature selection methods and it is applicable even for high-dimensional small-sample-size data.

Index Terms—feature selection, dimensionality reduction, neural networks, high-dimensional data

Abstrakt—Redukcia dimenzionality dát pomocou výberu príznakov je dôležitou súčasťou spracovania dát v strojovom učení. Vedie k zlepšeniu výkonnosti predikčných modelov, šetreniu výpočtových zdrojov a skráteniu trvania výpočtov. Neurónové siete dosiahli obrovské úspechy pri riešení mnohých nelineárnych problémov učenia a výskumy ukazujú, že môžu byť základom pre vývoj metód výberu príznakov. Práca predstavuje novú metódu výberu príznakov založenú na použití riedkej vrstvy doprednej neurónovej siete. Metóda zavádza dve skupiny podmienok, ktorých splnenie vedie k riedkej výberovej vrstve. Navrhovaná metóda je hodnotená na syntetických a reálnych dátach z hľadiska úspešnosti pri identifikovaní relevantných príznakov a vplyvu na presnosť predikcie. Výsledky experimentov ukazujú, že dosahuje porovnateľnú alebo vyššiu výkonnosť ako populárne tradičné metódy výberu príznakov a je použiteľná dokonca aj pre vysokorozmerné dáta s malým počtom pozorovaní.

Kľúčové slová—výber príznakov, redukcia dimenzionality, neurónové siete, vysokorozmerné dáta

I. ÚVOD

V súčasnosti sme zaplavení obrovským množstvom dát pochádzajúcich z rôznych zdrojov. Dáta generované v oblastiach ako zdravotná starostlivosť, bioinformatika, doprava, sociálne médiá alebo online vzdelávanie sa často vyznačujú veľkým počtom atribútov (príznakov, premenných). Extrémne vysoká dimenzionalita dát predstavuje výzvu nielen pre ich efektívnu obsluhu, ale aj pre automatizované objavovanie znalostí

v dátach rôzneho druhu [1], [2]. Pri aplikovaní algoritmov strojového učenia na vysokorozmerné dáta, často s desiatkami alebo stovkami tisícov atribútov, sa prejavuje fenomén známy ako prekliatie dimenzionality (*curse of dimensionality*) [1]. Spôsobuje výrazné zníženie výkonnosti algoritmov, ktoré boli pôvodne navrhnuté pre dáta nízkych rozmerov. Okrem toho vysoká dimenzionalita kladie zvýšené požiadavky na výpočtové zdroje a predlžuje čas výpočtu.

Výber príznakov (*feature selection*) je jednou z techník redukcie dimenzionality, ktorej cieľom je vybrať menšiu podmnožinu z pôvodnej množiny atribútov, odstránením tých, ktoré sú považované za irelevantné alebo redundantné, bez významnejšej straty informácie alebo zníženia výkonnosti algoritmov strojového učenia [1]. Z vyššie uvedených dôvodov môže výber príznakov viesť dokonca k zvýšeniu výkonnosti algoritmov strojového učenia. Okrem zníženia požiadaviek na výpočtové zdroje a skrátenia času výpočtu je jej prínosom aj lepšia zrozumiteľnosť a interpretovateľnosť dát a modelov vytváraných metódami strojového učenia [3]. Výber príznakov zjednodušuje modely, a tak je účinným nástrojom na zabránenie ich pretrénovaniu.

V ostatných rokoch dosiahli umelé neurónové siete a hlboké učenie [4] bezprecedentné úspechy pri riešení mnohých nelineárnych problémov strojového učenia, ako sú rozpoznávanie a spracovanie obrazu a reči, strojové preklady alebo medicínska diagnostika. Zaujímavou otázkou je vzťah neurónových sietí a výberu príznakov.

Neurónové siete v procese učenia konštruujú nové abstraktné vysokoúrovňové príznaky z pôvodných nízkoúrovňových vstupných premenných, a tak prinášajú novú reprezentáciu dát. Napriek tomu môže byť aplikovanie výberu premenných pred použitím neurónových sietí užitočné. Vo všeobecnosti irelevantné premenné spomaľujú tréning neurónovej siete a zvyšujú požiadavky na výpočtové zdroje. Navyše zvyšujú riziko pretrénovania. Výber premenných znižuje zložitosť modelu na vstupnej úrovni, a tým vedie k lepšiemu učeniu sa hlbokých neurónových sietí a k šetreniu zdrojov [5]. Redukuje počet parametrov neurónovej siete, čo vedie k eliminácii pretrénovania. Na druhej strane výskumy ukazujú, že neurónové siete a hlboké učenie môžu byť základom pre vývoj metód výberu premenných [6]. Napriek úspechom, ktoré neurónové

siete dosahujú v rôznych oblastiach, ich modely sú ťažko vysvetliteľné a nie je známe, na základe ktorých vstupných premenných dospeli k výsledku. V prípade ich použitia na výber premenných je však výsledok dobre interpretovateľný, pretože vyberajú dôležité pôvodné premenné, ktorých význam je jasný.

V tejto práci je navrhnutá nová metóda výberu premenných, ktorá vychádza zo súčasných trendov v tejto oblasti. Základom pre jej použitie je dopredná neurónová sieť vhodná na riešenie danej klasifikačnej alebo regresnej úlohy. Dôležitosť jednotlivých vstupných premenných je určovaná pomocou špeciálnej vrstvy na výber premenných vlozenej medzi vstupnú a prvú skrytú vrstvu tejto siete.

V kapitole II sú analyzované súčasné trendy vo výbere príznakov a v kapitole III sú popísané niektoré spôsoby využitia neurónových sietí pri vývoji metód na výber premenných. Kapitola IV bližšie predstavuje navrhovanú metódu a kapitola V prezentuje dosiahnuté experimentálne výsledky. Kapitola VI obsahuje závery a námety pre ďalšie pokračovanie výskumu.

II. SÚČASNÉ TRENDY VO VÝBERE PRÍZNAKOV

V posledných rokoch bolo vyvinutých mnoho metód na výber príznakov založených na rôznych princípoch. Publikované práce dokazujú prínos ich použitia pri predspracovaní dát [3], [5]. Avšak neexistuje „najlepšia“ univerzálna metóda, pretože výber vhodnej metódy závisí od vlastností konkrétneho súboru dát a riešenej úlohy [2]. Preto sa ďalej vyvíjajú lepšie metódy pre špecifické problémy. Výber príznakov zároveň čelí novým výzvam, ktoré súvisia s meniacim sa charakterom získavaných dát [1], [5], [6].

A. Výber príznakov pre HDSSS dáta

Problém výberu príznakov pre vysokorozmerné dáta s malým počtom pozorovaní (*high dimensionality small sample size data* – HDSSS data) je skúmaný v súvislosti s analýzou génov, ktorá pracuje s obrovským počtom premenných, ale malým počtom pozorovaní. Problém súvisí s témami ako stabilita výberu premenných, riedka reprezentácia dát a využitie viacerých zdrojov pri výbere príznakov [5].

Stabilita znamená necitlivosť výsledku výberu príznakov na variácie tréningovej množiny. Pri génovej analýze hľadajú biológovia malý počet génov alebo proteínov, ktoré by vysvetlili mechanizmy správania vzoriek. Nestabilita vo výsledkoch výberu však vedie k nedôvere doménových expertov. Riešením by mohli byť skupinové (*ensemble*) metódy výberu príznakov, ktoré v prvom kroku produkujú skupinu množín vybraných príznakov a v druhom kroku ich agregujú do stabilnejšieho výstupu [7], [8], [9]. Ďalšie riešenie ponúka váženie pozorovaní v tréningovej množine, pričom váha vyjadruje vplyv pozorovania na odhad relevancie príznakov. Hlavná myšlienka spočíva v oslabení vplyvu odľahlých bodov na výber príznakov. Iná stratégia by mohla byť založená na spájaní príznakov do skupín podľa ich vzťahov, ktoré obvykle vo vysokorozmerných dátach existujú. Takéto skupiny sú odolnejšie voči variáciám v tréningovej množine [5].

V poslednej dobe sa veľká pozornosť venuje metódam výberu príznakov založeným na riedkom učení (*sparse learning*), ktoré môžu byť efektívnym nástrojom na výber príznakov z HDSSS dát [10]. Tieto metódy minimalizujú chybu predikcie spolu s pridaným riedkym penalizačným členom, čo vedie k zmenšeniu, ba dokonca k vynulovaniu mnohých váh premenných, a následne k ich vynechaniu z výberu.

Malý počet pozorovaní v HDSSS dátach má negatívny vplyv na spoľahlivosť štatistických analýz. Jedným zo spôsobov riešenia problému je hľadať ďalšie informačné zdroje, ktoré pomôžu pochopiť dáta. Extrakcia a konverzia heterogénnych znalostí z rôznych zdrojov tak, aby boli užitočné pre výber príznakov a zo všetkých zdrojov poskytovali výsledky výberu rovnakého typu, je kľúčovým problémom výberu príznakov z viacerých zdrojov (*multi-source feature selection*) [11].

B. Výber príznakov pre big data

Vysokorozmerné dáta s veľkým počtom pozorovaní (*big data*) sú bežné v oblasti rozpoznávania a spracovania textu, reči a obrazu. Môžu ohroziť škálovateľnosť súčasných metód výberu príznakov [5]. Zlepšenie škálovateľnosti sa dá dosiahnuť použitím distribuovanej stratégie výpočtov. Dostupnosť množstva informácií poskytuje možnosti pre viacpohľadový výber príznakov (*multi-view feature selection*).

Keď sú dáta distribuované na veľkom počte uzlov v sieti, distribuovaný algoritmus na výber príznakov musí umožniť výmenu informácií medzi uzlami a pomocou určitého mechanizmu fúzie zabezpečiť, aby všetky uzly pracovali spoločne na riešení optimalizačnej úlohy [12]. V súčasnosti distribuované metódy výberu príznakov pracujú s rovnakými množinami príznakov na každom uzle, avšak je potrebné zaoberať sa aj takými metódami, ktoré predpokladajú, že jednotlivé uzly majú k dispozícii rôzne reprezentácie príznakov.

Viacpohľadový výber príznakov [13] sa odlišuje od výberu príznakov z viacerých zdrojov (kapitola II-A), pri ktorom sa selekcia robí nezávisle pre každý zdroj a čiastkové výsledky sa agregujú do konečného výsledku. Výber z viacerých zdrojov predpokladá, že každý zdroj je postačujúci na naučenie sa cieľového modelu, a ignoruje vzťahy medzi zdrojmi. Na druhej strane pri viacpohľadovom učení môžu jednotlivé pohľady poskytovať navzájom komplementárne informácie a reprezentovať rôzne priestory príznakov, ktoré sú navzájom závislé. Viacpohľadový výber príznakov nepredpokladá, že samostatné pohľady sú postačujúce na naučenie sa cieľového modelu, a berie do úvahy vzťahy medzi pohľadmi, čo môže viesť k zlepšeniu výkonnosti v reálnych aplikáciách.

Ďalším problémom je viacznačkový výber príznakov (*multi-label feature selection*). V tradičnom koncepte výberu príznakov má každé pozorovanie priradenú jednu triedu (značku – *label*). V reálnych úlohách je však niekedy jedno pozorovanie asociované s viacerými triedami. Súčasná metóda viacznačkového výberu príznakov sú zväčša rozšírením známych jednoznačkových metód, ako napríklad v [14], [15]. Ďalším cieľom je navrhnúť metódy, ktoré budú priamo pracovať s viacznačkovými dátami, bez potreby transformácie, uvažujúc vzájomné vzťahy medzi značkami.

C. Výber príznakov v dynamickom priestore príznakov

Tradičné metódy výberu príznakov sú založené na statickom priestore príznakov, teda vstup pre výber príznakov sa nemení. Avšak v mnohých oblastiach, ako je sťahovanie sieťových dát alebo videa, online obchodovanie alebo medicínske testovanie, sú príznaky generované dynamicky, dokonca všetky príznaky ani nie sú dopredu známe. Selektcia príznakov v dynamickom priestore príznakov sa nazýva online výber príznakov (*online feature selection*) [6].

Všeobecne algoritmy pre online výber príznakov pracujú tak, že s príchodom nového príznaku vždy vyhodnocujú, či má byť tento príznak pridaný do existujúcej množiny vybraných príznakov a či má byť po jeho pridaní niektorý príznak z tejto množiny vynechaný [5]. V praxi však príznaky nemusia prichádzať jednotlivo, ale v skupinách. Vtedy treba uvažovať skupinové informácie, ako napríklad v [16], [17].

Iný problém predstavuje situácia, keď je priestor príznakov síce známy a prístupný pred výberovým procesom, ale postupne pribúdajú pozorovania do tréningovej množiny. V takom prípade sa pri výbere príznakov používa online učenie (*online learning*), ktorého cieľom je aktualizovať model s prichádzajúcimi novými pozorovaniami. Aktuálnym problémom pre online výber príznakov je simultánna zmena priestoru príznakov aj objemu sťahovaných dát.

Výkonnosť metód online výberu príznakov možno ďalej zlepšovať tromi spôsobmi: stabilitou metód, analýzou redundancie a časovou efektívnosťou. Časová efektívnosť metód je zvlášť dôležitá pre systémy, ktoré pracujú v reálnom čase [6].

D. Výber príznakov pre dáta so známou štruktúrou

Mnoho existujúcich algoritmov pre výber príznakov je navrhnutých pre generické dáta, ignorujúc prirodzenú štruktúru príznakov a vzťahy medzi nimi [1]. Pritom pre lepší výber príznakov by mohlo byť výhodné túto znalosť využiť.

Viaceré biologické štúdie ukázali, že gény majú tendenciu pracovať v skupinách podľa ich biologických funkcií. Zároveň medzi nimi existujú určité regulačné vzťahy. Využitie štruktúry disjunktných a prekrývajúcich sa skupín príznakov alebo ich stromovej štruktúry pri výbere príznakov je prezentované v prácach [18], [19], [20]. Vzťahy medzi jednotlivými príznakmi môžu zachytávať grafové štruktúry. Príznaky sú reprezentované uzlami grafu, pričom dva uzly sú prepojené hranou práve vtedy, ak je medzi zodpovedajúcimi príznakmi závislosť.

V oblasti spracovania prirodzeného jazyka možno pri výbere príznakov zohľadniť existujúce hierarchie na základe podradených a nadradených slov (príznakov). Cieľom hierarchického výberu príznakov (*hierarchical feature selection*) je vybrať podmnožinu príznakov bez hierarchickej redundancie [21].

V mnohých sieťach sú prítomné prepojené dáta (*linked data*). Dáta v sociálnych sieťach, kde sú používatelia prepojení so svojimi priateľmi, alebo dáta v biologických sieťach, kde existujú interakcie medzi proteínmi, poskytujú významné informácie vo forme prepojení, ktoré môžu byť rôznych typov. Tieto dáta predstavujú nové možnosti a výzvy pre výber príznakov.

III. UMELE NEURÓNNOVÉ SIETE A VÝBER PREMENNÝCH

Úspechy umelých neurónových sietí a hlbokého učenia pri riešení mnohých problémov strojového učenia často vedú k názoru, že je možné robiť pokročilé strojové učenie aj bez výberu príznakov. V niektorých prípadoch, hlavne keď nie je počet pozorovaní dostatočne veľký, by sa však hlboké učenie malo kombinovať s výberom premenných [5]. Tento prístup je v literatúre označovaný ako hlboké učenie založené na výbere premenných (*feature-selection-based deep learning*). Výber premenných na úrovni vstupu do siete znižuje komplexnosť modelu a uľahčuje jeho interpretáciu. Vedie k úsporám času a zdrojov a môže prispieť k zlepšeniu presnosti predikcie. [6]. Zároveň výskum ukazuje, že aj neurónové siete a hlboké učenie môžu byť základom pre vývoj nových metód výberu príznakov, čo sa označuje ako výber príznakov založený na hlbokom učení (*deep-learning-based feature selection*).

V literatúre možno nájsť mnoho algoritmov na výber príznakov založených na neurónových sieťach. Prehľad nedávno publikovaných metód poskytuje práca [22].

Viaceré metódy využívajú spätnú elimináciu. Štartujú trénovaním siete s úplnou množinou premenných. Potom vypočítajú ich dôležitosť (*saliency*) a odstránia premennú s najmenšou dôležitosťou. Ďalej pokračujú trénovaním podsiete so zostávajúcou množinou premenných a cyklus opakujú dovtedy, kým nie je splnené určité kritérium zastavenia. Na konci procesu poskytnú najlepšiu zo skúmaných podmnožín premenných. V [23] je dôležitosť premennej určená hodnotou účelovej funkcie vypočítanou po jej odstránení zo siete. Metóda [24] odstraňuje premenné na základe presnosti predikcie (*accuracy*) podsietí, ktoré vzniknú po ich vynechaní. Iná metóda založená na klasifikačnej presnosti [25] využíva doprednú selektciu.

Ďalšie miery dôležitosti zahŕňajú deriváciu účelovej funkcie, váhy hrán siete alebo oboje [26]. Príkladom takej miery je súčet druhých mocnín váh hrán medzi vstupným neurónom (premennou) a neurónmi prvej skrytej vrstvy. Dôležitosť premenných sa určí po natrénovaní siete iba s celou množinou premenných. Cieľom trénovania siete je minimalizácia účelovej funkcie, ktorá obvykle obsahuje aj regularizačné členy, napríklad kombinácie ℓ_1 a ℓ_2 regularizácie [27], [28] alebo $L_{2,p}$ regularizáciu [29], a to kvôli dosiahnutiu väčšieho počtu nulových váh a zabráneniu pretrénovaniu siete.

Príklad metódy výberu premenných na báze hlbokej neurónovej siete (*deep neural network* – DNN) možno nájsť v práci [27]. Medzi vstupnú a prvú skrytú vrstvu doprednej DNN je vložená riedka 1–1 lineárna vrstva, kde je vstupný uzol zodpovedajúci premennej X_i spojený len s i -tým uzlom tejto vrstvy a prepojenie má váhu w_i . Po natrénovaní siete s použitím regularizácie *elastic net* určuje vektor váh w dôležitosť jednotlivých premenných. Metóda bola použitá pri analýze DNA. Podobný prístup je uplatnený v prácach [30], [22], kde je použitá 1–1 vrstva s aktivačnou funkciou sigmoid, resp. ReLU (*rectified linear unit*). V práci [22] slúžia vybrané premenné na predikciu pozitívnych prípadov nákazy COVID-19. Ďalšia metóda na báze DNN [31] bola použitá pri rozpoznávaní akcií na videozázname. Výber premenných

je v tomto prípade založený na sieťovom pozitívnom príspevku premenných k aktivácii neurónov v prvej skrytej vrstve. Metóda [32] realizuje výber premenných na základe minimálnej rekonštrukčnej chyby v DBN (*deep belief network*) a bola použitá pri rozpoznávaní scén na satelitných snímkoch. V práci [33] autori navrhli metódu založenú na rekurentnej neurónovej sieti (RNN) a použili ju pri klasifikácii buniek pomocou premenných vybraných z údajov DNA čipov.

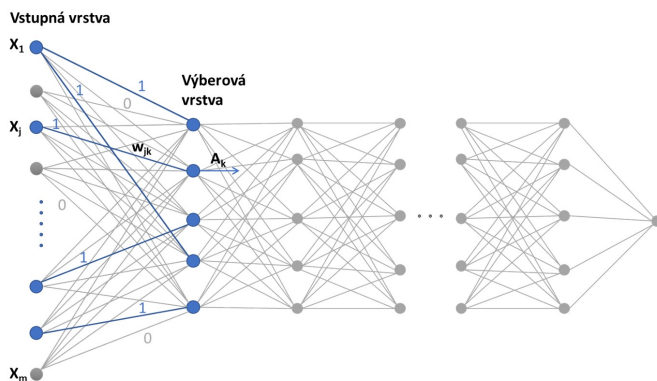
IV. NÁVRH METÓDY VÝBERU PREMENNÝCH ZALOŽENEJ NA RIEDKEJ VRSTVE NEURÓNOVEJ SIETE

V tejto práci je navrhnutá nová metóda výberu premenných nazvaná Výber premenných založený na riedkej vrstve neurónovej siete (SNeL-FS) s normalizačnými ohraničeniami [34].

Uvažujme doprednú neurónovú sieť, ktorá rieši danú klasifikačnú či regresnú úlohu. Za účelom výberu dôležitých vysvetľujúcich premenných (príznakov) bude medzi vstupnú a prvú skrytú vrstvu tejto siete vložená hustá skrytá vrstva, ktorá sa bude nazývať výberovou vrstvou. Výberová vrstva má dim neurónov, kde dim označuje požadovaný počet vybraných premenných, a ďalej je plne prepojená s prvou skrytou vrstvou pôvodnej neurónovej siete.

Cieľom učenia sa neurónovej siete s vloženou výberovou vrstvou pri výbere premenných je nájsť optimálne riešenie, ktoré súčasne vyhovuje nasledovným dvom podmienkam:

- Váhy hrán medzi vstupnou a výberovou vrstvou nadobúdajú iba hodnoty 1 a 0, kde 1 znamená, že zodpovedajúca premenná je vybraná príslušným neurónom výberovej vrstvy, a 0 znamená opak.
- Pre každý neurón výberovej vrstvy je súčet váh hrán vstupujúcich do tohto neurónu rovný 1.



Obr. 1. Výber premenných pomocou riedkej vrstvy neurónovej siete

Ak sú tieto podmienky splnené, každý neurón výberovej vrstvy vyberie práve jednu premennú. Teda váhy výberovej vrstvy možno interpretovať ako výber premenných. Poznamenajme, že dva neuróny výberovej vrstvy môžu vybrať tú istú premennú, a preto po eliminácii premenných s nulovými váhami bude pôvodný priestor premenných transformovaný do priestoru s najviac dim dimenziami. Príklad použitia výberovej vrstvy na výber premenných je na obrázku 1.

Zatiaľ čo podmienka ii) obmedzuje veľkosť váh, podmienka i) požaduje, aby boli hodnoty váh iba 0 alebo 1, čo sa zdá byť kritické, pretože taká podmienka nemôže byť splnená pomocou spojitých metód používaných v neurónových sieťach. Túto podmienku možno aproximovať požiadavkou na riedkosť riešenia vo výberovej vrstve. V porovnaní s inými publikáciami [27], [28], [29] je v tejto práci prezentovaný nový spôsob, ako riedke riešenie dosiahnuť.

A. Definícia a vlastnosti výberovej vrstvy

Uvažujme súbor dát (\mathbf{X}, \mathbf{y}) s vysvetľujúcimi premennými X_1, X_2, \dots, X_m a s cieľovou premennou \mathbf{y} . Predpokladáme, že súbor \mathbf{X} je štandardizovaný, to znamená, že pre každé $j \in \{1, \dots, m\}$ je priemer alebo stredná hodnota $E(X_j) = 0$ a štandardná odchýlka, ako aj rozptyl $\text{Var}(X_j) = 1$. Ďalej uvažujme doprednú neurónovú sieť vhodnú na riešenie daného klasifikačného alebo regresného problému. Navrhovaná metóda SNeL-FS s normalizačnými ohraničeniami modifikuje sieť vložením hustej výberovej vrstvy s dim neurónmi medzi jej vstupnú a prvú skrytú vrstvu. Výberová vrstva nepoužíva žiadnu nelineárnu aktivačnú funkciu, a preto reprezentuje lineárnu transformáciu vstupných premenných. Neuróny výberovej vrstvy nepoužívajú žiadnu prahovú hodnotu ($bias = 0$), takže parametre výberovej vrstvy sú len váhy hrán medzi vstupnou a výberovou vrstvou. Výstupy (aktivačné hodnoty) neurónov výberovej potom reprezentujú nové premenné definované takto:

$$A_k = \sum_{j=1}^m w_{jk} X_j \quad \text{pre } k = 1, \dots, dim,$$

kde w_{jk} označuje váhu hrany medzi j -tým neurónom vstupnej vrstvy a k -tým neurónom výberovej vrstvy.

Prezentovaná metóda výberu premenných vyberá dim dôležitých premenných dátového súboru \mathbf{X} podľa optimálnych váh výberovej vrstvy získaných tréningom modifikovanej neurónovej siete. Cieľom učenia sa neurónovej siete je minimalizovať danú účelovú funkciu, pričom váhy a aktivačné hodnoty výberovej vrstvy majú spĺňať nasledovné dva súbory podmienok:

$$\sum_{j=1}^m |w_{jk}| \leq 1 \quad \text{pre } k = 1, \dots, dim, \quad (1)$$

$$\text{Var}(A_k) \geq 1 \quad \text{pre } k = 1, \dots, dim. \quad (2)$$

V súbore podmienok (1) sú uvažované absolútne hodnoty váh, nakoľko vo všeobecnosti sú prípustné aj záporné váhy. Hodnota $\text{Var}(A_k)$ v súbore podmienok (2) je rozptyl (variancia) premennej A_k vypočítaný cez všetky pozorovania vstupného súboru dát \mathbf{X} alebo, v prípade použitia *minibatchov*, cez všetky pozorovania v každom *minibatchi* spracovanom počas procesu tréningu.

Dá sa dokázať [34], že ak je súbor dát \mathbf{X} štandardizovaný, a žiadne dve vysvetľujúce premenné nie sú úplne korelované, potom pre každý neurón výberovej vrstvy je riešením nerovností (1) a (2) množina vektorov váh hrán vstupujúcich do tohto neurónu, ktoré majú práve jednu nenulovú zložku

s absolútnou hodnotou 1. Z toho vyplýva, že sústava nerovností (1) a (2) v istom zmysle nahrádza podmienky i) a ii). Výberová vrstva teda realizuje výber premenných. V ideálnom prípade každý neurón výberovej vrstvy vyberie presne jednu premennú, ktorá prislúcha hrane s nenulovou váhou.

B. Model pre výber premenných

Uvažujme doprednú neurónovú sieť, ktorá tvorí základ pre použitie navrhovanej metódy výberu premenných. Predpokladajme, že model obsahuje H skrytých vrstiev a jeho parametre θ tvoria množinu $\{\mathbf{W}^1, \mathbf{b}^1, \dots, \mathbf{W}^{H+1}, \mathbf{b}^{H+1}\}$, kde \mathbf{W}^h je matica váh hrán medzi $(h-1)$ -ou a h -tou vrstvou a \mathbf{b}^h je vektor prahových hodnôt h -tej vrstvy pre $h = 1, \dots, H+1$. Nech minimalizovaná účelová funkcia má tvar

$$C(\theta) = C_0(\theta) + \lambda \sum_{h=1}^{H+1} \Omega(\mathbf{W}^h), \quad (3)$$

kde C_0 je funkcia vyjadrujúca priemernú chybu predikcie a $\sum_{h=1}^{H+1} \Omega(\mathbf{W}^h)$ je regularizačný člen vynásobený regularizačným parametrom $\lambda \in R_0^+$. Typ použitej funkcie chyby závisí od riešeného problému. Vo všeobecnosti môže byť aplikovaný aj iný spôsob regularizácie, ktorý zabráni možnému pretrénovaniu modelu.

Po vložení výberovej vrstvy medzi vstupnú a prvú skrytú vrstvu sa model mierne zmení, pričom jeho nové parametre sú $\tilde{\theta} = \{\mathbf{W}, \tilde{\mathbf{W}}^1, \tilde{\mathbf{b}}^1, \dots, \tilde{\mathbf{W}}^{H+1}, \tilde{\mathbf{b}}^{H+1}\}$, kde $\mathbf{W} = (w_{jk})$ je matica váh hrán prepájajúcich vstupnú a výberovú vrstvu. Nakoľko pridaná výberová vrstva má spĺňať podmienky (1) a (2), pôvodná optimalizačná úloha sa zmení na nasledovný optimalizačný problém s ohraničeniami:

$$\begin{aligned} \underset{\tilde{\theta}}{\text{minimalizuj}} \quad & C(\tilde{\theta}) = C_0(\tilde{\theta}) + \lambda \sum_{h=1}^{H+1} \Omega(\tilde{\mathbf{W}}^h) \text{ za podmienok} \\ & \sum_{j=1}^m |w_{jk}| - 1 \leq 0 \quad \text{pre } k = 1, \dots, \text{dim}, \quad (4) \\ & 1 - \text{Var}(A_k) \leq 0 \quad \text{pre } k = 1, \dots, \text{dim}. \end{aligned}$$

Problém (4) znamená minimalizovať účelovú funkciu C vzhľadom na parametre $\tilde{\theta}$, ktoré majú spĺňať vyššie definované dve množiny ohraničení. Pretože riešiť takýto problém je extrémne náročné, úloha je transformovaná na optimalizačný problém bez ohraničení, ktorého riešenie aproximuje riešenie pôvodného problému (4). Táto myšlienka je založená na Karush-Kuhn-Tuckerovom prístupe [35], ktorý zovšeobecňuje metódu Lagrangeových multiplikátorov.

Optimalizačný problém (bez ohraničení), riešenie ktorého bude použité na výber najdôležitejších premenných, bude po určitých úpravách definovaný nasledovne:

$$\begin{aligned} \underset{\tilde{\theta}}{\text{minimalizuj}} \quad & F(\tilde{\theta}) = C(\tilde{\theta}) + \lambda_S \Omega_S(\mathbf{W}) + \lambda_A \Omega_A(\mathbf{W}), \text{ kde} \\ & \Omega_S(\mathbf{W}) = \sum_{k=1}^{\text{dim}} \max(0, \sum_{j=1}^m |w_{jk}| - 1), \quad (5) \\ & \Omega_A(\mathbf{W}) = \sum_{k=1}^{\text{dim}} \max(0, 1 - \text{Var}(A_k)), \end{aligned}$$

kde $\lambda_S, \lambda_A \in \mathbb{R}_0^+$ sú dva multiplikátory, jeden pre každú množinu ohraničení.

Možno vidieť, že penalizačný člen Ω_S predstavuje slabšiu formu ℓ_1 regularizácie. Neurón výberovej vrstvy prispieva k penalizácii len vtedy, ak súčet absolútnych hodnôt váh hrán vstupujúcich do tohto neurónu prevyší hodnotu 1. Penalizačný člen Ω_A posilňuje regularizačný účinok člena Ω_S . Príspevok neurónu výberovej vrstvy k penalizácii Ω_A je kladný, iba ak je rozptyl jeho aktivačných hodnôt menší ako 1.

Proces tréningu siete minimalizuje pôvodnú účelovú funkciu C definovanú vztahom (3) vzhľadom na parametre $\tilde{\theta}$ a vyvažuje podmienky (1) a (2). Ak je pokuta Ω_S v určitej epoche tréningu siete kladná, v snahe minimalizovať ju, budú počas ďalších epoch znižované absolútne hodnoty príslušných váh výberovej vrstvy. To bude zároveň viesť k znižovaniu rozptylov aktivačných hodnôt dotknutých neurónov a k zvyšovaniu pokuty Ω_A . Naopak znižovanie pokuty Ω_A zapríčiňuje zvyšovanie absolútnych hodnôt váh výberovej vrstvy (teda aj pokuty Ω_S). Súčasne však musí byť minimalizovaná priemerná chyba predikcie, takže sa pravdepodobne budú zvyšovať hlavne váhy zodpovedajúce relevantným premenným. Počas tréningu siete teda dochádza k vyvažovaniu pokút Ω_A a Ω_S .

C. Miery dôležitosti premenných

Optimálny model získaný tréningom neurónovej siete s vloženou výberovou vrstvou poskytuje optimálne váhy $\tilde{\mathbf{W}}$ výberovej vrstvy. V ideálnom prípade matica váh $\tilde{\mathbf{W}} = (\tilde{w}_{jk})$ obsahuje dim nenulových váh, jednu pre každý neurón výberovej vrstvy. Potom každý neurón výberovej vrstvy vyberie presne jednu dôležitú vysvetľujúcu premennú, ktorej prislúcha nenulová váha. Poznamenajme, že viac neurónov môže vybrať tú istú premennú.

Vo všeobecnosti platí, že aj keď sa optimálne riešenie blíži k riedkemu riešeniu, počet nenulových váh matice $\tilde{\mathbf{W}}$ nemusí byť dostatočne malý na to, aby bolo možné jednoznačne vybrať nanajvýš dim premenných. Z toho dôvodu budú definované dve (alternatívne) miery dôležitosti premenných, ktoré umožnia výber požadovaného počtu najdôležitejších premenných. Obe miery sú odvodené z optimálnych váh výberovej vrstvy. Prvá využíva súčet váh a druhá je založená na maximálnej váhe.

1) *Dôležitosť určená sumou váh:* Najjednoduchšou definíciou dôležitosti premennej je súčet absolútnych hodnôt váh hrán vystupujúcich zo vstupného neurónu, ktorý danej premennej zodpovedá. Túto mieru možno upraviť na základe faktu, že vydelením všetkých váh hrán vstupujúcich do nejakého neurónu výberovej vrstvy konštantou a súčasným vynásobením všetkých váh hrán vystupujúcich z tohto neurónu v ďalšej vrstve tou istou konštantou sa riešenie z pohľadu predikcie nezmení.

Keďže nemožno vylúčiť prípady, že sa v optimálnom modeli štandardná odchýlka aktivačných hodnôt niektorého neurónu výberovej vrstvy významne odlišuje od hodnoty 1, na výpočet dôležitosti budú použité normalizované váhy. Teda pre každé $k \in \{1, \dots, \text{dim}\}$ budú váhy hrán vstupujúcich do k -teho neurónu výberovej vrstvy vydelené hodnotou $\text{std}(\hat{A}_k)$. Navyše, aby bola vylúčená závislosť hodnoty dôležitosti premennej od

počtu neurónov výberovej vrstvy, súčet normalizovaných váh bude vydelený hodnotou dim . Uvažujme premennú X_j pre $j \in \{1, \dots, m\}$. Dôležitosť určená sumou váh (*sum-weight saliency*) premennej X_j podľa váh \widehat{W} je definovaná takto:

$$S_{SW}(X_j, \widehat{W}) = \frac{1}{dim} \sum_{k=1}^{dim} \frac{|\widehat{w}_{jk}|}{std(\widehat{A}_k)}. \quad (6)$$

2) *Dôležitosť určená maximálnou váhou*: Druhý prístup je založený na predpoklade, že každý neurón výberovej vrstvy vyberie jednu premennú. Nezáleží na tom, či bola premenná vybraná raz alebo viackrát. Potom váhy hrán vstupujúcich do nejakého neurónu výberovej vrstvy môžu byť normalizované delením súčtom všetkých takýchto váh. Dôležitosť určená maximálnou váhou (*max-weight saliency*) premennej X_j podľa váh \widehat{W} je definovaná ako maximum zo všetkých normalizovaných váh hrán vychádzajúcich z neurónu vstupnej vrstvy, ktorý prislúcha premennej X_j :

$$S_{MW}(X_j, \widehat{W}) = \max_{k=1}^{dim} \frac{|\widehat{w}_{jk}|}{\sum_{i=1}^m |\widehat{w}_{ik}|}. \quad (7)$$

3) *Porovnanie mier dôležitosti*: Je ľahké vidieť, že za podmienok (1) a (2) sú hodnoty oboch definovaných mier dôležitosti premenných pre každú premennú súboru dát \mathbf{X} z intervalu $\langle 0, 1 \rangle$. Pre premenné, ktoré sú jasne vybrané neurónmi výberovej vrstvy, sú hodnoty dôležitosti určenej maximálnou váhou blízke 1. Hodnoty dôležitosti určenej sumou váh sú obyčajne menšie. Pri použití optimálnych váh výberovej vrstvy získaných tréningom neurónovej siete miera dôležitosti určená sumou váh detailnejšie rozlišuje dôležitosť jednotlivých premenných, zatiaľ čo druhá miera, založená na maximálnej váhe, významnejšie oddeľuje relevantné premenné od irelevantných.

V. VYHODNOTENIE EXPERIMENTÁLNYCH VÝSLEDKOV

Výkonnosť navrhutej metódy SNeL-FS s normalizačnými ohraničeniami bola skúmaná z dvoch dôležitých hľadísk, a to najprv z hľadiska ich schopnosti identifikovať známe vysvetľujúce premenné dôležité pre určenie cieľovej premennej a potom z hľadiska vplyvu výberu premenných na presnosť predikcie. Schopnosť určiť známe relevantné premenné bola testovaná na ôsmich umelo generovaných súboroch dát. Experimenty na hodnotenie zlepšenia predikčnej výkonnosti po aplikovaní navrhnutých metód boli realizované na dvanástich verejne dostupných reálnych vysokorozmerných súboroch dát. Pomocou numerických experimentov bola demonštrovaná použiteľnosť metódy pre binárnu klasifikáciu, pre klasifikáciu do viacerých tried, ako aj pre regresné úlohy.

Experimentálne výsledky navrhutej metódy boli porovnané s výsledkami troch často používaných tradičných metód výberu premenných, a to mRMR s kritériom vzájomná informácia (s označením mRMR/MI) [36], ReliefF [37] a f-score [38].

Prezentované výsledky metódy SNeL-FS s normalizačnými ohraničeniami boli získané pomocou implementácie realizovanej s použitím efektívneho výpočtového rámca TensorFlow [39]. Implementačné aspekty sú podrobne popísané v práci [34].

A. Experimentálne výsledky pre syntetické dáta

Porovnanie metód výberu premenných z hľadiska ich schopnosti objaviť relevantné premenné bolo realizované na syntetických súboroch dát. Súborny Madelon a XOR predstavujú binárne klasifikačné úlohy a súborny LinReg a Friedman regresné úlohy [40]. V prezentovaných experimentoch boli tieto súborny použité v dvoch variantoch, ktoré sa líšia počtom pozorovaní. Súborny dát s malým počtom pozorovaní (označené ako Mad, Xor, Reg a Fri) pozostávajú z 200 pozorovaní a súborny s veľkým počtom pozorovaní (Mad5k, Xor5k, Reg5k a Fri5k) obsahujú 5 000 pozorovaní. Všetky súborny dát majú 500 vysvetľujúcich premenných. Ich základné charakteristiky sú popísané v tabuľke I. V stĺpcoch tabuľky sú postupne uvedené: názov súboru, jeho skratka, počet pozorovaní, počet vysvetľujúcich premenných a za ním v zátvorke počet relevantných premenných.

Tabuľka I
CHARAKTERISTIKY SYNTETICKÝCH SÚBOROV DÁT

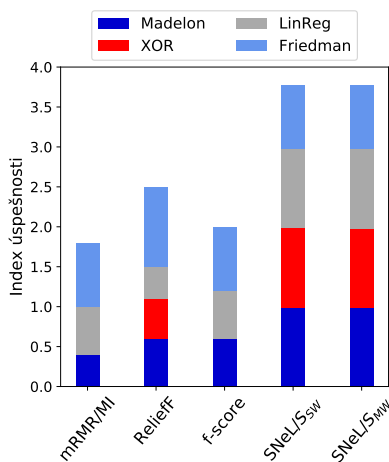
Názov	Skratka	Pozorovania	Premenné (N_{inf})
Madelon	Mad	200	500 (5)
XOR	Xor	200	500 (2)
LinReg	Reg	200	500 (5)
Friedman	Fri	200	500 (5)
Madelon 5k	Mad5k	5 000	500 (5)
XOR 5k	Xor5k	5 000	500 (2)
LinReg 5k	Reg5k	5 000	500 (5)
Friedman 5k	Fri5k	5 000	500 (5)

Na hodnotenie postupností vybraných premenných bol použitý index úspešnosti (*Suc.* skóre), ktorý je definovaný podľa [41] vzťahom:

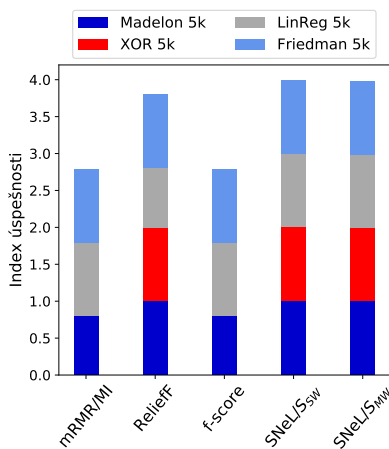
$$Suc. = \frac{R_s}{R_t} - \alpha \frac{I_s}{I_t}. \quad (8)$$

Táto miera vyjadruje podiel počtu známych relevantných premenných vybraných danou metódou R_s a celkového počtu známych relevantných premenných R_t , ktorý je znížený o počet za výber irelevantných premenných, určenú podielom počtu vybraných irelevantných premenných I_s a celkového počtu irelevantných premenných v súbore dát I_t . Koeficient $\alpha = \min\left\{\frac{1}{2}, \frac{R_t}{I_t}\right\}$ zabezpečí lepšie hodnotenie pre prípad, že je do výberu zahrnutá irelevantná premenná, ako pre prípad chýbajúcej relevantnej premennej. Ak je výber premenných danou metódou reprezentovaný ich zoradením podľa dôležitosti (*ranking*) a všetky TF sú vybrané na prvých miestach, potom je metóda ohodnotená maximálnou hodnotou indexu $Suc. = 1$, teda 100%.

Hodnoty indexu úspešnosti pre hodnotené metódy výberu premenných pri výbere 15 premenných sú graficky znázornené na obrázku 2 pre súborny s malým počtom pozorovaní a na obrázku 3 pre súborny s veľkým počtom pozorovaní. Pre metódu SNeL-FS s normalizačnými ohraničeniami sú uvádzané výsledky pre obe miery dôležitosti, a to dôležitosť určenú sumou váh (tento variant metódy je označený ako SNeL/ S_{SW}) a dôležitosť určenú maximálnou váhou (SNeL/ S_{MW}).



Obr. 2. Index úspešnosti pre syntetické súbory s malým počtom pozorovaní



Obr. 3. Index úspešnosti pre syntetické súbory s veľkým počtom pozorovaní

Pre oba typy súborov dosiahli najlepšie výsledky varianty metódy SNeL-FS s normalizačnými ohraničeniami SNeL/S_{SW} a SNeL/S_{MW}. Spomedzi tradičných metód sa k výsledkom SNeL-FS najviac priblížil filter ReliefF. Pre súbory s väčším počtom pozorovaní sú hodnoty indexu úspešnosti všeobecne vyššie ako jeho hodnoty pre súbory s malým počtom pozorovaní, nakoľko väčší počet pozorovaní pomáha metódam odhaliť skutočné závislosti v dátach. Detailnejšia analýza odhaľuje slabé miesta univariančnej metódy f-score, ktorá zlyhala najmä pri identifikovaní relevantných premenných v súboroch XOR. Takisto filter mRMR/MI nedokázal objaviť dve premenné, ktoré spolu určujú hodnotu cieľovej premennej, ale samostatne sa javia ako irelevantné. Všeobecne možno konštatovať, že výsledky navrhutej metódy na syntetických dátach preukázali jej použiteľnosť pre klasifikačné aj regresné úlohy. Oba jej varianty identifikovali všetky relevantné premenné vo všetkých súboroch okrem súboru Fri, kde odhalili 80% zo všetkých relevantných premenných.

B. Experimentálne výsledky pre reálne HDSSS dáta

Vplyv navrhutej metódy výberu premenných na predikčný výkon algoritmov strojového učenia bol hodnotený na verejne dostupných údajoch z DNA čipov (DNA *microarrays*), ktoré reprezentujú vysokorozmerné klasifikačné úlohy charakteristické malým počtom pozorovaní, nevyváženými triedami a zložitou dát [42]. DNA čipy obsahujú veľké množstvo údajov zo vzoriek tkanív a buniek, ktoré môžu odhaliť zmeny v sekvencii DNA. Na základe týchto znalostí možno diagnostikovať rôzne choroby, rozlišovať typy nádorov a podobne.

Experimenty boli realizované na ôsmich súboroch pre binárnu klasifikáciu a štyroch súboroch pre klasifikáciu s viacerými triedami. Ich základné vlastnosti sú popísané v tabuľke II. Prvé tri stĺpce obsahujú: názov súboru vrátane zdroja, počet pozorovaní (pacientov, od ktorých pochádzajú vzorky) a počet premenných (skúmaných génov). Posledný stĺpec obsahuje počet tried a v zátvorke počet pozorovaní (pacientov) v jednotlivých triedach.

Tabuľka II
CHARAKTERISTIKY REÁLNYCH SÚBOROV DÁT

Názov [Zdroj]	Pacienti	Gény	Triedy (pacienti)
Alon [43]	62	2 000	2 (40, 22)
Burczynski [44]	127	22 283	2 (85, 42)
Chin [45]	118	22 215	2 (43, 75)
Chowdary [46]	104	22 283	2 (62, 42)
Golub [47]	72	7 129	2 (47, 25)
Gordon [48]	181	12 533	2 (94, 87)
Singh [49]	102	12 600	2 (52, 50)
Tian [50]	173	12 625	2 (36, 137)
Khan [51]	83	2 309	4 (29, 11, 18, 25)
Nutt [52]	50	12 625	4 (14, 7, 14, 15)
Armstrong [53]	72	12 533	3 (24, 20, 28)
Bhattacharjee [54]	203	12 600	5 (139, 17, 6, 21, 20)

Pri hodnotení boli v experimentoch použité štyri všeobecne známe algoritmy strojového učenia, ktoré pracujú na rôznych princípoch, a to: naivný bayesovský klasifikátor (*Gaussian Naive Bayes* – NB), metóda podporných vektorov (*Support Vector Machine* – SVM) s radiálnym kernelom (*radial basis function* – RBF) a penalizačným parametrom $C = 1$, náhodný les (*Random Forest* – RF) zložený z 1 000 rozhodovacích stromov používajúcich entropiu ako kritérium kvality rozdelenia a metóda k najbližších susedov (*k-nearest neighbors* – k-NN) s hodnotou parametra $k = 5$. Pri implementácii boli použité algoritmy strojového učenia z knižnice *scikit-learn* [40].

Pri porovnávaní metód výberu premenných na reálnych súboroch dát bol aplikovaný experimentálny protokol, ktorý zabraňuje vychýleniu odhadu presnosti predikcie klasifikátorov pri využití metód výberu premenných. Bola použitá stratifikovaná k -násobná krížová validácia (*cross validation* – CV), kde k bolo pre binárne súbory dát nastavené na 10 a pre súbory dát s viacerými triedami len na 5 kvôli malej početnosti niektorých tried. Vyhodnocované metódy výberu premenných boli zahrnuté do CV slučky. Pre každú iteráciu CV slučky bola metóda výberu premenných spustená na príslušnej tréningovej

množine a klasifikátor bol trénovaný na tréningovej množine zúženej na vybrané premenné. Na vyhodnotenie presnosti predikcie bola použitá príslušná testovacia množina. Celkový odhad presnosti celého procesu bol určený ako priemer presností získaných z jednotlivých iterácií.

Pretože problémom väčšiny skúmaných súborov dát je nevyváženosť tried, na meranie presnosti predikcie bolo použité F_1 skóre. Táto miera je definovaná ako harmonický priemer precision a recall, t. j. $F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ a vyjadruje rovnováhu medzi nimi. Pri klasifikačných úlohách s viacerými triedami bolo aplikované vážené F_1 skóre.

Porovnávaný bol variant metódy SNeL-FS s normalizačnými ohraničeniami pre dôležitosť určenú maximálnou váhou SNeL/ S_{MW} s tromi tradičnými metódami. Hodnoty F_1 skóre dosiahnuté použitím 30 premenných vybraných aplikovaním týchto metód sú prezentované v tabuľke III. Pre každý súbor je pre každý klasifikátor zobrazené F_1 skóre vypočítané ako priemer so štandardnou odchýlkou z CV. Zároveň je v riadkoch s označením avg zobrazený priemer získaný z výsledkov všetkých štyroch klasifikátorov. Najlepšie výsledky pre každý súbor sú zvýraznené. Kvôli porovnaniu sú uvedené aj hodnoty F_1 skóre dosiahnuté bez výberu premenných.

V poslednom riadku tabuľky III sú prezentované výsledky WTL (*wintie/loss*) štatistiky, ktoré predstavujú počet súborov dát, pre ktoré je priemerné F_1 skóre získané po aplikácii metódy SNeL/ S_{MW} väčšie ako, rovné alebo menšie ako priemerné F_1 skóre dosiahnuté pomocou príslušnej tradičnej metódy. Výsledky ukazujú, že SNeL-FS výrazne prevyšuje porovnávané tradičné metódy výberu premenných. Použitím znamienkového testu [55] je nulová hypotéza, že SNeL-FS a akákoľvek porovnávaná metóda sú ekvivalentné, zamietnutá na hladine významnosti $\alpha = 0,05$.

Súhrnná štatistika WTL v poslednom stĺpci posledného riadku tabuľky III porovnáva priemerné F_1 skóre metódy SNeL-FS s najlepším priemerným skóre spomedzi troch klasických metód FS pre každý konkrétny súbor údajov. Tabuľka ukazuje, že metóda SNeL-FS je takmer vždy lepšia ako konvenčné metódy. Výnimku predstavujú súbory Chin a Gordon. Avšak výsledky pre súbor Gordon sú veľmi vyrovnané a všetky metódy dosiahli priemerné F_1 skóre približne 99%. Najlepšie priemerné skóre získané pomocou metódy mRMR prekonal priemerné skóre SNeL-FS o menej ako 0,1%. Na druhej strane metóda SNeL-FS dosiahla na tomto súbore dát najlepšie skóre s klasifikátormi SVC a kNN. V prípade súboru Chin sú výsledky tiež vyrovnané. Metódy SNeL-FS a f-score získali rovnaké najlepšie priemerné F_1 skóre 91,57%.

Možno konštatovať, že navrhovaná metóda fungovala pri použitej konfigurácii [34] lepšie z hľadiska F_1 skóre ako ostatné hodnotené metódy. To je v súlade s výsledkami získanými na syntetických dátach, kde metóda SNeL-FS vykazovala vyššiu úspešnosť pri detekcii relevantných premenných.

Pri porovnaní výsledkov z hľadiska dvoch navrhovaných mier dôležitosti boli pre reálne súbory dát dosiahnuté lepšie hodnoty F_1 skóre použitím miery založenej na maximálnej váhe. Pokiaľ ide o syntetické súbory dát, boli výsledky *Suc.* skóre pre obe miery takmer rovnaké.

VI. ZÁVER

Výber premenných (príznakov) je dôležitým krokom pri príprave dát na spracovanie metódami strojového učenia. Redukciou dimenzionality vysokorozmerných súborov prispieva k šetreniu výpočtových zdrojov, skráteniu trvania výpočtu a v neposlednom rade aj k zlepšeniu presnosti predikcie pri použití algoritmov učenia. Neurónové siete majú prirodzene zabudovanú schopnosť redukovat dimenzionalitu vstupných dát, avšak na rozdiel od metód výberu premenných neposkytujú priamo relevantné premenné spomedzi pôvodných premenných. Na druhej strane môžu byť použité ako základ pre metódy výberu premenných, a tým zachovávajú dôležité originálne premenné, čo zlepšuje čitateľnosť a interpretovateľnosť modelov.

Táto práca predstavila novú metódu výberu premenných pre kontrolované učenie, SNeL-FS, ktorá využíva dopredné neurónové siete na výber dôležitých vstupných premenných. SNeL-FS konštruje špeciálnu sieťovú vrstvu, ktorá je prostredníctvom sústavy normalizačných ohraničení donútená v procese tréningovania siete mnoho svojich váh vynulovať. Na vyhodnotenie dôležitosti premenných boli definované dve nové miery dôležitosti založené na váhach výberovej vrstvy.

Numerické experimenty boli realizované na dvoch typoch údajov: na ôsmich syntetických súboroch údajov a dvanástich vysokorozmerných reálnych súboroch dát. Výsledky ukazujú, že navrhovaná metóda je vhodná pre klasifikačné problémy s rôznym počtom tried, ako aj pre regresné problémy. Na synteticky generovaných súboroch metóda SNeL-FS účinne identifikovala relevantné premenné. Výsledky na reálnych súboroch dát dokázali, že metóda je schopná efektívne znížiť dimenzionalitu a dosiahnuť najlepší predikčný výkon z hľadiska priemerného skóre F_1 . Všetky experimenty boli vykonané použitím efektívnej implementácie metódy s využitím GPU.

V prezentovaných experimentoch tvoria základ pre použitie metódy SNeL-FS dopredné neurónové siete so štandardnými hustými vrstvami. Metóda umožňuje ako základ využiť aj reziduálne siete. Rozšírenie metódy na ďalšie špecifické typy neurónových sietí, ako napríklad konvolučné neurónové siete (CNN), nie je úplne priamočiare. Všeobecne sa výber premenných (pixelov) v obrázkoch, kde sa objekty záujmu môžu vyskytovať na ľubovoľnom mieste, nejaví ako prínosný. Užitočný by mohol byť vtedy, keď sú obrázky rovnakej veľkosti a objekty sú centrovane, napríklad v medicínskom zobrazovaní. Avšak ani v tomto prípade sa vloženie hustej výberovej vrstvy do CNN nezdá byť vhodné, pretože CNN pracuje so špecifickou štruktúrou premenných v tvare dvoj- alebo viacrozmernej mapy a vynechávanie premenných môže túto štruktúru, kľúčovú pre CNN, narušiť.

Všeobecne sa dá SNeL-FS považovať za východisko pre návrh ďalších metód výberu premenných pomocou nových spôsobov zabezpečenia riedkosti výberovej vrstvy. Zároveň je možné optimalizovať výkonnosť navrhovanej metódy zmenou architektúry použitej základnej neurónovej siete a spôsobu tréningovania modifikovanej siete, zmenou kritérií výberu optimálneho modelu a podobne.

Tabuľka III
 F_1 SKÓRE ZÍSKANÉ NA REÁLNYCH SÚBOROCH DÁT PRE 30 VYBERANÝCH PREMENNÝCH

Súbor dát	Klasif.	bez výberu	mRMR/MI	reliefF	f-score	SNeL/S _{MW}
Alon	NB	0.5867 ± 0.1384	0.7624 ± 0.1992	0.7957 ± 0.1632	0.7850 ± 0.1624	0.7790 ± 0.1829
	SVC	0.5300 ± 0.3761	0.6924 ± 0.2838	0.7957 ± 0.1632	0.7757 ± 0.1486	0.8290 ± 0.1674
	RF	0.6867 ± 0.2810	0.8090 ± 0.1575	0.7757 ± 0.1486	0.7900 ± 0.1620	0.7757 ± 0.1486
	kNN	0.4267 ± 0.3518	0.7257 ± 0.1488	0.6957 ± 0.2753	0.7757 ± 0.1486	0.7790 ± 0.1829
	avg	0.5575 ± 0.2868	0.7474 ± 0.1973	0.7657 ± 0.1876	0.7816 ± 0.1554	0.7907 ± 0.1704
Burczynski	NB	0.7887 ± 0.2007	0.9294 ± 0.0813	0.7654 ± 0.1214	0.9020 ± 0.0923	0.9126 ± 0.0845
	SVC	0.8403 ± 0.2276	0.9603 ± 0.0612	0.8909 ± 0.1096	0.9524 ± 0.0590	0.9635 ± 0.0564
	RF	0.8873 ± 0.1459	0.9181 ± 0.1117	0.8083 ± 0.1329	0.9210 ± 0.0859	0.9159 ± 0.1029
	kNN	0.8644 ± 0.0863	0.9429 ± 0.0700	0.8667 ± 0.1116	0.9496 ± 0.0836	0.9635 ± 0.0564
	avg	0.8452 ± 0.1651	0.9377 ± 0.0810	0.8328 ± 0.1189	0.9313 ± 0.0802	0.9389 ± 0.0750
Chin	NB	0.8895 ± 0.0616	0.9055 ± 0.0606	0.9058 ± 0.0749	0.9124 ± 0.0748	0.9124 ± 0.0748
	SVC	0.8941 ± 0.0567	0.9150 ± 0.0594	0.9215 ± 0.0521	0.9215 ± 0.0521	0.9163 ± 0.0525
	RF	0.9163 ± 0.0525	0.9249 ± 0.0542	0.9198 ± 0.0607	0.9073 ± 0.0556	0.9139 ± 0.0553
	kNN	0.8876 ± 0.0605	0.8955 ± 0.0525	0.9138 ± 0.0451	0.9215 ± 0.0521	0.9201 ± 0.0531
	avg	0.8969 ± 0.0578	0.9102 ± 0.0567	0.9152 ± 0.0582	0.9157 ± 0.0586	0.9157 ± 0.0589
Chowdary	NB	0.8816 ± 0.0984	0.9201 ± 0.0707	0.9657 ± 0.0698	0.9556 ± 0.1018	0.9746 ± 0.0513
	SVC	0.9460 ± 0.0667	0.9455 ± 0.0716	0.9464 ± 0.0864	0.9639 ± 0.0786	0.9746 ± 0.0513
	RF	0.9496 ± 0.0836	0.9746 ± 0.0513	0.9464 ± 0.0864	0.9746 ± 0.0513	0.9746 ± 0.0513
	kNN	0.9103 ± 0.0989	0.9460 ± 0.0667	0.9492 ± 0.0630	0.9746 ± 0.0513	0.9635 ± 0.0564
	avg	0.9219 ± 0.0869	0.9466 ± 0.0651	0.9519 ± 0.0764	0.9672 ± 0.0708	0.9718 ± 0.0526
Golub	NB	0.9800 ± 0.0600	0.9217 ± 0.1234	0.9350 ± 0.1001	0.9350 ± 0.1001	0.9350 ± 0.1001
	SVC	0.6233 ± 0.3426	0.9514 ± 0.0756	0.9514 ± 0.0756	0.9057 ± 0.1821	0.9657 ± 0.0698
	RF	0.9800 ± 0.0600	0.9067 ± 0.1200	0.9657 ± 0.0698	0.9200 ± 0.1833	0.9600 ± 0.0800
	kNN	0.6300 ± 0.2747	0.9200 ± 0.1833	0.9200 ± 0.1833	0.8867 ± 0.1956	0.9200 ± 0.1833
	avg	0.8033 ± 0.1843	0.9249 ± 0.1256	0.9430 ± 0.1072	0.9118 ± 0.1653	0.9452 ± 0.1083
Gordon	NB	0.9901 ± 0.0212	1.0000 ± 0.0000	0.9825 ± 0.0286	0.9894 ± 0.0227	0.9931 ± 0.0138
	SVC	0.9864 ± 0.0166	0.9933 ± 0.0134	0.9968 ± 0.0097	0.9935 ± 0.0129	0.9968 ± 0.0097
	RF	0.9968 ± 0.0097	0.9968 ± 0.0097	0.9935 ± 0.0129	0.9968 ± 0.0097	0.9935 ± 0.0129
	kNN	0.9471 ± 0.0271	0.9903 ± 0.0148	0.9808 ± 0.0210	0.9935 ± 0.0129	0.9935 ± 0.0129
	avg	0.9801 ± 0.0187	0.9951 ± 0.0095	0.9884 ± 0.0180	0.9933 ± 0.0146	0.9942 ± 0.0123
Singh	NB	0.7120 ± 0.0825	0.9075 ± 0.1058	0.9018 ± 0.0985	0.9184 ± 0.0996	0.9184 ± 0.0996
	SVC	0.8826 ± 0.1043	0.9070 ± 0.0876	0.9181 ± 0.0916	0.9206 ± 0.0926	0.9290 ± 0.0829
	RF	0.9226 ± 0.0921	0.9305 ± 0.0953	0.9290 ± 0.0829	0.9215 ± 0.0877	0.9270 ± 0.0836
	kNN	0.8277 ± 0.1043	0.9247 ± 0.0538	0.9124 ± 0.0838	0.9270 ± 0.0836	0.9290 ± 0.0829
	avg	0.8362 ± 0.0958	0.9174 ± 0.0856	0.9153 ± 0.0892	0.9219 ± 0.0909	0.9259 ± 0.0873
Tian	NB	0.8472 ± 0.0624	0.8465 ± 0.0561	0.7854 ± 0.0696	0.8243 ± 0.0603	0.8427 ± 0.0507
	SVC	0.8843 ± 0.0115	0.8736 ± 0.0406	0.8435 ± 0.0539	0.8658 ± 0.0454	0.8907 ± 0.0324
	RF	0.8843 ± 0.0115	0.8796 ± 0.0474	0.8849 ± 0.0246	0.8782 ± 0.0311	0.8885 ± 0.0198
	kNN	0.8794 ± 0.0267	0.8720 ± 0.0220	0.8786 ± 0.0305	0.8782 ± 0.0415	0.8830 ± 0.0272
	avg	0.8738 ± 0.0280	0.8679 ± 0.0415	0.8481 ± 0.0447	0.8616 ± 0.0446	0.8762 ± 0.0325
Khan	NB	0.9726 ± 0.0338	0.9482 ± 0.0264	0.7971 ± 0.1188	0.9526 ± 0.0447	0.9868 ± 0.0264
	SVC	0.9279 ± 0.0525	0.9511 ± 0.0247	0.8319 ± 0.1314	0.9769 ± 0.0283	0.9869 ± 0.0261
	RF	1.0000 ± 0.0000	0.9767 ± 0.0285	0.8576 ± 0.0988	0.9887 ± 0.0226	0.9887 ± 0.0226
	kNN	0.7938 ± 0.0699	0.9399 ± 0.0635	0.8170 ± 0.0855	1.0000 ± 0.0000	0.9756 ± 0.0300
	avg	0.9236 ± 0.0390	0.9540 ± 0.0358	0.8259 ± 0.1086	0.9795 ± 0.0239	0.9845 ± 0.0263
Nutt	NB	0.6863 ± 0.0752	0.6235 ± 0.1354	0.5620 ± 0.1408	0.6655 ± 0.1764	0.7174 ± 0.1093
	SVC	0.6131 ± 0.0815	0.6300 ± 0.1394	0.6428 ± 0.1579	0.6758 ± 0.1011	0.7552 ± 0.0746
	RF	0.6474 ± 0.1307	0.7104 ± 0.1737	0.6024 ± 0.1551	0.6498 ± 0.0805	0.7753 ± 0.0969
	kNN	0.5711 ± 0.1014	0.6028 ± 0.1971	0.7438 ± 0.0415	0.7226 ± 0.1300	0.7583 ± 0.0542
	avg	0.6295 ± 0.0972	0.6417 ± 0.1614	0.6378 ± 0.1238	0.6784 ± 0.1220	0.7515 ± 0.0837
Armstrong	NB	0.9598 ± 0.0328	0.9714 ± 0.0352	0.9714 ± 0.0352	0.9264 ± 0.0492	0.9868 ± 0.0264
	SVC	0.8913 ± 0.0545	0.9430 ± 0.0554	0.9321 ± 0.0422	0.9587 ± 0.0337	0.9598 ± 0.0328
	RF	0.9868 ± 0.0264	0.9576 ± 0.0348	0.9593 ± 0.0333	0.9587 ± 0.0337	0.9730 ± 0.0330
	kNN	0.9072 ± 0.0670	1.0000 ± 0.0000	0.9450 ± 0.0276	0.9431 ± 0.0287	0.9550 ± 0.0590
	avg	0.9363 ± 0.0452	0.9680 ± 0.0313	0.9519 ± 0.0346	0.9468 ± 0.0363	0.9687 ± 0.0378
Bhattacharjee	NB	0.8847 ± 0.0450	0.9166 ± 0.0355	0.8076 ± 0.0309	0.7946 ± 0.0598	0.9262 ± 0.0452
	SVC	0.8589 ± 0.0566	0.9289 ± 0.0508	0.7824 ± 0.0618	0.7481 ± 0.0492	0.9162 ± 0.0325
	RF	0.8835 ± 0.0682	0.9315 ± 0.0276	0.7968 ± 0.0705	0.8686 ± 0.0528	0.9363 ± 0.0235
	kNN	0.8767 ± 0.0471	0.8996 ± 0.0641	0.8293 ± 0.0752	0.8572 ± 0.0418	0.9163 ± 0.0430
	avg	0.8759 ± 0.0542	0.9191 ± 0.0445	0.8040 ± 0.0596	0.8171 ± 0.0509	0.9238 ± 0.0360
WTL	avg		11/0/1	12/0/0	11/1/0	10/1/1

LITERATURA

- [1] J. Li and H. Liu, "Challenges of feature selection for big data analytics," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 9–15, 03 2017. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7887649>
- [2] V. Bolón-Canedo, N. Sánchez-Maróño, and A. Alonso-Betanzos, "Feature selection for high-dimensional data," *Progress in Artificial Intelligence*, vol. 5, no. 2, pp. 65–75, 05 2016. [Online]. Available: <https://doi.org/10.1007/s13748-015-0080-y>
- [3] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research (JMLR)*, vol. 3, no. 3, pp. 1157–1182, 2003.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016, [Online] [cit. 2023-03-28] Dostupné na internete: <http://www.deeplearningbook.org>.
- [5] Y. Li, T. Li, and H. Liu, "Recent advances in feature selection and its applications," *Knowledge and Information Systems*, vol. 53, no. 3, pp. 551–577, 12 2017. [Online]. Available: <https://doi.org/10.1007/s10115-017-1059-8>
- [6] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70 – 79, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231218302911>
- [7] Y. Li, J. Si, G. Zhou, S. Huang, and S. Chen, "FREL: A stable feature selection algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 7, pp. 1388–1402, 07 2015.
- [8] B. Pes, N. Dessi, and M. Angioni, "Exploiting the ensemble paradigm for stable feature selection: A case study on high-dimensional genomic data," *Information Fusion*, vol. 35, pp. 132 – 147, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1566253516300847>
- [9] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, "Ensemble feature selection: Homogeneous and heterogeneous approaches," *Knowledge-Based Systems*, vol. 118, pp. 124–139, 02 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950705116304749>
- [10] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, 12 2017. [Online]. Available: <https://doi.org/10.1145/3136625>
- [11] Z. Zhao, J. Wang, S. Sharma, N. Agarwal, H. Liu, and Y. Chang, *An integrative approach to identifying biologically relevant genes*, 2010, pp. 838–849. [Online]. Available: <https://epubs.siam.org/doi/abs/10.1137/1.9781611972801.58>
- [12] X. Wu, X. Zhu, G. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97–107, 01 2014.
- [13] B. Cao, L. He, X. Kong, P. S. Yu, Z. Hao, and A. B. Ragin, "Tensor-based multi-view feature selection with applications to brain diseases," in *2014 IEEE International Conference on Data Mining*, 12 2014, pp. 40–49.
- [14] G. Lastra, O. Luaces, J. R. Quevedo, and A. Bahamonde, "Graphical feature selection for multilabel classification tasks," in *Advances in Intelligent Data Analysis X*, J. Gama, E. Bradley, and J. Hollmén, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 246–257.
- [15] W. Chen, J. Yan, B. Zhang, Z. Chen, and Q. Yang, "Document transformation for multi-label feature selection in text categorization," in *Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, ser. ICDM '07. USA: IEEE Computer Society, 2007, p. 451–456. [Online]. Available: <https://doi.org/10.1109/ICDM.2007.18>
- [16] J. Wang, M. Wang, P. Li, L. Liu, Z. Zhao, X. Hu, and X. Wu, "Online feature selection with group structure analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, p. 3029–3041, 11 2015. [Online]. Available: <http://dx.doi.org/10.1109/TKDE.2015.2441716>
- [17] K. Yu, X. Wu, W. Ding, and J. Pei, "Scalable and accurate online feature selection for big data," *ACM Trans. Knowl. Discov. Data*, vol. 11, no. 2, pp. s. 16:1–16:39, 12 2016. [Online]. Available: <http://doi.acm.org/10.1145/2976744>
- [18] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [19] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *Journal of Machine Learning Research*, vol. 12, 04 2009.
- [20] S. Kim and E. Xing, "Tree-guided group lasso for multi-task regression with structured sparsity," *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 543–550, 01 2010.
- [21] P. Ristoski and H. Paulheim, "Feature selection in hierarchical feature spaces," in *Discovery Science*, S. Džeroski, P. Panov, D. Kocov, and L. Todorovski, Eds. Cham: Springer International Publishing, 2014, pp. 288–300.
- [22] N. E. El-Attar, S. F. Sabbeh, H. Fasihuddin, and W. A. Awad, "An improved deepnn with feature ranking for covid 19 detection," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 2249–2269, 2022. [Online]. Available: <http://www.techscience.com/cmc/v71n2/45833>
- [23] E. Romero and J. M. Sopena, "Performing feature selection with multilayer perceptrons," *IEEE Transactions on Neural Networks*, vol. 19, no. 3, pp. 431–441, 03 2008.
- [24] R. Setiono and H. Liu, "Neural-network feature selector," *IEEE Transactions on Neural Networks*, vol. 8, no. 3, pp. 654–662, 05 1997.
- [25] V. Onnia, M. Tico, and J. Saarinen, "Feature selection method using neural network," in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, vol. 1, 10 2001, pp. 513–516.
- [26] L. M. Belue and K. W. Bauer, "Determining input features for multilayer perceptrons," *Neurocomputing*, vol. 7, no. 2, pp. 111–121, 03 1995. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0925231294E0053T>
- [27] Y. Li, C.-Y. Chen, and W. W. Wasserman, "Deep feature selection: Theory and application to identify enhancers and promoters," in *Research in Computational Molecular Biology*, T. M. Przytycka, Ed. Cham: Springer International Publishing, 2015, pp. 205–217.
- [28] N. Challita, M. Khalil, and P. Beausery, "New feature selection method based on neural network and machine learning," in *2016 IEEE International Multidisciplinary Conference on Engineering Technology*, 11 2016, pp. 81–85.
- [29] R. Wang, J. Bian, F. Nie, and X. Li, "Nonlinear feature selection neural network via structured sparse regularization," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2022.
- [30] V. Borisov, J. Haug, and G. Kasneci, "Cancelout: A layer for feature selection in deep neural networks," in *International conference on artificial neural networks*. Springer, 2019, pp. 72–83.
- [31] D. Roy, K. S. R. Murty, and C. K. Mohan, "Feature selection using deep neural networks," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 07 2015, pp. 1–6.
- [32] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2321–2325, 11 2015.
- [33] S. Chowdhury, X. Dong, and X. Li, "Recurrent neural network based feature selection for high dimensional and low sample size micro-array data," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 4823–4828.
- [34] P. Bugata and P. Drotár, "Feature selection based on a sparse neural-network layer with normalizing constraints," *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 161–172, 2023.
- [35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [36] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005. [Online]. Available: <https://ieeexplore.ieee.org/document/1453511>
- [37] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, no. 1, pp. 23–69, 2003.
- [38] B. Weir, "Estimating f-statistics: A historical view," *Philosophy of Science*, vol. 79, no. 5, pp. 637–643, 09 2012.
- [39] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <http://tensorflow.org/>

- [40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [41] V. Bolón-Canedo, N. Sánchez-Maróño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and Information Systems*, vol. 34, no. 3, pp. 483–519, 2013.
- [42] V. Bolón-Canedo, N. Sánchez-Maróño, and A. Alonso-Betanzos, "Data classification using an ensemble of filters," *Neurocomputing*, vol. 135, pp. 13–20, 01 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231213011405>
- [43] U. Alon, N. Barkai, D. Notterman, K. Gish, S. Ybarra, D. Mack, and A. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences*, vol. 96, no. 12, pp. 6745–6750, 06 1999.
- [44] M. E. Burczynski, R. L. Peterson, N. C. Twine, K. A. Zuberek, B. J. Brodeur, L. Casciotti, V. Maganti, P. S. Reddy, A. Strahs, F. Immermann, W. Spinelli, U. Schwertschlag, A. M. Slager, M. M. Cotreau, and A. J. Dornier, "Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells," *The Journal of Molecular Diagnostics*, vol. 8, no. 1, pp. 51–61, 02 2006.
- [45] K. Chin, S. DeVries, J. Fridlyand, P. T. Spellman, R. Roydasgupta, W.-L. Kuo, A. Lapuk, R. M. Neve, Z. Qian, T. Ryder, F. Chen, H. Feiler, T. Tokuyasu, C. Kingsley, S. Dairkee, Z. Meng, K. Chew, D. Pinkel, A. Jain, B. M. Ljung, L. Esserman, D. G. Albertson, F. M. Waldman, and J. W. Gray, "Genomic and transcriptional aberrations linked to breast cancer pathophysiology," *Cancer Cell*, vol. 10, no. 6, pp. 529–541, 12 2006.
- [46] D. Chowdary, J. Lathrop, J. Skelton, K. Curtin, T. Briggs, Y. Zhang, J. Yu, Y. Wang, and A. Mazumder, "Prognostic gene expression signatures can be measured in tissues collected in RNAlater preservative," *The Journal of Molecular Diagnostics*, vol. 8, no. 1, pp. 31–39, 02 2006.
- [47] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, 10 1999.
- [48] G. J. G. Gordon, R. V. R. Jensen, L.-L. L. Hsiao, S. R. S. Gullans, J. E. J. Blumenstock, S. S. Ramaswamy, W. G. W. Richards, D. J. D. Sugarbaker, and R. R. Bueno, "Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma," *Cancer Research*, vol. 62, no. 17, pp. 4963–4967, 09 2002.
- [49] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub, and W. R. Sellers, "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203 – 209, 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1535610802000302>
- [50] E. Tian, F. Zhan, R. Walker, E. Rasmussen, Y. Ma, B. Barlogie, and J. D. Shaughnessy, Jr., "The role of the Wnt-signaling antagonist DKK1 in the development of osteolytic lesions in multiple myeloma," *New England Journal of Medicine*, vol. 349, no. 26, pp. 2483–2494, 12 2003.
- [51] J. Khan, J. S. Wei, M. Ringnér, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, and P. S. Meltzer, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, no. 6, pp. 673–679, 2001.
- [52] C. L. Nutt, D. R. Mani, R. A. Betensky, P. Tamayo, J. G. Cairncross, C. Ladd, U. Pohl, C. Hartmann, M. E. McLaughlin, T. T. Batchelor, P. M. Black, A. Von Deimling, S. L. Pomeroy, T. R. Golub, and D. N. Louis, "Gene expression-based classification of malignant gliomas correlates better with survival than histological classification," *Cancer Research*, vol. 63, no. 7, pp. 1602–1607, 2003.
- [53] S. A. Armstrong, J. E. Staunton, L. B. Silverman, R. Pieters, M. L. den Boer, M. D. Minden, S. E. Sallan, E. S. Lander, T. R. Golub, and S. S. J. Korsmeyer, "M11 translocations specify a distinct gene expression profile that distinguishes a unique leukemia," *Nature Genetics*, vol. 30, pp. 41–47, 2002.
- [54] A. Bhattacharjee, W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson, "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses," *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13 790–13 795, 2001.
- [55] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 12 2006. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1248547.1248548>

Vybrané aspekty analýzy kvality povrchových úprav dosiek plošných spojov

Daniel Dzivý

Katedra technológií v elektronike
Fakulta elektrotechniky a informatiky, TUKE
Košice, Slovensko
Email: daniel.dzivyuke.sk

Alena Pietriková

Katedra technológií v elektronike
Fakulta elektrotechniky a informatiky, TUKE
Košice, Slovensko
Email: alena.pietrikovauke.sk

Abstrakt—Článok poskytuje komplexný prehľad o najmodernejších poznatkoch v oblasti povrchových úprav dosiek plošných spojov, povrchových úprav a rozhraní, diskusie o najnovších pokrokoch a súčasných problémoch a poukazuje na perspektívny vývoj v tejto oblasti. Článok sa zameriava na procesy povrchových úprav a povlakov a analýzu ich vlastností v oblasti elektroniky. Vybrané aspekty posudzovania kvality povrchových úprav DPS boli realizované so zreteľom na zmáčanie, spájkovateľnosť, vývoj mikroštruktúry na medzifázovom rozhraní a mechanické, krehkolomové a elektrické vlastnosti spájkovaných spojov. Okrem materiálovo-technologických problémov výroby štandardných povrchových úprav DPS boli v rámci článku analyzované novo-vyvinuté povrchy DPS na báze SnAg.

KLúčové slová—povrchové úpravy, doska plošného spoja, SnAg zliatina, zmáčanie, shear test, kontaktný uhol

Abstract— The article provides a comprehensive overview of the state-of-the-art knowledge in the field of surface finishes of printed circuit boards, surface treatments and interfaces, discusses the latest advances and current problems, and points to prospective developments in this area. The work focuses on the processes of surface finishes and coatings and the analysis of their properties in the field of electronics. Selected aspects of assessing the quality of PCB surface finishes were realized regarding to wetting, solderability, microstructure development at the interphase interface and mechanical, brittle, and electrical properties of soldered joints. In addition to the material-technological problems of the production of standard PCB surface treatments, newly developed PCB surfaces based on SnAg were analyzed in the work.

I. ÚVOD

Na kvalitu spájkovaného spoja (mechanické vlastnosti, elektrické vlastnosti, spoľahlivosť) vplyva množstvo faktorov, z ktorých za najdôležitejšie považujeme:

- typ aplikovanej spájky z pohľadu chemického zloženia použitej zliatiny a z pohľadu typu aplikovaného materiálu (kompaktný kov, guľôčka spájky alebo spájkovacia pasta),
- typ povrchovej úpravy z pohľadu chemického zloženia a z pohľadu technológie výroby a spracovania,

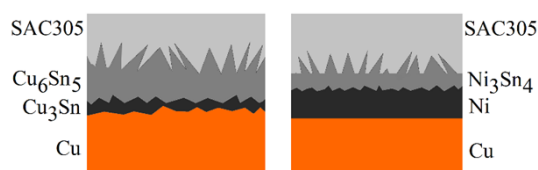
- zmáčavosť povrchových úprav DPS a tiež súčiastok, ktorá v konečnom dôsledku ovplyvňuje spájkovateľnosť.

Typ povrchovej úpravy DPS ovplyvňuje schopnosť DPS odolávať vonkajším vplyvom (vlhkosť, teplota, vibrácie, prach, agresívne plyny) a vplyva na vznik potenciálnych defektov spájkovaného spoja (voidy, viskre, odzmáčanie), na jeho dlhodobú kvalitu a spoľahlivosť a tiež na jeho mechanické a elektrické vlastnosti. Typ povrchovej úpravy DPS a typ aplikovanej spájkovacej zliatiny (spájkovacej pasty) spolu súvisia a pri posudzovaní vlastností spájkovaného spoja ich nemožno od seba oddeliť, nakoľko pri spájkovaní dochádza na medzifázovom rozhraní k metalurgickým reakciám. Jedným z kritérií pri výbere typu povrchovej úpravy DPS a tiež typu aplikovanej spájkovacej zliatiny, resp. spájkovacej pasty, je jej cena. Pri výbere povrchovej úpravy alebo typu spájkovacej zliatiny je potrebné dbať na fakt, že použitie zliatiny s vyšším obsahom drahších kovov má síce výhodu v lepšej zmáčavosti, lepšej odolnosti voči nárastu intermetalických zlúčenín (IMC), lepšej skladovateľnosti a lepšej odolnosti voči korózii, ale to sa odrazí vo vyššej cene. Jedným z problémov, ktorý vzniká v súvislosti s typom povrchovej úpravy DPS, je signálová strata (útlm signálu) v oblasti vysokých frekvencií. Strata a skreslenie signálu v oblasti vysokých frekvencií (min na úrovni GHz) súvisí s počtom a hrúbkou vrstiev aplikovaných pri povrchovej úprave DPS. Z tohto dôvodu je povrchová úprava ENIG, aj keď je najpoužívanejšia, nevhodná pre tieto aplikácie. Použitie ostatných homogénne nanášaných povrchových úprav DPS (OSP, ImSn, Im Ag) je pre účely použitia v oblasti vysokých frekvencií možné, avšak sprevádza ich množstvo ďalších materiálovo-technologických problémov, ktoré je potrebné v dnešnej dobe riešiť. Nezanedbateľným faktom je aj hrúbka a planarita nanesej povrchovej úpravy, ktorá má význam najmä v oblasti miniaturizovaných modulov elektroniky. Úlohou je nájsť optimálne riešenie pre konkrétnu aplikáciu pri výrobe [1–4].

A. Rast vrstiev intermetalických zlúčenín

Medzi základné problémy, ku ktorým dochádza počas výroby ale aj počas života spojov elektroniky patrí rýchlosť tvorby a rastu IMC na rozhraní spájky a povrchovej úpravy. IMC priamo ovplyvňujú mechanické vlastnosti a spoľahlivosť elektronických obvodov, nakoľko tvoria mechanické a

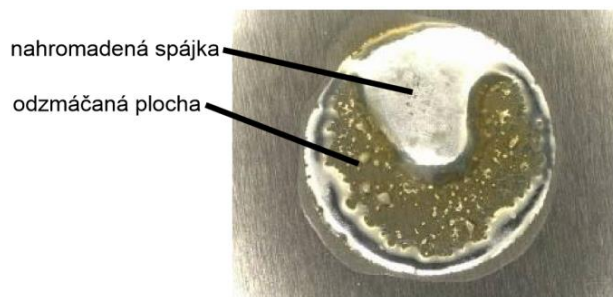
elektrické spojenie medzi elektronickými súčiastkami a prepojovacími cestami (DPS). Problémom je, že väčšina IMC počas doby života elektronických zariadení rastie, následkom čoho môžu vzniknúť praskliny a trhliny na rozhraní spájky a DPS, čím sa zvýši náchylnosť na krehké porušenie spájkovaného spoja. Rast IMC je podmienený zvýšenou teplotou a následnou difúziou. Teplotné cyklovanie vzniká bežným používaním elektroniky, takže mu nie je možné zabrániť. Niektoré typy IMC na báze Sn-Cu (Cu_6Sn_5) zvyčajne rastú rýchlo, pričom sa vytvorí krehká vrstva IMC Cu_3Sn (Obr. 1), ktorá rastie na úkor Cu_6Sn_5 počas celej doby pôsobenia teploty, v dôsledku čoho spájkovaný spoj krehne a to najmä po opakovanom procese pretavenia. Jedným z riešení potlačenia ich rastu je vytvorenie difúznej bariéry a to pridaním vrstvy kovu, ktorého aktivačná energia je menšia ako aktivačná energia medi. Najčastejšie používaným kovom, ktorý pôsobí v úlohe difúznej bariéry, je nikel, ktorý sa využíva pri povrchovej úprave ENIG [5–11]. Rozdiel v zložení IMC a usporiadaní daných vrstiev spájkovaného spoja po pretavení je na Obr. 1.



Obr. 1 Zloženie spájkovaného spoja na rozhraní spájky a medi (vľavo), na rozhraní spájky a niklu (vpravo)

B. Nezmáčanie/odzmáčanie

Problém s odzmáčaním alebo s nezmáčaním priamo súvisí s vytváraním spoja na rozhraní spájky povrchovej úpravy, resp. spájkovaného povrchu. Hlavným viditeľným znakom tohto problému je, že spájka nezmáča podkladový materiál či už na súčiastke, alebo na povrchu DPS (Obr. 2).



Obr. 2 Spájkovacia plôška, na ktorých nastal problém s odzmáčaním spájky na väčšine jej plochy

Problém s odzmáčaním môže nastať počas procesu opakovaného spájkovania, avšak častejšie vzniká počas opráv elektronických zariadení, kedy je potrebné opätovne prispájkovať porušené spájkované spoje. Najčastejším prejavom je vytváranie guľôčok spájky a taktiež stmavnutie spájkovacích plôšok. V takom prípade nedôjde k adhéznemu spojeniu súčiastky a substrátu a dochádza k vytvoreniu studeného spoja. Ako je možné vidieť na Obr. 2, pri správnom vytvorení spoja je povrch spájky hladký s malým uhlom zmáčania, no pri nezmáčaní alebo pri odzmáčaní vyzerá

spájkovaný nedokonalou a kontaktný uhol medzi spájkou a povrchom je veľký. Pre účely porovnávania kvality zmáčavosti spájkovaných spojov slúži norma J-STD-001, ktorou sú kladené požiadavky na výrobcov pre spájkované spoje. Medzi najčastejšie dôvody tohto problému patrí korózia, ktorá vzniká počas procesu spájkovania z dôvodu vysokej teploty a nedostatočnej čistiacej schopnosti taviva odstrániť oxidy z povrchu DPS počas spájkovania. Je potrebné poznamenať, že výber povrchovej úpravy taktiež podmieňuje vznik oxidov a korózie na povrchu DPS. Z tohto dôvodu sa na povrchy aplikujú ochranné kovové a nekovové vrstvy z dôvodu potlačenia vytvárania oxidov a z dôvodu korózie, ktoré vznikajú počas skladovania alebo prepravy DPS alebo počas viacnásobného pretavenia. Ďalšou príčinou je chemická kontaminácia spôsobená zásahom do výrobného procesu (zlá manipulácia, zlé skladovacie podmienky), ktorá zabráňuje spájke zmáčať povrch DPS, alebo taktiež zle nastavený výrobný proces (black pad). Pre tento účel boli vykonané testy zmáčania a spájkovateľnosti guľôčok spájky a porovnaný vplyv viac násobného procesu pretavenia [12–16].

C. Tvorba voidov

Jedným z problémov, ktoré sa často rieši pro analýze kvality spájkovaných spojov sú voidy, t.j. prázdne miesta v spájkovanom spoji, ktoré majú vplyv na mechanickú pevnosť a prípadný vznik krehkého porušenia, na elektrické vlastnosti, prestup tepla, spoľahlivosť, starnutie a pod. Príčina ich vzniku, ich množstvo, poloha v objeme spájkovaného spoja ako aj ich veľkosť môžu byť rôzne. Najčastejšou príčinou ich vzniku je prítomnosť taviva v spájkovacej paste. Povrchová úprava môže tiež vplývať na tvorbu voidov. Ide o malé mikrovoidy, ktoré vznikajú na rozhraní medzi povrchovou úpravou DPS a spájkou ako dôsledok medzifázovej reakcie spojenej s tvorbou intermetalických zlúčenín. Adhéznym mechanizmus pri tvorbe spoja je výrazne ovplyvnený typom povrchovej úpravy DPS ako aj spôsobom tepelného spracovania. Jedným z dôvodov vzniku voidov z pohľadu povrchových úprav sú anomálie vzniknuté pri výrobe planárnych povrchových úprav (ENIG, ImAg, OSP) spojené s nedokonalosťami v ich výrobnom procese (zvyšky fosforu pri nanášaní niklu, zvyšky kyseliny sírovej pri nanášaní medi alebo zlé odparená vrstva OSP) ale aj nedostatočné čistenie medi pred nanosením povrchových úprav. Ako je možné vidieť na Obr. 3, planárne mikro-voidy vznikajú na rozhraní medzi spájkou a povrchovou úpravou, čím sa zhoršia mechanické a elektrické vlastnosti vytvoreného spájkovaného spoja. Vznik voidov môže taktiež nastať, keď difúzia jedného kovu do druhého je rýchlejšia než opačne (meď difunduje do cínu rýchlejšie ako cín do medi). Tieto Kirkendallove voidy tak vznikajú najmä na rozhraní povrchovej úprava/spájka (viď. Obr. 3) a to aj dôsledkom pohybu rozhrania medzi kovmi alebo zliatinou a kovom, no najmä na miestach, kde je rozdielna rýchlosť difúzie. V elektronike dochádza k vzniku Kirkendallových voidov aj v dôsledku elektromigrácie. Kirkendallove voidy predstavujú riziko spoľahlivosti spájkovaných spojov v oblasti krehkých rôznych binárnych a ternárnych vrstiev IMC. Nakoľko ich priemer môže dosahovať až 1-3 μm , predstavujú kritický problém pri použití súčiastok s jemnými rozstupmi. Vzniku

voidov úplne nie je možné zabrániť a vyskytuje sa v prípade olovnatých aj bezolovnatých spájkok. Maximálna plocha voidov pri hornom pohľade na spájkovaný spoj je stanovený na 25% z celej plochy spájkovaného spoja podľa normy IPC-A-610 (Class 2) alebo IPC-7095. Požiadavky pre náročné alebo vesmírne aplikácie avšak môžu byť ešte vyššie. Výber povrchovej úpravy vplýva na výskyt voidov maximálne na úrovni 8% z celej plochy spájkovaného spoja. Na vytváranie voidov majú najväčší vplyv spájkovacie pasty, ktorých zloženie výrobcovia často prispôbujú typu povrchovej úpravy. V každom prípade ich tvorba súvisí s rýchlosťou tvorby IMC počas spájkovania. Problém voidov nie je v oblasti mikroelektroniky (na rozdiel od výkonovej elektroniky) až taký vážny, avšak medzi základné dôsledky voidov patrí zhoršená tepelná vodivosť spájkovaných spojov, znížená schopnosť viesť elektrický prúd (najmä pre malé spoje) a zhoršená spoľahlivosť a znížená mechanická pevnosť v prípade niektorých typov voidov [17–19].



Obr. 3 Ukážka výskytu planárnych mikro-voidov na rozhraní spájky a povrchovej úpravy [19]

V predloženom článku sa sústreďujeme komparatívnu analýzu štandardných povrchových úprav s neštandardnými, doteraz nepoužívanými a to z pohľadu analýzy:

- zmáčavosti
- elektrických vlastností,
- mikroštruktúry a tvorby IMC,
- mechanických a krehkolomových vlastností.

II. POUŽITÉ MATERIÁLY

V rámci článku sú analyzované bežne dostupné povrchové úpravy (Cu, ENIG, ImSn, LFHASL, PbHASL) s novovyvinutými povrchovými úpravami na báze zliatiny SnAg (SnAg7, SnAg7DOC, SnAg5, SnAg5ENi). Povrchové úpravy sú rozdeľované podľa použitých materiálov a z pohľadu typu vytvorenej IMC na medzifázovom rozhraní. Spájkované spoje s povrchovou úpravou Cu, ImSn, LFHASL, PbHASL a SnAg7DOC obsahujú IMC na báze Cu_6Sn_5 a Cu_3Sn a spájkované spoje s povrchovou úpravou ENIG, SnAg7, SnAg5 a SnAg5ENi obsahujú IMC na báze $(Cu,Ni)_6Sn_5$, $(Ni,Cu)_3Sn_4$ a Ni_3Sn_4 .

A. Bežne dostupné povrchové úpravy

Nakoľko základným materiálom pre vodivé prepojenia na DPS tvorí meď, tak povrchová úprava Cu je teda len samotná meď bez ďalších vrstiev kovu.

Povrchová úprava ImSn pozostáva z vrstvy chemicky nanoseného cínu na povrchu medi. Cín je vylučovaný

chemicky a to ponorom DPS do zlučiny cínu a kyseliny sírovej.

Pri povrchovej úprave LFHASL a PbHASL je naopak proces nanášania úplne odlišný. Tieto povrchové úpravy boli vyrobené ponorením DPS do roztavenej spájky, ktorej prebytočné množstvo bolo následne ofúknuté horúcovzdušnými nožmi z dôvodu menších rozdielov v hrúbke. Tieto povrchové úpravy sa líšia len materiálovým zložením (PbHASL – olovnatá spájka, LFHASL – bezolovnatá spájka).

Povrchová úprava ENIG pozostáva z dvoch vrstiev kovov: chemicky nanosená vrstva Ni-P a chemicky nanosená vrstva zlata. Význam vrstvy Ni-P spočíva v zamedzení tvorby nežiadúcich typov IMC počas procesu spájkovania. Zlato slúži ako ochranná vrstva pre vrstvu Ni-P.

B. Povrchové úpravy na báze SnAg

Zliatina SnAg sa v dnešných časoch využíva ako povrchová úprava nanášaná spôsobom HASL. V rámci našich analýz sme avšak použili nanášanie vrstvy SnAg pomocou riadenej elektrolyzy. Tento postup umožňuje pomocou zmeny parametrov meniť zloženie vylučovanej zliatiny. Povrchové úpravy na báze SnAg sú vyhotovené v štyroch rôznych prevedeniach:

- SnAg7 pozostáva z vrstvy Ni-P s hrúbkou 2,5 μm a z vrstvy zliatiny SnAg7 s hrúbkou 1,5 μm ,
- SnAg7DOC pozostáva len z vrstvy zliatiny SnAg7 priamo na medi,
- SnAg5 pozostáva z vrstvy Ni-P s hrúbkou 2,5 μm a z vrstvy zliatiny SnAg5 s hrúbkou 1,5 μm ,
- SnAg5ENi pozostáva z vrstvy elektrolyticky nanoseného Ni s hrúbkou 2,5 μm a z vrstvy zliatiny SnAg5 s hrúbkou 1,5 μm .

Každá z týchto povrchových úprav má svoje výhody a nevýhody v závislosti od spôsobu vyhotovenia, ktoré sa prejavia počas analýz.

III. METÓDY RIEŠENIA

Ako už bolo spomenuté, naša analýza bola zameraná na zmáčanie, mikroštruktúru, elektrické vlastnosti a na mechanické a krehkolomové vlastnosti.

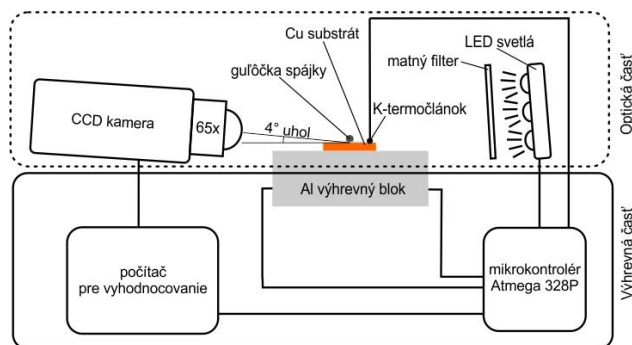
A. Analýza zmáčania

V rámci tejto analýzy sme vyvinuli zariadenie na meranie kontaktných uhlov, ktorým je možné merať kontaktný uhol počas procesu pretavenia guľôčky spájky (in-situ). Na Obr. 4 je bloková schéma zariadenia a jeho hlavné súčasti. Zariadenie pozostáva z optickej a výhrevnej časti.

Optickú časť tvorí CCD kamera so zväčšením 65x smerovaná na guľôčku spájky pod uhlom 4°. Oproti kamere je nasmerovaný zdroj svetla pomocou LED z dôvodu dosiahnutia jasnejšej kontúry obrysu guľôčky spájky. Aby svetlo bolo rovnomerne rozložené, pred LED svetlami sa nachádza matný filter. Na základe dostupných vedeckých publikácií v oblasti merania kontaktných uhlov je stanovená hodnota presnosti pre meranie kontaktného uhla $\pm 1^\circ$ ([20, 21]).

Výhrevnú časť zariadenia obsahuje hliníkový výhrevný blok, ktorého stred tvorí kanthalový odporový drôt izolovaný keramikou. Maximálny výkon ohrevu je 50W a postačuje na účely nízko-teplotného spájkovania. Ohrev je riadený pomocou mikrokontroléra Atmega 328P, ktorý spína relé. Vzorka má z hornej časti umiestnený termočlánok z dôvodu merania teploty. Pri dosiahnutí požadovanej teploty sa ohrev vypne. Zariadenie je kompatibilné s olovnatými aj bezolovnatými spájkami, nakoľko v nastavení je možné meniť maximálnu teplotu. Pri použití spájky SAC305 je teplota nastavená na 245 °C a pri použití spájky SnPb je maximálna teplota nastavená na 210 °C.

Hlavnou výhodou vyvinutého zariadenia je meranie kontaktného uhla priamo in-situ počas procesu pretavenia guľôčky spájky. Z bočného pohľadu je okrem toho možné vyhodnotiť aj koeficient zmáčania na základe výšky a priemeru pretavenej guľôčky spájky. Výhodou merania dynamickej zmeny kontaktného uhla v reálnom čase počas procesu pretavovania je určenie času roztekania v oblasti TAL (Time Above Liquidus). Pre tento účel sme definovali čas pre aktívne zmáčanie, ktorý reprezentuje interval od počiatku tavenia až kým koeficient zmáčania nedosiahne 80% svojej maximálnej hodnoty.



Obr. 4 Schematické znázornenie usporiadania aparatury pre in-situ meranie kontaktného uhla

B. Analýza mikroštruktúry

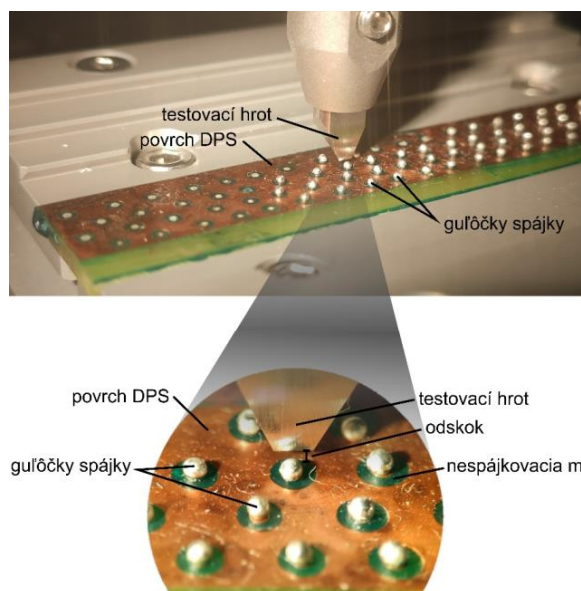
Na rozhraní povrchová úprava DPS - spájka prebiehajú pri spájkovaní procesy, ktoré ovplyvňujú kvalitu a spoľahlivosť spájkovaného spoja, rozhodujú o spájkovateľnosti povrchov a tvorbe IMC na rozhraní povrchu DPS a spájky. Povrchová úprava DPS prichádza do styku s roztavenou spájkou ako prvá. Hrúbky vrchných vrstiev planárnych povrchových úprav sú zvyčajne veľmi tenké. Tieto vrstvy (zlato, striebro, cín, zliatina SnAg) sa aplikujú nielen najmä za účelom zlepšenia spájkovateľnosti, ale slúžia tiež na ochranu spájkovaného povrchu voči korózii a iným vplyvom počas doby skladovania. Počas prvej fázy spájkovania dochádza k predifundovaniu tejto ochrannej vrstvy (povrchovej úpravy) do objemu spájky. Následne počas procesu spájkovania sa na rozhraní spájky a spájkovaného povrchu vytvára reakčná zóna, výsledkom ktorej sú v závislosti od typu povrchovej úpravy rôzne typy a hrúbky IMC. Vrstvy IMC sa nielen v priebehu procesu spracovania menia, ale menia sa aj dlhodobým používaním elektronického zariadenia, t.j. starnutím môže dôjsť k ich degradácii. IMC zásadným spôsobom vplyvajú na mechanické a elektrické vlastnosti spájkovaných spojov, ako aj na jeho kvalitu z

pohľadu spoľahlivosti a starnutia. Je známe, že tvrdosť IMC je väčšia ako tvrdosť samotnej spájky či základového materiálu, preto je veľmi dôležité poznať ich hrúbku, a teda vedieť posúdiť, do akej miery vrstva IMC ovplyvňuje kvalitu spájkovaného spoja.

Analýza typu, tvorby a hrúbky IMC bola v našej práci realizovaná prostredníctvom mikroskopických analýz. Pomocou týchto výsledkov analýzy tvorby IMC bolo možné zmerať hrúbky jednotlivých typov IMC na rozhraní spájky a základového materiálu (typu povrchovej úpravy). Hrúbka a morfológia IMC je dôležitým ukazovateľom ako zmáčavosti, tak aj mechanických a elektrických vlastností.

C. Analýza mechanických a krehkolomových vlastností

V našej práci sme sa zamerali na mechanickú pevnosť spájkovaných spojov, analýzu energie spotrebovanej na porušenie spoja a na mikroskopické SEM analýzy lomových plôch a to z pohľadu vplyvu jednotlivých typov povrchových úprav DPS. Výsledky mechanických vlastností sme porovnávali s výsledkami testov zmáčavosti a výsledkami hrúbok IMC vykonaných pomocou analýzy mikroštruktúry spájkovaného spoja a to pred a po umelom starnutí. Ako už bolo viackrát uvedené, povrchové úpravy DPS ovplyvňujú kvalitu adhézneho spoja. Každý typ povrchovej úpravy vplyva na tvorbu IMC a tým pádom aj na mechanické vlastnosti spoja iným spôsobom.



Obr. 5 Detail uchopenia vzorky a detail shear testu

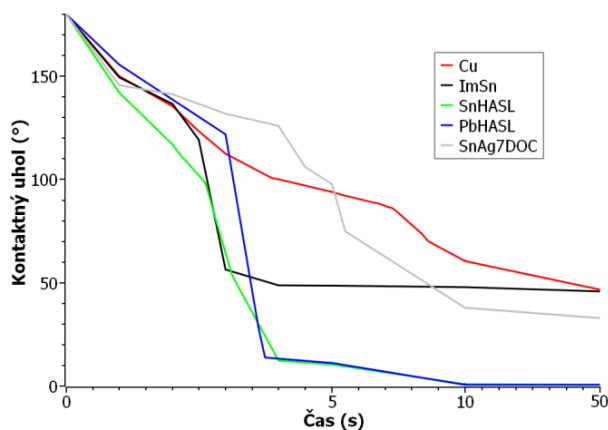
Pri realizácii shear testu prístrojom DAGE 4000 dochádzalo len k húževnatým, alebo kvázi-húževnatým porušeniam a to najmä z dôvodu použitia malej rýchlosti posunu vzorky voči testovaciemu hrotu, ktorá bola 100 $\mu\text{m/s}$. Prístroj DAGE 4000 síce umožňoval použiť aj väčšiu zaťažovaciu rýchlosť (600 $\mu\text{m/s}$), avšak naše predbežné testy ukázali, že pri vyššej rýchlosti zaťažovania dochádzalo k príliš veľkému rozptylu nameraných hodnôt. Da sa predpokladať, že pri väčšej rýchlosti zaťažovania sa náchylnosť na krehké

porušenie prejaví v omnoho vyššej miere. Záber na detail uchopenia je na Obr. 5. Na citlivé meranie sily počas shear testu bola použitá meracia kazeta S20KG, ktorá obsahovala testovací hrot prispôbený pre testovanie shear testom. Výsledné priebehy sily v závislosti od posunutia sa zobrazovali na monitore ihneď po meraní. V okolí guľôčky spájky je zelená nespájkovacia maska na vymedzenie plochy spájkovaného spoja, čím sa zabezpečila opakovateľnosť experimentu. Veľkosť odsokku bola nastavená na 80 μm .

IV. NAMERANÉ VÝSLEDKY A ICH DISKUSIA

A. Kontaktný uhol a zmäčavosť povrchov

Hlavný dôraz v našich experimentoch sme kládli na analýzu časového priebehu zmeny kontaktného uhla (in-situ merania), pričom sme porovnávali medzi sebou jednotlivé povrchové úpravy. Ako prvé sme porovnávali povrchové úpravy, resp. povrchy, na ktorých sa nenachádza vrstva niklu ako bariéra pre difúziu: Cu, ImSn, SnHASL, PbHASL, SnAg7DOC. Priebeh zmeny kontaktného uhla v závislosti od času a typu povrchových úprav, na ktorých vzniká spájkovaný spoj s meďou, sú na Obr. 6. Je možné vidieť, že povrchové úpravy na báze LFHASL a PbHASL vykazujú omnoho kratší čas zmäčania než ostatné povrchové úpravy. K ustáleniu hodnôt kontaktného uhla dôjde takmer okamžite (do 4 sekúnd). Príčinu môžeme hľadať v identických povrchových napätiach oboch stýkajúcich sa materiálov. Je možné predpokladať, že dôjde k úplnému rozliatiu spájky po povrchu. V prípade povrchovej úpravy ImSn došlo spočiatku k rýchlemu poklesu kontaktného uhla (čas pre aktívne zmäčanie do 4 s), nakoľko došlo k úplnému predifundovaniu tenkej vrstvy cínu (2 μm) do spájky. To, že výsledný kontaktný uhol dosiahol úroveň len 45° svedčí o tom, že po predifundovaní vrstvy cínu do spájky, začína pri tvorbe spoja prebiehať difúzia medi za súčasnej tvorby IMC.

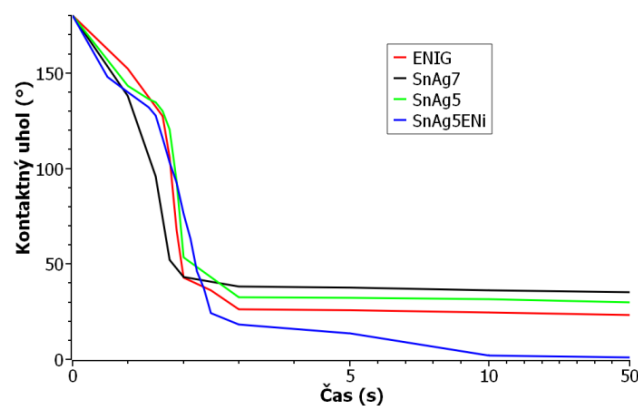


Obr. 6 Časový priebeh zmeny kontaktného uhla pre povrchové úpravy bez niklovej difúznej bariéry

Takýto výsledok merania považujeme za veľmi dôležitý, nakoľko ho bolo možné zmerať len na princípe in-situ merania. Podobný výsledok bol dosiahnutý v prípade povrchovej úpravy na báze Cu, avšak na rozdiel od ImSn, kontaktný uhol klesal po roztavení podstatne pomalšie a ustálil sa až po 50 s. Čas pre aktívne zmäčanie v oblasti TAL bol v tomto prípade 7,5 s. Z

grafu na Obr. 6 vidieť, že povrchová úprava na báze SnAg7DOC dosiahla nižší kontaktný uhol než povrchová úprava ImSn, avšak za dlhší čas. V prípade SnAg7DOC došlo k postupnému (pomalému) poklesu kontaktného uhla, pričom k jeho ustáleniu na úroveň cca 30° došlo po 10 sekundách. V prípade ImSn sme pozorovali okamžitý pokles kontaktného uhla (už po 4 sekundách), avšak jeho hodnota sa ustálila na úrovni 40°.

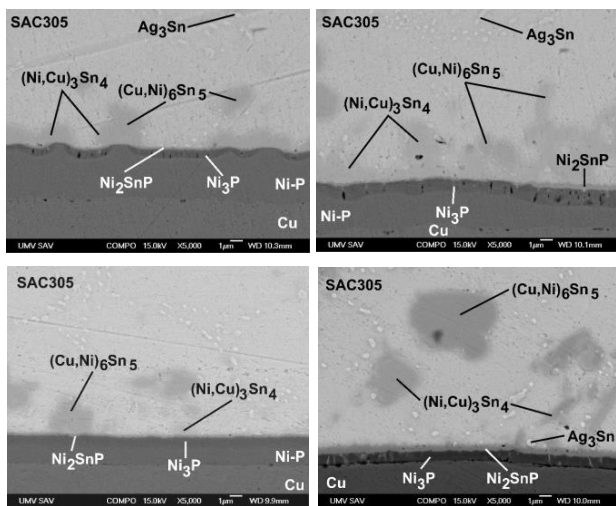
V prípade povrchových úprav DPS, u ktorých bola prítomnosť niklovej difúznej vrstvy ako súčasťou povrchovej úpravy, je situácia odlišná. Spájkovaný spoj sa tvorí priamo s niklovou vrstvou a nie s meďou. Na Obr. 7 sú uvedené časové priebehy kontaktných uhlov pre povrchové úpravy s niklovou difúznou bariérou (ENIG, SnAg7, SnAg5, SnAg5ENi). Je možné vidieť, že rýchlosť klesania kontaktného uhla bola pri všetkých štyroch povrchových úpravách skoro rovnaká, pričom čas pre aktívne zmäčanie bol do 2,6 s. Z grafu vyplýva, že povrchová úprava na báze SnAg5ENi dosiahla najnižší kontaktný uhol a to z dôvodu rozdielneho spôsobu nanášania niklovej vrstvy. Pri tejto povrchovej úprave bol použitý elektrolytický nikel, ktorý je pravdepodobne čistejší a dosahuje lepšiu difúznú rýchlosť než nikel použitý pri povrchovej úprave SnAg5, kde bol nikel vytvorený chemicky. Táto povrchová úprava (SnAg5) dosiahla menší kontaktný uhol než SnAg7 a to z dôvodu nižšej teploty tavenia zliatiny SnAg5 (245 °C). Tento efekt je možné vidieť v približne desiatej sekunde, kedy došlo k zmenšeniu kontaktného uhla až na hranicu merateľnosti. Pri porovnaní s povrchovou úpravou ENIG došlo k zlepšeniu kontaktného uhla len pri povrchovej úprave SnAg5ENi. V prípade prítomnosti niklovej difúznej vrstvy ako súčasťou povrchovej úpravy možno konštatovať, že celkové časy pre aktívne zmäčanie sú veľmi krátke, pričom výsledný kontaktný uhol (na úrovni 23 – 35°) dáva predpoklad pre vynikajúcu spájkovateľnosť. Za zmienku stojí aj fakt, že povrchové úpravy na báze SnAg, ktoré obsahujú niklovú bariéru, majú porovnateľné vlastnosti so štandardne používanou povrchovou úpravou na báze ENIG, pričom v prípade použitia elektrolytického Ni (SnAg5ENi) je koeficient zmäčania a kontaktný uhol takmer ideálny. In-situ meranie kontaktného uhla v priebehu pretavovania guľôčky spájky umožnilo analyzovať priebeh tvorby spoja a poukázať na pozitívny vplyv niklovej bariéry pri jeho tvorbe.



Obr. 7 Časový priebeh kontaktného uhla pre povrchové úpravy s niklovou difúznou bariérou

B. Mikroštruktúrne analýzu a hrúbky IMC

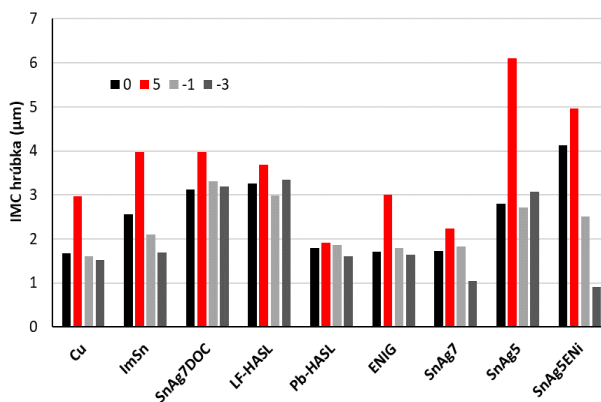
Po prvom pretavení (označenie vzoriek 0) sa najväčšia hrúbka IMC (viac ako 3 μm) vytvorila u vzoriek SnAg7DOC, LFHASL a SnAg5ENi a najmenšia hrúbka (menej ako 2 μm) bola pri povrchoch Cu, PbHASL, ENIG a SnAg7. V prípade Cu nebola po prvom pretavení dosiahnutá hrúbka na úrovni ostatných povrchových úprav a to z dôvodu, že spoj sa vytvoril len na základe difúzie Sn prítomného v SAC guľôčke. V prípade povrchovej úpravy na báze SnAg7DOC bola dosiahnutá väčšia hrúbka než v prípade povrchovej úpravy ImSn a to z dôvodu lepšieho zmáčania. Vo všeobecnosti možno tvrdiť, že v prípade veľmi dobrého zmáčania (malé kontaktné uhly), došlo k rýchlemu vytvoreniu medzifázového rozhrania za súčasnej rýchlej tvorby vrstiev IMC. V prípade malých kontaktných uhlov je možné očakávať väčšie hrúbky IMC na medzifázovom rozhraní. V prípade povrchovej úpravy LFHASL je väčšia hrúbka IMC (viac ako 3 μm) z dôvodu spôsobu výroby povrchovej úpravy LFHASL (ponorom do roztavenej spájky), takže prvotná IMC vznikla už pri tomto procese. Okrem toho, v tomto prípade musíme zobrať do úvahy aj väčšiu hrúbku nanosennej povrchovej úpravy (10 μm). Takáto hrúbka nanosennej povrchovej úpravy bola aj v prípade PbHASL, avšak teplota reflow profilu bola o 30 °C menšia. Okrem toho v prípade PbHASL bol obsah cínu v SnPb guľôčke spájky nižší, preto nedošlo k rýchlejšiemu rastu IMC ako pri procese reflow pre SAC spájky a hrúbka IMC sa pri PbHASL pohybovala na úrovni do 2 μm. V prípade povrchových úprav ENIG a SnAg7 je malá hrúbka IMC (do 2 μm, viď. Obr. 8) zapríčinená prítomnosťou difúzne bariéry na báze Ni-P, ktorá slúži na potlačenie difúzie Sn prítomného v guľôčke spájky do medi a teda na potlačenie tvorby IMC. V prípade povrchovej úpravy SnAg5ENi je použitá niklová bariéra, avšak ukázalo sa, že elektrolytický nikel nepotláča, ale naopak aktivuje tvorbu IMC, ako to bude spomenuté aj v nasledujúcej kapitole.



Obr. 8 Vytvorené IMC v medzifázovom rozhraní spájkovaného spoja s povrchovou úpravou ENIG po jednom pretavení (vľavo hore), po piatom pretavení (vpravo hore) a s povrchovou úpravou SnAg7 po jednom pretavení (vľavo dole), po piatom pretavení (vpravo dole)

Po umelom starnutí (5 cykloch reflow procesu) došlo k nárastu IMC hrúbok u všetkých typov povrchových úprav

avšak v rozdielnej miere, pričom najväčší nárast hrúbok IMC bol zaznamenaný u vzoriek SnAg5 a tiež SnAg5ENi a tiež Cu a ImSn. V prípade SnAg5 a tiež SnAg5ENi predpokladáme veľkú rýchlosť difúzie striebra do spájky s následnou rýchlou tvorbou IMC po opakovanom zaťažení reflow procesom. V prípade povrchovej úpravy SnAg5 je rast zapríčinený nedostatočnou hrúbkou vrstvy Ni-P a teda počas umelého starnutia došlo k difundovaniu medi do objemu spájky. Podobný výsledok bol dosiahnutý aj pri povrchovej úprave SnAg5ENi, avšak v tomto prípade nedošlo až k takému veľkému nárastu hrúbky IMC, i keď aj v tomto prípade bola hrúbka IMC po umelom starnutí v porovnaní s ostatnými povrchovými úpravami veľká, až na úrovni 5 μm. V prípade povrchových úprav ImSn a Cu je nárast hrúbky IMC spôsobený veľmi rýchlou difúziou medi do spájky. Pri všetkých povrchových úpravách došlo k zhomogenizovaniu morfológie vrstvy IMC (nie je taká rozdielna morfológia jednotlivých typov IMC, ako bola po prvom reflow procese). V najmenšej miere narástla hrúbka IMC u vzoriek LFHASL, PbHASL, SnAg7, ENIG a SnAg7DOC. Tento výsledok len potvrdzuje stabilitu a odolnosť týchto povrchových úprav voči tepelnému starnutiu. Z Obr. 9 vyplýva, že v prípade povrchových úprav LFHASL, PbHASL a SnAg7DOC došlo celkovo len k minimálnym zmenám (menej ako 1 μm) celkovej hrúbky IMC aj napriek rýchlej difúzii medi do spájky. Tento výsledok je zapríčinený zvoleným spôsobom starnutia (viacnásobným pretavením), nakoľko pri tomto spôsobe dochádza k roztaveniu spájky a k migrácii veľkých častí IMC do jej objemu. Okrem precipitátov Cu₆Sn₅ sa v objeme spájky objavujú jemné častice Ag₃Sn (popísané v nasledujúcej kapitole).

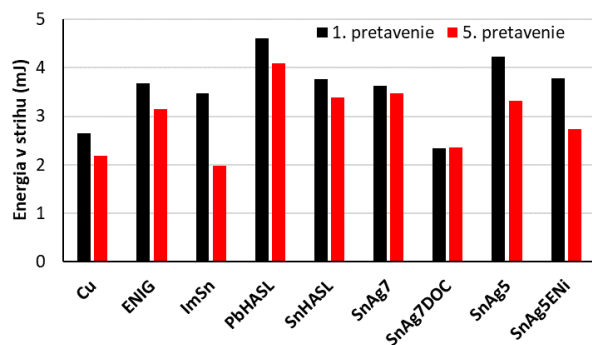


Obr. 9 Hrúbka IMC v závislosti od povrchovej úpravy, počtu pretavení (čierna farba -1. pretavenie, červená farba 5. pretavenie) a počtu teplotných zaťažení (svetlosivá farba – zaťaženie 1 cyklom, tmavosivá farba zaťaženie 3 cyklami)

Veľký nárast hrúbky IMC má negatívny vplyv na mechanické vlastnosti (popísané v nasledujúcej kapitole). Pri tomto hodnotení je potrebné zobrať do úvahy, že jednotlivé vrstvy IMC majú rozdielnu morfológiu a teda hrúbky vrstiev IMC vykazujú rozdielne „rozptyly“. Bližší popis morfológie a jej vplyvu na mechanické a lomové vlastnosti je tiež uvedený v nasledujúcej kapitole. Hrúbka IMC a jej morfológia sú dôležitým ukazovateľom kvality a spoľahlivosti spájkovaného spoja.

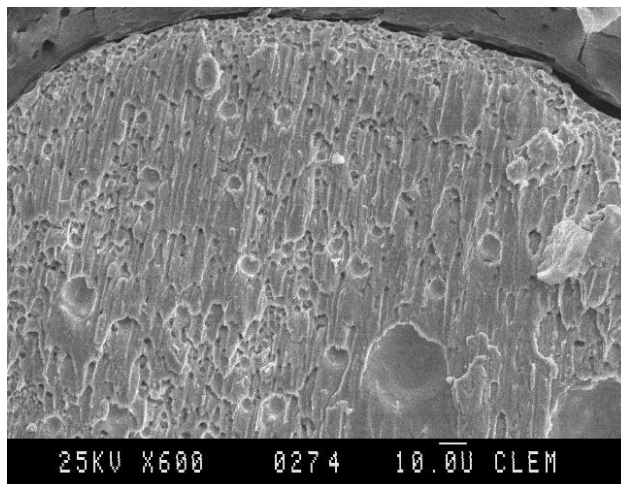
C. Mechnické a krehkolomové vlastnosti

Maximálne pevnosti spojov u vzoriek bez Ni úpravy majú nižšie hodnoty v porovnaní s povrchovými úpravami, ktoré Ni vrstvu obsahujú. Po 5 násobnom reflow procese nedošlo k poklesu maximálnej pevnosti pri vzorkách typu HASL, Cu a SnAg7, naopak, maximálna pevnosť mierne stúpla. Pri vzorkách SnAg7DOC a ENIG došlo k miernemu poklesu maximálnej pevnosti. K najväčšiemu poklesu maximálnej pevnosti došlo pri vzorkách ImSn, SnAg5 a SnAg5ENi. Vysvetlenie k týmto hodnotám uvádzame pri podrobnejšej analýze priebehov závislosti pevnosť v šmyku / posunutie v porovnaní s morfológiou lomovej plochy. Na Obr. 10 sú priemerné hodnoty energie potrebnej pre porušenie spájkovaného spoja v závislosti od použitej povrchovej úpravy. Z týchto výsledkov vyplýva, že vzorky s povrchovou úpravou PbHASL majú najväčšiu hodnotu energie a to aj po piatom pretavení. Najmenší pokles energie potrebnej pre porušenie spoja bol zaznamenaný pri vzorkách LFHASL, SnAg7, SnAg7DOC, Cu a ENIG. Je teda možné tvrdiť, že tieto vzorky sú odolnejšie voči náchylnosti na krehké porušenie. Najväčší pokles bol zaznamenaný pri vzorkách ImSn, SnAg5 a SnAg5ENi, čo značí výrazný pokles húževnatosti spájkovaných spojov.



Obr. 10 Energia v strihu po jednom a po piatich pretaveniach v závislosti od použitej povrchovej úpravy DPS

Výsledky lomových plôch po shear teste ukázali, že vo všetkých prípadoch bola zabezpečená dokonalá adhézia. Pre porušenie bol charakteristický kvázi-húževnatý alebo húževnatý lom, pričom vo všetkých prípadoch došlo k porušeniu vo vnútri objemu spájkovacej guľôčky (nie na rozhraní guľôčka – povrch). Náchylnosť na krehké porušenie (priestorovo členitý lom) je príznakom prítomnosti IMC, ktoré významným spôsobom ovplyvňujú pevnostné vlastnosti, húževnatosť spájkovaných spojov ako aj energiu potrebnú na porušenie. Typické pre lom boli parabolické priestorové útvary príznačné pre húževnaté alebo kvázi-húževnaté porušenie (Obr. 11), ktoré boli doplnené v najhlbšej časti parabolických útvarov hlbokými jamkami, na dne ktorých sú prítomné IMC.



Obr. 11 Detailný snímok porušenia jamky s krehkým porušením pre povrchovú úpravu Cu

Tieto výsledky sú v súlade s očakávaniami podľa literárnych údajov, pričom je možné vysloviť predpoklad, že v prípade aplikácie vyššej rýchlosti zaťažovania počas shear testu, by došlo výhradne k porušeniam krehkým a to vo vrstvách IMC, ktoré sa tvoria na rozhraní povrch – guľôčka spájky.

V. ZÁVER

V článku boli okrem klasických komerčne dostupných povrchových úprav DPS (Cu, ENIG, ImSn, LFHASL, PbHASL) analyzované aj vlastnosti novovyvinutých neštandardných typov na báze SnAg (SnAg7, SnAg7DOC, SnAg5, SnAg5ENi). Tento nový typ povrchovej úpravy DPS je jedinečný, dáva technologom v oblasti montážnych technológií ďalšie možnosti využitia kvalitnej povrchovej úpravy, ktorá svojimi vlastnosťami pripomína náročnú povrchovú úpravu na báze zlata a to najmä pre prípad, kedy je požiadavka na dokonalú planaritu a zmáčavosť, na opakovaný reflow, odolnosť voči korózii, vibráciám a starnutiu aj požiadavka na miniaturizáciu a tým aj vysokú hustotu montáže bez nebezpečenstva porúch, starnutia a následného negatívneho vplyvu na kvalitu výrobkov elektroniky. Pri riešení dizertačnej práce boli dosiahnuté nasledovné vedecké prínosy dizertačnej práce:

- Bol vyvinutý jedinečný prístroj pre analýzu zmáčavosti: in-situ meranie kontaktného uhla, stanovenie času pre aktívne zmáčanie povrchov v reálnom čase.
- Najdynamickejšia zmena kontaktného uhla prebehla počas pri pretavovaní guľôčky spájky počas prvých 10 (15) s. V závislosti od typu povrchovej úpravy následne dochádza v rôznej miere k ustáleniu hodnôt kontaktného uhla. Najrýchlejší priebeh ustálenia kontaktného uhla nastal pri vzorkách, ktoré obsahovali difúznú bariéru na báze ENi alebo Ni-P.
- Boli realizované mikroskopické analýzy na medzifázovom rozhraní povrch / spájka:
 - V závislosti od typu povrchových úprav bola zaznamenaná prítomnosť binárnych a ternárnych IMC, pričom použitím niklovej difúznej bariéry sa

vytvárajú teplotne stabilnejšie ternárne IMC na báze (Cu,Ni)₆Sn₅ a (Ni,Cu)₃Sn₄, pri povrchových úpravách bez difúznej vrstvy (ImSn, SnAg7DOC, Cu, LFHASL a PbHASL) vznikajú len binárne IMC na báze Cu₃Sn a Cu₆Sn₅, pričom charakteristickým znakom najmä u Cu₆Sn₅ je výrazný nárast vplyvom teploty a starnutia,

- Prítomnosť Ni vrstvy vyrobenej elektrolyticky podporuje tvorbu IMC vo zvýšenej miere v porovnaní s vrstvou Ni-P, vyrobenou chemicky.
- Povrchové úpravy s niklovou difúznou bariérou sú zmáčané roztavenou spájkou rýchlejšie než povrchové úpravy bez niklovej difúznej bariéry.
- Na potlačenie rastu nepriaznivých typov IMC je vhodné v prípade SnAg₇ a SnAg₅ aplikovať hrúbku Ni-P difúznej vrstvy až na úroveň 4 μm.
- Prítomnosť Ni-P difúznej vrstvy v prípade SnAg povrchov potvrdzuje pozitívny vplyv na všetky sledované vlastnosti aj po starnutí.

- Aplikácia shear testu je vhodnou metodikou na posúdenie náchylnosti na krehké porušenie. Bol potvrdený vplyv vzniku a rastu IMC na mechanické a krehkolomové vlastnosti spájkovaných spojov po 1 a opakovanom procese reflow na rozdielnych typoch povrchoch.

- Experimenty potvrdili, že okrem nižšej ceny povrchové úpravy typu SnAg sa vyznačujú lepšou zmáčavosťou, lepšimi mechanickými a elektrickými vlastnosťami, lepšou stabilitou vlastností po starnutí v porovnaní s ostatnými štandardnými povrchovými úpravami DPS a v mnohých prípadoch aj s povrchom ENIG.

- Bola realizovaná komparatívna analýza jednotlivých typov analyzovaných povrchových úprav z pohľadu technológie spracovania a odporúčania pre prax.

POĎAKOVANIE

Práca bola podporená MŠVVaŠ v rámci projektu KEGA 011TUKE-4/2023.

ZOZNAM POUŽITEJ LITERATÚRY

- [1] A. Siewiorek *et al.*, “Effects of PCB Substrate Surface Finish and Flux on Solderability of Lead-Free SAC305 Alloy”, *J. Mater. Eng. Perform.*, roč. 22, č. 8, s. 2247–2252, aug. 2013, doi: 10.1007/s11665-013-0492-4.
- [2] T. Lentz, “How Does Surface Finish Affect Solder Paste Performance?”, jan. 2021.
- [3] S. F. M. Amlı, M. A. A. M. Salleh, Rita. M. Said, N. R. A. Razak, J. A. Wahab, a M. I. I. Ramli, “Effect of surface finish on the wettability and electrical resistivity of Sn-3.0Ag-0.5Cu solder”, *IOP Conf. Ser. Mater. Sci. Eng.*, roč. 701, s. 012029, dec. 2019, doi: 10.1088/1757-899X/701/1/012029.
- [4] H. Lee, C. Kim, C. Heo, C. Kim, J.-H. Lee, a Y. Kim, “Effect of solder resist dissolution on the joint reliability of ENIG surface and Sn–Ag–Cu solder”, *Microelectron.*

Reliab., roč. 87, s. 75–80, aug. 2018, doi: 10.1016/j.microrel.2018.05.017.

- [5] X. Ma, F. Wang, Y. Qian, a F. Yoshida, “Development of Cu–Sn intermetallic compound at Pb-free solder/Cu joint interface”, *Mater. Lett.*, roč. 57, č. 22–23, s. 3361–3365, júl. 2003, doi: 10.1016/S0167-577X(03)00075-2.
- [6] Z. Moser, W. Gašior, J. Pstruś, a A. Dębski, “Wettability Studies of Pb-Free Soldering Materials”, *Int. J. Thermophys.*, roč. 29, č. 6, s. 1974–1986, dec. 2008, doi: 10.1007/s10765-007-0363-z.
- [7] J.-W. Yoon, S.-W. Kim, a S.-B. Jung, “IMC morphology, interfacial reaction and joint reliability of Pb-free Sn–Ag–Cu solder on electrolytic Ni BGA substrate”, *J. Alloys Compd.*, roč. 392, č. 1–2, s. 247–252, apr. 2005, doi: 10.1016/j.jallcom.2004.09.045.
- [8] J. M. Koo, Y. H. Lee, S. K. Kim, M. Y. Jeong, a S. B. Jung, “Mechanical and Electrical Properties of Sn-3.5Ag Solder/Cu BGA Packages during Multiple Reflows”, *Key Eng. Mater.*, roč. 297–300, s. 801–806, nov. 2005, doi: 10.4028/www.scientific.net/KEM.297-300.801.
- [9] Y. Sohn, “Effect of Morphological Change of Ni₃Sn₄ Intermetallic Compounds on the Growth Kinetics in Electroless Ni-P/Sn-3.5Ag Solder Joint”, *Metall. Mater. Trans. A*, roč. 51, č. 6, s. 2905–2914, jún. 2020, doi: 10.1007/s11661-020-05739-8.
- [10] S.-Y. Chang, Y.-C. Huang, a Y.-M. Lin, “Mechanical property and fracture behavior characterizations of 96.5 Sn–3.0 Ag–0.5 Cu solder joints”, *J. Alloys Compd.*, roč. 490, č. 1–2, s. 508–514, feb. 2010, doi: 10.1016/j.jallcom.2009.10.071.
- [11] J. Feng, C. Hang, Y. Tian, B. Liu, a C. Wang, “Growth kinetics of Cu₆Sn₅ intermetallic compound in Cu-liquid Sn interfacial reaction enhanced by electric current”, *Sci. Rep.*, roč. 8, č. 1, s. 1775, dec. 2018, doi: 10.1038/s41598-018-20100-1.
- [12] A. W. Mortensen a R. M. Devaney, “Characterization of Black Pad Defect on Electroless Nickel-Immersion Gold (ENIG) Plated Circuits”, *Int. Symp. Microelectron.*, roč. 2010, č. 1, s. 000608–000614, jan. 2010, doi: 10.4071/isom-2010-WP3-Paper1.
- [13] L. M. Toscano a E. Long, “A new surface finish for the electronics industry: Electroless nickel/immersion silver”, v *2014 9th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT)*, Taipei, Taiwan: IEEE, okt. 2014, s. 372–378. doi: 10.1109/IMPACT.2014.7048458.
- [14] L. Toscano a D. Cullen, “The Study, Measurement, and Prevention of Tarnish on Immersion Silver Board Finishes”, dec. 2022.
- [15] C. M. Melton, S. M. Yarlın, a C. J. Raleigh, “Continuous Wettability Measurement for Solder Reflow”, *MRS Proc.*, roč. 226, s. 29, 1991, doi: 10.1557/PROC-226-29.
- [16] “Implementing cleaner printed wiring board technologies: surface finishes”, Environmental

- Protection Agency, Washington, DC. Design for the Environment Branch., EPA 744-R-00-002, 2000. [Online]. Available at: https://www.epa.gov/sites/production/files/2013-12/documents/pwb_surface_finishes_guide.pdf
- [17] D. Bernard a K. Bryant, “DOES PCB PAD FINISH AFFECT VOIDING LEVELS IN LEAD-FREE ASSEMBLIES?”, jan. 2004.
- [18] T. D. Ewald, N. Holle, a K.-J. Wolter, “Void formation on PCB surface finish during reflow soldering”, v *Proceedings of the 2011 34th International Spring Seminar on Electronics Technology (ISSE)*, Trtatanska Lomnica, Slovakia: IEEE, máj. 2011, s. 153–161. doi: 10.1109/ISSE.2011.6053569.
- [19] R. Aspandiar, “VOIDS IN SOLDER JOINTS”, prezentované na SMTA Boise Expo and Tech Forum, 20. marec 2018. [Online]. Available at: <https://smtnet.com/library/files/upload/Voids-in-Solder-Joints.pdf>
- [20] E. Bormashenko, “Apparent contact angles for reactive wetting of smooth, rough, and heterogeneous surfaces calculated from the variational principles”, *J. Colloid Interface Sci.*, roč. 537, s. 597–603, mar. 2019, doi: 10.1016/j.jcis.2018.11.068.
- [21] M.-W. Yang a S.-Y. Lin, “A method for correcting the contact angle from the $\theta/2$ method”, *Colloids Surf. Physicochem. Eng. Asp.*, roč. 220, č. 1–3, s. 199–210, jún. 2003, doi: 10.1016/S0927-7757(03)00064-5.

Capitalizing on Unlabeled Data for Improved Chest X-ray Interpretation: A Self-Supervised Learning Perspective

Matej Gazda

Faculty of Electrical Engineering and Informatics
Technical University of Kosice
Letná 1/9, 040 01 Košice
Email: matej.gazda@tuke.sk

Jan Plavka

Faculty of Electrical Engineering and Informatics
Technical University of Kosice
Letná 1/9, 040 01 Košice
Email: jan.plavka@tuke.sk

Abstract—The healthcare industry faces a shortage of qualified personnel and an aging population, creating a need for high-quality healthcare solutions. Convolutional Neural Networks (CNNs) can automate medical image analysis tasks, but annotated data is scarce, and patient privacy is a concern. This paper describes innovative AI models that utilize labeled and unlabeled data and self-supervised learning to address these challenges. Utilizing unannotated data for medical imaging, and diagnosing illnesses from chest X-rays was studied. This work proposes new approaches to overcome data limitations in medical imaging and suggests possible future research directions.

Zdravotnícky priemysel čelí nedostatku kvalifikovaných pracovníkov a starnúcej populácii, čo vytvára potrebu kvalitných zdravotníckych riešení. Konvolučné neurónové siete (CNN) môžu automatizovať analýzu lekárskeho obrazov, ale anotované dáta sú vzácne a súkromie pacienta iba prehľbuje daný problém. Táto práca popisuje inovatívne AI modely, ktoré využívajú označené a neoznačené dáta a učenie sa pod vlastným dohľadom na riešenie týchto problémov. Bolo preskúmané využitie neanotovaných dát pre lekárske obrázky a diagnostikovanie chorôb z röntgenových obrazov hrudníka. Táto práca navrhuje nové prístupy na prekonanie limitovaných dát v medicínskom odvetví a naznačuje možné budúce smery výskumu.

I. INTRODUCTION

The World Health Organization (WHO) anticipates a global shortage of 15 million healthcare workers by 2030, with underdeveloped nations expected to bear the brunt of this shortage [1]. Deploying healthcare personnel to rural and remote areas introduces additional challenges. Tasks critical to healthcare, such as the analysis of medical images, demand efficient workflow management. Given the impending healthcare workforce shortage, it becomes crucial to investigate alternative methods to improve healthcare efficiency and affordability.

Artificial Intelligence (AI) has demonstrated significant potential in addressing various tasks across medical and non-medical domains. It has proven particularly useful in interpreting Chest X-rays (CXRs) [2] and tumor segmentation [3]. With the expansion of telemedicine, there is a growing demand for advanced AI models that can provide second opinions to healthcare providers and patients.

Several AI models have recently received approval from regulatory authorities such as the Food and Drug Administration (FDA) and European Medicines Agency (EMA). For example, Siemens Healthineers has developed an FDA-approved AI model capable of rapidly quantifying and visualizing calcified coronary lesions from CT scans. Qure AI has also developed a model that effectively detects abnormalities in CXRs images.

However, despite these promising advancements, the integration of AI models in healthcare continues to face significant challenges. Data collection and annotation, in particular, pose hurdles. The process of annotating large volumes of data is not only costly but also time-consuming. Further, the acquisition of data for commercial use can present legal challenges, including issues with unclear data ownership and difficulty in obtaining patient consent for retrospective studies. It is becoming increasingly clear that data is a vital and valuable commodity [4].

AI models must account for variability in pathology appearance, image acquisition devices, and patient demographics to perform effectively within the medical field. As the quality of medical imaging continues to improve, the need for models to handle greater variability during training and inference becomes more critical. Numerous institutions and reviews have identified dataset limitations as one of the most significant challenges to overcome in the application of AI in healthcare.

In AI, the effective performance of various tasks remains a persistent challenge. With the increasing incorporation of AI in medical decision-making, patients must have access to a variety of AI models, as not all companies have equal access to data or offer similar prices. Companies with limited data resources may need to explore alternative strategies that require less data. Such competition could potentially lead to the developing of more advanced and cost-effective AI models for patients.

The contributions are as follows:

- Exploration of self-supervised methods for COVID-19 and pneumonia diagnosis from CXRs, contributing to the expanding field of AI diagnostic tools and potentially alleviating the burden on healthcare professionals.

- Demonstration that self-supervised pretraining can outperform pretraining on large annotated datasets from different domains, such as the ImageNet database, potentially opening new doors to more cost-effective and efficient AI model training.
- Investigation of AI models that effectively address challenges of data collection and annotation, taking into account patient privacy and legal constraints.

A. Literature Review

CXRs have been instrumental in diagnosing various respiratory conditions, including pneumonia and COVID-19. Several studies have been carried out to examine the effectiveness and accuracy of these imaging techniques in such contexts.

In the early stages of the COVID-19 pandemic, CXRs served as a valuable tool for diagnosing and monitoring the progression of the disease. A study by Wong et al. [5] highlighted the importance of CXR imaging in providing immediate diagnostic imaging data for patients suspected of having COVID-19. The study noted the common CXR findings in COVID-19 patients included bilateral and peripheral ground-glass opacities.

The utilization of AI in interpreting CXRs for diagnosing pneumonia and COVID-19 has also been the focus of many recent studies. A study by Rajpurkar et al. [6] developed a deep learning model, CheXNet, which outperformed radiologists in diagnosing pneumonia using CXRs. Similarly, Apostolopoulos and Mpesiana [7] developed a COVID-19 detection model using CXRs, which achieved an accuracy of 96.78%.

AI-based models have also been used to distinguish between COVID-19 and other types of pneumonia. For instance, a study by Oh et al. [8] used a deep learning model to differentiate between COVID-19 and bacterial pneumonia with a sensitivity of 92.9% and a specificity of 94.7%.

Despite these advancements, some challenges persist. A review by Roberts et al. [9] highlighted the limitations of CXRs in detecting early-stage COVID-19 due to the disease's tendency to initially present in the peripheral regions of the lung, which can be difficult to visualize with CXR.

In summary, CXRs have proven to be a valuable tool in the diagnosis and management of respiratory diseases such as pneumonia and COVID-19. The integration of AI has further enhanced the diagnostic capabilities of CXRs, although challenges remain, particularly in detecting early-stage diseases.

B. Data

In this research, several CXR datasets were employed. To begin, the CheXpert dataset [10] was used for network pretraining due to its requirement for a large-scale dataset. This dataset includes 224,316 chest radiographs involving 65,420 patients. The samples were labeled based on data from radiology reports between October 2002 and July 2017, covering 14 common conditions such as pneumonia, pneumothorax, and cardiomegaly. These labels were not used during pretraining as the proposed model operates unsupervised.

Table I
DATASETS USED IN THIS STUDY

Dataset	# samples	# class 0	# class 1	source
pretext				
CheXpert	224,316	na	na	[10]
target				
CELL	5,323	3,883	1,349	[11]
ChestX-ray-14	2,706	1,353	1,353	[2]
C19-Cohen	807	564	243	[12]
COVID-GR	852	426	426	[13]

Four public datasets were obtained to test the model's adaptability to an external target dataset. The Cell dataset [11] was the first, containing 5,323 CXR images from children. It includes 3,883 cases of viral and bacterial pneumonia and 1,349 normal images. Two expert physicians provided the labels, and a third physician confirmed them. The ChestX-ray14 dataset [2] was the second dataset used. It comprises 112,120 CXR images from 30,805 patients, with eight disease labels. A subset of this dataset was used, focusing on patients with pneumonia and a comparable number of healthy controls.

The final two datasets are recent compilations intended for COVID-19 detection. The C19-Cohen dataset [12] (accessed 21.10.2020) includes various types of pneumonia (viral, bacterial, and fungal). We selected two classes: 304 patients with COVID-19 and 114 patients with other types of pneumonia. Lastly, the COVID-GR dataset [13] was also used to evaluate the proposed model. It comprises 426 CXR images from patients with COVID-19 across four severity levels, along with an equal number of control subjects. Notably, 76 out of the 426 COVID-19 patients tested positive via PCR, yet their CXRs were assessed as normal, making the classification task more complex. An overview of all the datasets used in this research is presented in Tab. I.

II. METHODS

A. The Learning Process of Convolutional Neural Networks

Optimizing a neural network operates on minimizing an objective function $J(\theta)$, understood as the average loss over all training instances. This can be expressed as

$$J(\theta) = \mathbb{E}_{(\mathbf{x}, y) \sim \hat{p}_{data}} L(f(\mathbf{x}; \theta), y), \quad (1)$$

where \hat{p}_{data} signifies the empirical distribution, L symbolizes a per-example loss function, $f(\cdot)$ is the function creating the prediction when the input is \mathbf{x} , and y is the true label [15].

The minimization of the objective function $J(\theta)$ is achieved by adjusting the parameters θ against the direction of its gradient. Three well-known methods exist for optimizing an objective function: batch gradient descent, stochastic gradient descent, and mini-batch gradient descent [16]. The selection of a specific method is driven by the quantity of training data. Batch gradient descent operates simultaneously on the entire training dataset in a large batch. Stochastic gradient descent works with one example at a time. Most deep learning algorithms employ mini-batch gradient descent, which handles

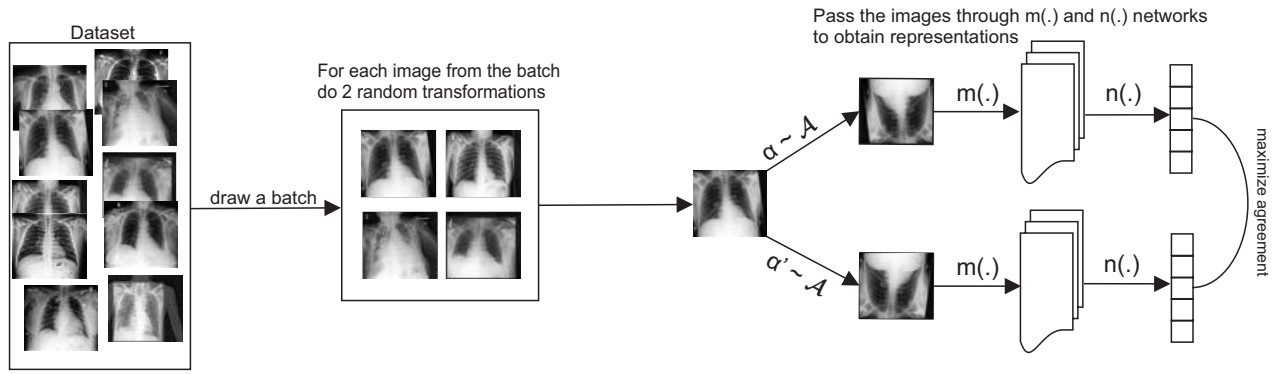


Figure 1. Proposed self-supervised architecture for CXR images [14].

a small dataset segment at a time. The parameter update in mini-batch gradient descent can be calculated as

$$\theta^{(t+1)} = \theta^{(t)} - \eta \frac{1}{B} \sum_{i=Bt+1}^{B(t+1)} \frac{\partial L(f(\mathbf{x}_i; \theta), y_i)}{\partial \theta}, \quad (2)$$

L is a loss function of one instance from the batch sampled at time t , B is the batch size, and η is a hyperparameter known as the learning rate. Note that if B is equal to the number of samples in the training dataset, it signifies batch gradient descent, and if $B = 1$ it signifies stochastic gradient descent. The batch size B profoundly affected the training procedure and was extensively discussed in [17].

Several gradient descent optimization algorithms are available and often treated as hyperparameters from the model's perspective. In this research, we employed the Adadelta [18] optimization algorithm, an extension of Adagrad [19]. Adadelta presents several beneficial features, such as resilience to architectural choice, insensitivity to hyperparameters, and no manual learning rate setting requirement.

B. Contrastive Self-supervision in Convolutional Neural Networks

In this study, we introduce a neural network that addresses pretext tasks by adopting an approach akin to contrastive instance discrimination, as seen in methodologies like SimCLR [20] and MoCo [21]. The data augmentation module in our network forms two variants of each image in a given batch, thus generating both positive and negative pairs.

Specifically, a positive pair emerges from two differently augmented versions of the same image. In contrast, a negative pair is constituted by two images derived from different source images through augmentation.

The neural network is trained to differentiate between these pairs by amplifying the congruence between two instances of the same image, thereby distinguishing between positive and negative pairs.

To compare the representations, we utilize cosine similarity, which is described for two vectors \mathbf{u}, \mathbf{v} as follows:

$$\text{sim}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u}^T \mathbf{v}}{\|\mathbf{u}\|_2 \cdot \|\mathbf{v}\|_2}. \quad (3)$$

Alternative similarity functions, such as the Euclidean distance or dot product, could also be implemented.

Our proposed learning architecture is tripartite: a backbone neural network $m(\cdot)$, a projection head $n(\cdot)$, and a stochastic data augmentation module \mathcal{A} . The backbone network $m(\cdot)$, constructed as a ResNet50 Wide network, extracts representations from the augmented data samples. The projection head $n(\cdot)$ then transmutes these outputs into a latent space where contrastive loss is applied. The output vector's size is a flexible hyperparameter, allowing for the final size to be appropriately adjusted to match the original image's dimensions. The data augmentation module \mathcal{A} generates random augmentations such as resized crop, horizontal flip, rotation, Gaussian blur, and color jitter. Fig. 1 provides a comprehensive visual representation of the entire learning architecture.

The optimization process commences with a minibatch sampled at time t , which consists of images $\mathcal{X}_t = \{\mathbf{x}_{Bt+1}, \mathbf{x}_{Bt+2}, \dots, \mathbf{x}_{Bt+B}\}$ of size B drawn from the dataset samples \mathcal{X} , $\mathcal{X}_t \subseteq \mathcal{X}$, akin to supervised learning. Each image in the minibatch is then augmented twice to form a positive pair, resulting in $2B$ images. These images are subsequently encoded via the backbone network $m(\cdot)$ to derive the representation vectors, which are processed through the projection head $n(\cdot)$ to acquire projection vectors, denoted as \mathcal{Z}_t . The model error is computed using the NT-Xent loss (*normalized temperature-scaled cross entropy loss*) introduced in [22]. The loss for a positive pair (z_i, z_j) drawn from the set of projections of augmented images, \mathcal{Z}_t , is calculated as follows:

$$l(z_i, z_j) = -\log \frac{e^{\text{sim}(z_i, z_j)/\tau}}{\sum_{z_k \in \mathcal{Z}_t - z_i} e^{\text{sim}(z_i, z_k)/\tau}}, \quad (4)$$

where sim is the similarity function, $\mathcal{Z}_t - \{z_i\}$ represents the $2B$ projections of augmented images (excluding the projection z_i), and τ is a hyperparameter known as temperature.

Once the loss is computed, we employ backpropagation to optimize the weights of the backbone neural network $m(\cdot)$ and the projection head $n(\cdot)$. Upon completing the training process, we obtain the features from the last layer of the

Table II

WE PRESENT THE PREDICTIVE PERFORMANCE OF OUR PROPOSED METHOD ACROSS FOUR CHEST X-RAY (CXR) DATASETS, INCLUDING PREVIOUSLY PUBLISHED RESULTS FOR CONTEXT AND COMPARISON. * DUE TO THE INTEGRATION OF THE C19-COHEN DATASET WITH OTHER CXR DATASETS IN PRIOR STUDIES, NO DIRECT COMPARISON RESULT IS AVAILABLE. FURTHERMORE, THE 19-COHEN DATASET IS CONTINUALLY UPDATED WITH NEW DATA. ** THE RESULT FOR PNEUMONIA PREDICTION IS DERIVED FROM A MULTICLASS CLASSIFICATION FRAMEWORK.

Dataset	Fraction of dataset												ImageNet pretraining				Published result
	1%				10%				100%				ACC	AUC	SEN	SPE	ACC/AUC
	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE					
Cell	85.6	96.6	99.5	62.4	86.9	96.9	99.2	66.2	91.5	97.7	98.7	79.5	83.2	94.9	98.2	58.1	92.8/96.8 [11]
C19-Cohen	84.9	89.2	90.6	73.6	81.8	91.5	85.8	73.6	81.1	88.2	83	77.4	69.8	77.0	80.2	49.1	na*
COVID-GR	79.5	86.6	83.5	75.6	77.8	86.0	80	75.6	78.4	87.1	83.5	73.3	71.3	77.7	69.4	73.2	76.16/na [13]
ChestX-ray-14	71.5	79.1	72.8	71.4	71.2	78.1	72.0	71.2	71.4	78.4	71.3	71.5	69.6	75.1	67.0	69.6	na/65.8** [2]

backbone neural network $m(\cdot)$, while the projection head $n(\cdot)$ is discarded.

III. RESULTS

In line with MoCo and SimCLR approaches, we utilize transfer learning to measure the predictive performance of the embeddings extracted by the feature extractor part, while the embedding size was arbitrarily selected to 128. As a final classifier, logistic regression was used to test four different CXR datasets. We also tried other classifiers, such as linear head or different types of machine learning classifiers. However, the trend was similar. Thus we omit them for clarity. To verify the model outcomes, 5-fold cross-validation was used, while 70% of data was used for training, 10% for validation, and 20% was left for the testing part.

The performance metrics of the trained models, including Accuracy (ACC), Area Under the Curve (AUC), Sensitivity (SEN), and Specificity (SPE), are summarized in Table II.

We note that the predictive accuracy of the models varies across the different datasets. More specifically, we observe a higher degree of variation in prediction accuracy for the pneumonia classification task (Cell, ChestX-ray-14) than for the COVID-19 classification task (C19-Cohen, COVID-GR). This discrepancy is largely attributable to the diverse nature of the dataset compositions. Each dataset derives from distinct sources and is captured using varied devices, which subsequently affects image attributes. Additionally, it's worth noting that the Cell dataset, consisting of CXR images from children aged four to six years, presents a unique classification challenge.

Analyzing the Cell dataset, we find a trend of increasing AUC and ACC as the fraction size of the dataset used for the pretext task grows. Our model achieves a peak AUC of 97.7%, surpassing the 96.6% reported in [11], which used an ImageNet-based transfer learning approach. This suggests that self-supervised learning on smaller, semantically closer datasets can yield more beneficial results than supervised pretraining on large but semantically distant datasets, such as ImageNet.

In contrast, the ChestX-ray-14 dataset model yielded substantially lower results than the Cell dataset model, with performance being independent of the dataset fraction used for training the pretext task. Table II reveals that our proposed model surpasses the score reported in [2]. However, this comparison may not be entirely equitable, as [2] tackled a

multiclass problem, which could have adversely affected the model's accuracy compared to binary classification.

The models trained on the COVID-GR and C19-Cohen datasets produced similar outcomes. Remarkably, one model trained on the C19-Cohen dataset achieved an AUC of up to 91.5

A. CNN explainability

In an attempt to elucidate the decision-making process of Convolutional Neural Networks (CNNs), we utilize gradient-weighted class activation mapping (Grad-CAM) to emphasize the pertinent sections of the Chest X-Ray (CXR) image that correlate with a particular decision. Grad-CAM, a class-discriminative localization method, provides visual justifications for the choices made by CNNs [23]. In Figure 2, CXRs from six distinct patients, accurately diagnosed with pneumonia, are presented. These images exhibit a mix of ground-glass opacities, consolidations, and air bronchograms.

Air bronchograms, seen as dark, thread-like lines on radiographic images due to air-filled bronchi becoming visible amidst the opacified surrounding alveoli (also known as "white lung field"), serve as another pathological indicator commonly associated with pneumonia. The visibility of air bronchograms is enhanced when the lung tissue is more edematous, but this is only true for bronchi that contain a certain amount of residual air.

In the current analysis, the attention map's highlighted lung areas include zones exhibiting noticeable pulmonary opacification and air bronchograms, providing substantiation for an accurate diagnosis. In the clinical setting, these are the same radiological markers a radiologist would seek, similar to the neural network employed for decision-making here.

However, some areas illuminated by the attention map fall outside the lung region. These areas, unable to be connected to any pneumonia-induced pathology, likely represent regions inaccurately assessed either by the visualization algorithm or the CNN itself.

In order to gain insight into the learned representations, we selected models from four distinct checkpoints and obtained their corresponding features. These models underwent training for 10, 25, 50, and 100 epochs. Figure 3 depicts t-SNE visualizations of the extracted features derived from the Cell and COVID-GR datasets, using these four models. The visualization reveals that, although the Cell dataset exhibits a discernible yet modest enhancement in the distinguishability

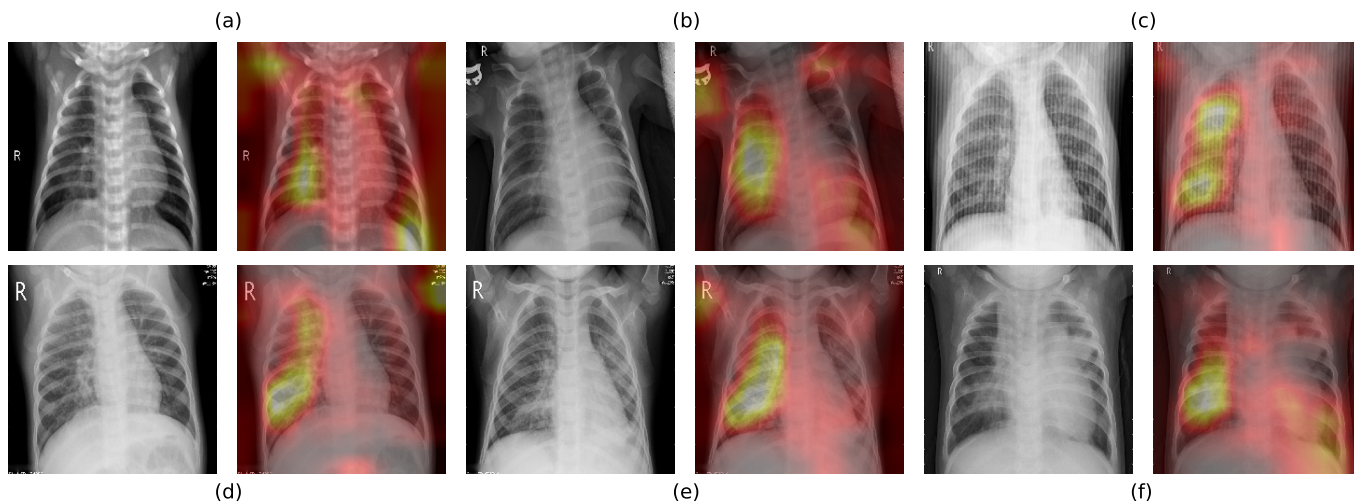


Figure 2. Grad-CAM CNN Decision Visualization [14]

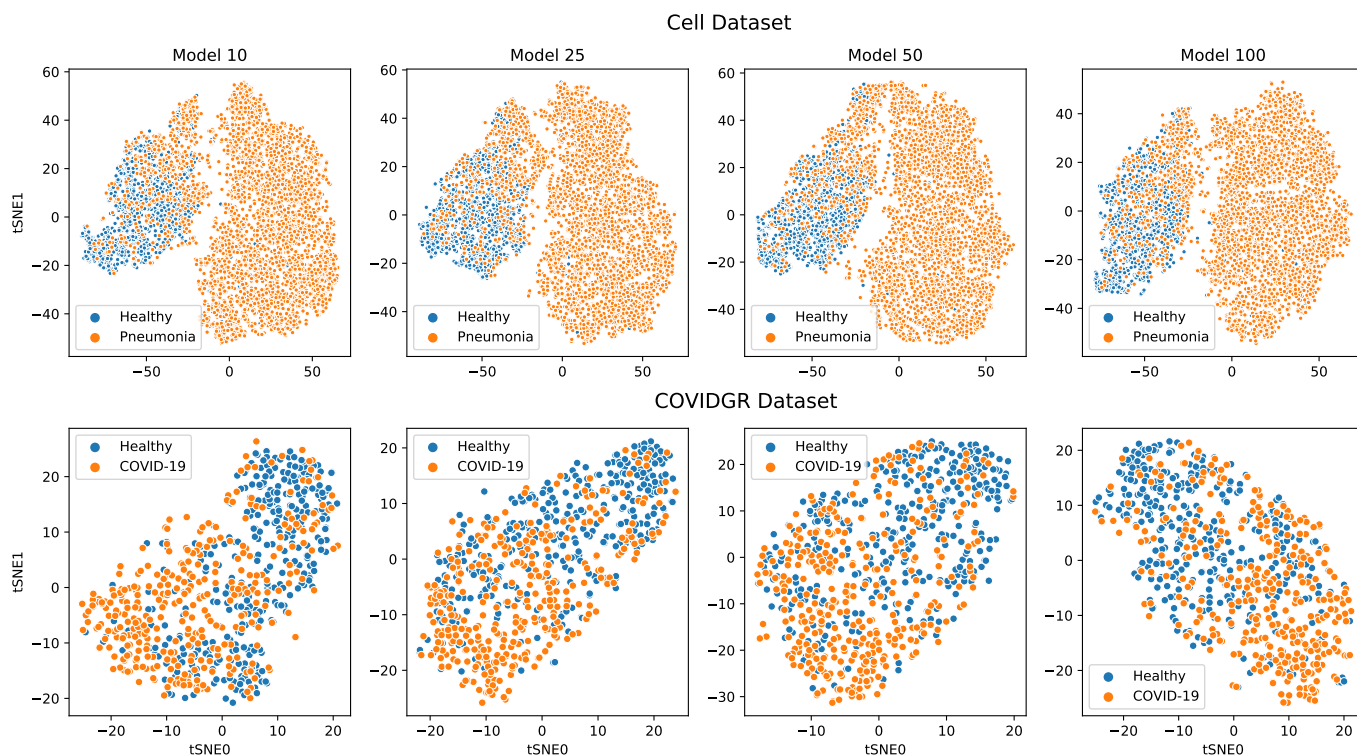


Figure 3. tSNE representation of extracted embeddings [14].

of its two classes, the two classes in the COVID-GR dataset appear to be intertwined across all four images.

IV. CONCLUSION

The last pandemic has underscored the importance of integrating diagnostic decision-support systems into clinical decision-making processes. The effective implementation of these systems into modern medical devices could streamline various tasks and mitigate the burden on healthcare personnel. These automated solutions also play a pivotal role in non-

crisis periods by affording medical experts more time for tasks demanding meticulous attention or specialized expertise.

In the context of enhancing medical expert systems, we presented a novel solution for classifying CXR images. Our proposed method employs a CNN pretrained on an unlabeled dataset of CXR images. This approach circumvents the necessity for labeled data, which are often limited and costly in the medical field, thereby opening up new avenues for using CNNs. Our findings challenge the belief that CNNs must only be trained on natural images, like those in the

ImageNet dataset. Instead, our study demonstrates that they can be trained on images more semantically aligned with the target task. For instance, a network pre-trained on the CheXpert dataset effectively learned significant representations and extracted pertinent features for detecting pneumonia and COVID-19. The performance of our unsupervised model stands strong when compared with its supervised counterparts. This approach harbors immense potential due to the novelty of self-supervised contrastive learning for visual representations. Future methodological enhancements could further improve its performance, making it an even more powerful tool in medical imaging diagnostics.

ACKNOWLEDGMENT

This work was supported in part by the Slovak Research and Development Agency under Contract APVV-16-0211, and in part by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under Contract VEGA 1/0327/20.

REFERENCES

- [1] "WHO – health-workforce." https://www.who.int/health-topics/health-workforce#tab=tab_1. Accessed: 2022-10-24.
- [2] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.
- [3] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, *et al.*, "The liver tumor segmentation benchmark (LITS)," *arXiv preprint arXiv:1901.04056*, 2019.
- [4] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nature Medicine*, vol. 28, no. 1, pp. 31–38, 2022.
- [5] H. Y. F. Wong, H. Y. S. Lam, A. H.-T. Fong, S. T. Leung, T. W.-Y. Chin, C. S. Y. Lo, M. M.-S. Lui, J. C. Y. Lee, K. W.-H. Chiu, T. W.-H. Chung, *et al.*, "Frequency and distribution of chest radiographic findings in patients positive for covid-19," *Radiology*, vol. 296, no. 2, pp. E72–E78, 2020.
- [6] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the cheXnext algorithm to practicing radiologists," *PLoS medicine*, vol. 15, no. 11, p. e1002686, 2018.
- [7] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and engineering sciences in medicine*, vol. 43, pp. 635–640, 2020.
- [8] Y. Oh, S. Park, and J. C. Ye, "Deep learning covid-19 features on cxr using limited training data sets," *IEEE transactions on medical imaging*, vol. 39, no. 8, pp. 2688–2700, 2020.
- [9] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, *et al.*, "Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 199–217, 2021.
- [10] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghighi, R. Ball, K. Shpanskaya, *et al.*, "Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 590–597, 2019.
- [11] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [12] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 image data collection: prospective predictions are the future," *arXiv 2006.11988*, 2020.
- [13] S. Tabik, A. Gómez-Ríos, J. L. Martín-Rodríguez, I. Sevilano-García, M. Rey-Area, D. Charte, E. Guirado, J. L. Suárez, J. Luengo, M. A. Valero-González, P. García-Villanova, E. Olmedo-Sánchez, and F. Herrera, "Covidgr dataset and covid-sdnet methodology for predicting COVID-19 based on chest X-ray images," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3595–3605, 2020.
- [14] M. Gazda, J. Plavka, J. Gazda, and P. Drotar, "Self-supervised deep convolutional neural network for chest x-ray classification," *IEEE Access*, vol. 9, pp. 151972–151982, 2021.
- [15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [16] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.
- [17] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural networks: Tricks of the trade*, pp. 437–478, Springer, 2012.
- [18] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.
- [19] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.
- [20] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," *arXiv preprint arXiv:2002.05709*, 2020.
- [21] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3733–3742, 2018.
- [22] K. Sohn, "Improved deep metric learning with multi-class n-pair loss objective," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 1857–1865, 2016.
- [23] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

Advanced Data Processing Methods for Parkinson's Disease Detection from Voice

Máté Hireš

*Department of Computers and Informatics
Faculty of Electrical Engineering and Informatics
Technical University of Košice, Slovak Republic
mate.hires@tuke.sk*

Peter Drotár

*Department of Computers and Informatics
Faculty of Electrical Engineering and Informatics
Technical University of Košice, Slovak Republic
peter.drotar@tuke.sk,*

Abstract—The automatized voice analysis in Parkinson's disease (PD) is a powerful tool making the detection and monitoring process easier which is beneficial to both patients and healthcare providers. PD affects an individuals motor and non-motor system often causing dysarthria, which results in a weak, harsh, breathy voice or mumbling. The objective of this study is to provide a deep learning framework for PD detection which does not require manual feature extraction. We propose an ensemble of multiple fine-tuned CNN networks for this purpose. The multiple fine-tuning utilizes transfer learning, while aims to bridge the semantical gap between the source and target task. The Xception and ResNet50 models are employed to validate the proposed approach on the PC-GITA dataset, and validated for each vowel phonation. The voice samples are converted to log-frequency spectrograms which are used as the input of the networks. Our models were able to differentiate between the voice of healthy and PD-affected subjects. The Xception network yielded the best of 99% accuracy, which was achieved in case of the /a/ vowel. This shows the potential of deep learning based approaches or the diagnosis and monitoring process of PD.

Automatizovaná hlasová analýza pri Parkinsonovej chorobe (PD) poskytuje nástroj, ktorý uľahčuje proces detekcie a monitorovania, čo je výhodné pre pacientov aj poskytovateľov zdravotnej starostlivosti. PD postihuje motorický a nemotorický systém jednotlivca, pričom často spôsobuje aj dysartriú, ktorá vedie k slabému, drsnému, dýchavičnému hlasu alebo mumľaniu. Cieľom tejto štúdie je poskytnúť rámec hlbokého učenia pre detekciu PD, ktorý nevyžaduje manuálnu extrakciu príznakov. Na tento účel navrhujeme skupinový klasifikátor viacnásobne ladených sietí CNN. Viacnásobné ladenie využíva prenositeľné učenie, pričom jeho cieľom je preklenúť sémantickú medzeru medzi zdrojovou a cieľovou úlohou. Modely Xception a ResNet50 sa používajú na validáciu navrhovaného prístupu na dátovej sade PC-GITA, a validovaná pre každú samohlásku. Vzorky hlasu sú konvertované na log-frekvenčné spektrogramy, ktoré sa používajú ako vstup pre siete. Naše modely dokázali rozlíšiť hlas zdravých jedincov a jedincov. Xception model dosiahol presnosť 99% v prípade samohlásky /a/. To ukazuje potenciál prístupov založených na hlbokom učení pre diagnostiku a monitorovanie PD.

Index Terms—convolutional neural networks, multiple fine-tuning, Parkinson's disease, pathological voice analysis

I. INTRODUCTION

Parkinson's disease (PD) is a common neurodegenerative disorder, which affects the brains neuromediator system. PD can cause motor- and non-motor disorders [1].

PD also causes voice impairments, which are connected to rigidity and bradykinesia [2]. The changes in voice appear in

early stages of PD, therefore offering important biomarkers for early-stage PD diagnosis. PD can result in soft, monotonous speech with reduced vocal tract loudness or dysfluency [3]. The collective naming of voice disorders caused by PD is Parkinsonian dysarthria.

There is no cure for PD, however treatments such as deep brain stimulation with speech therapy can help manage the symptoms or improve the communication skills of an individual [4]. The traditional diagnosis includes uncomfortable observations of the vocal folds, which is a very complex and expensive process, also taking to account the self-reported functional impairments according to the Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS-III) [5]. However, there is also some clinical bias, leading to misdiagnosed cases [6].

Computerized detection of PD has received the attention of researchers in the past years mainly thanks to their non-invasive nature. Various machine learning and deep learning methods were proposed recently for detection and monitoring of PD, most often assessing the individuals gait [7], handwriting [8] or voice [9]. While the analysis of handwriting and gait requires special devices for measurements, the voice assessment can be performed even from home, since it only requires a recording device.

The aim of computerized voice analysis in PD is to reduce the need for manual work for the detection of the disease and therefore automatize the diagnosis and monitoring process. Most often, the analysis of Parkinsonian voice includes the assessment of phonation, articulation, prosody and linguistic [10]. Rusz, et al. proposed quantitative measurements for Parkinsonian voice and speech characterization [11]. Based on their study, the linguistic and cognitive conditions of an individual may complicate the diagnosis process, however the phonation showed to be more suitable for diagnosing and monitoring PD. Phonation is usually measured in sustained vowel phonations [12].

Studies [13], [14] provide an overview of commonly used approaches and feature sets in the field of PD diagnosis and monitoring.

Deep neural networks, especially convolutional neural networks (CNNs) got very popular in the past years, due to the fact that they do not require manual feature extraction. They

were successfully employed also in the field of Parkinsonian voice detection achieving promising results.

However, one limitation of the CNNs is their need for large amount of training data in order to be properly trained. In such cases augmentation techniques can be utilized to artificially inflate the quantity of training data. The augmentations are very helpful to create more diverse data and therefore help to avoid the models overfitting and improve their generalizability and robustness [15], [16], [17]. Data warping utilizes label-preserving image transformations to create new data, such as zooming, flipping, rotating, while by augmenting the feature space of the data synthetic samples are created, which is called oversampling [18], [19]. Synthetic data can be created by applying kernel filters to sharpen or blur the images [20]. Voice-specific data augmentation techniques also showed to be a promising oversampling tool in speech processing [21], [22], [23].

In this study we propose multiple-fine-tuning for CNN training and an ensemble of such trained CNN networks for Parkinson's disease detection from sustained phonation of vowels. The CNN networks take log-frequency power spectrograms as input while also utilizing data augmentation techniques to prevent the model from overfitting. The aim of this approach is to improve the accuracy of voice-based Parkinson's disease detection.

II. DATA

A. PC-GITA dataset

The PC-GITA dataset contains recordings from 50 healthy controls (HC) and 50 patients diagnosed with Parkinson's disease (PD). The individuals in the HC group had no symptoms associated to PD. The dataset is well balanced in terms of age and gender. There are 25 men and 25 women participants in each group, the mean age of the OD group is 62.2 and the mean age of the HC group is 61.2. All the recordings were captured at 44 100 Hz rate and 16 bit resolution in noise-controlled conditions. One session contains several different recordings. The provided speech tasks are the following:

- sustained phonation of vowels /a/, /e/, /i/, /o/, /u/ (three repetitions);
- reading of a set of different Spanish words;
- a dialog between a doctor and a subject;
- reading of sentences with different emphasis in some marked words;
- a monologue (Spontaneous speech).

The tasks were selected to analyze several aspects of the voice of the subjects, such as phonation, articulation and prosody. The detailed description of all the tasks can be found in [24].

B. Saarbruecken voice database

The Saarbruecken voice database (SVD) contains voice recordings of 2042 German speakers, from which 687 is healthy (428 females and 259 males) and 1355 of them

are patients (727 females and 629 males) with 71 different pathological voice disease [25].

All the samples are recorded in the same environment, at 50 kHz frequency with 16 bit resolution. Every session consists of the following recordings:

- Recordings of the vowels /a/, /i/, /u/ produced at normal, high and low intonation,
- Recordings of the vowels /a/, /i/, /u/ with rising-falling intonation,
- Recording of the sentence "Guten Morgen, wie geht es Ihnen?" ("Good morning, how are you?").

C. Vowels dataset

The vowels dataset contains recordings of sustained vowels /a/, /e/, /i/, /o/, /u/ provided by Open Data Commons¹. This datasets purpose is vowel recognition. Therefore, this dataset does not focus on pathological disease detection, however it is used in our experiments.

There are 113 to 121 recordings for each vowel with one to three samples per recording [26].

III. PROPOSED APPROACH

A. Convolutional neural networks

Convolutional neural network (CNN) is a special type of deep neural networks consisting of multiple hidden blocks between the input and output layers. These blocks contain alternating convolutional and pooling layers. The objective of the convolutional layers is to learn certain features of the input grid-like data such as images. These layers return feature-maps, this process is often called feature extraction [27], [28]. The pooling layers (usually applying max- or average pooling) reduce the connections between the convolutional layers, decreasing the size of the feature map [29], [30].

The training process of the convolutional neural networks is based on minimizing an objective loss function, which can be defined as

$$\mathcal{L} = \frac{1}{N} \sum_{n=1}^N \ell(\theta; \mathbf{y}^{(n)}, \mathbf{o}^{(n)}). \quad (1)$$

Here θ defines all the parameters of the CNN, $\mathbf{y}^{(n)}$ is the target label of the n -th input sample and $\mathbf{o}^{(n)}$ is the output for the n -th sample.

The selection of the most appropriate objective function is very important. Cross-entropy is the most commonly used objective function and it is suitable for binary classification tasks such as pathological voice detection. The binary cross-entropy loss can be defined as

$$\mathcal{L} = -[y \log(p) + (1 - y) \log(1 - p)], \quad (2)$$

where y is the target label and p is the approximated probability [27].

The optimization process is most frequently performed utilizing the stochastic gradient descent (SGD) algorithm [31].

¹<https://opendatacommons.org/>

The SGD algorithm estimates the gradient with respect to a mini-batch (\mathbf{x}, \mathbf{y}) of n training examples from the training set and updates each parameter, leading to a more stable convergence. The selection of the learning rate is important while using SGD, since it regulates the speed of the convergence [32].

B. Data processing and augmentation

To prevent model overfitting due to a limited number of training samples, we artificially extend our training set by adding spectrograms of modified input voice data. To achieve this, we applied two custom augmentations to the input samples, resulting in three times as many training samples.

The first augmentation, time-shifting involved using a random signal rolling technique to prevent the model from fitting to the time localization of a given sample. For the second augmentation, we applied a high-pass filter to remove lower frequencies that might be irrelevant and prevent the network from adapting solely to low-frequency features, such as hoarseness. These augmentations were applied to the signal waveforms, and the modified data was used to generate spectrograms. Figure 1 presents a comparison of the original and modified waveforms.

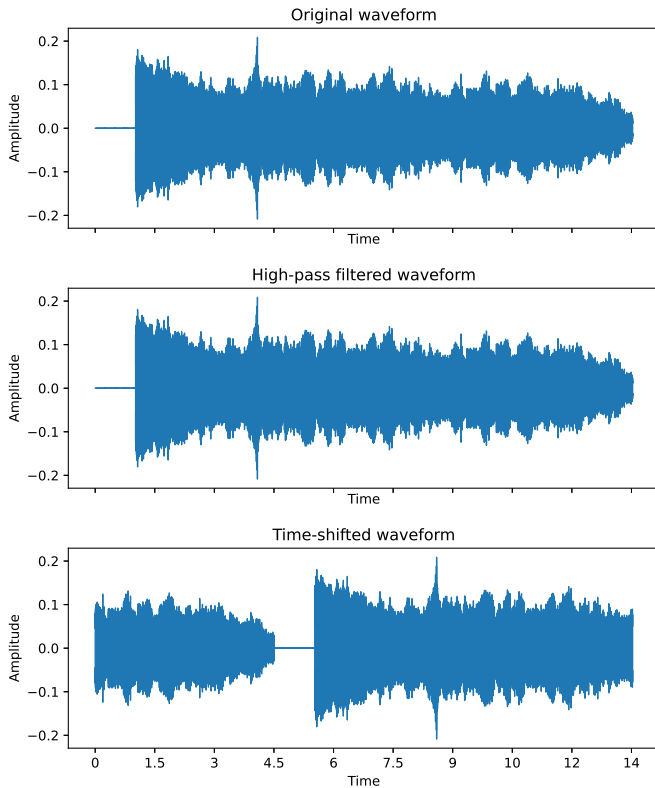


Fig. 1: Comparison of the time-series representation of the modified voice samples (Random vowel sample from the PC-GITA dataset)

To explore the feasibility of employing a CNN model, which was initially designed for image classification, for detecting

PD using voice signals, we transformed our voice data into image format. The conversion from audio to image was accomplished by generating log-frequency power-spectrograms of the input voice signals. To preserve the information from the time- and frequency domain, the audio signals were converted into the time-frequency domain using short-time Fourier transform (STFT). We utilized 40 ms frames with 20 ms overlap to calculate the spectrogram features. Subsequently, the log-spectrum of each signal was computed as $20 \log_{10}(X(h, k))$ and plotted to generate the spectrograms.

To further improve the models performance, we enhance the spectrogram images by applying Gaussian blurring to the spectrograms. The Gaussian blurring is a morphological image transformation method that aims to even out pixel values, thereby removing any extreme outliers. To perform this smoothing operation on the spectrogram image, 3×3 kernel was convolved with the spectrogram image. A comparison of the original and blurred spectrograms are shown in Fig. 2.

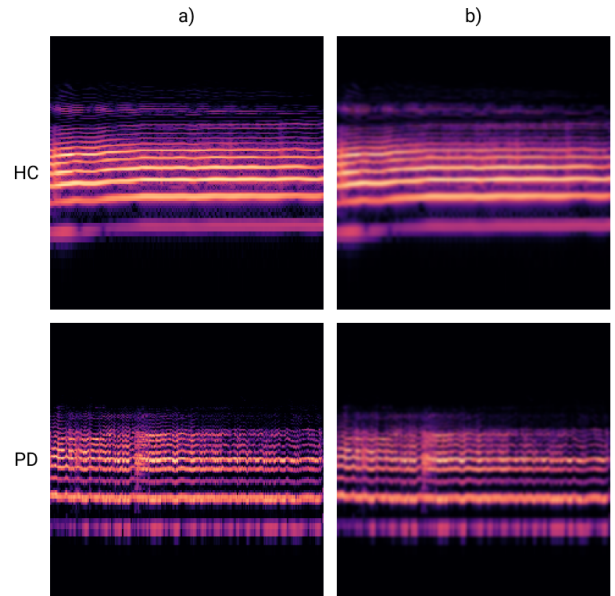


Fig. 2: Comparison of spectrograms of a healthy control (HC) and patient with PD: a) no additional preprocessing, b) Gaussian blurring (PC-GITA dataset, /a/ vowel) [33]

C. Multiple fine-tuned convolutional neural networks

One limitation of the CNN networks is their dependency on the data. To properly train a CNN model, large amount of training data is required. Collecting new data may be very difficult and time-consuming. To ease the need for high amount of training data, transfer learning (TL) can be utilized. Its aim is to overcome the requirement for the model to be trained from scratch in the target domain.

Let us first consider a domain $D = \{\chi, P(X)\}$, while χ represents the features and $P(X)$ is the marginal distribution, $X = \{x_1, x_2, \dots, x_n\} \in \chi$. $T = \{Y, f(\cdot)\}$ is a task, where $Y = \{y_1, y_2, \dots, y_n\}$ represents the labels and $f(\cdot)$ is a loss function to be learned by pairs $\{x_i, y_i\}$ [34].

Definition 1 (Transfer learning [34]). — Given a source domain D_s and a learning task T_s , a target domain D_t and a target task T_t , transfer learning aims to help improve the learning of the target predictive function $f_t(\cdot)$ in D_t using the knowledge in D_s and T_s , where $D_s \neq D_t$ or $T_s \neq T_t$.

Transfer learning can be employed to help when there is lack of data for neural network training. The pre-trained models have their weight set, which are fine-tuned on the target task. Usually the weights of the classification layers are fine-tuned, but any layer of the network can be fine-tuned if required.

We propose a novel approach, multiple fine-tuning, which aims to reduce the semantical gap between the source and the target task by introducing a training on a mediator dataset.

Definition 2 (Multiple fine-tuning [33]). — Let us consider a multiple fine-tuning (MFT) task $\mathcal{T}_{MFT} = (\mathcal{D}_s, \mathcal{T}_s, \mathcal{D}_m, \mathcal{T}_m, \mathcal{D}_t, \mathcal{T}_t, f_t(\cdot))$, where \mathcal{D}_m is a mediator domain and \mathcal{T}_m is the corresponding learning task. MFT aims to improve the learning of the target predictive function $f_t(\cdot)$ in \mathcal{D}_t by the knowledge in \mathcal{D}_m and \mathcal{T}_m by first aiming to improve it in \mathcal{D}_m by the knowledge in \mathcal{D}_s and \mathcal{T}_s . Here, $\mathcal{D}_s \neq \mathcal{D}_m$ or $\mathcal{T}_s \neq \mathcal{T}_m$, $\mathcal{D}_s \neq \mathcal{D}_t$ or $\mathcal{T}_s \neq \mathcal{T}_t$, and $\mathcal{D}_m \neq \mathcal{D}_t$ or $\mathcal{T}_m \neq \mathcal{T}_t$ respectively.

Practically, the process consists of three phases. First a CNN network is trained on a large dataset in the source domain, therefore learning the lower-level features allowing the model to achieve better convergence. In the second phase, the model is fine-tuned on a mediator dataset, which should be semantically closer to the target tasks and it could contain enough data to achieve convergence. On this task, the weights of all the layers of the CNN are updated. In the third phase, the classification layers (or all the layers if desired) are updated when fine-tuning the model on the target task.

The process of the MFT is illustrated on Fig. 3 considering

vowel classification in the Vowels dataset as mediator task.

D. Ensemble of multiple fine-tuned CNNs

One of the benefits of the MFT is to enhance the models performance and robustness. Another advantage of the MFT tuned network is that it creates diversity when trained on different mediator datasets. We can take advantage of this diversity to construct an ensemble of MFT trained CNNs. An ensemble’s goal is to reduce variance and decrease error rate. It integrates the predictions of multiple classifiers and consolidates their results into a unified output [35].

Definition 3 (Ensemble of multiple fine-tuned CNNs [33]). — Let us consider n MFT tasks, where $\mathcal{T}_{MFT}^i = (\mathcal{D}_s, \mathcal{T}_s, \mathcal{D}_m^i, \mathcal{T}_m^i, \mathcal{D}_t, \mathcal{T}_t, f_t^i(\cdot))$ is the i -th MFT task for $i \in \{0, 1, \dots, n\}$. Let $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$ be the vector of predictions made by the n multiple-fine-tuned base models, while \hat{y}_i is the label predicted by the i -th classifier. The final classification is obtained by majority voting, where the overall predicted label of the model ensemble can be calculated as $\hat{y}_e = \text{mode}(\hat{\mathbf{y}})$.

Fig. 4 illustrates the concept of the proposed ensemble.

IV. EXPERIMENTS AND RESULTS

We utilized two state-of-the-art CNN architectures, the Xception [36] and ResNet50 [37] networks. Both networks are pre-trained on the ImageNet dataset [38], which contains more than 14 million images, therefore the source task is image classification.

Two different mediator datasets are considered in this study to bridge the gap between the source and target tasks, the Vowels dataset and the SVD dataset (see II-A).

In all of our experiments, we resized the spectrograms images to 224×224 size, which is the default input shape

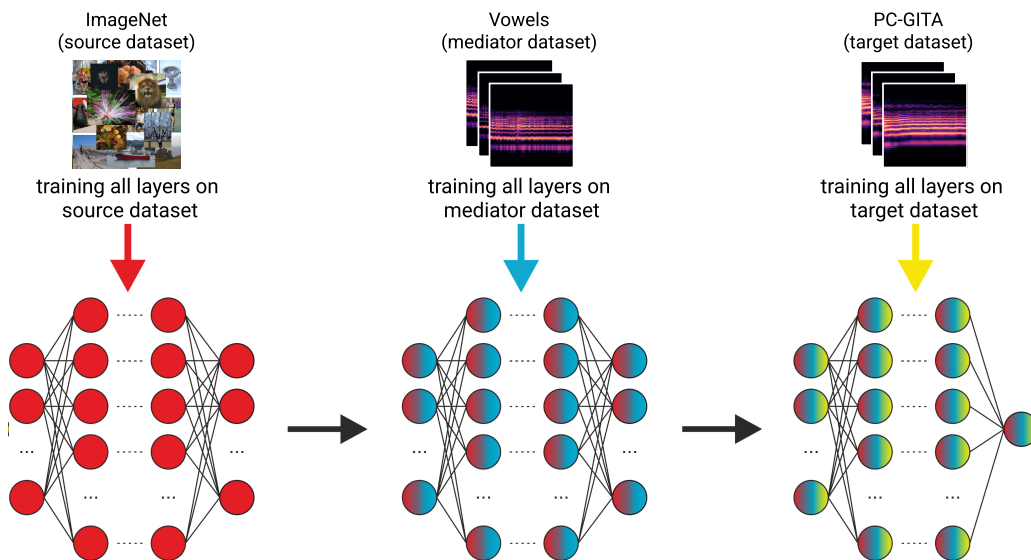


Fig. 3: Illustration of the process of multiple fine-tuning with Vowels dataset as mediator [33]

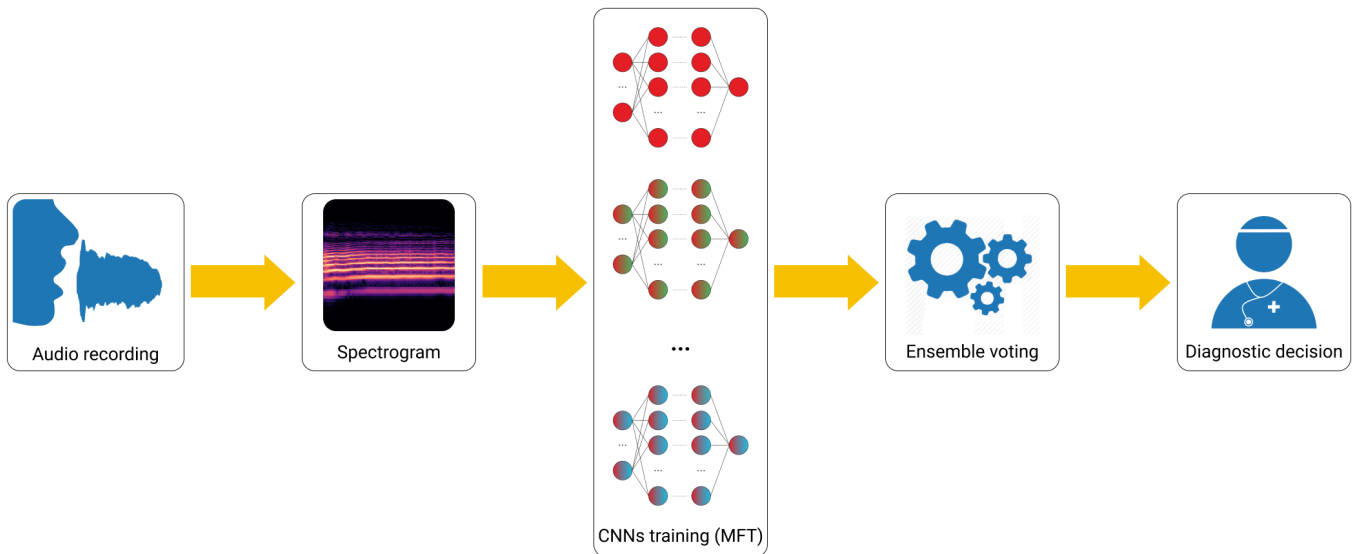


Fig. 4: Illustration of the concept of the proposed ensemble of multiple-fine-tuned CNNs [33]

of the ResNet50 model. They were additionally re-scaled to convert the pixel values into the range [0, 1].

In the first experiment, we fine-tuned our pre-trained models on the Vowels dataset, considering vowel classification as mediator task. Since this task does not focus on pathological voice data, in the second experiment we utilized the SVD as mediator dataset, where we labelled each disease together as pathological and classified healthy and pathological voice. In this case, we considered all the vowel subsets of the given dataset.

In both cases, we removed the models classification head and replaced with a new classifier block, consisting of three dense layers. A dropout layer was added between the last two layers of this block. The models were trained on the mediator tasks with all the layers unfreezed, therefore updating the weights of all the layers so it is able to learn voice-specific features and therefore we expect them to better adapt to the target task.

In the last phase, we fine-tuned the models on the target PC-GITA dataset, where the target task is PD detection in sustained vowel phonations. Again, we replaced the classification head of the networks and trained the models with all the layers unfreezed.

We employed the SGD optimizer with mini-batches and batch size of 16 to optimize the model. The learning rate was set to 0.0005 for the mediator and 0.005 for the target task.

For each experiment we employed a 10-fold cross validation for model validation, where data from one subject was used either in the training or the testing dataset. We measure the model’s performance according to accuracy (ACC), specificity (SP), sensitivity (SE), area under the ROC curve (AUC).

A. Results

In Tab. I we present the baseline results achieved without MFT training and the application of Gaussian blurring.

Network	Results				Vowel
	ACC (%)	SP (%)	SE (%)	AUC (%)	
Xception	88 ± 4	85.3 ± 10.24	90.7 ± 9.04	88 ± 4	/a/
Xception	93.99 ± 2.91	91.3 ± 5.21	96.7 ± 3.33	94 ± 2.91	/e/
Xception	76.67 ± 2.98	66.7 ± 7.3	86.7 ± 8.94	76.7 ± 2.98	/i/
Xception	76 ± 0.12	88 ± 9.8	64 ± 13.73	76 ± 3.59	/o/
Xception	92.99 ± 3.48	97.3 ± 4.42	88.7 ± 8.46	93 ± 3.48	/u/
ResNet50	86 ± 3.27	81.3 ± 11.08	90.7 ± 9.52	86 ± 3.27	/a/
ResNet50	87.67 ± 3.67	86 ± 8.67	89.3 ± 7.42	87.7 ± 3.67	/e/
ResNet50	80 ± 4.94	73.3 ± 11.93	86.7 ± 12.29	80 ± 4.94	/i/
ResNet50	80.67 ± 4.42	88 ± 8.84	73.3 ± 11.93	80.7 ± 4.42	/o/
ResNet50	94.33 ± 4.73	96.7 ± 4.47	92 ± 6.53	94.3 ± 4.73	/u/

TABLE I: Baseline results achieved on PC-GITA dataset without considering MFT and Gaussian blurring

Table II shows the prediction results of differently trained CNN networks while Gaussian blurring was applied on the input spectrograms. Since the application of Gaussian blurring showed a slight but consistent improvement in the classification results, we only present the results achieved considering blurred spectrograms in the case MFT approaches as well.

As the results reflect, both of the MFT approaches showed similar trends. In case of the Xception network the MFT with the blurred spectrograms improved the models accuracy in three out of five cases. The ResNet50 model was able to achieve improvements in four out of five cases.

To further demonstrate the effect of the proposed MFT, we illustrate the extracted features by the CNN network with and without the use of MFT. We employed the t-distributed stochastic neighbor embedding (t-SNE) [39] for this purpose, and the results are shown on Fig. 5 considering the /o/ vowel subset of the PC-GITA dataset and MFT with SVD as mediator dataset. Fig. 5(a) shows the extracted features by the model pre-trained on the ImageNet dataset, while on Fig. 5(b) are presented the features extracted by the same model multiple fine-tuned on the SVD dataset as mediator.

Network	No MFT				MFT - Vowels				MFT - SVD				Vowel
	ACC (%)	SE (%)	SP (%)	AUC (%)	ACC (%)	SE (%)	SP (%)	AUC (%)	ACC (%)	SE (%)	SP (%)	AUC (%)	
Xception	89.33 ± 4.2	84	94.7	89.3	90.67 ± 6.3	86	95.3	90.7	89.33 ± 7.4	88.7	90	89.3	/a/
Xception	93.67 ± 3.5	91.3	96	93.7	89.33 ± 5.1	90	88.7	89.3	86.67 ± 6.5	85.3	88	86.7	/e/
Xception	78.33 ± 2.2	68.7	88	78.3	85.33 ± 7.8	87.3	83.3	85.3	85.67 ± 6.5	89.3	82	85.7	/i/
Xception	73.33 ± 5.6	87.3	59.3	73.3	86 ± 8.1	82.7	89.3	86	88.67 ± 9.8	92.7	84.7	88.7	/o/
Xception	92 ± 3.1	96	88	92	88.33 ± 6.6	89.3	87.3	88.3	86.33 ± 7.4	86	86.7	86.3	/u/
ResNet50	87 ± 5	84	90	87	89 ± 5.2	92.7	85.3	89	87 ± 7.7	90	84	87	/a/
ResNet50	91.33 ± 4.8	90	92.7	91.3	89.67 ± 5.1	91.3	88	89.7	87 ± 7.2	87.3	86.7	87	/e/
ResNet50	81 ± 7.5	74.7	87.3	81	87.67 ± 8.7	87.3	88	87.7	88.67 ± 4.8	90	87.3	88.7	/i/
ResNet50	90.99 ± 2.1	93.3	88.7	91	86.33 ± 9.2	89.3	83.3	86.3	87 ± 6.6	86	88	87	/o/
ResNet50	94.33 ± 2.3	96	91.3	93.7	88 ± 8.3	86	90	88	86 ± 7.5	88.7	83.3	86	/u/

TABLE II: Prediction results of differently trained CNNs considered in this study

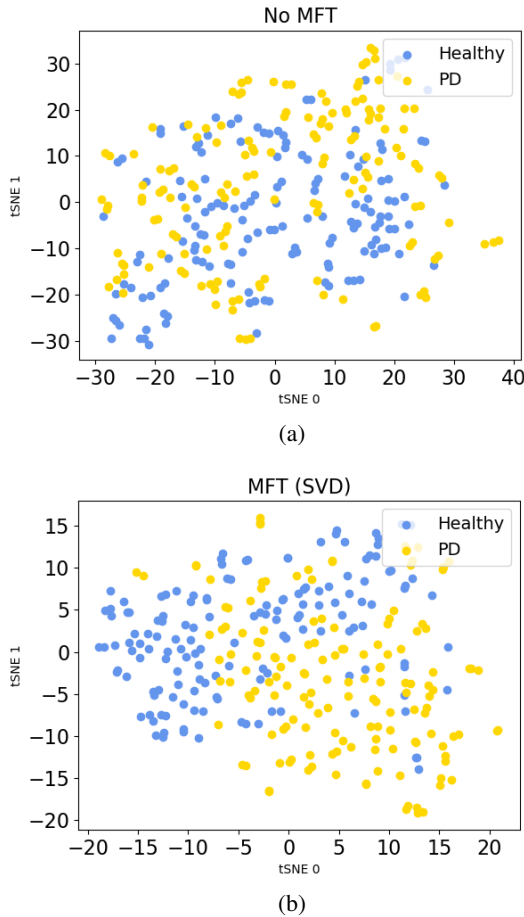


Fig. 5: Comparison of t-SNE visualizations of features extracted by ResNet50 with and without MFT (PC-GITA dataset - /o/ vowel subset) [33]

The separability of the predicted HC (yellow dots) and PD (blue dots) classes is improved in the case of features extracted by the MFT model.

Since an advantage of the MFT-trained CNNs is to create diversity on the classifier level, we leverage this benefit to create an ensemble of differently trained CNNs. We create

the ensemble using three base CNN models, each trained differently to achieve diversity:

- 1) a CNN network pretrained on the ImageNet dataset and then fine-tuned on the PC-GITA dataset;
- 2) a CNN network pretrained on the ImageNet dataset, fine-tuned on the Vowels dataset, and then fine-tuned on the PC-GITA dataset;
- 3) a CNN network pretrained on the ImageNet dataset, fine-tuned on the SVD dataset, and then fine-tuned on the PC-GITA dataset.

The proposed ensemble model uses the voting principle. The final diagnostic decision is made by majority voting, while each base classifier have equal weights. The achieved results are presented in Tab. III. The use of the ensemble model enabled us to achieve highly accurate results above 90% in all of the vowel tasks, which is a significant boost in the models performance. Thanks to the diversity of the base classifiers, the Xception model in case of the /a/ vowel task was able to achieve the best results, which represents a 99% accuracy.

Network	Results				Vowel
	ACC (%)	SP (%)	SE (%)	AUC (%)	
Xception	99 ± 2.13	86.23 ± 1.93	93.33 ± 2.37	89.77 ± 0.66	/a/
Xception	96.67 ± 3.33	88.87 ± 2.58	90.9 ± 3.62	89.9 ± 2.89	/e/
Xception	92 ± 7.78	81.77 ± 9.28	84.43 ± 2.58	83.1 ± 3.4	/i/
Xception	91.99 ± 6.18	87.57 ± 4.09	77.77 ± 13.19	83.77 ± 7.41	/o/
Xception	91.33 ± 7.02	88.23 ± 5.65	87.33 ± 0.53	88.53 ± 2.49	/u/
ResNet50	91.67 ± 7.92	87.77 ± 2.68	86.43 ± 2.58	86.43 ± 0.8	/a/
ResNet50	94.33 ± 7.67	90 ± 2.2	89.13 ± 2.58	88.77 ± 1.84	/e/
ResNet50	89 ± 5.78	85.33 ± 7.54	87.53 ± 0.33	85.9 ± 3.48	/i/
ResNet50	90.67 ± 5.93	88.87 ± 3.18	87.33 ± 2.9	87.1 ± 3.14	/o/
ResNet50	91 ± 5.78	91.33 ± 3.31	89.1 ± 2.26	89.9 ± 3.14	/u/

TABLE III: Results of the ensemble of MFT tuned CNNs for PD detection

V. CONCLUSION

In this paper we present an ensemble of multiple fine-tuned convolutional neural networks for Parkinson’s disease detection from voice recordings. Sustained vowel phonations are used as input converted to log-frequency power spectrograms. The CNNs work as feature extractors and therefore no manual feature engineering is required. Furthermore, the use of sustained vowel phonations is makes the model language

independents. The multiple fine-tunings objective is to bridge the gap between the source task and the target task extending the training process with a mediator task. Moreover, the MFT-trained CNN models create diversity on a classifier level. As the achieved results reflect, the diversity of the base CNN models enabled the ensemble classifier to achieve highly accurate results and thus improving the accuracy of the PD diagnosis process. The best results were achieved with the Xception model on the /a/ vowel subset of the PC-GITA dataset, which represents a 99% accuracy, 86.2% sensitivity, 93.3% specificity and 89.6% AUC.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under contract No. APVV-16-0211.

REFERENCES

- [1] O.-B. Tysnes and A. Storstein, "Epidemiology of parkinson's disease," *Journal of Neural Transmission*, vol. 124, no. 8, pp. 901–905, 2017.
- [2] M. Trail, C. Fox, L. O. Ramig, S. Sapis, J. Howard, and E. C. Lai, "Speech treatment for parkinson's disease," *NeuroRehabilitation*, vol. 20, no. 3, pp. 205–221, 2005.
- [3] D. G. Hanson, B. R. Gerratt, and P. H. Ward, "Cinegraphic observations of laryngeal function in parkinson's disease," *The Laryngoscope*, vol. 94, no. 3, pp. 348–353, 1984.
- [4] L. O. Ramig, C. Fox, and S. Sapis, "Speech treatment for parkinson's disease," *Expert review of neurotherapeutics*, vol. 8, no. 2, pp. 297–309, 2008.
- [5] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.
- [6] A. J. Hughes, S. E. Daniel, L. Kilford, and A. J. Lees, "Accuracy of clinical diagnosis of idiopathic parkinson's disease: a clinico-pathological study of 100 cases," *Journal of neurology, neurosurgery & psychiatry*, vol. 55, no. 3, pp. 181–184, 1992.
- [7] P. Ghaderyan and S. M. Ghoresi Beyrami, "Neurodegenerative diseases detection using distance metrics and sparse coding: A new perspective on gait symmetric features," *Computers in Biology and Medicine*, vol. 120, p. 103736, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482520301189>
- [8] M. Gazda, M. Hireš, and P. Drotár, "Multiple-fine-tuned convolutional neural networks for Parkinson's disease diagnosis from offline handwriting," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–12, 2021.
- [9] J. R. Orozco-Arroyave, J. D. Arias-Londono, J. F. Vargas-Bonilla, and E. Nöth, "Analysis of speech from people with parkinson's disease through nonlinear dynamics," in *International conference on nonlinear speech processing*. Springer, 2013, pp. 112–119.
- [10] S. Skodda, W. Visser, and U. Schlegel, "Short-and long-term dopaminergic effects on dysarthria in early parkinson's disease," *Journal of Neural Transmission*, vol. 117, no. 2, pp. 197–205, 2010.
- [11] J. Ruzs, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease," *The journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 350–367, 2011.
- [12] J. R. Duffy, *Motor speech disorders e-book: Substrates, differential diagnosis, and management*. Elsevier Health Sciences, 2019.
- [13] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, and J. I. Godino-Llorente, "Advances in parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects," *Biomedical Signal Processing and Control*, vol. 66, p. 102418, 2021.
- [14] Q. C. Ngo, M. A. Motin, N. D. Pah, P. Drotár, P. Kempster, and D. Kumar, "Computerized analysis of speech and voice for parkinson's disease: A systematic review," *Computer Methods and Programs in Biomedicine*, p. 107133, 2022.
- [15] V.-V. Eklund, "Data augmentation techniques for robust audio analysis," Master's thesis, 2019.
- [16] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A simple data augmentation method for automatic speech recognition," *arXiv preprint arXiv:1904.08779*, 2019.
- [17] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates *et al.*, "Deep speech: Scaling up end-to-end speech recognition," *arXiv preprint arXiv:1412.5567*, 2014.
- [18] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: A survey," *arXiv preprint arXiv:2204.08610*, 2022.
- [19] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [20] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognition*, p. 109347, 2023.
- [21] N. Jaitly and G. E. Hinton, "Vocal tract length perturbation (vtlp) improves speech recognition," in *Proc. ICML Workshop on Deep Learning for Audio, Speech and Language*, vol. 117, 2013, p. 21.
- [22] S. Uhlich, M. Porcu, F. Giron, M. Enekl, T. Kemp, N. Takahashi, and Y. Mitsufoji, "Improving music source separation based on deep neural networks through data augmentation and network blending," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 261–265.
- [23] N. Kanda, R. Takeda, and Y. Obuchi, "Elastic spectral distortion for low resource speech recognition with deep neural networks," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*. IEEE, 2013, pp. 309–314.
- [24] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rativa, and E. Nöth, "New spanish speech corpus database for the analysis of people suffering from parkinson's disease." in *LREC*, 2014, pp. 342–347.
- [25] W. J. B. M. Pützer, "Saarbruecken voice database." [Online]. Available: <http://www.stimmdatenbank.coli.uni-saarland.de>
- [26] D. A. R. Venegas, *DATASET OF VOWELS*, 2018, <https://www.kaggle.com/darubiano57/dataset-of-vowels>.
- [27] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [28] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.
- [29] A. Hyvärinen and U. Köster, "Complex cell pooling and the statistics of natural images," *Network: Computation in Neural Systems*, vol. 18, no. 2, pp. 81–100, 2007.
- [30] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*. IEEE, 2012, pp. 3304–3308.
- [31] M. Zinkevich, M. Weimer, L. Li, and A. J. Smola, "Parallelized stochastic gradient descent," in *Advances in neural information processing systems*, 2010, pp. 2595–2603.
- [32] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.
- [33] M. Hireš, M. Gazda, P. Drotár, N. D. Pah, M. A. Motin, and D. K. Kumar, "Convolutional neural network ensemble for parkinson's disease detection from voice recordings," *Computers in biology and medicine*, vol. 141, p. 105021, 2022.
- [34] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [35] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, "Practical machine learning tools and techniques," *Morgan Kaufmann*, p. 578, 2005.
- [36] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [38] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [39] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

Využitie algoritmov strojového učenia na detekciu malvéru v operačnom systéme Android

Michal Hulič

Katedra počítačov a informatiky Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach
Slovensko
michal.hulic@tuke.sk

Liberios Vokorokos

Katedra počítačov a informatiky Fakulta elektrotechniky a informatiky Technická univerzita v Košiciach
Slovensko
liberios.vokorokos@tuke.sk

Abstrakt— Tento článok sa zaoberá problematikou z oblasti počítačovej bezpečnosti. Táto publikácia sa konkrétne venuje podobe škodlivého kódu, ktorý napáda predovšetkým mobilný operačný systém Android. Ako hlavná oblasť výskumu bolo využitie viacerých algoritmov strojového učenia pri klasifikácii vzoriek na maligne a benigne za použitia statickej, dynamickej a hybridnej analýzy aplikácií a ich zdrojového kódu pri využití viacerých dátových súb (datasetov). Na dosiahnutie cieľa sme využili dve známe frameworky pre klasifikáciu vzoriek a výstupné dáta sme analyzovali pomocou algoritmov strojového učenia. Ako najideálnejší algoritmus pre tento typ úloh sa javil AdaBoost a ako typ analýzy vzoriek mal najlepšie výsledky hybridný model analýzy.

KLúčové slová — Analýza vírusových vzoriek, Strojové učenie, Framework, Počítačové vírusy, Operačný systém Android

I. ÚVOD

Písanie tejto publikácie bolo inšpirované hlavne knižnou literatúrou, ktorá formovala aj stanovené oblasti výskumu a to najmä oblasť počítačovej bezpečnosti. Jedná sa o literatúru uvedenú v citáciách [1–3]. Naším cieľom je preskúmať súčasný stav prístupov k detekcii mobilného malvéru, pričom zavedieme aj nové metodiky detekcie na zabezpečenie optimálnych bezpečnostných protipatrení. Predostreté ciele realizujeme v nasledujúcej postupnosti krokov. Analyzovať stav problematiky detekcie malvéru v zariadeniach s operačným systémom Android. K dosiahnutiu tohto zámeru bolo potrebné najprv preskúmať hlavné komponenty modelu bezpečnosti systému Android. Následne je možné pristúpiť k charakteristike rôznych typov škodlivého mobilného malvéru, vybraných zahmlievacích techník, ako aj najznámejších techník prieniku mobilného malvéru. Realizovať rozbor funkčnosti, významu, vzniku a potenciálneho vplyvu škodlivého kódu v procese statickej, dynamickej a hybridnej analýzy škodlivého softvéru, pričom jednotlivé typy analýz budú porovnané. S týmto krokom úzko súvisí špecifikácia techník statickej analýzy, techník dynamickej analýzy, techník hybridnej analýzy a v neposlednom rade rozbor technológií na detekciu škodlivého malvéru pre Android. Preskúmať použitie algoritmov strojového učenia pri klasifikácii malvéru. Skúma sa využitie troch frameworkov na detekciu malvéru pre export metriky pre osem známych algoritmov strojového učenia. Implementujeme a vyhodnocujeme využitie frameworku MalwareAnalyst, ktorý predstavuje automatizovaný a

konfigurovateľný hybridný analytický nástroj, kombinujúci statickú analýzu s dynamickou inštrumentáciou na analýzu správania aplikácií na platforme Android.

II. BEZPEČNOSŤ ANDROIDU

Oprávnenia aplikácií chránia súkromie používateľov tým, že ochraňujú prístupové údaje, ako napríklad kontakty používateľov, a obmedzujú vykonávanie akcií, ako napríklad spustenie aplikácií alebo uskutočnenie hovoru. Android kategorizuje povolenia do rôznych typov, vrátane povolení na inštaláciu dočasných aplikácií, povolení na beh aplikácií a špeciálnych povolení [4]. Povolenia pre dočasné práva poskytujú aplikácii obmedzený prístup k údajom a akciám. Systém automaticky udeľuje povolenie na inštaláciu aplikácie po jej nainštalovaní. Povolenia počas behu programu, sa tiež klasifikujú ako nebezpečné povolenia, poskytujú aplikácii prístup k obmedzeným údajom a obmedzeným akciám. Aplikácia musí požiadať o povolenie na beh skôr, ako získa prístup k obmedzeným údajom alebo vykoná obmedzené akcie. Od API úrovne 23, aplikácie, ktoré používajú run time povolenia, vyzývajú používateľov na prijatie povolenia počas používania aplikácie a nie pri jej inštalácii. Od API úrovne 29 sú používatelia vyzývaní k povoleniu alebo zamietnutiu povolenia pomocou dialógového okna pre runtime povolenia. Existuje celkom 30 nebezpečných povolení uvedených v rozhraní Android API.

Android vyžaduje, aby všetky aplikácie boli pred inštaláciou do zariadenia podpísané certifikátom alebo aktualizované s podpísaným platným certifikátom. Aplikácie, ktoré sa pokúšajú nainštalovať bez podpisu, sú zamietnuté operačným systémom pri inštalácii balíka v zariadení s Androidom [5]. Jednou z dôležitých výhod podpisu kódu je možnosť kontroly operačným systémom, či nainštalovaná aplikácia bola aktualizovaná zo zdroja s rovnakým podpisom certifikátu.

Android využíva bezpečnostne vylepšený Linux (SELinux - Security-Enhanced Linux) na presadzovanie povinného riadenia prístupu (MAC - Mandatory Access Control) vo všetkých procesoch, aj v tých, ktoré sú spustené s privilégiami root/superužívateľa [6]. SELinux pracuje na princípe predvoleného odmietnutia, čo znamená, že všetko, čo nie je výslovne povolené, je zamietnuté. SELinux môže byť spustený v dvoch režimoch: povoleným (ang. permissive), v ktorom sa

odmietnutie povolenia zaznamenáva, ale nie je presadzované, a vynúteným (ang. enforced), v ktorom sú povolenia odmietnuté. Android prevádzkuje SELinux v režime presadzovania (ang. enforcing mode).

Okrem SELinuxu používa Android aj SECComp (SECure COMputing) filter na ďalšie obmedzenie prístupu k jadrú blokováním prístupu k určitým systémovým volaniam. Od Androidu 7.0 sa SECComp používa pre procesy v mediálnom frameworku [7]. Od systému verzie Android 8.0, je do verzie Zygote nainštalovaný SECComp filter, ktorý sa správa ako hlavný proces, z ktorého sú odvodené všetky aplikácie, teda aplikácie sa spúšťajú ako podproces. Tento proces blokuje prístup k určitým systémovým volaniam, ktoré boli použité pri niektorých bezpečnostných útokoch. SECComp FILTER blokuje tiež volania systému riadenia kľúčov, ktoré nie sú užitočné pre aplikácie.

Overené zavádzanie (ang. booting) zabezpečuje, že všetok spustený kód pochádza z dôveryhodného zdroja, napríklad od výrobcu zariadenia, a nie od útočníka alebo škodlivého kódu [8]. Funguje tak, že vytvára úplný reťazec dôveryhodnosti, ktorý sa začína v hardvérovo chránenom koreni dôveryhodnosti, pokračuje cez zavádzač až po zavádzací oddiel a ďalšie overené oddiely. V tejto kapitole sme sa podrobne venovali komponentom operačného systému Android, ktoré zabezpečujú bezpečnosť operačného systému voči malvéru. Táto komplexná analýza nám umožňuje venovať sa v ďalšej kapitole podrobnejšie rozboru samotného mobilného malvéru.

III. MOBILNÝ MALVÉR

Napriek dobre premyslenému a prepracovanému bezpečnostnému modelu systému Android poskytuje jeho veľký podiel na trhu, ako aj jeho otvorenosť, zaujímavú príležitosť pre útočníkov. V priebehu rokov sa škodlivý softvér pre Android rozšíril a vyvinul viac ako v prípade akejkoľvek inej platformy. Malvér môže byť označený ako akýkoľvek typ softvéru, ktorý vedie k používaniu zariadenia bez vedomia majiteľa. Existuje mnoho rôznych typov mobilného malvéru [8], vrátane trójskych koní, červov, botnetov, spyvéru a ransomvéru, ktoré sa riadia rôznymi modelmi správania. Najnovšie typy mobilného malvéru popisujeme v nasledujúcich podkapitolách. Podľa poslednej správy o hrozbách, vydané spoločnosťou Symantec, bolo v posledných rokoch zistených oveľa viac vzoriek malvéru ako v predchádzajúcich rokoch. V jednom roku bolo vytvorených viac ako 317 miliónov nových vzoriek malvéru, čo znamená, že každý deň sa do siete internetu uvoľnil takmer milión nových hrozieb. V tejto kapitole vysvetľujeme, čo je malvér a prečo predstavuje relevantnú oblasť výskumu. Opisujeme rôzne typy malvéru a techniky zahmlievania (obfuskácie). Následne predstavujeme rôzne metódy analýzy malvéru. Rôzne typy vírusov prezentujeme v nasledujúcich podkapitolách.

A. Vírusy

Termín vírus v súvislosti s počítačmi prvýkrát zaviedol Cohen v roku 1987 [9]. Vírus je softvér, ktorý môže infikovať iné programy ich úpravou. Prípadajú sa k iným programom a vykonávajú sa na pozadí, zatiaľ čo hosťiteľský program robí

to, na čo bol pôvodne naprogramovaný. Skladá sa z troch častí:

- mechanizmus infekcie (ang. mechanism) - je spôsob, akým sa vírus „rozмноžuje“ alebo šíri.
- spúšťač (ang. trigger) - je podmienka, za ktorej sa vírus aktivuje alebo doručí svoje užitočné zaťaženie.
- užitočné zaťaženie (ang. payload) - je škodlivá činnosť, ktorú vykonáva.

B. ČERV

Funkčne sú vírusy a červy podobné. Červy sú infekčné a šíria sa. Útočníci navrhujú červy tak, aby sa mohli replikovať ako vírusy. Červ sa však replikuje bez cieľenia a infikovania konkrétnych súborov, ktoré sú už na počítači prítomné. Na šírenie využívajú počítačovú sieť a spoliehajú sa na zlyhania zabezpečenia na cieľovom počítači, aby sa k nemu dostali a kradli alebo vymazávali údaje. Červy sú sieťové vírusy, ktoré sa môžu šíriť po sieti tým, že sa duplikujú. Nepremieňajú ani nezničia súbory používateľa, ale uložia sa do hlavnej pamäte, kde sa duplikujú a môžu tak spôsobiť, že systém a sieť prestanú reagovať [11].

Stallings et al. [1] uvádzajú, že "červ je program, ktorý sa dokáže replikovať a posielat' kópie z počítača do počítača cez sieťové pripojenia. Po príchode môže byť červ aktivovaný, aby sa znova replikoval a šíril". Zvyčajne sa to robí jedným z dvoch spôsobov: využitím zraniteľností v sieťovej službe alebo cez e-mail. V dnešnej dobe môžeme bezpečne predpokladať, že ako tretia možnosť slúžia rôzne sociálne siete. Podobne ako vírusy aj červy pracujú podľa nasledujúcich troch krokov:

1. Vyhľadajte ďalšie systémy na infekciu.
2. Vytvorte spojenie so vzdialeným systémom.
3. Skopírujte sa do vzdialeného systému a spustíte kópiu na novom systéme. Okrem toho sa môže pokúsiť zistiť, či je nový systém už infikovaný predtým, ako sa sám infikuje.

C. ADVÉR

Advér (ang. adware) je označovaný ako reklamný softvér [12]. Advér je označenie pre tie programy, ktoré sú navrhnuté tak, aby zobrazovali reklamy na počítači pri prezeraní kybernetického priestoru a potom presmerovali požiadavky vyhľadávania na reklamné webové stránky a zároveň pritom zhromažďovali informácie. Advér sa považuje za škodlivý, pretože zhromažďuje údaje bez súhlasu. Ide o typ malvéru, ktorý automaticky distribuuje reklamy aj do operačného systému. Softvér s podporou reklamy sa často dodáva spolu so softvérom a aplikáciami a väčšina z nich slúži ako nástroj na získanie príjmu.

D. SPYVÉR

Spyvér je skratka pre špiónážny softvér, ktorý sleduje prácu počítačového systému. Je naprogramovaný tak, aby monitoroval a zaznamenával údaje o prehliadaní, ako aj dôverné informácie a iné činnosti [13]. Ide o typ malvéru, ktorý špehuje a sleduje akúkoľvek aktivitu a odosiela ju na vzdialený server. Možnosti spyvéru môžu zahŕňať akumuláciu stlačenia klávesov, zhromažďovanie finančných údajov alebo monitorovanie aktivity.

E. TRÓJSKY KÔŇ

Trójsky kôň alebo trojan je škodlivý program, ktorý oklame užívateľa spôsobom, že sa javí ako použiteľný neškodný súbor alebo ako legitímny softvér. Útočníci distribuujú trójske kone ako rutinný softvér, hru alebo nástroj, ktorý presvedčí používateľa, aby si ho nainštaloval do svojho počítača [14]. Názov trójsky kôň použil staroveký grécky básnik Homér vo svojej básni Ilias. Odyseus, jeden z gréckych bojovníkov vymyslel lešť: Gréci predstierali odchod z Tróje a na pobreží zanechali ako dar dreveného obrovského koňa. Trójanovia sa potešili a vtiahli koňa do mesta. Vo vnútri koňa boli ukrytí grécki bojovníci, ktorí v noci vyšli z koňa a otvorili brány mesta Grékom. Trójske kone sú rovnako škodlivé na počítačoch a považujú sa za deštruktívne. Experti na kybernetickú bezpečnosť považujú trójske kone za jeden z najnebezpečnejších typov malvéru, konkrétne najnebezpečnejšie trójske kone sú určené na získavanie finančných informácií od používateľov.

Zadné vrátka (ang. backdoor) sú tajnou cestou do programu, ktorá obchádza bežné bezpečnostné postupy. Takto umožňuje neoprávnený prístup do systému [15]. Môže byť spustená špeciálnou sériou vstupov, ktoré sú spustené so špeciálnym užívateľským menom.

F. ZAPISOVAČ STLAČENIA KLÁVESOV

Zapísač stlačenia klávesov (ang. key logger) zachytáva prenikanie stlačenia klávesov na počítači užívateľom, často bez vedomosti užívateľa o tejto aktivite [16]. Keyloggery majú legitímne využitie ako profesionálny nástroj na monitorovanie informačných technológií. Zapísač stlačenia klávesov sa však bežne používa na škodlivé účely, a teda zachytávanie citlivých informácií, ako sú používateľské mená, heslá, odpovede na bezpečnostné otázky a finančné informácie.

G. Rootkit

Rootkit je súbor softvérových nástrojov [16], zvyčajne škodlivých, ktoré poskytujú neautorizovanému používateľovi privilegovaný prístup k počítaču. Po nainštalovaní rootkitu môže ovládač rootkitu na diaľku spúšťať súbory a konfigurácie transmutačného systému na hostiteľskom počítači [16]. Rootkity sa nemôžu samy šíriť ani replikovať. Musia byť nainštalované na zariadení.

H. Zadné vrátka

Zadné vrátka (ang. backdoor) sú tajnou cestou do programu, ktorá obchádza bežné bezpečnostné postupy. Takto umožňuje neoprávnený prístup do systému [16]. Môže byť spustená špeciálnou sériou vstupov, ktoré sú spustené so špeciálnym užívateľským menom.

I. Ransomvér

Ransomware je typ malvéru, ktorý uzamkne údaje v počítači obeť, zvyčajne šifruje nimi. Kyberzločinec ukrytý za malvérom požaduje platbu pred dešifrovaním uzamknutých údajov a po zaplatení vráti obeť prístup k dátam. [17] Motív útokov ransomvéru je takmer vždy peňažný a na rozdiel od iných typov útokov je obeť bežne upozornená, že došlo k

zneužitiu a dostane pokyny na uskutočnenie platby, aby sa údaje obnovili do normálu. Ide o typ malígneho softvéru, ktorý v podstate obmedzuje prístup k počítaču zašifrovaním súborov alebo uzamknutím systému a zároveň vyžaduje výkupné. Používatelia sú nútení zaplatiť autorovi škodlivého softvéru za odstránenie obmedzení a získanie prístupu k svojim počítačom.

IV. ZAHMLIEVACIE TECHNIKY

Zahmlievacie techniky z anglického výrazu obfuskácia, je technika ktorá robí zdrojové kódy ťažšie zrozumiteľné a lepšie ukryté. Autori malvéru používajú rôzne techniky zahmlievania. Robia tak preto, aby bolo ťažšie ich odhaliť antivírusovými programami a aby bolo pre reverzných inžinierov náročnejšie pochopiť, čo program robí [18]. Zahmlený program bude robiť to isté ako pôvodný kód, ale bude pre nový kód ťažšie zrozumiteľný. Niektoré spoločnosti a vývojári takisto používajú zahmlený kód na ochranu svojho duševného vlastníctva (zdrojového kódu). Avšak autori malvéru sa snažia sťažiť detekciu svojho malvéru a preto antivírusoví tvorcovia museli vyvinúť nové metódy na jeho odhalenie.

Ďalej uvádzame rôzne metódy používané na zahmlievanie škodlivého softvéru.

A. Šifrovanie

Aby sa skryla škodlivá časť kódu, táto časť je často šifrovaná. Ako už bolo uvedené, pre každú infekciu sa používa nový náhodný kľúč. Týmto spôsobom bude podpis súboru zakaždým iný. Zašifrovaná časť kódu bude dešifrovaná iba pri spustení súboru. Nevýhodou tejto metódy je, že časť kódu, ktorá vykonáva dešifrovanie, bude vo všetkých verziách rovnaká. Z tohto dôvodu môže byť vírus stále rozpoznávaný na základe vzoru dešifrovacieho kódu.

B. Baliaci softvér – Komprimovanie

Okrem šifrovania je možné použiť baliace programy na zmenu podpisu súboru a vyhnúť sa detekcii. Napriek skutočnosti, že programy na komprimovanie boli pôvodne vytvorené na zmenšenie veľkosti súboru z dôvodu obmedzení miesta na disku a šírky pásma, v súčasnosti ich často používajú autori škodlivého softvéru. Packery alebo komprimovače sú známe ako dobré nástroje na chaotické usporiadanie programu, pretože malá zmena v jednom z balených súborov bude mať za následok veľmi odlišný podpis. Je tiež ľahké zabaliť rovnaký malvér do niekoľkých rôznych baliacich programov, aby bolo opäť možné získať rôzne podpisy.

C. Polymorfizmus

Namiesto zmeny runtime kódu (bežiacemu procesu), ako je to v prípade softvéru na komprimovanie, sa zmení statický binárny kód. Účelom je opäť zmeniť podpis súboru.

Postup je nasledovný:

- (a) Najprv malvér zapne polymorfický spúšťač.
- (b) Zmutovaný malvér sa dešifruje do strojových inštrukcií.
- (c) Tieto strojové inštrukcie sú zapísané do pamäte.
- (d) Spustí sa dešifrovaný malvér.
- (e) Škodlivá činnosť je vykonaná.

(f) Vygeneruje sa nový kľúč.

(g) Inštrukcie stroja sa transformujú späť pomocou nového kľúča

Inštrukcie stroja načítané v pamäti sú rovnaké pre každú verziu malvéru. V dôsledku toho môže byť podpis rozpoznaný pri spustení kódu v karanténe.

D. Metamorfizmus

Zatiaľ čo strojové inštrukcie v pamäti sú rovnaké v každej verzii mutovaného malvéru používajúceho polymorfny prístup, nie je to tak v prípade metamorfneho prístupu. Pri tomto prístupe sa zmenia inštrukcie načítané v pamäti a zapisujú sa späť do infikovaného súboru.

E. Všeobecné techniky zahmlievania

Nasledujúce techniky zahmlievania sa často používajú v predtým opísaných polymorfických a metamorfnych kategóriách malvéru.

- Zahmlievanie reťazcov Zahmlievanie reťazcov sa používa na sťaženie pochopenia obsahu premenných v kóde. Dá sa to dosiahnuť rozdelením reťazcov na menšie časti alebo ich zakódovaním.

- Randomizácia mien Samotné premenné a názvy funkcií môžu byť tiež obfuskované. Použitím náhodných názvov neposkytuje informácie o tom, na čo sa používa, preto ho treba chápať z kontextu, ktorý vyžaduje vyššiu úroveň zručností testera.

- Vloženie mŕtveho kódu Vkladanie mŕtveho kódu znamená pridať kód, ktorý neničí schopnosti programu. Kód je tu len na to, aby zmiatol každého, kto sa do neho pozerá, alebo aby zmenil podpis súboru. Môžu to byť jednoduché inštrukcie, ktoré nerobia nič podobné ako inštrukcia nop, alebo to môže byť séria inštrukcií, ktoré skončia v počítačnom bode. Niektorí autori tiež pridávajú zložitý kód, ktorý sa nikdy nespustí.

- Preradenie registrov Pri opätovnom priradení registrov sa používané registre prepínajú v novej verzii malvéru. Napríklad register EAX možno znova priradiť k registru EBX. Nezavádza žiadne časové oneskorenie.

- Zmena poradia podprogramov Zmenou poradia podprogramov v kóde sa podpis súboru ľahko zmení. Pre program s n podprogramami možno vytvoriť $n!$ variantov súboru

- Náhrada inštrukcie Ďalšou technikou je nahradenie inštrukcie. Cieľom je použiť rôzne pokyny, ktoré poskytujú rovnaké výsledky. Príkladom je nahradenie inštrukcie mov inštrukciami push a pop.

F. Selecting a Template (Heading 2) Transpozícia kódu

Zmenou poradia sekvencií môže dôjsť k ďalšiemu zmätku. Je možné buď zamiešať inštrukcie a použiť skoky na získanie správneho príkazu na vykonanie, alebo je možné analyzovať, ktoré inštrukcie sú na sebe závislé, a prepínať tie, ktoré nie sú.

G. Anti debugging

Na skomplikovanie procesu ladenia možno použiť rôzne techniky proti ladeniu. Branco et. al [19] poskytuje dôkladný

prehľad niekoľkých možných techník. Uvádzam niekoľko z nich v tejto časti:

- PEB NtGlobalFlag Prítomnosť hodnoty v tomto poli môže naznačovať ladiaci program.

- IsDebuggerPresent Funkcia kernel32, ktorá používa pole BeingDebugged v PEB (Process Environment Block) na detekciu ladiaceho nástroja.

- Príznaky haldy Niekoľko príznakov haldy označuje prítomnosť ladiaceho nástroja.

- Vlastné ladenie Proces môže vytvoriť svoju kópiu a pripojiť kópiu k sebe ako debugger. Týmto spôsobom sa k nemu nemôže pripojiť skutočný debugger, pretože k procesu môže byť naraz pripojený iba jeden debugger.

V. ANALÝZA ŠKODLIVÉHO SOFTVÉRU

Analýza škodlivého softvéru je proces kontroly a rozboru funkčnosti, významu, vzniku a potenciálneho vplyvu škodlivého kódu. Tento proces sa vykonáva pomocou statickej, dynamickej alebo hybridnej analýzy s využitím nástrojov, techník a metód špecifických pre každý typ analýzy. Cieľom analýzy je extrakcia informácií z kódu škodlivého softvéru a tieto informácie môžu byť rôzneho charakteru [20]. Niektoré z nich môžu byť jednoduché, napríklad ako je typ súboru, kým iné môžu byť zložitejšie, ako napríklad účely zneužitia a detailná inšpekcia kódu. Analýza malvéru znamená kontrolu binárnych súborov malvéru, aby sme pochopili jeho funkčnosť a metódy na identifikáciu a klasifikáciu súborov. Tento proces zahŕňa zhromažďovanie funkcií z binárnych súborov škodlivého softvéru, aby sme získali prvotný prehľad o analyzovaných údajoch. Atribúty alebo vlastnosti údajov/vzoriek sa analyzujú, aby sme získali relevantné informácie.

A. Tradičný prístup

Vzťahuje sa na analýzu alebo vyšetrovanie malígneho programu bez jeho vykonania. Je to proces získavania informácií z malvéru, ktorý sa nespúšťa. Statická analýza môže byť vykonaná priamo so skutočným kódom (ak je prítomný) a ak nie, môže byť aplikovaná na rôzne reprezentácie spustiteľných súborov. Statická analýza sa považuje za najjednoduchšiu, najrýchlejšiu a menej neistý proces analýzy

. Je to najjednoduchšie a najrýchlejšie riešenie, pretože na proces analýzy nie sú potrebné žiadne špeciálne podmienky a požiadavky.

Malvér je jednoducho podrobený analytickým nástrojom. Nie je to nebezpečná metóda, pretože malvér sa počas analýzy nespustí. V dôsledku toho následne neexistuje žiadne nebezpečenstvo, že by sa infekcia počas podrobenia analýzy rozšírila, a neexistuje obava z vytvárania bezpečného prostredia pre statickú analýzu. Vzory zistené pri tomto druhu analýzy zahŕňajú signatúru reťazcov, bajtové sekvencie alebo operačné kódy (operačné kódy), frekvenčnú distribúciu, bajtovú sekvenciu n-gramov alebo operačné kódy n-gramy, volania API, štruktúru rozloženého programu atď. Cieľom analýzy je identifikovať malvér predtým, ako prebehne dynamická analýza. Demontáž škodlivých programov je potrebná na odhalenie vzorov, niektoré bežné demontážne nástroje (disasembly) sú objdump, IDA Pro atď. Statická

analýza sa považuje za menej presnú metódu analýzy malvéru oproti dynamickej, pretože údaje extrahované zo statickej analýzy nemusia byť úplné, údaje sa zhromažďujú, keď je malvér v pasívnom režime (nevykonáva sa). Extrahované údaje sú obmedzené a neodhalia veľa dôležitých informácií o malvéri. Prevládajúce techniky používané v statickej analýze sú:

B. Statická analýza

Vzťahuje sa na analýzu alebo vyšetrenie malígneho programu bez jeho vykonania. Je to proces získavania informácií z malvéru, ktorý sa nespúšťa. Statická analýza môže byť vykonaná priamo so skutočným kódom (ak je prítomný) a ak nie, môže byť aplikovaná na rôzne reprezentácie spustiteľných súborov. Statická analýza sa považuje za najjednoduchšiu, najrýchlejšiu a menej neistý proces analýzy [21]. Je to najjednoduchšie a najrýchlejšie riešenie, pretože na proces analýzy nie sú potrebné žiadne špeciálne podmienky a požiadavky. Malvér je jednoducho podrobený analytickým nástrojom. Nie je to nebezpečná metóda, pretože malvér sa počas analýzy nespustí. V dôsledku toho následne neexistuje žiadne nebezpečenstvo, že by sa infekcia počas podrobenia analýzy rozšírila, a neexistuje obava z vytvárania bezpečného prostredia pre statickú analýzu. Vzory zistené pri tomto druhu analýzy zahŕňajú signatúru reťazcov, bajtové sekvencie alebo operačné kódy (operačné kódy), frekvenčnú distribúciu, bajtovú sekvenciu n-gramov alebo operačné kódy n-gramy, volania API, štruktúru rozloženého programu atď [22].

Cieľom analýzy je identifikovať malvér predtým, ako prebehne dynamickej analýze. Demontáž škodlivých programov je potrebná na odhalenie vzorov, niektoré bežné demontážne nástroje (disasembly) sú objdump, IDA Pro atď. Statická analýza sa považuje za menej presnú metódu analýzy malvéru oproti dynamickej, pretože údaje extrahované zo statickej analýzy nemusia byť úplné, údaje sa zhromažďujú, keď je malvér v pasívnom režime (nevykonáva sa). Extrahované údaje sú obmedzené a neodhalia veľa dôležitých informácií o malvéri. Prevládajúce techniky používané v statickej analýze sú:

- toková analýza - analýza funkcií
- reťazcová analýza - analýza konkrétnych reťazcov
- analýza podpisov - statická analýza na základe hash kódu

C. Dynamická analýza

Dynamická analýza je proces získavania údajov z malvéru počas jeho vykonávania. Vzťahuje sa na analýzu škodlivého programu počas jeho vykonávania v kontrolovanom prostredí (virtuálny počítač, emulátor, karanténa atď.) s cieľom identifikovať nepriateľské aktivity po vykonaní programu. Správanie sa monitoruje pomocou nástrojov, ako je proces monitor, prieskumník procesov, Wireshark alebo analyzátor škodlivej aktivity [22]. Tento druh analýzy sa snaží monitorovať systémové volania, príkázanú autorizáciu, volania funkcií a API, sieť, tok informácií atď [22]. Dynamická analýza má väčší prehľad o funkciách a správaní malvéru oproti statickej, pretože analýza hromadí informácie počas vykonávania malvéru. Na vykonanie dynamickej analýzy potrebujeme dve veci:

- prostredie, v ktorom môžeme spustiť malvér. Je to kontrolovaný spôsob pre účely analýzy.
- analytické nástroje, ktoré monitorujú a zaznamenávajú prostredie spôsobené škodlivým softvérom jeho cieľovému systému. Na rozdiel od statickej analýzy sa dynamickej analýze považuje za vysoko rizikový, ale vysoko ziskový proces pri analýze škodlivého kódu [23]. Nebezpečenstvo infekcie, šírenia alebo škodlivého správania na hosťovskom systéme, je vysoké, pretože sa malvér spúšťa a preto je potrebné izolovať hosťovské prostredie. Zisk je však vysoký, pretože údaje extrahované z malvéru počas vykonávania odhaľujú viac detailov a umožňujú sa prejaviť škodlivému kódu. V dynamickej analýze skúmame zmeny v aktivite registra, aktivite sieťovej prevádzky, procesu a aktivite súborov. Hlavné nástroje dynamickej analýzy sú monitor procesov, Wireshark, Anubis, Capture Bat atď.

D. Hybridná analýza

Technika hybridnej analýzy zahŕňa konsolidáciu statických a dynamických funkcií nahromadených pri skúmaní aplikácie a vykresľovaní zmien údajov, keď je aplikácia spustená. Táto metóda zvyšuje presnosť identifikácie. Hlavná nevýhoda hybridnej analýzy je že spotrebovávajú systémové zdroje a trvá dlho, kým sa analýza vykoná.

Hybridná analýza spája vlastnosti statickej a dynamickej analýzy pre rýchlu analýzu a lepšie výsledky je preto najpresnejšia.

E. Porovnanie

Nevýhodou statickej analýzy je že nie vždy dokáže odhaliť neznámy malvér a jeho varianty. V porovnaní so statickou analýzou je dynamickej analýza efektívnejšia a nevyžaduje rozsiahlu analýzu spustiteľného súboru, ale na druhej strane zaberie viac času a spotrebuje viac zdrojov ako statická analýza, keďže je náročnejšia na výpočtový výkon [24]. Ďalším problémom je, že kontrolované prostredie, v ktorom je malvér monitorovaný, môže byť odlišné od toho cieľového, program sa môže správať odlišne, pretože niektoré distribúcie malvéru môžu byť spustené len za určitých podmienok, napríklad prostredníctvom konkrétneho príkazu, v konkrétny systémový dátum alebo iba na konkrétnom hardvéri, napr. na konkrétnom type mobilného telefónu a v dôsledku toho, nie je možné zistiť niektoré aktivity vo virtuálnom prostredí. Pri statickej analýze je extrakcia údajov účinná iba vtedy, ak malvér neobsahuje žiadny typ šifrovania alebo zahmlievania.

Dynamická analýza má výhodu v tom, že malvér vykonáva v kontrolovanom prostredí a nemôže teda napáchať rozsiahle škody v prostredí ktoré nechceme infikovať. Rozbaľovanie a dešifrovanie súborov na analýzu je taktiež potrebné v boji proti malvéru no opäť je nutné vynaložiť na tento proces hardvérové prostriedky a vyžaduje aj čas. Manuálna analýza malvéru je namáhavá úloha, navyše si od analytikov vyžaduje odborné znalosti v oblasti softvérových interných prvkov a reverzného inžinierstva. Ťažba údajov a strojové učenie sa ukázali ako sľubné pri automatizácii určitých komponentov analýzy malvéru, ale tieto metódy sa stále vo veľkej miere spoliehajú na extrakciu hlavných funkcií z údajov, čo je netriviálna úloha, ktorá si neustále vyžaduje odborníkov so

špecializovanými súbormi zručností. Ako počet zariadení pripojených do kyberpriestoru neustále rastie, útoky sa navyše exponenciálne zvyšujú. Ako najvhodnejší typ analýzy malvéru považujeme hybridnú analýzu, je však náročná na čas a výpočtový výkon a zároveň je z časti manuálna a z časti automatizovaná, v závislosti od použitej technológie na analýzu malvéru.

VI. VYUŽITIE DETEKcie NA ZÁKLADE SIGNATÚRY A DETEKcie NA ZÁKLADE ANOMÁLIÍ AKO VSTUP PRE KLASIFIKÁCIU ALGORITMOV STROJOVÉHO UČENIA

Schémy detekcie mobilného malvéru sa dajú rozdeliť do dvoch veľkých tried na základe signatúr a anomálií. Prvá trieda zhromažďuje vzory a signatúry pochádzajúce z identifikovaných malvérov a porovnáva ich s neznámymi časťami kódu na určenie ich stavu. Druhá trieda využíva voľnejší prístup, ktorý pozoruje normálne správanie časti kódu počas určitého času a porovnáva ho s metrikami tohto normálneho modelu v porovnaní s akýmkoľvek odchýlkovým správaním. Analýza aplikácií sa na druhej strane vykonáva buď pomocou statickej, dynamickej alebo hybridnej metódy. Statická analýza znamená skúmanie aplikácie bez jej skutočného spustenia, na rozdiel od dynamického prístupu. V posledných niekoľkých rokoch bola vynaložená významná práca na vylepšenie týchto troch typov analýzy aplikácií s cieľom efektívne odhaliť škodlivý softvér v systéme Android [25]. Avšak hybridné prístupy na detekciu malvéru s dôrazom na strojové učenie doteraz dostali málo pozornosti. Funkcie vyplývajúce len zo statickej analýzy nemusia byť dostatočné pri skúmaní súbežných mobilných malvérových aplikácií, preto je potrebné zaviesť zložitejšie funkcie.

A. Navrhovaná metodika

V tejto časti sú uvedené podrobnosti o našej metodike, t.j. o implementovanom nástroji, procese extrakcie príznakov a použitých súboroch. Framework Django Django je rámec na vývoj webových aplikácií v jazyku Python. Tento rámec umožňuje rýchly vývoj webových aplikácií pri zachovaní čistoty a štruktúrovanosti kódu spolu s používaním moderných návrhových vzorov a metodík. Ďalšou dôležitou výhodou je poskytnutie webového servera rámcom Django, čo vám pri implementácii nástroja umožňuje sústrediť sa na vývoj samotnej aplikácie namiesto zbytočného trávenia času vytváraním vlastného webového servera. Okrem toho je Django populárny, otvorený zdrojový kód, ktorý je dobre podporovaný komunitou a stále sa vyvíja. Najmä z týchto dôvodov bol zvolený ako základ pre vývoj nástroja v závislosti od požiadavky, aby výstup aplikácie vo forme technickej správy bol vo formáte HTML a nástroj bol dostupný ako webová služba.

Aplikácie implementované pomocou frameworku Django sú štruktúrované do návrhového vzoru inšpirovaného vzorom Model View Controller (často skrátene MVC). Django však implementuje tento návrhový vzor v mierne upravenom stave Model Template View (MTV). Modelom je zdroj údajov webovej aplikácie, vo väčšine prípadov napríklad databázový systém. Šablóna sa stará o vykreslenie spracovaných údajov do konečnej podoby - teda webovej stránky HTML.

A napokon zobrazenie zabezpečuje prepojenie predchádzajúcich dvoch komponentov do jedného celku. Časť šablóny Model, ako ju implementuje a poskytuje framework Django, musela byť v implementácii vynechaná.

V predvolenom nastavení Django automaticky vytvára a udržiava databázu na základe definovaného triedy a ich atribúty. Všetky triedy, ktoré dedia triedu `django.db.models.Model`, sú teda automaticky používané frameworkom na vytváranie tabuliek a správu obsahu databázy všeobecne. Táto časť frameworku Django bola preto vynechaná a nahradená vlastnou implementáciou vzoru Model, ktorá nededí už spomínanú triedu `django.db.models.Model`, a teda nepodlieha súvisiacej automatickej manipulácii s databázou. Napriek tejto zásadnej zmene si novo implementovaný model stále zachováva pôvodnú logiku založenú na návrhovom vzore Model Template View, t. j. oddelenie logickej časti kódu, ktorá poskytuje zdroj údajov, ktoré sa majú v aplikácii zobraziť - v prípade tohto nástroja sú zdrojom údaje získané analýzou vzorky škodlivého softvéru. Inými slovami, je to tento vlastný špeciálny Model, ktorý poskytuje analýzu vzorky a potom získané údaje odovzdáva Zobrazeniu, ktoré sa potom vykresľuje pomocou Šablóny.

Knižnica pefile

Pri implementácii Malwareanalyst bola použitá aj knižnica `pefile2`, ktorej cieľom je analyzovať súbory vo formáte Portable Executable a poskytnúť výstup tejto analýzy, najmä údaje extrahovateľné z hlavičky Portable Executable, v štruktúrovanom štýle ako inštanciu triedy PE, ktorej atribútmi sú hodnoty extrahované z hlavičky PE. Okrem toho táto knižnica podporuje viac typov súborov PE než len spustiteľné súbory platformy Windows 32bit (t. j. najbežnejšie spustiteľné súbory na tejto platforme), ako napríklad súbory Windows 64bit alebo binárne súbory EFI a iné. Okrem toho táto knižnica podporuje analýzu binárnych súborov PE pre rôzne architektúry a je nezávislá od endianity analyzovaných súborov. Knižnica úspešne vyriešila problém nájdenia atribútu pri správnom odsadení v binárnom súbore (teda podľa technickej dokumentácie formátu PE nebolo potrebné sledovať potrebné atribúty a ich umiestnenie v hlavičke PE), ktorý navyše mohol byť dynamický - závislý od predchádzajúceho obsahu, napríklad dĺžky programu v hlavičke DOS. Okrem toho sa knižnica postarala o správne načítanie hodnoty daného atribútu v závislosti od endianity a všeobecne cieľovej architektúry súboru. Hodnoty získané knižnicou `pefile` používa nástroj vo viacerých analytických moduloch, najmä v tých, ktoré sa týkajú statickej analýzy.

B. Cuckoo Framework

Cuckoo je popredným open source nástrojom na dynamickú analýzu malvéru, veľmi dobre implementuje správu virtualizovanej infraštruktúry a analýzu softvérových vzoriek na tejto infraštruktúre. Možnosti nástroja Cuckoo Sandbox sú veľmi rozsiahle a nástroj dokáže efektívne spravovať infraštruktúru jedného virtuálneho stroja na otvorenej virtualizačnej platforme VirtualBox až po obrovskú decentralizovanú infraštruktúru rôznych virtuálnych strojov a sietí na komerčných virtualizačných platformách, ako je VMWare. Jeho výstup vo forme webovej stránky je stručný a

obsahuje len zlomok informácií vzhľadom na to, koľko informácií získa jeho automatická dynamická analýza.

Okrem toho vykonáva statickú analýzu na pomerne základnej úrovni. Keďže Cuckoo Sandbox je veľmi dobre implementovaný na správu virtualizačnej infraštruktúry a na dynamickú analýzu malvéru, potom vzhľadom na skutočnosť, že dynamická analýza malvéru (vo svojej podstate spúšťanie škodlivého kódu) je veľmi nebezpečná, rozhodli sme sa použiť Cuckoo Sandbox na dynamickú analýzu malvéru, najmä preto, že ide o dlhodobý fungujúci open source projekt, ktorý je tiež dobre známy a má už niekoľko rokov veľmi dobrú podporu komunity. Preto sa rozhodlo, že je lepšie použiť tento existujúci, zjavne dobrý komunitný nástroj.

Cuckoo Sandbox ponúka niekoľko rôznych komunikačných rozhraní. Malwareanalyst používa na komunikáciu so serverom Cuckoo API rozhranie REST API, ktoré beží popri hlavnom serveri Cuckoo, ktorý sa stará o správu virtuálnych strojov a analýzu zaslaných vzoriek škodlivého softvéru. Po odoslaní vzorky škodlivého softvéru Malwareanalyst pravidelne používa rozhranie REST API na dopytovanie servera Cuckoo API, aby zistil, či je analýza dokončená, a keď je, moduly Malwareanalyst potom z výsledkov analýzy vytvoria webovú technickú správu (alebo jej časť). Nevýhodou pridania dynamickej analýzy do procesu analýzy Malwareanalyst je však jednoznačne predĺženie trvania analýzy malvéru. Zatiaľ čo statická analýza malvéru vždy trvala len niekoľko sekúnd, pri dynamickej analýze malvéru je potrebné pripočítať čas potrebný na analýzu malvéru pomocou programu Cuckoo. To znamená najmä čas, ktorý je potrebný na to, aby sa vzorka malvéru spustila a pozorovalo sa jej správanie. Ak je tento čas príliš krátky, mohli by sme analýzu ukončiť skôr, ako malvér vykoná nejakú dôležitú škodlivú činnosť, a potom by sme mohli vzorku falošne identifikovať ako neškodnú. Naopak, príliš dlhý čas by nadmerne predĺžil čas analýzy a analýza pomocou Malwareanalyst by potom bola veľmi zdĺhavá.. Tento čas sa dá ľubovoľne meniť v konfiguračných súboroch Cuckoo a v čase testovania Malwareanalyst bol nastavený najprv na 30, potom na 20 sekúnd (skrátene na 20 sekúnd nevedelo k žiadnemu výraznému zhoršeniu analýzy, ale bolo by vhodné vykonať v tomto ohľade ďalšie dôkladné merania a pred ďalším skrátením na menej ako 20 sekúnd). K času samotnej analýzy je však potrebné pripočítať aj čas spojený so správou virtuálnych strojov.

Najskôr je potrebné identifikovať voľný virtuálny stroj, vytvoriť čistou inštanciu (existujúca používaná inštancia môže byť infikovaná predtým analyzovanou vzorkou a analýza na takomto stroji by mohla priniesť nesprávne výsledky), spustiť novovytvorenú čistou inštanciu a nahrat vzorku škodlivého softvéru na analýzu prostredníctvom siete. Pridanie dynamickej analýzy do procesu spolu mnohonásobne predĺži čas potrebný na analýzu vzorky škodlivého softvéru systémom Malwareanalyst na celkový čas niekoľkých desiatok sekúnd.

C. Datasetsy

V súvislosti s detekciou mobilného malvéru bolo vytvorených niekoľko súborov údajov (datasetov), ktoré pokrývajú rozmanité a mimoriadne prípady škodlivého softvéru. Tieto

korpusy vzoriek škodlivého softvéru sa považujú za referenčné súbory údajov na vyhodnotenie nových schém, ktoré sa pokúšajú rozlišovať medzi vzorkami škodlivého a dobrého softvéru. Na začiatku sa v tejto kapitole použili tri známe súbory údajov, ktoré sa vo veľkej miere používajú v súvisiacej literatúre. Súbor údajov Drebin z rokov 2010 až 2012, zbierka (podskupina) z VirusShare z rokov 2014 až 2017, ako aj zbierka z Androzo z rokov 2017 až 2020. Predovšetkým počet vzoriek škodlivého softvéru odobratých z každej z uvedených zbierok údajov bol 5 560, 10 000 a 1 520, t. j. spolu 17 080 vzoriek. Po analýze 17 080 vzoriek mobilného škodlivého softvéru z troch súborov údajov (datasetov) bolo celkovo zhromaždených viac ako 50 000 rôznych vektorov.

Každý vektor je binárnou reprezentáciou každej odlišnej funkcie. Prirodzene, ak odhliadneme od kapacity nástroja na dolovanie údajov (datamining tool), celkový počet 50 000 rôznych vektorov, čo znamená 50 000 rôznych stĺpcov (príznačov), presahuje schopnosť nástroja MalwareAnalyst a prakticky aj prístupú proof-of-concept ich získať v primeranom čase pomocou dotazov SQL. Pre potreby experimentov sa teda nakoniec použila menšia podmnožina 100 až 120 vzoriek škodlivého softvéru na dataset. Konkrétne sme náhodne vybrali 100 vzoriek malvéru z každej zo súborov údajov Drebin a VirusShare a 120 z - náročnejšieho - Andro-Zoo.

Konkrétne bežné rodiny malvéru môžu viesť k nadhodnoteným výsledkom. Preto sme odfiltrovali spoločné informácie medzi rodinami malvéru, ako napríklad rok vzorky a balík názov z každého súboru údajov, aby sa získali rovnomerne rozložené vzorky z rôznych rodín škodlivého softvéru. Je dôležité spomenúť, že všetky tri (redukované) súbory údajov sú vyvážené a obsahujú rovnaký počet náhodne vybraných škodlivých a neškodných aplikácií. Benígne aplikácie boli zozbierané z obchodu Google Play a sú spoločné pre všetky tri súbory údajov okrem súboru AndroZoo, ktorý obsahuje 20 ďalších.

D. Klasifikátory a metriky

Prevažná väčšina najmodernejších metód zameraných na detekciu škodlivých vzoriek bola hodnotená na základe vyvolania a presnosti správnych odpovedí v kombinácii F1 meraním. Pre každú inštanciu zvlášť sa tieto miery správnosti sústreďujú na poskytnutie binárnej odpovede, buď pozitívnej (trieda škodlivého softvéru), alebo negatívnej (trieda neškodlivého softvéru). To naznačuje, že na stanovenie tejto hranice sú potrebné informácie o rozdelení pozitívnych a negatívnych inštancií. Pre každý súbor údajov sa súbor získaných skóre na základe testovacích prípadov normalizuje v intervale [0,1] na klasifikačný model na skúmanú metódu. V tomto smere je odhad prahu nastavený na hodnotu 0,5. Okrem toho ako hlavnú hodnotiacu mieru používame AUC krivky operatívnej charakteristiky. Vo všeobecnosti AUC kvantifikuje účinnosť každého skúmaného prístupu pre všetky možné prahové hodnoty skóre. Hodnota AUC sa spravidla získava skúmaním poradia skóre, a nie ich presných hodnôt, ktoré vznikajú pri aplikácii metódy na súbor údajov. Je pozoruhodné, že odhad miery AUC je založený na všetkých

možných prahových hodnôtach. Okrem toho táto miera hodnotenia nezávisí od rozdelenia pozitívnych a negatívnych vzoriek.

V tomto článku sa používa značný počet známych a populárnych ML algoritmov učenia pod dohľadom (supervised). To znamená, že pre každý súbor údajov uvažujeme sedem najpoužívanejších binárnych klasifikačných modelov, a to logistickú regresiu (LR), Naive Bayes, Random Forest, AdaBoost, Support-vector machine (SVM), k-nearest neighbors (k-NN) a Stochastic Gradient Descent (SGD). Väčšina použitých klasifikačných algoritmov spadá pod učenie s dohľadom. V tejto kategórii sa algoritmy učenia pod dohľadom snažia vytvoriť všeobecný model škodlivých prípadov na základe trénujúcich údajov. Je zrejmé, že výkonnosť takýchto klasifikátorov silne závisí od veľkosti, kvality a reprezentatívnosti tréningových údajov. Na druhej strane, k-NN je slabý učiaci sa (známy ako lenivý učiaci sa), pretože nepoužíva tréningové údaje na vytvorenie všeobecného modelu, ale rozhoduje sa na základe informácií získaných pre každú vzorku zvlášť. Pre každý použitý klasifikátor sa používajú predvolené hodnoty nastavení parametrov. Všeobecný model každého klasifikátora sa zostavuje podľa techniky 10-násobného krížového overovania. Nakoniec sa pre každý súbor údajov vytvorí jednoduchý meta-model, t. j. súhrnná funkcia, ktorá kombinuje všetky základné klasifikátory použité buď hybridnými, alebo statickými metódami samostatne. Tento heterogénny súbor je založený na priemernom skóre, ktoré je výsledkom všetkých siedmich binárnych klasifikačných modelov pre každú vzorku. Pokiaľ ide o literatúru týkajúcu sa škodlivého softvéru pre Android, "ensemble learning" je k dispozícii nasledujúca literatúra [26].

VII. HODNOTENIE

Náš prístup hodnotíme pomocou detekcie založenej na signatúre aj na anomáliách. Pokiaľ ide o detekciu založenú na signatúre, MalwareAnalyst má schopnosť automaticky generovať detekčné signatúry kombináciou statických a dynamických vlastností. V podkapitole najprv uvádzame výsledky založené na signatúre a potom sa podrobne venujeme výsledkom klasifikácie po natrénovaní ML modelu s identickým súborom príznakov.

A. Detekcia na základe signatúry

Všeobecne platí, že detekcia zneužitia, známa aj ako detekcia založená na signatúrach, sa spolieha na známe signatúry, teda detekčné pravidlá zamerané na rozlišovanie medzi neškodnými a škodlivými časťami kódu. Tieto systémy sú schopné identifikovať škodlivý softvér, s ktorým sa už predtým stretli a môžu mať vysokú mieru prenosnosti medzi platformami. Avšak, chýba im schopnosť rozpoznať nové prípady škodlivého softvéru alebo variácie známeho softvéru. Detekčná schopnosť systému detekcie zneužitia, ktorý je skúmaný v tejto podkapitole, teda závisí hlavne od čo najnovších detekčných pravidiel, ktorými bol systém nakonfigurovaný. V rámci systému MalwareAnalyst tento typ detekcie zahŕňa porovnávanie vlastností nových ("neznámych") aplikácií, ktoré boli zozbierané počas statickej a dynamické analýzy, s vlastnosťami všetkých škodlivých

aplikácií, ktoré boli predtým uložené v databáze. Inými slovami, pre vyhodnotenie detekcie založenej na signatúrach na troch súboroch údajov bolo potrebné vopred analyzovať dostatočné množstvo vzoriek škodlivého softvéru na vytvorenie signatúr.

Zvolené bolo použitie techniky redukcie vzoriek, aby sa dosiahlo lepšie zobrazenie výsledkov. Týmto spôsobom bolo vybraných 1902 náhodných vzoriek škodlivého softvéru (zodpovedajúcich 499 vzoriek z databázy Drebin a 1403 vzoriek z databázy VirusShare), ktoré boli staticky aj dynamicky analyzované s cieľom získať ich vlastnosti. Tento počet sa týka iba vzoriek malvéru, ktoré sa nachádzajú v databáze. (a) nie sú zahrnuté v redukovaných súboroch údajov (b) boli filtrované na vyradenie duplicitných rodín malvéru pomocou názvu balíka a hash podpisu každej aplikácie. Odvozené signatúry obsahujú kombinácie znakov získaných zo statickej aj dynamické analýzy. Kombinácia zahŕňa aspoň jeden znak z minimálne dvoch rôznych kategórií znakov, konkrétne oprávnenia, zámery, volania API, sieťovú prevádzku, triedy java a spustené zámery. Keď sa analyzuje nová aplikácia, MalwareAnalyst kontroluje možnú zhodu signatúr. Ak analyzovaná aplikácia poskytne výsledky, ktoré sa zhodujú s dvoma alebo viacerými signatúrami, aplikácia sa označí ako škodlivá. Okrem toho sú v správe o analýze MalwareAnalystu ako podozrivé označené aj jednotlivé funkcie, ktoré sa pravidelne objavujú medzi malvérom, ale zriedkavo medzi goodvérom. Konkrétne, na ilustráciu, v ľavom stĺpci tabuľky III. je uvedených 15 najčastejších oprávnení zistených v použitom súbore 1902 vzoriek malvéru. Na druhej strane pravý stĺpec obsahuje ďalších 15 oprávnení, ktoré boli pozorované v tom istom súbore vzoriek, ale v súbore 120 goodwarových aplikácií vybraných z obchodu Google Play boli zriedkavé alebo sa vôbec nevyskytovali. Oba tieto stĺpce sú zoradené podľa percenta výskytu v súbore 1902 kusov škodlivého softvéru.

TABLE I. SKÓRE DETEKcie NA ZÁKLADE SIGNATÚRY (%)

Dataset	True Positive Rate (TPR)		Improvement Rate
	Static	Hybrid	Hybrid
AndroZoo	85.6	94.1	+8.5
Drebin	87	94	+7
VirusShare	95	99	+4

Dynamická inštrumentácia navyše odhaľuje určitý vzor v triedach jazyka Java používaných medzi škodlivým softvérom. Najmä niekoľko malvérových aplikácií, t. j. zodpovedajúco 9 %, 2 % a 1 % inštancií v našich súboroch údajov AndrooZoo, VirusShare a Drebin, inštanciuje triedu DexClassLoa. Táto trieda sa používa na načítanie iných tried zo súborov typu jar a apk obsahujúcich položku classes.dex. Tento class loader môže byť zneužitý na vykonanie kódu, ktorý nie je nainštalovaný ako súčasť samotnej aplikácie. Ďalším zaujímavým zistením je, že trieda java.net.Socket [27] sa v prípadoch malvéru vyskytovala pomerne často, a to 15 % v prípade AndrooZoo a 10 % v prípade VirusShare a Drebin.

TABLE II. 30 PODOZRIVÝCH POVOLENÍ.

Najčastejšie využívané služby malvérom	%	Najmenej využívané goodvérom	%
android.permission.INTERNET	96.2	com.android.browser.permission.READ_HISTORY_BOOKMARKS	12.2
android.permission.READ_PHONE_STATE	91.8	com.android.browser.permission.WRITE_HISTORY_BOOKMARKS	11.9
android.permission.WRITE_EXTERNAL_STORAGE	85.7	android.permission.WRITE	11.4
android.permission.ACCESS_NETWORKSTATE	59	android.permission.ACCESS	9.8
android.permission.ACCESS_WIFI_STATE	46.3	android.permission.INSTALL_PACKAGES	5.8
android.permission.ACCESS_COARSE_LOCATION	42.6	com.android.launcher.permission.READ_SETTINGS	5.5
android.permission.ACCESS_FINE_LOCATION	37.7	android.permission.MOUNT_UNMOUNT_FILESYSTEMS	5.4
android.permission.WAKE_LOCK	26.5	android.permission.RESTART_PACKAGES	4.5
android.permission.VIBRATE_25.3	25.3	android.permission.WRITE_APN_SETTINGS	4.1
android.permission.RECEIVE_BOOT_COMPLETED	24.2	android.permission.CHANGE_CONFIGURATION	2.2
com.android.launcher.permission.INSTALL	23.3	android.permission.WRITE_SECURE_SETTING	2.2
android.permission.CHANGE_WIFI_STATE	19.9	android.permission.ACCESS_COARSE_UPDATES	1.7
android.permission.CALL_PHONE	16.9	android.permission.DELETE_PACKAGES	1.5
android.permission.GET_TASKS	14.9	android.permission.READ_SETTINGS	0.8
android.permission.SEND_SMS	14.8	android.permission.RECEIVE_WAP_PUSH	0.5

B. Detekcia na základe anomálií

Na detekciu anomálií na základe statickej aj dynamickej analýzy súborov údajov opísaných v podkapitole sa použili vektory znakov. Na tento účel sme využili open source nástroj na dolovanie údajov, s názvom Orange. Ako je podrobne uvedené v podkapitole, použili sme sedem ML algoritmov, ktoré sa vo veľkej miere používajú v súvisiacej literatúre. Zvolenou technikou výberu bola 10-násobná krížová validácia. V tabuľke III. je uvedený porovnávací prehľad výsledkov klasifikácie získaných zo statickej analýzy s výsledkami získanými z hybridnej analýzy. Zdá sa, že navrhovaný prístup založený na hybridnej analýze neustále prekonáva statickú metódu, keď sa berie do úvahy výkonnosť AUC, ako aj účinnosť binárnych mier (CA, P, R a F1). Ako možno pozorovať, takmer vo všetkých prípadoch všetkých troch datasetov je statická analýza efektívnejšia iba v Derbinovom korpuse, kde prekonáva statický model ten hybridný, keď sa použijú modely Naive Bayes a LR. Z toho vyplýva, že navrhovaná hybridná analýza je jednoznačne lepšia ako statická.

TABLE III. VÝSLEDKY PODĽA DATASETU VIRUSSHARE A METRIKY VÝKONNOSTI KLASIFIKÁCIE

Dataset	ML Classifier	AUC		CA		F1		Precision		Recall	
		Static	Hybrid	Static	Hybrid	Static	Hybrid	Static	Hybrid	Static	Hybrid
VirusShare	Logistic Regression	0.973	0.98	0.955	0.98	0.955	0.98	0.955	0.98	0.955	0.98
	Naïve Bayes	0.955	0.977	0.89	0.945	0.89	0.945	0.89	0.947	0.89	0.945
	Random Forest	0.974	0.98	0.955	0.975	0.955	0.975	0.955	0.975	0.955	0.975
	AndroZoo	0.97	0.975	0.96	0.97	0.96	0.97	0.961	0.97	0.96	0.97
	AdaBoost	0.945	0.98	0.945	0.98	0.945	0.98	0.945	0.98	0.945	0.98
	SGD	0.935	0.97	0.935	0.97	0.935	0.97	0.937	0.97	0.935	0.97
	SVM	0.963	0.98	0.93	0.975	0.93	0.975	0.935	0.975	0.93	0.975

Pre zobrazenie uvádzame iba výsledky pre dataset VirusShare ktoré mali najlepšie skóre. Logistická regresia a hybridné modely AdaBoost sa zdajú byť celkovo najvýkonnejšími modelmi pre dataset AndroZoo. Okrem toho väčší počet skúmaných klasifikátorov, konkrétne LR, Random Forest, AdaBoost a SVM, pomáha hybridným dosiahnuť výnimočné výsledky v porovnaní so statickou metódou pre súbor údajov VirusShare. Zdá sa však, že prevažná väčšina skúmaných hybridných modelov je výnimočne úspešná (jedinou výnimkou je model Naive Bayes) v datasete Drebin. Vo všeobecnosti hybridný prístup dosahuje stabilnejší výkon vo všetkých troch súboroch údajov v porovnaní s metódou založenou na statickej analýze. Je tiež pozoruhodné, že najväčšie zlepšenie hybridných modelov sa dosahuje v korpuse AndroZoo. Je pozoruhodné, že zlepšenie výkonnosti v

zmysle AUC najlepších hybridných modelov vzhľadom na výkonnosť najlepších statických modelov je pri korpuse AndroZoo vyššie ako 10%. Všetky tieto skutočnosti naznačujú, že hybridný prístup je oveľa spoľahlivejší a účinnejší v náročnejších prípadoch škodlivého softvéru, kde sú nové a sofistikovanejšie podmienky škodlivého softvéru. Verzia našej metódy založená na statickej analýze je tiež veľmi efektívna a dosahuje najlepšie výsledky v korpuse VirusShare a Drebin. Pokiaľ ide o súbor údajov AndroZoo, statická analýza dosahuje slabé výsledky. Na obrázkoch je znázornená účinnosť kriviek ROC pri použití LR klasifikátora na jednotlivé súbory údajov pre prezentovanú verziu hybridnej metódy, ako aj verziu na základe statickej analýzy.

VIII. ZÁVER

Detekcia mobilného malvéru je veľmi aktívnou oblasťou výskumu, pretože popularita mobilných zariadení neustále rastie, čo je podporené aj šírením mobilných sietí novej generácie, t. j. 5G a ďalších. So zameraním na populárnu platformu Android sme v priebehu nášho výskumu skúmali súčasný stav prístupov k detekcii mobilného malvéru, pričom sme skúmali metodiky detekcie statickej ale predovšetkým dynamickej a hybridnej analýzy na zabezpečenie optimálnych bezpečnostných protipatrení. V publikácii sme sa venovali bezpečnosti operačného systému Android a možnostiam jeho obrany proti škodlivému kódu. Vysvetľujeme rozdiely medzi kategóriami mobilného malvéru. Predstavujeme spôsoby analýzy škodlivého kódu a predpokladáme, ktorá metóda bude najefektívnejšia. Skúmali sme prístupy k detekcii škodlivého kódu na základe statickej, dynamickej či hybridnej analýzy. Vysvetľujeme základné pojmy strojového učenia a kategórie algoritmov strojového učenia. Detekcia mobilného malvéru je oblasť výskumu, ktorá zahŕňa rôzne techniky. Predostreli sme možnosti využitia frameworkov ako klasifikátorov pri možnostiach detekcie malvéru pri dynamickej analýze. Zamerali sme sa na využitie detekcie na základe signatúry a detekcie na základe anomálií. V tejto kapitole bol predstavený MalwareAnalyst, automatizovaný a konfigurovateľný hybridný analytický nástroj, ktorý kombinuje statickú analýzu s dynamickou instrumentáciou na analýzu správania aplikácií na platforme Android. Prostredníctvom rozsiahlej analýzy sme odhalili škálu otvorených výskumných výziev a uvádzame budúce smery a osvedčené postupy na vytvorenie spoľahlivých metodík klasifikácie mobilného malvéru. V tejto publikácii boli predstavené metódy na významné zlepšenie súčasného stavu v oblasti detekcie mobilného malvéru.

REFERENCES

- [1] Stallings, William, et al. Computer security: principles and practice. Upper Saddle River: Pearson, 2018.
- [2] Stallings, William. Cryptography and Network Security: Principles and Practice. 7th ed. Upper Saddle River N.J: Prentice Hall, 2017
- [3] Sikorski, Michael; HONIG, Andrew. Practical malware analysis: the hands-on guide to dissecting malicious software. no starch press, 2012.
- [4] Müller, Andreas C.; GUIDO, Sarah. Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.", 2016. [
- [5] Conrad, Eric, Seth Misener, and Joshua Feldman. CISSP® Study Guide. Syngress, 2023.

- [6] Deisenroth, Marc Peter; FAISAL, A. Aldo; ONG, Cheng Soon. *Mathematics for machine learning*. Cambridge University Press, 2020.
- [7] Shalev-Shwartz, Shai; BEN-DAVID, Shai. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- [8] J. Z. Kolter and M. A. Maloof. Learning to detect and classify malicious executables in the wild. *Journal of Machine Learning Research*, page 2721-2744, 2006.
- [9] D. Damopoulos, S. A. Menesidou, G. Kambourakis, M. Papadaki, N. Clarke, and S. Gritzalis. Evaluation of anomaly-based ids for mobile devices using machine learning classifiers. *Security and Communication Networks*, 5:3-14, 2012.
- [10] D. J. Wu, C. H. Mao, T. E. Wei, H. M. Lee, and K. P. Wu. Droidmat: Android malware detection through manifest and api calls tracing. *AsiaJCIS*, pages 62- 69,2012.
- [11] Li, Wenjie, Jinchen Ji, and Lihong Huang. Dynamics of a controlled discontinuous computer worm system. *Proceedings of the American Mathematical Society* 148.10 (2020): 4389-4403.
- [12] Joy, Jestin, Anita John, and James Joy. Rootkit detection mechanism: A survey. *Advances in Parallel Distributed Computing: First International Conference on Parallel, Distributed Computing Technologies and Applications, PDCTA 2011, Tirunelveli, India, September 23-25, 2011. Proceedings*. Springer Berlin Heidelberg, 2011.
- [13] Branco, Rodrigo Rubira, Gabriel Negreira Barbosa, and Pedro Drimel Neto. "Scientific but not academical overview of malware anti-debugging, antidisassembly and anti-vm technologies." *Black Hat 1.2012* (2012): 1-27.
- [14] Roundy, Kevin A., and Barton P. Miller. Hybrid analysis and control of malware. *Recent Advances in Intrusion Detection: 13th International Symposium, RAID 2010, Ottawa, Ontario, Canada, September 15-17, 2010. Proceedings 13*. Springer Berlin Heidelberg, 2010.
- [15] Chen, Tianqi, and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.
- [16] Torgo, Luis, and Rita Ribeiro. Precision and recall for regression. *Discovery Science: 12th International Conference, DS 2009, Porto, Portugal, October 3-5, 2009 12*. Springer Berlin Heidelberg, 2009.
- [17] Altuwaijri, Haya, and Sanaa Ghouzali. "Android data storage security: A review." *Journal of King Saud University-Computer and Information Sciences* 32.5 (2020): 543-552.
- [18] W. Wang, X. Wang, D. Feng, J. Liu, Z. Han, and X. Zhang. Exploring permission-induced risk in android applications for malicious application detection. *IEEE Transactions on Information Forensics and Security*, 9(11):1869-1882, 2014.
- [19] F. Idrees, M. Rajarajan, M. Conti, R. Rahulamathavan, and T. Chen. Pindroid: A novel android malware detection system using ensemble learning. *Computers and Security*, 68:36-46, 2017.
- [20] H. Kang, J.-w. Jang, A. Mohaisen and H. K. Kim, Detecting and Classifying Android Malware Using Static Analysis along with Creator Information, *International Journal of Distributed Sensor Networks*, vol. 11, p. 479174, 2015.
- [21] K. Tam, A. Feizollah, N. B. Anuar, R. Salleh and L. Cavallaro, The Evolution of Android Malware and Android Analysis Techniques, *ACM Computing Surveys*, vol. 49, pp. 1-41, 2017.
- [22] H.-J. Zhu, Z.-H. You, Z.-X. Zhu, W.-L. Shi, X. Chen and L. Cheng, DroidDet: Effective and robust detection of android malware using static analysis along with rotation forest model, *Neurocomputing*, vol. 272, pp. 638-646, 2018.
- [23] J. Jiang, S. Li, M. Yu, G. Li, C. Liu, K. Chen, H. Liu and W. Huang, Android Malware Family Classification Based on Sensitive Opcode Sequence, *IEEE Symposium on Computers and Communications (ISCC)*, pp. 1-7, 2019.
- [24] C. Uppin and G. George, Analysis of Android Malware Using Data Replication Features Extracted by Machine Learning Tools, *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 5, no. 5, 2019.
- [25] A. Bacci, A. Bartoli, F. Martinelli, E. Medvet, F. Mercaldo and C. A. Visaggio, Impact of Code Obfuscation on Android Malware Detection based on Static and Dynamic Analysis, in *Proceedings of the 4th International Conference on Information Systems Security and Privacy, Funchal, Madeira, Portugal, 2018*.
- [26] F. Martinelli, F. Mercaldo, A. Saracino and C. A. Visaggio, I find your behavior disturbing: Static and dynamic app behavioral analysis for detection of android malware., in *14th International Conference on Privacy, Security and Trust (PST 2016)*, IEEE, 2016.
- [27] R. Sihwail, K. Omar, K. Zainol Ariffin and S. Al Afghani, Malware Detection Approach Based on Artifacts in Memory Image and Dynamic Analysis, *Applied Sciences*, vol. 9, no. 18, p. 3680, Sep 2019.

Simulation model of small hydropower plant

Richard Olexa

Katedra elektrotechniky a informatiky
Fakulta elektrotechniky a informatiky, TUKE
Košice, Slovensko
Email: richard.olexa@tuke.sk

Pavol Fedor

Katedra elektrotechniky a informatiky
Fakulta elektrotechniky a informatiky, TUKE
Košice, Slovensko
Email: pavol.fedor@tuke.sk

Abstract—This scientific paper is dedicated to the field of modelling of small hydropower plant systems with an installed asynchronous generator, in order to create a tool for simple, accurate and material-saving design of the small hydropower plants, their control structures, or their possible modifications, respectively. In the overview section, we present the current knowledge and we describe the whole design process of the aforementioned simulation model, which we followed. Next chapter presents the modifications of the model we have developed, together with an explanation of their intended functionality. The final part represents a description of the developed laboratory model of the hydromechanical part of a small hydropower plant, which was physically constructed in the premises of one of our laboratories, too.

Keywords: *simulation model of small hydropower plant; hydraulic turbine model; hydraulic turbine efficiency; asynchronous generator*

I. INTRODUCTION

Today's events are increasingly highlighting the need to consider climate change as one of the sensitive and crucial topics that is no longer just a basis for a purely technical debate, but is also becoming a topic of everyday conversation among ordinary people. The increasingly frequent occurrence of weather disturbances, i.e. extremely severe storms, prolonged droughts or even the recent coronavirus pandemic, are also contributing to this field. All of this puts pressure on governments, but also on supranational groups such as the European Union, which are allocating considerable resources to improving the current level of environmental stress caused by human activity. Of course, one of the main themes is the promotion of electromobility as a tool for gradually replacing vehicles that use the combustion of oil products as a source of energy. On the other hand, it is also necessary to consider the way in which electricity is produced, so that emissions are not simply transferred from the roads and their surroundings to the areas of power stations, which still use fossil fuels as a means of producing their product. For this reason, there are interests in promoting the production of electricity from renewable sources, one of which is water.

Hydropower plants can thus, with appropriate site selection, management and system design, provide electricity in an environmentally clean way, even in areas that are not yet electrified. An example of such areas could be remote locations in the mountains, where the presence of running water is very frequent, the kinetic energy of which, and its subsequent

conversion into electricity, can, if the system is properly selected, ensure the availability of this commodity in the area in question.

For this reason, small hydropower plants in particular present a challenge to scientists and practitioners in the field, particularly in terms of proper control design, modelling and sizing. By creating a functional simulation model, it is possible to make the control design as well as the actual dimensioning of the individual parts of the hydropower systems more efficient.

The energy produced by hydropower plants is defined as a renewable energy because the energy source, the potential or kinetic energy of water, is continuously renewed by its cycling in nature. Water has been used as a source of energy since time immemorial, just like wood. An example is the construction and subsequent use of the so-called waterwheel. This involved converting the kinetic energy of the water in the rivers on which these systems were built into the mechanical energy needed primarily for grinding grain in the mills. As the field of using the potential of this energy source has gradually developed, its deployment has expanded until today, when hydropower systems are even capable of supplying up to 90% of the country's electricity needs. One such example is the ITAIPU hydroelectric power plant in Paraguay, located on the Paraná River, with an installed capacity of 14GW. [1]

II. OVERVIEW OF THE CURRENT KNOWLEDGE IN THE FIELD OF MODELLING A SMALL HYDROPOWER PLANTS

There are number of currently available methods for modelling the behavior of small hydropower plants, with the biggest difference usually being the extent of their complexity. At the same time, according to the scientific literature and papers [2]-[3]-[4]-[5]-[6]-[7]-[8], it is quite clear that the most frequently differing sub-model of the overall hydropower plant simulation model is the submodel representing the hydraulic turbine and the processes associated with it. The difference lies in the authors' decision to proceed or not to with a linear, on otherwise strongly non-linear system, which has a not insignificant impact on the accuracy of the simulation model as well as on its scale.

Usually presented models of hydropower plant systems, made for simulation purposes, describe the procedure of design and creation of the model structures in question for hydropower

This research and paper was supported under the grant APVV-16-0206 and APVVV-19-0210.

plants of large capacity, to which a number of typical properties and processes are associated, which, due to the constructional simpler units of small hydropower plants, can be neglected in their modelling. Since our research was dedicated to small capacity hydropower plants, we propose a simplified alternative, which nevertheless satisfies the desired purposes by its properties. Assuming the requirement to design and build simulation models of large hydropower plants, it is possible to follow a rather detailed description of the individual steps, a summary of which is given in the publication *Modelling and Controlling Hydropower Plants* [9] by its authors Munoz-Hernandez, Jones and Mansoor.

The elaboration of the specifics related to the different parts of the simulated process parts of hydropower plant systems is similarly addressed by the authors Kishor and Fraile-Ardanuy [10], where they provide a summary of the chapters devoted to this issue in their publication.

As already mentioned, a rather detailed description of a creation of a simulation model of hydropower plants is described in the book publication [9], whereas the authors also present possible modifications of parts of the model, such as in case of several generators or hydraulic turbines installed in one hydropower plant, etc., together with an explanation of the changes related to these modifications, despite the blocks directly related to the subject of change. The authors of references [1]-[2]-[3]-[10] follow a similar approach, with more or less differences.

In the following chapters, we pay attention to our proposed modifications of the usually used simulation structures, designed in order to adapt the model to the specifics of small hydropower plants.

III. PROPOSED SIMULATION MODEL OF SMALL HYDROPOWER PLANT

The basic topology of a small hydropower plant simulation model composition could be divided into several parts or smaller subsystems, respectively:

- Subsystem of Hydraulic Turbine Control (Governor)
- Subsystem of Servodrive controlling the water supply to the hydraulic turbine (Servo)
- Subsystem of Hydraulic Turbine
- Subsystem of other Eletromechanical parts of the system (Electrical Subsystem)

The considered input variables to such a simulation model depicted in the Fig. 1 are: the grid frequency, defining the reference angular speed of the generator, or turbine in the case of a common shaft system; the reference or desired value of the generated power of the hydropower plant; or the simulated waveforms of the grid-connected electrical loads considered in the electrical subsystem.

The structure of the simulation schemes described in the following subchapters naturally follows the aforementioned distribution and the variables entering it. However, it should be noted that, as in case of all the studied so far published simulation models [9]-[2]-[3]-[4]-[5]-[6]-[7]-[8], our simulation model of a small hydropower plant, together with the parameters entering it, are in a normalized state (p.u.). This means that for example, the monitored values of the delivered electrical power, the angular velocity, or even the input values of the grid frequency, etc., are in the range of 0 to 1.

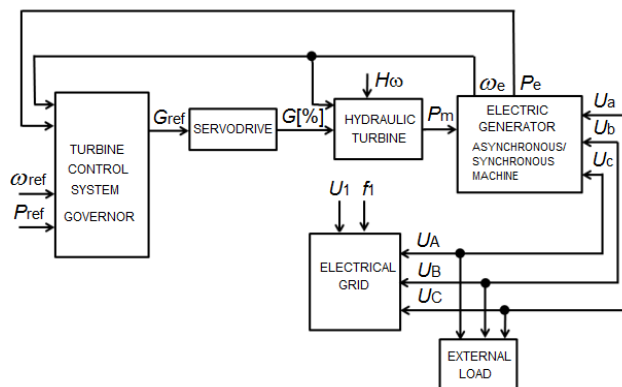


Fig. 1. Basic block diagram of hydropower plant simulation models [15]

A. Subsystem of Hydraulic Turbine Control (Governor)

For the purpose of developing our simulation model of a small hydropower plant, we have relied on the knowledge obtained from the description of different methods and procedures for the design of control structures for hydropower plants, published in the scientific literature of [9]-[2]-[3]-[4].

Our idea of the form of interaction of the proposed control with other parts of the hydropower plant (HPP) model (the case of an HPP with an installed synchronous machine) is as follows:

I. In a situation where the hydraulic turbine as well as the other parts of the hydropower plant are stationary, i.e. their angular velocity is zero, it is necessary, at the request of the operator, to ensure by suitable control that the hydraulic turbine starts up as quickly as possible to the so-called synchronous speed, defined by the frequency of the network and the design of the generator, in order to avoid undesirable effects on the mechanical system of the hydropower plant or on the network in the event of failure to comply with this requirement.

II. Once the desired reference mechanical speed of the electrical generator has been achieved and the other criteria have been met, the connection of the HPP to the grid can proceed. In this state, the small hydropower plant (SHPP) is already connected to the grid but does not supply any power to the grid.

III. After successful connection, the hydropower plant can automatically, or upon instruction from the operator, enter the power control mode. Again, in this case there are several variations of the definition of the level of power required to be delivered. As an example, operation with the operator-defined power requirement; or operation at the full power that is achievable under the current conditions of the available water head; or according to other influences determining the reference value of the required power generated by the HPP.

According to the abovementioned information, and taking into account the results published so far in this field [9]-[2]-[3]-[4]-[5]-[6]-[7]-[11]-[12], we have considered the use of a PID-type controller, and thus of the proportional-integral-derivative type, with a derivative component equal to zero, in the construction of this subsystem considered for the time being. Thus, we have neglected the derivative component for the time being on the grounds that its influence in the form of a rapid "boosting" of the output would be minimal, since the output of the downstream modeled part, the Servodrive, has a limitation (saturation block) placed in the line of its output value. For the sake of creating a model with universal applicability and potential preconditioning for the inclusion of currently neglected influences, we retain the derivative component in the control scheme, anyway with a zero derivative gain. We have chosen to control the delivered power of the HPP in the form of a proportional-integral controller, PI.

The described control structural diagram together with the added switch block and first-order filter are shown in Fig. 2. We avoid the influence of unwanted rapid changes of the desired (reference) value of the frequency or angular velocity of the hydraulic turbine by including the first-order filter in its line and thus optimizing the overall operation of the proposed control of the HPP. The filter in question is characterized by a first-order inertia and a defined time constant T_f . Its setup is described in more detail in the following parts of this subchapter.

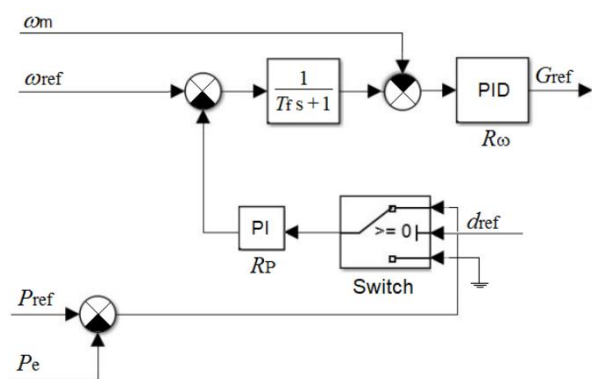


Fig. 2. Proposed control structure in the form of a block diagram of the Subsystem for Hydraulic System Control (Governor)

The block for switching on the control of the supplied power, shown in the diagram, is included in the model for the purpose of selecting the type of the controlled variable, while its activation by the input signal d_{ref} is the switching on of the boost to the desired or reference value of the angular velocity of the small hydropower plant. For the reasons mentioned in the introductory sections of this paper, we next consider the installation and modeling of an induction generator with a short penstock. Considering the characteristics resulting from its design and operating principle of an asynchronous generator, we put in a limitation of its "boost" to the output of the generated power controller, thus reflecting the slip limits of that particular asynchronous machine installed in the small hydro power plant.

As a consequence of the above, the start of the regulation of the generated power is only relevant after a successful connection of the small hydropower system to the power grid.

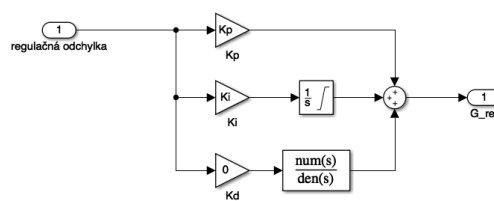


Fig. 3. The proposed Simulink control structure of the angular velocity of a small hydropower system, where: K_p - the gain of the proportional component; K_i - the gain of the integration component; and K_d - the gain of the derivative component of the controller in question + "regulačná odchylka" means control deviation signal and " G_{ref} " represents a demanded output of the Servodrive opening position

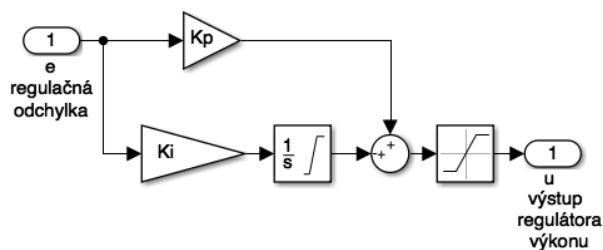


Fig. 4. Proposed Simulink control structure for power control of small hydropower plants with asynchronous generator, where: e - control deviation and u - output of the controller in question

In the Fig. 3 and 4, our proposed control structures for each of the controlled variables are being depicted. All the parameters have to be determined according to the specifics of the simulated system thus, we put all of them relevant for the model parametrization in a summary table mentioned later in this paper.

B. Subsystem of Servodrive controlling the water supply to the hydraulic turbine (Servo)

Our proposed simulation model of a small hydropower plant considers the use of a servodrive in the position of the actuator providing the change of position, or opening and closing of the water supply to the hydraulic turbine, with respect to the output signal from the above described subsystem of the hydraulic turbine’s Governor. Thus, the input of the Servo submodel is the desired value of its position, which is adjusted by the Governor in order to compensate for the mismatch between the desired and actual speed of the hydraulic turbine. (see the Fig. 1, 2, 3 and 4)

Thus, we proceeded to build this simulation subsystem by modelling a second-order dynamic system, whereas the state variable is the opening speed of the water inlet to the turbine, i.e. the opening speed of the servo in question, as well as its position. Due to the possible occurrence of the undesired phenomena of water hammer and cavitation due to rapid changes of the servo position in hydropower plant designs with higher water column levels, we proceed to include a limitation of the rate of change of the servo position, or the opening of the water inlet, which is usually included in the control also in the case of smaller hydropower plants where these phenomena are not common, especially from the safety point of view. Also, in the model depicted in the Fig. 5, a block of saturation placed in the line of resulting position of the opening of the water inlet to the hydraulic turbine is also present, in order to define its realistic working area. Both included constraints are strongly nonlinear in nature.

The parameters characterizing the subsystem of the servodrive are: the gain K_A and the time constant T_A defined by the characteristic of the servodrive installed. The speed variation limit is usually set to 0.01 when the speed increases and to -0.01 when the speed decreases.

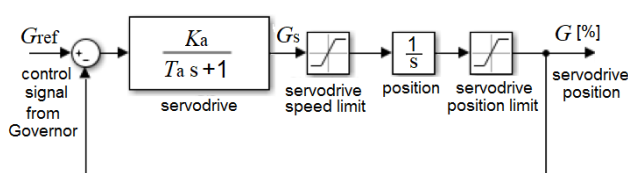


Fig. 5. Proposed subsystem for the servo drive controlling the water supply to the hydraulic turbine where: G_{ref} - desired value of the opening position, as output from the angular velocity control; G_s - opening speed of the servo drive; G - output actual opening position of the water supply to the hydraulic turbine

The determination of the limit values for the position of the servodrive is always dependent on the system being modelled. However, we have defined the bounds to be a closed interval $\langle 0.01; 0.975 \rangle$ for the purposes of the baseline setting.

C. Subsystem of Hydraulic Turbine

Similarly as the hydraulic turbine is referred to in most of the literature dealing with hydropower plants as a kind of heart of the whole system, its subsystem, or its simulation model, is also a crucial element of the overall model of the hydropower plant. It is therefore clear that an accurate and appropriate way of modelling this subsystem is a key element in order to achieve the desired goal, i.e. to create a complete simulation model of a small hydropower plant, suitable for accurate diagnosis and characterization of the responses to predefined inputs, in order to pre-design the tuning of the proposed systems.

From an research of available sources [9]-[2]-[3]-[4]-[5]-[7]-[13]-[14], there are several alternatives for the representation of the behavior of the hydraulic turbine section within the whole hydropower plant model, both in terms of the procedure for building the corresponding model and their mathematical description. The most common element of difference between the different alternatives is the way in which the efficiency of the hydraulic turbine is taken into account, or in other words, its influence on the mechanical power P_m achieved at the outlet.

The foundation for our design and development of the simulation model of an SHPP in is based on the flow of water through the turbine at its zero load, whereas the publications [2] and [5] describe the design process, as well as the variables used quite in adetail. All this results in the incorporation of the dynamic characteristics of hydraulic turbines in their dedicated simulation models, assuming the existence of: an inlet pipe (Penstock); and a very large or no surge chamber (Surge Chamber). The proposed simulation structure of this subsystem is thus presented in Fig. 6.

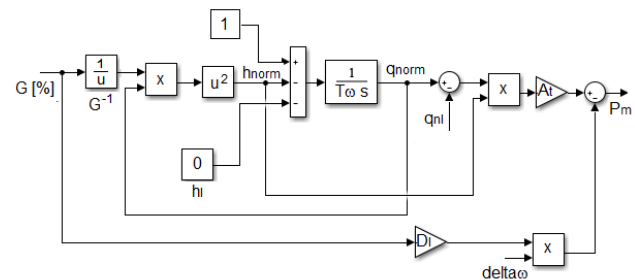


Fig. 6. Proposed subsystem that takes into account the characteristics of the hydraulic turbine (Hydraulic Turbine) based on the flow through the turbine at its zero load

However, during the design phase of the creation of the simulation model of an SHPP, we investigate other ways of incorporating the hydraulic turbine’s efficiency into the proposed model. This initiative led us into a research of the possible application of fuzzy logic model, as a submodel of this subsystem in question, that would continuously provide a model with varying value of the turbines efficiency, based on

hydraulic turbine's actual wear factor, as well as according to the its current angular speed, and the water flow, respectively.

Based on the aforementioned, when we were designing and selecting the method of considering the efficiency of the hydraulic turbine in our proposed simulation model of a small hydropower plant, we concluded that both of the mostly used approaches [2]-[3] are based on rather complicated calculations of the turbine efficiency depending on the detailed design parameters of the hydropower plant. All these data are often not available and only need to be estimated, while changing over time due to system operation. Moreover, combined with its the strongly nonlinear dependence, it is usually difficult to implement these effects in a simulation models of these systems. As an example, we present below in Fig. 7 and Fig. 8 the calculated efficiency maps of the same hydraulic turbine as a function of the water flow through the hydraulic turbine and its actual angular velocity, obtained through the method of calculating the efficiency of a hydropower turbine, described in [3].

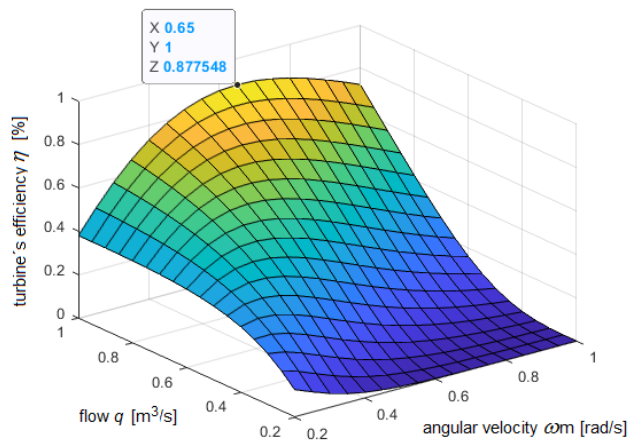


Fig. 7. Hydraulic turbine efficiency map assuming: $R_{blades}=1.9m$; $A_{blades}=9m^2$ - The maximum achievable turbine efficiency in this condition is **87.75%**

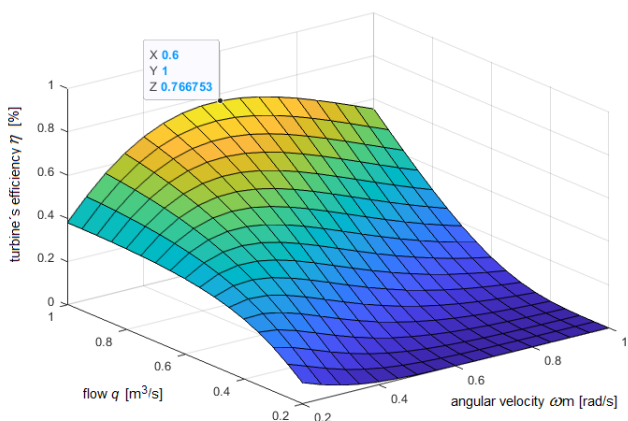


Fig. 8. Hydraulic turbine efficiency map assuming: $R_{blades}=1.9m$; $A_{blades}=8m^2$ - The maximum achievable turbine efficiency in this condition is **76.67%**

By comparing these hydraulic turbine efficiency maps, we demonstrate the significant difference in the maximum achievable hydraulic turbine efficiency due to the simple wear of the turbine blades and hence the change in the A_{blades} parameter. The relationship between efficiency and the change in the A_{blades} parameter is strongly nonlinear, which means that it is often quite challenging to implement in a model. [15]

However, by being able to determine the efficiency level of a hydraulic turbine based on measurements of the turbine water flow q and the mechanical angular velocity ω_m , assuming steady state, for parameter values $R_{blades}=1.9m$ and $A_{blades}=9m^2$ (see Table 1), it is possible to propose the use of a so-called fuzzy inference structure (FIS) for the efficiency of a hydropower turbine, describing the measured dependence between $[\omega_m, q] \rightarrow \eta$.

TABLE I. MEASURED (CALCULATED) EFFICIENCY VALUES AT DIFFERENT STEADY STATE OPERATING POINTS OF THE HYDRAULIC TURBINE

Efficiency [%]		ω_m				
		0.2	0.4	0.6	0.8	1
q	0.2	12.77	30.36	36.23	37.83	37.97
	0.4	2.74	26.40	49.60	64.41	73.24
	0.6	0.35	13.64	40.90	66.92	87.22
	0.8	0.04	5.68	27.20	56.26	84.31
	1	0.00	2.11	16.07	42.03	72.49

Based on such database (Table 1), a fuzzy model can be designed using standard analysis tools with adaptive approaches, resulting in both improved modelling quality and reduced development time. In terms of the adaptive approach used, we have implemented a hybrid setup based on the use of the "fuzzy inference tool" in conjunction with a neural network. In [16]-[17], from the large number of methods for developing adaptive fuzzy systems [17]-[18], we chose the adaptive neuro-fuzzy inference system ANFIS in the Matlab environment. The [19]-[20] basic feature of this analysis is the reduction of the number of fuzzy rules and the appropriate setting of their initial parameters. Thus, in the implementation, we set the parameters as follows: *Range of Influence* = 0.5; *Squash factor* = 1.1; and *Reject ration* = 0.005.

The result of the aforementioned procedure is a static Sugeno-type fuzzy system with twelve rules, shown in Fig. 9. [15]

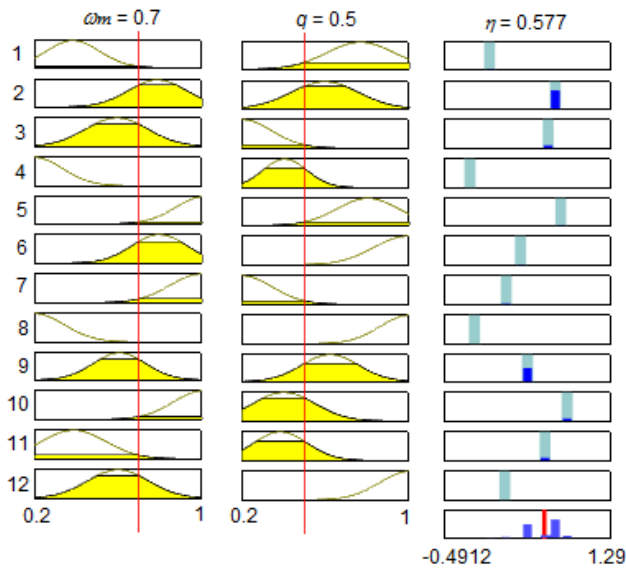


Fig. 9. Proposed fuzzy model of hydraulic turbine efficiency - Sugeno type with twelve rules

This means that our proposed fuzzy model of hydraulic turbine efficiency replaces the numerical calculations according to [3] and so avoiding the use of usually unavailable, and thus only estimated, technical parameters entering the efficiency calculation; and, at the same time, we also take into account, on the contrary, the changes in turbine wear in the simulation process. [15]

Comparing Fig. 7 and Fig. 10, and also comparing the values numerically calculated according to the steps described in [3] with the output values of the proposed fuzzy system for the efficiency of a hydraulic turbine of a small hydropower plant, we conclude that the fuzzy model achieves almost identical results to the analytically obtained values, which is graphically depicted in Fig. 11.

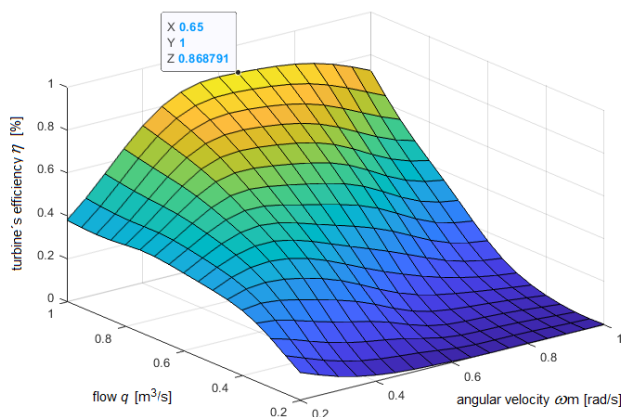


Fig. 10. Efficiency map of a hydraulic turbine, obtained on the basis of a fuzzy model of its efficiency

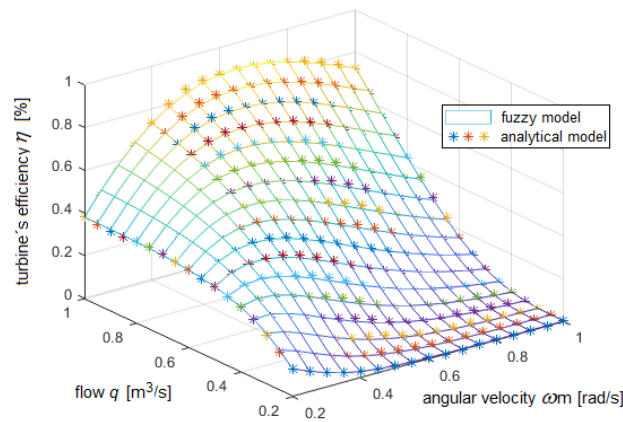


Fig. 11. Comparison of the resulting turbine efficiencies obtained analytically and through the proposed fuzzy model

However, for the purpose of modelling the behavior and operation of a small hydro power plant, despite the facts described above, we proceeded to retain a topology whereas the turbine efficiency is taken into account based on the water flow through the turbine at its zero load. This decision is based on the complexity and extent of the measured data required to create such a fuzzy model for the particular hydraulic turbine being modeled, as well as the challenging, and in some cases unrealistic, definitions of the volume of wear on the turbine blades and the hydraulic turbine itself.

D. Subsystem of other Eletromechanical parts of the system (Electrical Subsystem)

The chosen approach for creation of the electrical subsystem of a small hydropower plant can be divided into two basal sub-models, which in the result form a single entity and thus the desired subsystem for the Electrical Subsystem of the hydropower system.

- I. Submodel of the electrical network or load
- II. Submodel of the electro-mechanical part

This method is based on the following assumptions: the connection of a small hydropower plant to a “hard” grid, with a constant frequency and voltage amplitude; and an asynchronous machine installed at the position of the electrical generator. Since the processes simulated by this subsystem are mainly electrical and therefore considerably faster than the mechanical ones, taking place in the part controlling the water inflow to the turbine and in the hydraulic turbine itself, we propose to proceed in the first step to the implementation of the necessary model parts in the simplest possible form, since the speed of these processes is much higher than that of the aforementioned mechanical ones, which ultimately gives us the possibility to think of such simplification.

In accordance to the above-mentioned, we approach the modelling of the electricity network in a simplified form taking into account the above assumptions. The input of the

values of the frequency of the network and the amplitude of its voltage is already implemented in a normalized form (p.u.), which must be taken into account when it is intended to input real values in the corresponding units. Next, we consider the signal input, the meaning of which is to determine the time instant of the grid connection (for simulation purposes). The signal thus obtained is a necessary input for switching the unconnected/connected to grid status. For the purpose of the simulation outputs of the small hydropower plant model, we proceeded to include the so-called "clock" block, where we defined that at time 150s the connection of the HPP to the fictitious power grid should occur. This submodel thus has two outputs, namely the electric angular velocity ω_l , directly dependent on a grid frequency f_l and the grid voltage U_l (see Fig. 12).

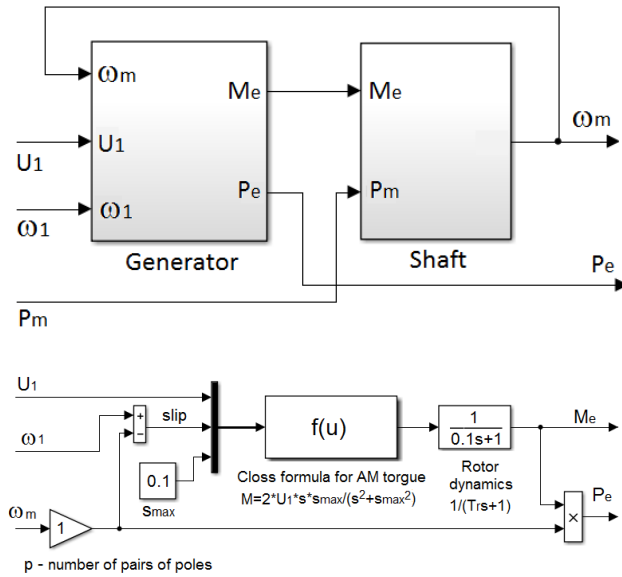


Fig. 12. Block diagram of the electrical submodel, where: the upper diagram represents an Interconnection of the generator and the turbine (the common shaft); and the lower part states for a proposed asynchronous generator model

The electromechanical subsystem is intended to be a part simulating the behavior of the purely electrical part, i.e. the asynchronous machine (generator); and simulating the shaft mechanically connecting the turbine body to the electrical generator. As can be seen in the upper and lower schemes in the Fig. 12, the inputs to this submodel are considered to be: P_m , the mechanical output power of the hydraulic turbine; f_l , the grid frequency in normalized form, or its corresponding electrical angular velocity ω_l ; and the voltage amplitude U_l , also in normalized form. Thus, the outputs of the subsystem in question are: the resulting generated (electrical) power of the small hydropower plant P_e ; and the resulting mechanical angular velocity of the system ω_m .

The principle according to which we proceeded to assemble the electrical generator subsystem, depicted in Fig. 12 (lower part), is described in more detail in paper [21]. By incorporating the associated blocks into the simulation scheme, based on the mathematical description of the Kloss formula (1), we express the generator torque M_e .

$$M_g/M_{max} = 2/(s_{max}/s + s/s_{max}) \quad (1)$$

where: we define M_{max} as the moment of reversal; s represents the current slip value of the asynchronous machine; and s_{max} expresses the slip reversal of the installed asynchronous generator. The breakdown torque is calculated based on the simplified equation (2):

$$M_{max} = m_l/\omega_e \cdot [U_l]^2 / [2X_\sigma] = k [U_l]^2 \quad (2)$$

In the formula (2) we introduce the constant k , the origin of which is due to the summation of the variables X_σ , ω_e and m_l , which we consider constant, and thus by summing them we introduce the so-called machine constant k . Since the installations in a small hydropower plants are installations of asynchronous generators of relatively large capacity, we proceed to calculate the break-down slip s_{max} and the actual slip s according to the relation (3) and (4): [21]

$$s = (\omega_e - \omega_m)/\omega_e \quad (3)$$

$$s_{max} = (p \pm \sqrt{(p^2 - 1)})s_n \quad (4)$$

whereas: s_n - slip at nominal machine torque (nominal angular velocity ω_n); p - asynchronous machine overload coefficient (m_{max}), which we calculate according to the relation (5).

$$p = m_{max} = M_{max}/M_n = (s_{max}/s + s/s_{max})/2 \quad (5)$$

In the Fig. 12, the first order inertia block, defined by the machine constant k described above, and the corresponding time constant of the rotor of the installed generator T_r , is also shown as a participation of the described model. This block has been included in the model in order to account for the slowest dynamic actions of the generator in the proposed simulation scheme of this submodel.

The method of normalization of the values presented in the proposed block diagram of this electrical generator subsystem depends on the back-end technical specification of that particular installed machine, as well as other related parts (e.g., in relation to the electrical grid subsystem, or hydraulic turbine, etc.).

The output variable of the electric generator submodel is both the actual torque of the installed asynchronous machine M_g and its electric power generated P_e . Thus, if the electrical power P_e

has a negative value we know that this is the power generated, and therefore supplied, by the small hydropower plant to the grid which it is connected to. Thus, in the generating mode of the asynchronous machine, its torque M_g also takes on negative values, since we know in general that power is defined as the product of the angular velocity and the actual torque value of the machine, while the angular velocity in the direction of the water flow being considered positive. Based on the above, and according to the formula (6), we proceed to create a submodel of the shaft (see Fig. 13).

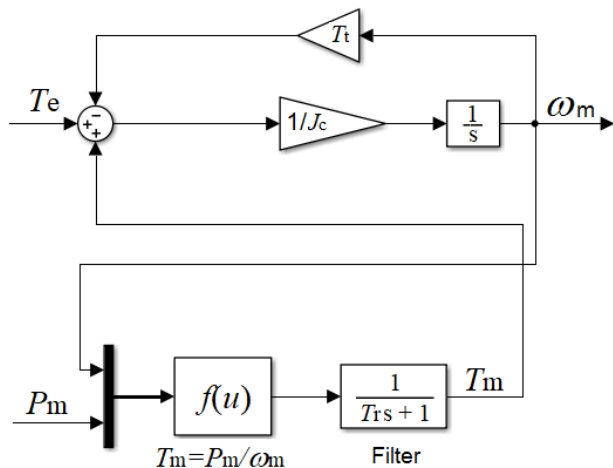


Fig. 13. Proposed structure of a subsystem simulating the behaviour of the common shaft of a small hydropower plant - the so-called generator-load fixed link

$$M - M_z = J \cdot d\omega/dt \quad (6)$$

In the light of the foregoing, we thus proceed to modify it (6) to:

$$M_g + M_t - M_{traci} = J_m \cdot (d\omega_{mech})/dt \quad (7)$$

whereas: ω_{mech} - mechanical angular velocity; J_c - total inertia on the shaft (electric generator + shaft + hydraulic turbine); M_{traci} - frictional torque equivalent, used to account for accounting a frictional losses; M_t - hydraulic turbine torque, defined as the ratio of the mechanical power of the turbine P_{mech} and its mechanical angular velocity ω_{mech} ; and M_g - the aforementioned actual torque of the electric generator of a hydropower plant of choice. Thus, based on the Laplace image of the relation (7), we proceeded to construct the submodel in question, shown in the form of a block diagram in the Fig. 13.

E. SHPP data used for the model design and verification

During the design phase of the creation of this simulation model of small hydropower plant, we used the information from the operation of the Dobšiná hydropower plants. However, the data obtained did not corresponded to our initial

target, i.e. the simulation of an asynchronous machine as a generator in the small hydropower plant. Therefore, for the process of initial design, as well as basic verification of the model, we proceeded with a data shown in the Table 2 and 3, whereas the asynchronous generator data were used from a datasheet of the asynchronous machine we chose, according to the parameters of the synchronous machine originally installed in the hydropower plant in question.

TABLE II. DATA OF A SMALL HYDROPOWER PLANT (DOBŠINA II.) USED FOR THE INITIAL DESIGN AND VERIFICATION OF THE PROPOSED MODEL

Parameter	Definition	Value used
P	mechanical power of the hydraulic turbine	2.625 MW
L	length of the penstock installed	1255 m
n_N	nominal velocity of the turbine	750 rpm
n_{max}	max. velocity of the turbine	1800 rpm
h_{base}	max. available water head	51.7 m
h_r	rated water head	43.3 m
q_{base}	water flow through the turbine for h_{base}	1 p.u. (4.6 m/s)
q_N	nominal water flow through the turbine	0.8686 p.u.
q_{NL}	no-load flow of the turbine	0,043 p.u.
T_w	water constant	7.391 s
A_t	turbine constant	1.1122
J_m	total system mechanical inertia	36.6 kgm ²

TABLE III. DATA CONSIDERED FOR AN ASYNCHRONOUS MACHINE OF CHOICE, AS A FIT FOR REPLACING THE INSTALLED SYNCHRONOUS MACHINE

Parameter	Definition	Value used
P_N	nominal power of the generator installed	2.3 MW
U_{1N}	nominal voltage	690 V (60Hz)
M_N	nominal torque	23 527 Nm
s_{max}	break-down slip	0,1
n_p	number of poles	8

IV. SIMULATION RESULTS AND VERIFICATION OF THE PROPOSED SIMULATION MODEL OF A SHPP WITH AN ASYNCHRONOUS GENERATOR

In order to explain the idea of how we verified the functionality of our proposed simulation model, it is necessary to describe in more detail the considered operation and the associated verification method. For the purpose of this model, we implement the operation of a small hydropower plant in such a way that the operator, by pressing a designated button, commands the control of the hydropower plant to start the start-up phase of the system at a speed corresponding to the frequency of the grid to which the hydropower plant is going to be connected. When the steady state is reached, i.e. when the turbine starts up at quasi-synchronous speed, the hydropower plant is automatically connected to the aforementioned power grid, but still with the required value of the supplied or generated power $P_{ref} = 0$. At the time when the operator (in our case the input signal) defines the value of the demanded generated power to a value different from zero, the SHPP switches to the mode of supplying electricity to the grid, using the so-called boost line additive from the regulation of the delivered power, according to the principle described earlier in this paper

Thus, we have defined three basic states of a small hydro power plant with an asynchronous generator, which are important for the purpose of verifying the correctness of the model. The first of these states is the start-up of the HPP to a defined reference speed, where the accuracy of the speed control of the small hydro power plant is crucial, since we need to achieve a value of mechanical angular velocity corresponding to the electrical one, in order to minimize potential current spikes when connecting such a machine to the power grid.

In Fig. 14 we depicted the achieved start-up or start-up control, respectively of the hydraulic turbine, where for the purpose of functionality check, the reference speed or the grid frequency was defined to $\omega_{ref} = 0.85$. As can be seen, in a time of approximately 310s we reach the required speed, thus putting the HPP in a state of readiness for the next step, or state, which is the connection of the system to the power grid.

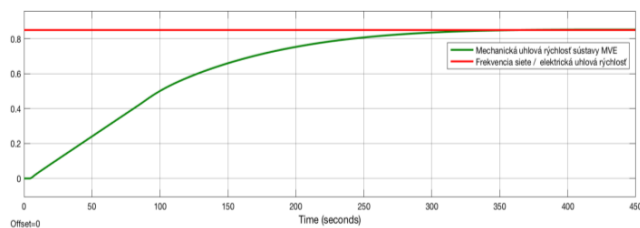


Fig. 14. The output of the mechanical angular velocity of a small hydropower plant during the start-up from the state $\omega_{mech} = 0$ to $\omega_{ref} = 0.85$

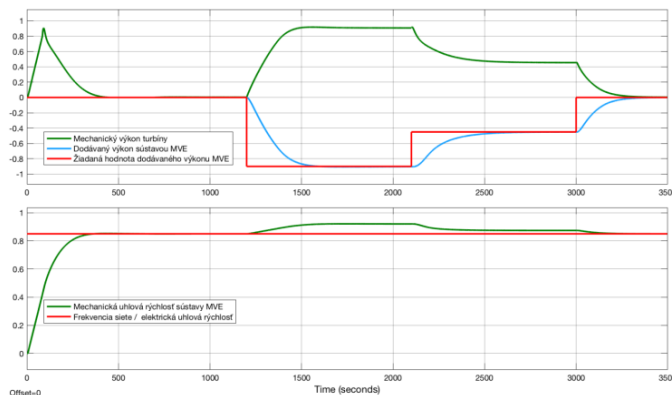


Fig. 15. The simulate outputs of the generated power of the system and its corresponding angular velocity at the defined profile of the required delivered power of the HPP to the power grid in time

Once our system is connected to the grid, it is up to the decision of the operator or other predefined logic to initiate the regulation of the delivered power. We verify this state by simulation, whose output received values are shown in Fig. 15. In the above simulation, we worked with a predefined profile of values of the required delivered power of the small hydro power plant system (red) to the power grid to which it is connected. From the above-depicted outputs, it can be seen that even at our defined value of nominal delivered power $P_{ref} = 0.9$, the modelled asynchronous motor achieves a slip of approximately 7.5%, which is still within the working range of the machine, since the break-down slip has been determined to be equal to 10%.

Accordingly, we considered the so far-described model, to be adequate and sufficient for basic simulation purposes associated with the problematics of small hydropower plants.

V. PROPOSED MODIFICATION OF THE BASIC DESIGN OF SIMULATION MODEL OF A SMALL HYDROPOWER PLANT

This chapter of the paper deals with the proposed modifications of the previously described simulation structure, especially in connection with the research direction defined at the beginning, and thus with the preparation of a simulation model, which will be the basis for the future design and simulation of the implementation of various new control structures and interventions in the technology, assuming the use of an asynchronous machine, as an electric generator of a small hydroelectric power plant.

In the available literature dealing with this issue, there are several dynamic models of an asynchronous machine, depending on the choice of its state variables and the defined rotating coordinate system [22]-[23]-[24]. By choosing the rotor flux and stator current as the model state variables, we can characterize the asynchronous machine by the system of five nonlinear differential equations (8) and (9) below.

$$\begin{bmatrix} \frac{di_{1x}}{dt} \\ \frac{di_{1y}}{dt} \\ \frac{d\psi_{2x}}{dt} \\ \frac{d\psi_{2y}}{dt} \end{bmatrix} = \begin{bmatrix} -\omega_0 & \omega_1 & -K_{12}\omega_g & -K_{12}\omega_m n_p \\ -\omega_1 & \omega_0 & K_{12}\omega_m n_p & -K_{12}\omega_g \\ M\omega_g & 0 & -\omega_g & \omega_2 \\ 0 & M\omega_g & -\omega_2 & \omega_g \end{bmatrix} \begin{bmatrix} i_{1x} \\ i_{1y} \\ \psi_{2x} \\ \psi_{2y} \end{bmatrix} + \begin{bmatrix} K_{11} & 0 \\ 0 & K_{12} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{1x} \\ u_{1y} \end{bmatrix} \quad (8)$$

$$n_p \frac{M}{L_2} (\psi_{2x} i_{1y} - \psi_{2y} i_{1x}) - T_{load} = J \frac{d\omega_m}{dt} \quad (9)$$

Thus, according to the above-mentioned, it is a current-flow description, or a model of an asynchronous machine with a squirrel cage rotor [25]-[26], where the expression of the individual quantities, i.e., the rotor flux vector ψ_2 and the stator current i_1 , is realized through their corresponding components expressed in an 90° coordinate system (x,y), which, with respect to the stator of the machine, rotates at the so-called stator field velocity (ω).

We express the variables and constants appearing in the formulas (8) and (9) using the defined motor parameters and converting them to the corresponding constants as given in equations of (10) to (15).

$$K_{11} = \frac{3}{2} \left(L_{s1} + \frac{L_{s2} L_h}{L_{s2} + L_h} \right)^{-1} \quad (10)$$

$$K_{12} = -\frac{3}{2} \left(L_{s1} + L_{s2} + \frac{L_{s1} L_{s2}}{L_h} \right)^{-1} \quad (11)$$

$$\omega_0 = K_{11} \left(R_1 + \left(\frac{M}{L_2} \right)^2 R_2 \right) \quad (12)$$

$$M = \frac{2}{3} L_h \quad (13)$$

$$\omega_g = \frac{R_2}{L_2} \quad (14)$$

$$L_2 = \frac{2}{3} (L_{s2} + L_h) \quad (15)$$

TABLE IV. DATA CONSIDERED FOR AN ASYNCHRONOUS MACHINE OF CHOICE, AS AN EXAMPLE FOR MODIFICATION TESTING AND COMPARISON

Parameter	Definition	Value used
P_N	nominal power of the generator	3 kW
U_{1N}	nominal voltage	220 V
	type of connection	Y
I_{1N}	nominal current	6.9 A
$T_N = M_N$	nominal torque	20 Nm
n_N	nominal velocity of the generator	1430 ot.min ⁻¹
n_p	number of poles	4
J	inertia of the generator	0.1 kgm ²
ω_0	calculated parameter (12)	207.9 s ⁻¹
L_2	calculated parameter	0.14 H
R_1	stator winding's resistance	1.8 Ω
R_2	rotor's resistance	1.85 Ω
L_h	main inductance	0.202 H
M	calculated parameter (13)	0.13 H
$L_{s1} = L_{s2}$	leakage inductance	0.0086 H
K_{11}	calculated parameter (10)	59.35 H ⁻¹
K_{12}	calculated parameter (11)	-56.93 H ⁻¹
ω_g	rotor winding time constant (14)	13.18 s ⁻¹

Applying equations (8) and (9), we proceeded to create a corresponding block and simulation scheme of the asynchronous generator (see Fig. 16), in which the output variables u_{1x} , u_{2x} represent the components of the stator voltage vector U_1 .

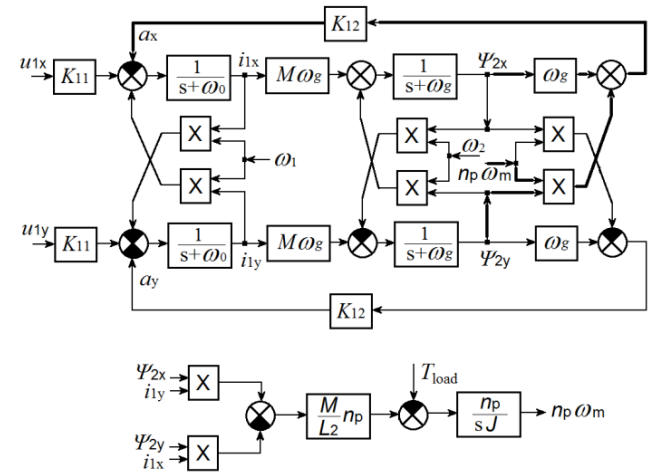


Fig. 16. Proposed simulation scheme of an asynchronous generator

Since in this subchapter we are dealing with the design and exploration of other possibilities of modelling of the given submodel of the asynchronous generator within the overall simulation model of the small hydropower plant, and at the same time since we do not have a detailed technical specification of the machine, which was chosen as a replacement of the synchronous machine installed in the Dobšiná II., we approach the modelling of a machine of diametrically different parameters (see Table IV.), which, however, does not mean any issue, since this section is devoted to the investigation of the behaviour of this particular part and not of the small hydropower plant as a whole.

The simulation model of the asynchronous machine thus created (see Fig. 16) has a strongly non-linear character, which

is also evident from its mathematical description, defined by the relations (8) and (9). At the same time, the system can be characterized as a dynamic system, containing multiple inputs and positive feedbacks.

As can be seen from the previous chapters, our proposed SHP simulation model considers the use of p.u. input and output values. On the other hand, the proposed modified, current-flow asynchronous generator model does not consider p.u. values but a real values instead. Hence, we proceed to locally convert the values input to the MVE simulation model into their real form and then normalize the output of the asynchronous machine simulation model thus modified. The aforementioned refers to the input values: voltage amplitude U_j ; network frequency f_l , or the corresponding electrical angular velocity ω_e ; the current mechanical angular velocity of the system ω_m ; and the output values: generated electrical power P_e ; and the current torque of the asynchronous machine $M_e(T)$.

Having implemented the above-mentioned modifications, we first proceeded to verify how the model based on a simple representation of the asynchronous generator behavior within the MVE model using the Kloss formula (1) and (2), approximates a significantly more accurate form of simulation through the flow-current model of the asynchronous machine described above. In this comparison, we have used an asynchronous machine with the parameters listed in Table IV. and applied these parameters in both the simplified model and the modified model. We did not change other parameters of the hydropower plant for the moment, since this is a local check of the functionality of a single, given subsystem. In Fig. 17 we present the comparison of the outputs of the two models of the asynchronous generator subsystem of the HPP. It is evident that the results obtained in steady-state operation for both models are satisfactory for the purposes for which the small hydropower plant simulation model is designed and are almost identical to each other, especially in the case of the generated power. The difference between the torque values of the simulated asynchronous machine is mainly attributed to the normalization of the signals as well as to the complexity of the modified model. Additionally, the feedback to the control loop of a small hydroelectric power plant was coupled to the model based on the use of the Kloss relation. At the same time, it is clear that for fast processes and their tracking it is appropriate to choose this modified model of the asynchronous generator subsystem.

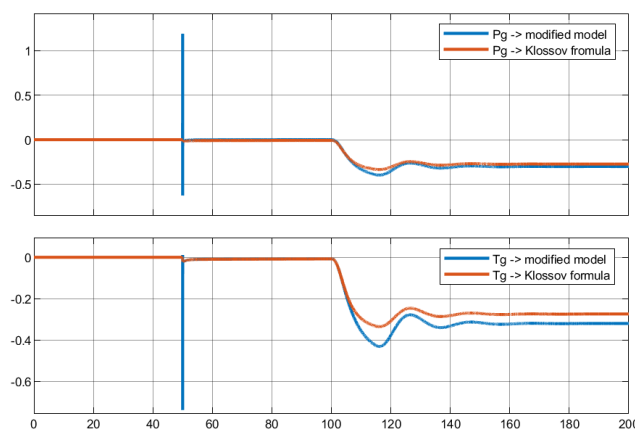


Fig. 17. Comparison of the output generated torque of an asynchronous machine as well as its generated power, i.e. the modified model vs. model based on the use of the Kloss formula (T_g - actual machine torque; P_g - actual generated power)

After implementing this model in a complete simulation scheme of a small hydroelectric power plant, we proceeded to use this modification for the purpose of tracking the internal fast transients primarily associated with the asynchronous generator part. Below we present the progress of the simulation performed by us when: in the time interval from 0 to 50 seconds, the start-up of the HPP is simulated to a defined angular velocity ω_{ref} , which corresponds to the frequency of the grid; then, at time $t = 50$ s, the connection of the HPP to the power grid occurs; then, at time $t = 100$ s, the start of the control of the power generated by the HPP occurs, or the start of the power supply to the power grid, respectively. In the Fig. 18 below, we see that after this point, the above-synchronous speed of the installed asynchronous machine is achieved by the system, which is therefore proportional to the delivered, and ultimately the reference generated power.

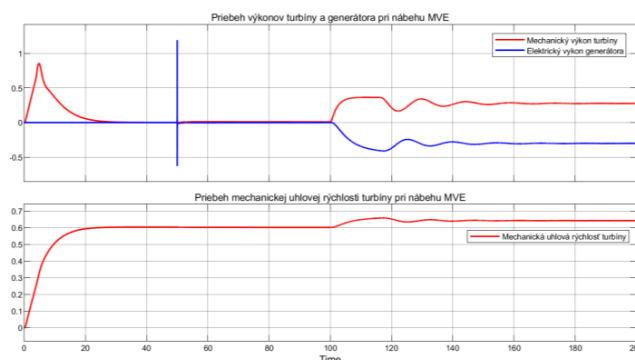


Fig. 18. Simulation of the above-mentioned states of a small hydropower plant with respect to the delivered power and the mechanical angular velocity of the system using the modified model of asynchronous generator

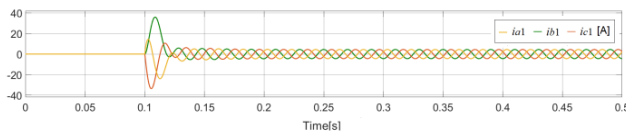


Fig. 19. Stator currents of the simulated asynchronous generator obtained from the proposed modified asynchronous machine submodel (the origin of the time axis in the graph represents the time $t = 50$ s in the previous graph and thus the moment of connection of the SHP to the grid after its start-up to synchronous speed)

The result of this modified simulation model of the HPP is a model ready for a simulation of various small hydro power plants with asynchronous generator in different operating states with the possibility to observe also fast transients occurring primarily at the specific moment when the system is being connected to the grid, or during its failure, etc. Such simulations consequently provide a strong basis for a correct and faster design of these systems, as well as a tool for the optimization of existing small hydropower plants, minimizing the necessary implementation and testing time for possible modifications directly in the SHPP.

VI. VERIFICATION OF THE PROPOSED MODIFIED SIMULATION MODEL OF A SMALL HYDROPOWER PLANT BASED ON THE REAL DATA FROM SHPP DOBŠINÁ III.

Naturally, the necessary part of designing such a simulation model of a small hydropower plant is also its comparison or verification based on the data of the existing system of the given specifics. Such a hydropower plant chosen by us is the small hydropower plant Dobšiná III, which consists of one Kaplan turbine of 275 kW directly coupled to an asynchronous squirrel cage generator of 315kW. This SHPP thus corresponds to the type of small hydropower plant whose simulation is dealt with in this paper.

According to the inspection and the data measured as well as defined by the provided drawings for the small hydropower plant Dobšiná III. (see Table V. and Table 6), we proceeded to the procedure of adapting the proposed simulation model, specified in more detail in previous chapter of this particular paper.

TABLE V. INSTALLED ASYNCHRONOUS GENERATOR DATA (SHPP DOBŠINÁ III.)

Parameter	Definition	Value used
P_N	nominal power of the generator	315 kW
U_{1N}	nominal voltage	400 V
	type of connection	D
I_{1N}	nominal current	510 A
M_N	nominal torque	3126 Nm

m_{max}	overload coefficient	2,7
s_{max}	Break-down slip	0,12592
n_N	nominal velocity of the generator	1011 ot.min ⁻¹
n_{max}	max. mechanical velocity of the generator	2600 ot.min ⁻¹
n_p	number of poles	6
J	inertia of the generator	7,3 kgm ²
$cos\varphi$	power factor	0,89
ω_0	calculated parameter	364,23 s ⁻¹
L_2	calculated parameter	0,00652 H
R_1	stator winding's resistance	0,01796 Ω
R_2	rotor's resistance	0,03656 Ω
L_h	main inductance	0,00933 H
M	calculated parameter	0,00622 H
$L_{s1}=L_{s2}$	leakage inductance	0,00045 H
K_{11}	calculated parameter	1711,4 H ⁻¹
K_{12}	calculated parameter	-1632,9 H ⁻¹
ω_g	rotor winding time constant	5,6106 s ⁻¹

TABLE VI. PARAMETERS OF THE HYDRAULIC PART OF THE SHPP DOBŠINÁ III.

Parameter	Definition	Value used
P	mechanical power of the hydraulic turbine	275 kW
L	length of the penstock installed	56.932 m
n_N	nominal velocity of the turbine	750 rpm.
n_{max}	max. velocity of the turbine	1800 rpm
h_{base}	max. available water head	10 m
h_r	rated water head	9.8 m
q_{base}	water flow through the turbine for h_{base}	1 p.u. (3.9 m/s)
q_N	nominal water flow through the turbine	0.8205 p.u.
q_{NL}	no-load flow of the turbine	0,07 p.u.
T_w	water constant	1.00594 s
A_t	turbine constant	1.1992
J_m	total system mechanical inertia	40 kgm ²

As it was unfortunately not possible to provide us with precise data on the control of the simulated small hydropower

plant Dobšiná III, or with the setting of its parameters, respectively, in the comparison below we mainly aim at verifying the behavior of the hydromechanical parts of the system of the small hydropower plant, which will be fully shown during its start-up. Of course, the setting of the angular speed controller of the hydraulic turbine plays an important role in this case as well, but the modelling of an inertia as well as other properties characteristic for the hydraulic turbine subsystem can be evaluated in this way.

In Fig. 20 we present the run-up of the small hydropower plant Dobšiná III, which was provided by its operation. From the above-mentioned graph it is clear that the start-up of the given hydropower plant was initiated at 8:28:55 with reaching the required level (100%), which corresponds to the grid frequency and the characteristics of the installed asynchronous generator, at 8:29:20. This means that, in our understanding, at a duration of 25 seconds, the Dobšiná III HPP was in a state ready to connect to the grid. Similarly, for comparison, we proceeded to simulate the power plant in question in terms of its characteristic data as well as calculated values, shown in Table V. and Table VI., below in Fig. 21. We can see that in our model we proceeded to a different definition of the maximum frequency value, or its corresponding angular velocity, for the purpose of normalizing this signal. However, the real values are identical and thus the 60% of the mechanical angular velocity modeled by us, corresponds to the 100% shown in Fig. 20.

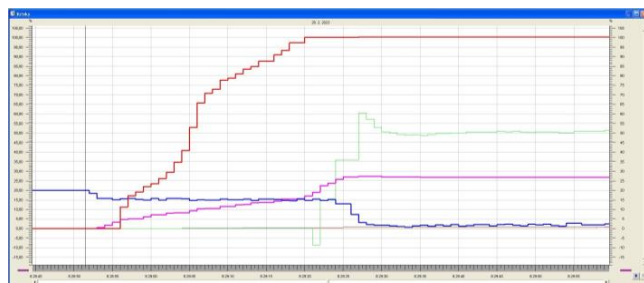


Fig. 20. Run-up of the small hydropower plant Dobšiná III. to the speed corresponding to the grid frequency - 100% (brown), obtained from its RiS (control information system)

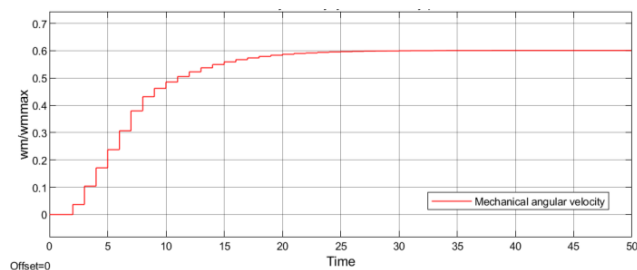


Fig. 21. Run-up of the small hydropower plant Dobšiná III. to the speed corresponding to the grid frequency - 60%, obtained from our proposed simulation model of the Dobšiná III. hydropower plant in the Matlab Simulink environment

We can see that both of the above-mentioned figures have almost the same time of reaching the desired value, which, since it is a verification of the correctness of the model, we evaluate as a correct way of modelling the hydromechanical part of the small hydropower plant Dobšiná III. At the same time, we attribute the difference in the shape of the angular velocity increase of the system to the measuring component, i.e. the output from the velocity sensor installed in the Dobšiná III hydropower plant, as well as to the already mentioned possible differences in the topology of the velocity control of the hydroelectric power plant.

VII. CONCLUSION

In this paper we described the complex process of the development of the small hydropower plant simulation model, while considering an asynchronous generator being used for the electricity generation. Moreover, we explain the other possibilities of approaching this problematic and explained our decisions for each of the above-described subsystems of our complex model.

After some basic verification procedures of the initially designed model, we proceeded with a possible modifications development, in order to fulfil our initially set target, i.e. the creation of the simulation model allowing the faster and more correct design of each of the technological parts included in the model, especially the control topologies relevant for such a system. We accomplished this with the modification described in the Chapter V of this paper.

At the end we incorporated the whole model, as well as the modification proposed, and proceeded with an absolute verification of the model, especially its hydromechanical part, based on the data obtained from real existing SHPP of such characteristics, i.e. SHPP Dobšiná III. The result proved that the model was set up and designed properly and corresponds to the real measured outputs from the simulated system.

In conclusion, we designed, created and verified quite complex simulation model of a small hydropower plant that gives its user the option of simple designing and verifying the intended changes and topologies for a real future SHPPs, as well as it represents an useful tool for the possible modifications verification of the already-existing systems, while lowering the necessary on site work at the minimum level, and thus keeping the costs low.

ACKNOWLEDGMENT

This work was supported by the APVV grant APVV-16-0206 and APVV-19-0210.

REFERENCES

- [1] Wagner, Herman-Josef a Mathur, Jyotirmay. Introduction to Hydro Energy Systems: Basics, Technology and Operation. Heidelberg : Springer, 2011. ISBN 978-3-642-20708-2.
- [2] Hydraulic Turbine and Turbine Control Models for System Dynamic Studies. Working Group on Prime Mover and Energy Supply Models for System Dynamic Performance Studies. 1. : IEEE, 1992, IEEE Transactions on Power Systems, Zv. VII, pp. 167-179. ISSN: 1558-0679.

- [3] A. Acakpovi, E.B. Hagan and F.X. Fifatin, "Review of Hydropower Plant Models," 18, , December 2014, *International Journal of Computer Applications*, Zv. 108, pp. 33-38.
- [4] A.A. Usman and R.A. Abdulkadir, "Modelling and Simulation of Micro Hydro Power Plant Using Matlab Simulink," New Delhi : ICSTM, 2015. 2nd International Conference on Science, Technology and Management. pp. 1121-1133.
- [5] M. Sattouf, "Simulation Model of Hydro Power Plant Using Matlab/Simulink," Brno : IJERA, 2014, *International Journal of Engineering Research and Applications*, Zv. 4, pp. 295-301. ISSN: 2248-9622.
- [6] M. Vinatoru, "Monitoring and Control of Hydro Power Plant," Cluj-Napoca : s.n., 2007. IFAC Workshop ICPS'07.
- [7] C. Jaliu, I. Visa, D. Diaconescu, et al., "Dynamic Model of a Small Hydropower Plant," 2010. 12th International Conference on Optimization of Electrical and Electronic Equipment.
- [8] H. Fang, Z. Shen., "Dynamic Real-time Simulator for Hydraulic Turbine Generating Unit Based on Programmable Computer Controller," Korea : IEEE, 2004. The 30th Annual Conference of the IEEE Industrial Electronics Society. pp.. 1502-1507.
- [9] G. Munoz-Hernandez, S. Mansoorand D. Jones, "Modelling and Controlling Hydropower Plants," 1st Edition. London : Springer London Ltd, 2013. pp. 299. ISBN: 978-1-4471-6221-6.
- [10] N. Kishor and Fraile-Ardanuy, "Modeling and Dynamic Behaviour of Hydropower Plants," London: The Institution of Engineering and Technology, 2017. ISBN: 978-1-78561-195-7
- [11] I. Salhi, S. Doubabi, and N.B. Essounbouli, "Fuzzy Control Of Micro Hydro Power Plants," 5th IET International Conference on Power Electronics, Machines and Drives (PEMD 2010), 2010, pp. 1-6, doi: 10.1049/cp.2010.0084.
- [12] Guangtao Zhang, Yuanchu, Cheng, Na, Lu and Qi, Guo, "Research of Hydro-Turbine Governor Supplementary Control Strategy for Islanding AC Grid at Sending Terminal of HVDC System," 4, s.l. : IEEE, 4th. December 2016, *IEEE Transaction on Energy Conversion*, Zv. 31, pp. 1229-1238
- [13] C.L.T. Borges and R.J. Pinto, "Small hydro-power plants energy availability modeling for generation reliability evaluation," *IEEE Trans Power Syst.*, vol.23, no.3, 2008, pp. 1125-1135. doi.10.1109/TPWRS.2008.926713.
- [14] E. Özbay and M.T. Gençoğlu, "Modeling of small hydro power plants," Proc. of National Conference on Electrical, Electronics and Computer Engineering, Bursa, Turkey, 2010, pp. 32-36.
- [15] R. Olexa, D. Perduková, and P. Fedor, "Fuzzy Approach of Modeling a Hydraulic Turbine Efficiency," *MM Science Journal*, vol. 2020, 2020, pp. 4086-4092. ISSN 1803-1269. doi.10.17973/MMSJ.2020_11_2020033.
- [16] D. Perduková, and P. Fedor, "Design rules of a fuzzy controller using genetic algorithms," Proc. of the 7th Int. Scientific Symposium on Electrical Power Engineering ELEKTROENERGETIKA 2013. Stará Lesná, 2013, pp. 513-516. ISBN 978-80-553-1441-9
- [17] P. Girovsky, J. Timko, J. Žilková, et al., "Neural Estimators for Shaft Sensorless FOC Control of Induction Motor," Proc. of 14th International Power Electronics and Motion Control Conference (EPE-PEMC), Ohrid, Macedonia, 2010. DOI: 10.1109/EPEPEMC.2010.5606907.
- [18] R.D. Gaceanu and F. Poph, "A fuzzy incremental clustering approach to hybrid data discovery," *Acta electrotechnica et informatica*, 2012, Vol.12, No.2, pp. 16-23.
- [19] P. Fedor, and D. Perduková, "Model Based Fuzzy Control Applied to a Real Nonlinear Mechanical System," *Iranian Journal of Science and Technology, Transactions of Mechanical Engineering*, 22016, Vol.40, No.2, pp. 113-124.
- [20] J.S.R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System," *IEEE Trans. Syst., Man, and Cyber.*, 1993, Vol.23, No.3, pp. 665-685.
- [21] F. Poliak, V. Fedák and L. Zobray, „Elektrické pohony,” Košice : Alfa, 1987. pp. 606. EPO: 063-568-87.
- [22] M. Lešo, J. Žilková and M. Pástor, „Pokročilé metódy riadenia DC-DC meničov,” *Elektrotechnické listy*, Vol. 1, No. 2, 2016, pp. 1-4. ISSN: 2453-8981
- [23] C.D. Tran, M. Kuchař, M. Sobek, V. Sotola and B.H. Dinh, "Sensor Fault Diagnosis Method Based on Rotor Slip Applied to Induction Motor Drive," *Sensors*, vol.22, no.22, 2022, art. no. 8636. doi.10.3390/s22228636.
- [24] P. Palacký, M. Sobek, P. Hudeček, D. Slivka and V. Sládeček, "Asynchronous motor working as generator with VSI using flux field oriented control," Proc. of 11th International Scientific Conference Electric Power Engineering (EPE), 2010, pp. 817 – 822.
- [25] I. Neborak, V. Sládeček, and M. Kuchař, "Modelling and simulation of induction machine and frequency converter considering power losses," 16th International Scientific Conference on Electric Power Engineering (EPE), 2015, pp. 251-255. doi.10.1109/EPE.2015.7161100
- [26] D. Perduková, P. Palacký, P. Fedor, P. Bober and V. Fedák, "Dynamic Identification of Rotor Magnetic Flux, Torque and Rotor Resistance of Induction Motor," *IEEE Access*, vol.8, 2020, pp. 142003-142015. ISSN 21693536. doi.10.1109/ACCESS.2020.3013944.

Detekcia škodlivého softvéru založená na algoritmoch strojového učenia

Jakub Palša

Technická univerzita v Košiciach
Fakulta elektrotechniky a informatiky
Košice, Slovenská Republika
Email: jakub.palsa@tuke.sk

Liberios Vokorokos

Technická univerzita v Košiciach
Fakulta elektrotechniky a informatiky
Košice, Slovenská Republika
Email: liberios.vokorokos@tuke.sk

Abstract—V práci sa venujeme trénovaniu modelov strojového učenia využívajúcich algoritmy rozhodovacie stromy (DT), náhodný les (RF), Naivná Baysová metóda (NB), metóda podporných vektorov (SVM), extrémne náhodné stromy (ET) a XGBoost (XGB) na dvoch dátových sádach, ktoré boli získané statickou a dy-namickou analýzou reálnych škodlivých a neškodných vzoriek. Ich úspešnosť následne porovnáваме medzi sebou pričom vyhodnotíme najlepší algoritmus. Najúspešnejšie klasifikačné modely, využívajúce algoritmus XGBoost, dosahujú presnosť detekcie 91,9% so senzitivitou 98,2% a 0,853 AUC v prípade dátovej sady statickej analýzy, respektíve presnosť 96,4% so senzitivitou 98,5% a 0,940 AUC v prípade dátovej sady dynamickej analýzy. Najúspešnejšie modely strojového učenia sú následne exportované a využité pri navrhnutom programe MLMD, ktorý automatizuje proces statickej a dynamickej analýzy a umožňuje využiť natrénované modely aj pri klasifikácii na nových vzorkách.

Abstract—This paper focuses on training machine learning models using Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, Extremely Randomized Trees and XGBoost algorithms on two datasets obtained using static and dynamic analysis of real malicious and benign samples. We then compare their success with each other and evaluate the best algorithm. The best performing classification models, using the XGBoost algorithm, achieved a 91.9% detection accuracy, a 98.2% sensitivity and 0.853 AUC on the static analysis dataset and a 96.4% accuracy, a 98.5% sensitivity and 0.940 AUC on the dynamic analysis dataset. Then, we exported the best performing machine learning models and used them in our proposed MLMD program, automating the process of static and dynamic analysis and allowing the trained models to be used for the classification on new samples.

I. ÚVOD

Digitálne zariadenia sú v súčasnosti značne prepojené so životmi ľudí a na uchovávanie informácií a prístup k nim ich už dnes používa takmer každý. Tieto zariadenia sú ale kvôli cennosti informácií čoraz častejšie cieľom rôznych typov útokov, pričom útočníci často využívajú škodlivý softvér. Škodlivý softvér predstavuje program, ktorý dokáže v napadnutom systéme vykonávať istý druh škodlivej činnosti [Monnappa(2018)]. Podľa štatistík spoločnosti Malwarebytes [Malwarebytes(2020)] sa počet hrozieb každým rokom zvyšuje a zvyšuje sa taktiež aj ich závažnosť. Napadnutie škodlivým softvérom môže viesť ku strate či už jednoducho nahraditeľných dát, ale aj k strate citlivých dát a peňazí. O nebezpečenstve škodlivého softvéru môžeme čoraz

častejšie počuť aj v správach, ktoré ukazujú jeho následky na spoločnostiach ale aj jednotlivcoch. Aj samotné spoločnosti si začínajú uvedomovať vážnosť situácie a bezpečnosti tak venujú veľkú časť rozpočtu [Elisan(2018)].

Riešenia sa snažia ponúkať tvorcovia bezpečnostných riešení, ktorí na neustály vývoj nového škodlivého softvéru pribúdajúceho každý deň musia reagovať a vyvíjať nové možnosti na detekciu škodlivého softvéru. Tradičné metódy detekcie už často nestačia na pokročilejšie techniky využívané škodlivým softvérom, pomocou ktorých sa dokáže vyhnúť detekcii bezpečnostným softvérom. Nevýhodou pri použití tradičných spôsobov na detekciu škodlivého softvéru je taktiež to, že vedia detegovať len škodlivý softvér, ktorý už bol analyzovaný a ktorého príznaky sú už známe. Potrebu sofistikovanejších techník na detekciu škodlivého softvéru sa tvorcovia softvéru na zabezpečenie systému snažia vyriešiť aj použitím strojového učenia, pri ktorom sa modely strojového učenia dokážu učiť zo vstupných dát a ktorého použitie sa s úspechom presadilo aj pri riešení iných súčasných problémov [Lawrynowicz(2014)].

Hlavnými prínosmi tejto práce sú:

- analýza a porovnanie algoritmov XGBoost a extrémne náhodné stromy pri detekcii škodlivého softvéru,
- získanie príznakov z výstupov statickej analýzy vzoriek programom Dependency Walker a z výstupov dynamickej analýzy vzoriek nástrojom Cuckoo Sandbox,
- automatizovanie procesov analýzy a klasifikácie nových vzoriek pri navrhnutom programe.

V kapitole II sa venujeme základnej teórii detekcie škodlivého softvéru a obsahuje prehľad aktuálneho stavu problematiky. V kapitole III opíšeme pôvod škodlivých a neškodných vzoriek, spôsob ich spracovania nástrojmi *Dependency Walker* a *Cuckoo Sandbox* do dátových sád a trénovanie samotných modelov, spolu s hľadaním najlepších hyperparametrov. Modely následne vyhodnotíme a najlepšie využijeme pri navrhovanom programe MLMD, ktorého základný návrh spolu s ukázkami grafického rozhrania sa taktiež nachádza v tejto kapitole. V kapitole IV porovnáme nami dosiahnuté výsledky s opísanými existujúcimi prácami a taktiež porovnáme úspešnosť nášho programu s nástrojom *VirusTotal*. V kapitole V je uvedený záver a zhrnutie článku.

II. TEÓRIA

Nasledujúca časť obsahuje stručný popis statickej a dynamickej analýzy a tradičných spôsobov na detekciu škodlivého softvéru, ich výhod a nevýhod. Následne opisuje niekoľko existujúcich prác venujúcich sa tejto problematike.

A. Analýza škodlivého softvéru

Cieľom analýzy škodlivého softvéru je zistenie funkcionality škodlivého softvéru, jeho štruktúry a zámerov útočníka. Spôsoby analýzy možno rozdeliť do nasledujúcich kategórií [Mohanta(2020)]:

- **statická analýza** - analýza programu bez jeho spúšťania. Pri základnej statickej analýze sa hľadajú statické informácie - ako napríklad reťazce, sieťové adresy, volané funkcie, informácie z hlavičky spustiteľného súboru. Pokročilá statická analýza zahŕňa využitie techník reverzného inžinierstva za pomoci rôznych špeciálnych nástrojov [Fedak(2020)].
- **dynamická analýza** - analýza programu počas vykonávania. Vykonáva sa v bezpečnom virtuálnom prostredí, v ktorom sa pozorujú aktivity spúšťaného programu, na základe ktorých je odvodený účel programu [Mohanta(2020)][Hisham(2015)]. V zabezpečenom prostredí sa pozorujú zmeny v registroch, sieťová aktivita, volanie funkcií, modifikácie súborov na disku a podobne.

Obidva typy analýz majú svoje výhody a nevýhody. Statická analýza je bezpečnejšia, keďže škodlivý kód nie je priamo spúšťaný a je možné pomocou nej odhaliť všetky škodlivé zámery. Nevýhodou statickej analýzy je však neúčinnosť proti pokročilým zahmlievacím technikám, ktorých cieľom je znemožnenie analýzy zdrojového kódu [Fedak(2020)]. Dynamická analýza umožňuje odhaliť funkcionality programu bez potreby analýzy kódu, ktorá môže byť zdĺhavá. Nevýhodou je však potreba zaistenia bezpečného prostredia na jeho vykonanie [Hisham(2015)], keďže spustenie škodlivej vzorky v skutočnom systéme by mohlo byť rizikom, a taktiež to, že pokročilejší škodlivý softvér často dokáže rozpoznať virtuálne prostredie a obmedzí vykonávanie škodlivej činnosti.

B. Detekcia škodlivého softvéru

Na detekciu škodlivého softvéru sa v súčasnosti využívajú hlavne nasledujúce spôsoby:

- **detekcia založená na signatúrach** - najčastejšie využívaný spôsob detekcie škodlivého softvéru [Damodaran(2020)]. Signatúry sú jedinečné postupnosti bajtov získavané pri analýze škodlivého softvéru, podľa ktorých možno identifikovať škodlivý softvér [Hisham(2015)]. Bezpečnostné programy obsahujú databázu signatúr už známeho škodlivého softvéru a pri kontrole nového súboru je tento súbor analyzovaný a porovnávaný s databázou [Damodaran(2020)]. V prípade, že analyzovaný súbor obsahuje signatúru z databázy, s veľkou presnosťou ho možno označiť ako škodlivý.
- **detekcia založená na správaní** - pri detegovaní sa skúma správanie programu a jeho aktivít [Damodaran(2020)].

Pokusy o vykonanie akcií, ktoré sú abnormálne alebo neoprávnené by mohli naznačovať, že program je škodlivý alebo minimálne podozrivý. Abnormálnymi akciami môžu byť napríklad modifikovanie iných súborov, pridávanie nových používateľov systému, zastavovanie bezpečnostného softvéru v systéme a podobne [Damodaran(2020)].

- **heuristická detekcia** - spočíva v hľadaní istých znakov indikujúcich škodlivé správanie využitím pravidiel a algoritmov, a to hľadaním príkazov a inštrukcií typických pre škodlivý softvér alebo sledovaním jeho správania a aktivít počas vykonávania, prípadne často ich kombináciou [Cisar(2020)]. Jednotlivé aktivity sú ohodnotené podľa nebezpečenstva a v prípade prekročenia stanovenej prahovej hodnoty je prijaté preventívne opatrenie [Seker(2021)].

Najväčšou nevýhodou detekcie založenej na rozpoznávaní signatúr je to, že pomocou nej možno detegovať len známy škodlivý softvér, ktorého signatúry sa nachádzajú v databáze signatúr [Hisham(2015)]. Tento problém čiastočne riešia spôsoby detekcie založené na správaní a heuristická detekcia, pomocou ktorých je možné detegovať aj nový škodlivý softvér, ktorého príznaky zatiaľ neboli nájdené. Ich nevýhodou je však vyššia miera falošnej pozitivity, teda nesprávne označenie neškodného programu ako škodlivého [Hisham(2015)].

C. Detekcia pomocou strojového učenia

Strojové učenie sa v posledných rokoch uplatňuje v mnohých oblastiach a inak tomu nie je ani pri detekcii škodlivého softvéru, pri ktorom sa snažia vyriešiť nevýhody zaužívaných spôsobov detekcie, a to najmä v prípade nových druhov škodlivého softvéru pribúdajúcich každý deň. S tým súvisí zvýšený záujem vedeckej komunity o túto oblasť a množstvo existujúcich prác. Tie sa odlišujú najmä použitými algoritmi a tým, aký typ analýzy sa využíva - statická alebo dynamická. Nasledujúca časť obsahuje opis niekoľkých existujúcich prác v tejto problematike, a to konkrétne využitím strojového učenia s učiteľom.

D. Existujúce práce

Detekcii škodlivého softvéru využitím techník strojového učenia sa ako prví vo svojej práci [Schultz(2001)] venovali Schultz a kol., a to využitím techník statickej analýzy. Použitými príznakmi v práci boli extrahované reťazce, informácie z hlavičiek spustiteľných súborov (volané a importované funkcie) a postupnosti bajtov dĺžky n (tzv. *n-gramy*). V práci celkovo použili 3 algoritmy strojového učenia, a to RIPPER, NB a Multi-NB, pričom pre porovnanie ich doplnili o tradičný spôsob detekcie škodlivého softvéru - detekcie využitím signatúr. Celková presnosť detekcie bola v ich experimentoch v rozmedzí 83,62% až 97,11% využitím algoritmov strojového učenia, čo oproti využitiu tradičného spôsobu detekcie pomocou signatúr s celkovou presnosťou 49,28% predstavovalo významný nárast.

Využitím techník statickej analýzy sa vo svojej práci [Bai(2014)] venovali aj Bai a kol. V práci využili 197

príznakov získaných z hlavičky spustiteľných súborov, a to konkrétne počty API volaní a počty použitých dynamicky spojených knižníc. Pre výber tých najvhodnejších príznakov, pomocou ktorý by bolo možné čo najlepšie rozlíšiť medzi škodlivou a neškodnou vzorkou, autori využili filtračné a obalovacie techniky. Využitými a porovnávanými algoritmi v tejto práci boli algoritmy založené na rozhodovacích stroch, a to algoritmy J48 a RF. Vykonaním niekoľkých experimentov autori dosiahli celkovú presnosť detekcie v rozmedzí 94,6% až 99,1%.

Informácie z hlavičiek spustiteľných súborov použili aj Kumar a kol. [Kumar(2019)], ktorí využili tzv. integrovaný set príznakov z celkovým počtom 68 príznakov. Autori v práci využili a porovnali celkovo 6 algoritmov, a to algoritmy RF, DT, LR, NB, LDA a KNN. Celkové dosiahnuté presnosti detekcie boli v rozmedzí 56,04% až 98,78%. Najmenej úspešným algoritmom v porovnaní bol algoritmus NB a naopak najúspešnejším algoritmom s najlepšou presnosťou detekciu bol algoritmus RF.

Príznačky z dekompilovaného strojového kódu vo forme sekvencií opcode rôznej dĺžky vo svojej práci využil Bragen [Bragen(2015)], ktorý taktiež využil a porovnal 6 algoritmov, a to algoritmy RF, NB, KNN, ANN (MLP BP), SVM a J48. Najúspešnejším algoritmom s najlepšou presnosťou detekcie bol opäť algoritmus RF, kde pri sekvenciách s dĺžkou 3 (3-gramy) dosiahol presnosť detekcie 95,58%.

Získaním sekvencií s dĺžkou 5 (5-gramy) a informácií o API volaniach z hlavičky spustiteľného súboru získali vo svojej práci Chowdhury a kol. [Chowdhury(2018)], ktorí využili a porovnali úspešnosť algoritmov NB, J48, RF, SVM, ANN (MLP BP). Vykonali niekoľko experimentov podľa typu využitých príznakov a najlepšiu presnosť detekcie 97,7% získali využitím algoritmu ANN with a feed-forward perceptron a skombinovaním obidvoch typov príznakov. Najmenej úspešným algoritmom v tejto práci bol algoritmus NB s presnosťou detekcie 87,5%.

Uvedené práce získavali príznaky len využitím techník statickej analýzy. Problém pri tomto postupe pri detekcii škodlivého softvéru vo svojej práci opisujú Moser a kol. [Moser(2007)], pričom za hlavný dôvod uvádzajú použitie zahmlievacích techník na skrytie škodlivého kódu. Odporúčaným spôsobom je tak podľa ich názoru využitie aj dynamickej analýzy, ktorá je voči tomuto problému menej zraniteľná.

Shijo a Salim [Shijo(2015)] tak okrem príznakov získaných extrahovaním reťazcov zo spustiteľných súborov využitím techník statickej analýzy využili aj dynamickú analýzu vykonávanú v izolovanom virtuálnom prostredí, a to na získanie informácií ako sú zmeny v registroch a počty volaní API funkcií. Porovnávanými algoritmi v tejto práci boli algoritmy RF a SVM. Niekoľkými experimentami vo svojej práci porovnávali tieto algoritmy a ich úspešnosti pri jednotlivých typoch analýz. Najlepší výsledok bol dosiahnutý využitím algoritmu SVM na príznakoch získaných kombináciou statickej a dynamickej analýzy, a to 98,7%. Na príznakoch získaných len dynamickou analýzou bola presnosť detekcie o niečo

horšia, a to 97,1%. Pri statickej analýze bola presnosť detekcie 95,88%. Dosiahnuté výsledky tak ukázali, že kombinácia statickej a dynamickej analýzy zvýšila presnosť detekcie oproti použitiu samostatnej statickej a dynamickej analýzy.

Použitiu dynamickej analýzy na získanie príznakov sa venovali aj Firdausi a kol. [Firdausi(2010)], ktorí na dynamickú analýzu využili voľne dostupný nástroj Anubis. Vo svojej práci na niekoľkých experimentov použili a porovnali celkovo 5 algoritmov strojového učenia, a to algoritmy NB, SVM, ANN (MLP BP), KNN a algoritmus J48. Ako najlepší algoritmus sa v ich práci ukázal algoritmus J48 s presnosťou detekcie 96,8% a ako najhorší sa ukázal algoritmus NB s presnosťou detekcie 62,8%.

Z tzv. výpisov pamäte (z ang. *memory dumps*) počas vykonávania vzoriek využitím nástroja Cuckoo Sandbox získavali príznaky Mosli a kol. [Mosli(2017)], ktorí zároveň porovnávali úspešnosť 3 algoritmov - KNN, SVM, RF. Najlepšiu presnosť detekcie získali využitím algoritmu RF, a to 91,4%. Najhoršie výsledky dosiahli použitím algoritmu KNN s presnosťou detekcie 91%.

Porovnanie existujúcich prác možno vidieť v tabuľke I.

TABLE I: Prehľad existujúcich prác

práca	typ analýzy	veľkosť dátovej sady (skladov / čísel)	porovnané algoritmy	najlepšia presnosť (%)	najlepšia senzitivita (%)
Schultz a kol.	statická	3 265 / 1 001	RIPPER, NB, Multi-NB	97,21	97,43
Bai a kol.	statická	10 521 / 8 592	J48, RF	95,1 - 99,1	91,3 - 99,1
Kumar a kol.	statická	2 722 / 2 488	RF, DT, LR, NB, LDA, KNN	98,78	99,9
Bragen	statická	962 / 771	RF, NB, KNN, SVM, J48, ANN	95,58	96,77
Chowdhury a kol.	statická	41 265 / 10 020	NB, J48, RF, SVM, ANN	97,7	93
Shijo a Salim	statická	967 / 490	SVM, RF	95,88	95,9
Shijo a Salim	dynamická	967 / 490	SVM, RF	97,16	97,2
Shijo a Salim	hambíbovaná	967 / 490	SVM, RF	96,71	96,7
Firdausi a kol.	dynamická	220 / 250	NB, SVM, MLP, KNN, J48	96,8	96,9
Mosli a kol.	dynamická	3 130 / 1 137	KNN, SVM, RF	91,4	91,1
Kumar, Geetha	Ember dataset	300K / 300K	Gaussian, NB, KNN, Linear SVC, DT, AdaBoost, RF, Extra Trees, GB, XGBoost	98,5	0,89-0,99
Dhanraj, Dhanraj	Open data source	4000 / 2709	NB, DT, RF, GB, XGBoost	99,95	99,95
Shibadi et al.	Open data source	984 / 172	KNN, SVM, Bernoulli NB, RF, DT, LR, HV	98,2	92

III. METODIKA NÁVRHU

Je zrejme, že použitie strojového učenia na detekciu škodlivého softvéru je výhodné s dobrou presnosťou detekcie. Existujúce práce taktiež naznačujú, že využitie rozhodovacích stromov k detekcii škodlivého softvéru sa zdá byť vhodným postupom vzhľadom k presnosti detekcie škodlivej vzorky. Presnosť detekcie je možné zvýšiť taktiež kombináciou viacerých modelov, čo je možné vidieť v prípade použitia algoritmu RF.

Z uvedených dôvodov boli pre náš program vybrané algoritmy využívajúce rozhodovacie stromy. V ďalšej časti tak budú opísané jednotlivé kroky strojového učenia s učiteľom, a to:

- 1) získanie dát a ich spracovanie do vhodnej podoby na trénovanie modelov strojového učenia,
- 2) trénovanie modelov na trénovacích dátach a hľadanie najlepších hyperparametrov modelu,
- 3) vyhodnotenie modelov na testovacích dátach a ich porovnanie.

Výsledkom týchto krokov budú natrénované modely schopné predikovať aj na doposiaľ neznámych vzorkách. Tie budú následne využité v programe na detekciu škodlivého softvéru.

A. Získanie vzoriek na trénovanie

Počet vzoriek určených na trénovanie a testovanie klasifikačných modelov obsahuje celkovo 3838 spustiteľných súborov - 3000 súborov je škodlivých a 838 tvoria neškodné súbory. Pôvod škodlivých vzoriek je repozitár *VirusShare* [VirusShare(2021)], ktorý obsahuje skutočne škodlivé vzorky určené pre výskumnú komunitu a podobne. Konkrétnym sťahovaným balíkom bol *VirusShare_00164.zip* s celkovým počtom 65 536 škodlivých vzoriek, z ktorého bolo vybraných 3000 vo formáte *exe*. Pôvod neškodných vzoriek sú repozitáre *PortableFreeware* [PortableFreeware(2021)] a *PortableApps* [PortableApps(2021)], z ktorých bolo stiahnutých celkovo 838 neškodných vzoriek vo formátoch *exe* a *msi*.

B. Statická analýza vzoriek

Statická analýza programu je vykonávaná bez spustenia jednotlivých vzoriek, a to programom *Dependency Walker*, ktorého výstupy sú uložené a následne spracovávané do podoby dátovej sady.

1) *Dependency Walker: Dependency Walker* [DependencyWalker(2022)] je voľne dostupný program vytvorený spoločnosťou Microsoft. Tento program je využívaný na analýzu spustiteľných súborov a umožňuje zobraziť detailné informácie o skenovanom súbore vrátane informácií o importovaných moduloch a funkciách [DependencyWalker(2022)].

2) *Spracovanie výstupu a vytvorenie dátovej sady*: Programom *Dependency Walker* bola vykonaná statická analýza všetkých vzoriek a výstupy jednotlivých analýz boli uložené v textových súboroch, do ktorých boli ukladané importované a exportované funkcie jednotlivých modulov, ako to zobrazuje obrázok 1. Použitými príznakmi je 112 vybraných funkcií najčastejšie importovaných v škodlivých programoch, ktorých názvy tak tvoria stĺpce dátovej sady. Počet importovaní každej funkcie bol zapísaný k odpovedajúcemu stĺpcu (s názvom danej funkcie) v dátovej sade. Ako posledný stĺpec dátovej sady bol doplnený stĺpec s názvom *malware*, ktorého hodnota v riadkoch bola nastavená na hodnotu 1 v prípade škodlivej vzorky a 0 v prípade neškodnej vzorky.

```
[ ^ ] API-MS-WIN-SECURITY-BASE-L1-1-0.DLL
```

Import	Ordinal	Hint	Function	Entry Point
[C]	N/A	0 (0x0000)	AccessCheck	0x0DCE9AF7
[C]	N/A	7 (0x0007)	AddAccessAllowedAce	0x0DCEE813
[C]	N/A	8 (0x0008)	AddAccessAllowedAceEx	0x0DCEAF14
[C]	N/A	11 (0x000B)	AddAccessDeniedAceEx	0x0DD16058
[C]	N/A	13 (0x000D)	AddAce	0x0DCEEDE7
[C]	N/A	19 (0x0013)	AdjustTokenPrivileges	0x0DCE9C5E
[C]	N/A	20 (0x0014)	AllocateAndInitializeSid	0x0DCEAAC1
[C]	N/A	24 (0x0018)	CheckTokenMembership	0x0DCEABA7
[C]	N/A	26 (0x001A)	CopySid	0x0DCE9A6F
[C]	N/A	31 (0x001F)	CreateWellKnownSid	0x0DCF089D
[C]	N/A	32 (0x0020)	DeleteAce	0x0DD02331
[C]	N/A	34 (0x0022)	DuplicateToken	0x0DCEA885
[C]	N/A	35 (0x0023)	DuplicateTokenEx	0x0DCEA812
[C]	N/A	38 (0x0026)	EqualSid	0x0DCE9A97
[C]	N/A	40 (0x0028)	FreeSid	0x0DCEA801
[C]	N/A	41 (0x0029)	GetAce	0x0DCE8A11
[C]	N/A	42 (0x002A)	GetAclInformation	0x0DCEED94

Fig. 1: Časť výstupu statickej analýzy

V niektorých prípadoch statická analýza programom *Dependency Walker* zlyhala kvôli nesprávnemu formátu testovaného

súboru a chýbajúcim modulom, a tak tieto výstupy analyzované neboli. Dátová sada na trénovanie a testovanie bola po spracovaní uložená vo formáte *csv* s celkovým počtom riadkov 3584. Zloženie dátovej sady zobrazuje tabuľka II.

TABLE II: Zloženie dátovej sady po statickej analýze

Trieda	Počet
škodlivé	2747
neškodné	837
spolu	3584

C. Dynamická analýza vzoriek

Dynamická analýza je vykonávaná spúšťaním a pozorovaním správania sa jednotlivých vzoriek v zabezpečenom virtuálnom prostredí, čo zabezpečuje nástroj *Cuckoo Sandbox*. Výsledok analýzy príslušnej vzorky je zaznamenaný do report súboru, z ktorého sú extrahované príznaky do dátovej sady. Príznakmi je počet volaní každej z 298 funkcií z Windows API počas vykonávania jednotlivých vzoriek. Vyhľadávané funkcie sú všetky funkcie, ktoré boli nájdené pri analýze škodlivých vzoriek.

1) *Cuckoo Sandbox: Cuckoo Sandbox* [Cuckoo(2022)] je voľne dostupný nástroj na analýzu spustiteľných ale aj iných súborov a web stránok v izolovanom prostredí. Po spustení analýzy prostredníctvom webového rozhrania alebo rozhrania REST API zbiera informácie počas behu, pričom zbieranými informáciami sú napríklad súbory vytvárané, odstraňované, sťahované počas vykonávania, práca s pamäťou, volané funkcie systému a sieťová aktivita a podobne. Zhromaždené informácie sú následne taktiež dostupné prostredníctvom grafického webového rozhrania alebo pomocou rozhrania REST API.

Nástroj obsahuje tzv. softvér centrálnej správy (z ang. *central management software*) [Cuckoo(2022)] riadiaceho spúšťanie a taktiež zbieranie výsledkov analýzy vzoriek. Tá je vykonávaná v izolovanom prostredí, ktoré môže byť virtuálnym ale aj fyzickým zariadením alebo viacerými zariadeniami.

2) *Testovacie prostredie pre dynamickú analýzu*: Pre vykonanie dynamickej analýzy je potrebné správne nakonfigurovať nástroj *Cuckoo Sandbox* a vytvoriť virtuálne zariadenia, na ktorých bude bežať dynamická analýza vzoriek. Testovacie prostredie je zobrazené na obrázku 2 a kľúčovými časťami sú:

- **hostiteľský systém** s operačným systémom *Ubuntu 18.04 LTS*. Dôvodom výberu tohto operačného systému je to, že obsahuje *Python* vo verzii 2.7, čo je podmienkou pre inštaláciu nástroja *Cuckoo Sandbox*. Tento systém obsahuje už spomenutý softvér centrálnej správy na riadenie dynamickej analýzy, ktorý je prístupný cez lokálny server *localhost* na porte 8090 pomocou REST API rozhrania.
- **hostovský systém** s operačným systémom Windows 7, v našom prípade 2 virtuálne zariadenia vytvorené nástrojom *VMCloak*, na ktorých je vykonávaná dynamická analýza vzoriek a ktorých čistý stav po inštalácii je uložený ako snímka prostredia. Po vykonaní analýzy je táto snímka prostredia automaticky obnovená, keďže beh škodlivej

vzorky mohol narušiť prostredie. Keďže operačný systém v základe obsahuje bezpečnostné mechanizmy, ktoré by mohli brániť vykonávať škodlivú aktivitu pri dynamickej analýze, bezpečnostné mechanizmy Windows Defender, Windows Firewall a Windows Update boli na oboch hosťovských systémoch vypnuté. Virtuálne zariadenia majú prístup k internetu pomocou tzv. *VirtualBox Host-only adaptéra*.

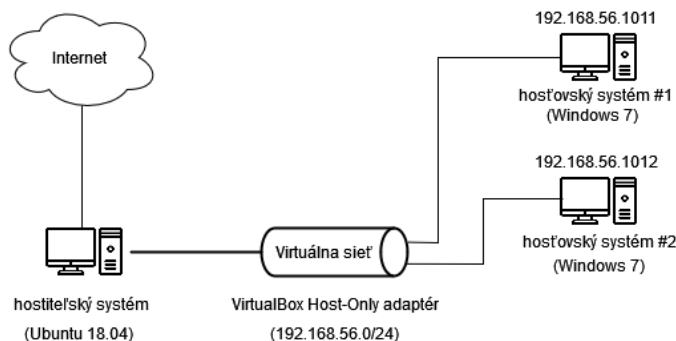


Fig. 2: Architektúra prostredia pre dynamicckú analýzu

3) *Získanie a spracovanie výstupu, vytvorenie dátovej sady*: Pre prácu s nástrojom Cuckoo Sandbox je využívaná Python knižnica *requests*, pomocou ktorej sú odosielané HTTP žiadosti na koncové body systému obsahujúcemu softvér centrálnej správy (hostiteľský systém), ktorý zhromažďuje výsledky analýzy. Využívané sú 4 metódy nami implementovanej triedy *CuckooAPI*, ktoré umožňujú pracovať s nástrojom Cuckoo Sandbox, a to:

- **metóda na odoslanie súboru na analýzu** (*submit_file*), ktorá vracia pridelené ID,
- **metóda na odoslanie URL adresy na analýzu** (*submit_url*), ktorá taktiež vracia pridelené ID,
- **metóda na zistenie stavu analýzy** (*get_status*), ktorá vracia stav analýzy s daným ID,
- **metóda na získanie výslednej správy o analýze** (*save_report*) na získanie výsledku analýzy vo formáte JSON.

Po získaní výsledku dynamickej analýzy vo formáte JSON je tento súbor spracovávaný skriptom v jazyku *Python* s cieľom nájdenia počtu volaní každej z 298 vybraných funkcií Windows API. Tie tak tvoria príznaky dátovej sady a ich názvy sú uložené v hlavičke dátovej sady. Názvy funkcií a počet ich volaní počas vykonávania programu obsahuje vo výstupnom JSON súbore objekt *behaviour* a v ňom vnorené objekty, ako to zobrazuje obrázok 3. V prípade nájdenia hľadanej funkcie, aktualizuje sa hodnota v príslušnom stĺpci dátovej sady počtom volaní danej funkcie, v opačnom prípade je uložená hodnota 0. Po dokončení hľadania počtu volaní všetkých funkcií je prvý stĺpec dátovej sady doplnený stĺpcem s názvom *malware*, ktorého hodnota v riadkoch bola nastavená na hodnotu 1 v prípade škodlivej vzorky a 0 v prípade neškodnej vzorky.

Podobne ako pri statickej analýze, objavili sa prípady, pri ktorých dynamiccká analýza kvôli nesprávne formátu

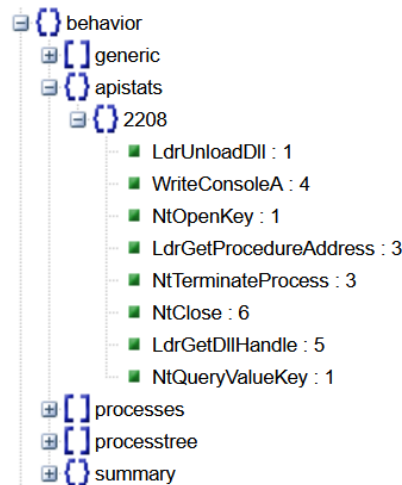


Fig. 3: Výstup dynamickej analýzy

analyzovanej vzorky programom *Cuckoo Sandbox* zlyhala, a tak tieto výstupy spracovávané neboli. Dátová sada na trénovanie a testovanie bola po spracovaní uložená vo formáte *csv* s celkovým počtom riadkov 3765. Zloženie dátovej sady zobrazuje tabuľka III.

TABLE III: Zloženie dátovej sady po dynamickej analýze

Trieda	Počet
škodlivé	2937
neškodné	828
spolu	3765

D. Trénovanie a testovanie klasifikačných modelov

V práci je analyzované použitie 2 algoritmov strojového učenia typu učenia s učiteľom. Obidva algoritmy sú založené na rozhodovacích stromoch, ktorých použitie sa zdá veľmi vhodné na základe výsledkov v už existujúcich prácach, kde najlepšie výsledky dosahovali algoritmy založené na rozhodovacích stromoch. Sú to algoritmy:

- **extrémne náhodné stromy** (z ang. *Extremely randomized trees*, v skratke *ET*), implementovaný pomocou triedy *ExtraTreesClassifier* z balíka *Scikit-learn*,
- **Extreme Gradient Boosting** (v skratke *XGBoost*), implementovaný pomocou triedy *XGBClassifier* z balíka *xgboost*.

Obidve dátové sady boli rozdelené metódou *train_test_split* z balíka *Scikit-learn* a 75% dát bolo určených na trénovanie a 25% na testovanie. Testovaných bolo niekoľko hodnôt hyperparametrov, zobrazených v tabuľke IV a V a pre každú kombináciu hyperparametrov bolo trénovanie a testovanie opakované 20 krát. Dosiahnuté výsledky boli zapisované do spoločného súboru vo formáte *csv*.

TABLE IV: Testované hyperparametre klasifikátora Extra-TreesClassifier

hyperparameter	hodnota
criterion	gini; entropy
n_estimators	10; 50; 100; 150; 200; 300
min_samples_split	2; 3; 4; 5
min_samples_leaf	1; 2; 3; 5
max_features	auto; sqrt; log2; None
class_weight	balanced; balanced_subsample; None
max_depth	3; 5; 8; None

TABLE V: Testované hyperparametre klasifikátora XGBClassifier

hyperparameter	hodnota
n_estimators	10; 50; 100; 150; 200; 250; 300
max_depth	3; 5; 7; 9; 11
learning_rate	0.001; 0.01; 0.1; 0.2; 0.3
colsample_bytree	0.3; 0.7; 1
subsample	0.5; 1
scale_pos_weight	0; 3
gamma	0; 1; 2; 3

E. Vyhodnotenie modelov

Jednotlivé výsledky z 20-ich opakovaní boli pre každú kombináciu hyperparametrov spriemerované a bola vypočítaná štandardná odchýlka, pričom boli zoradené od najlepšej po najhoršiu vzhľadom na metriku klasifikačnej presnosti. Doplnujúcimi uvedenými metrikami sú metriky špecifickosť a senzitivita.

1) *Metriky vyhodnotenia*: Použité metriky sú odvodené a možno ich vypočítať z tzv. konfúznej matice (z ang. *confusion matrix*) zobrazenej na obrázku 4, podľa ktorej je možné vhodným spôsobom zhodnotiť výsledok klasifikácie. Jej obsahom sú hodnoty [Hossin(2015)]:

- **TN** - počet správne predikovaných vzoriek negatívnej triedy (negatívna trieda - neškodné),
- **TP** - počet správne predikovaných vzoriek pozitívnej triedy (pozitívna trieda - škodlivé),
- **FP** - počet nesprávne predikovaných vzoriek negatívnej triedy predikovaných do pozitívnej triedy,
- **FN** - počet nesprávne predikovaných vzoriek pozitívnej triedy predikovaných do negatívnej triedy.

		skutočnosť	
		pozitívna	negatívna
predikcia	pozitívna	TP	FP
	negatívna	FN	TN

Fig. 4: Konfúzna matica

a) **Klasifikačná presnosť**: Klasifikačná presnosť vyjadruje koľko vzoriek z celkového počtu vzoriek bolo správne klasifikovaných, pričom je vyjadrená ako [Hossin(2015)]:

$$\text{presnosť} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

b) **Senzitivita**: Senzitivita vyjadruje schopnosť modelu správne klasifikovať pozitívne vzorky (v našom prípade škodlivé), teda vyjadruje koľko vzoriek bolo správne klasifikovaných ako pozitívne z celkového počtu pozitívnych vzoriek. Je vyjadrená ako [Hossin(2015)]:

$$\text{senzitivita} = \frac{TP}{TP + FN} \quad (2)$$

c) **Špecifickosť**: Špecifickosť vyjadruje schopnosť modelu správne klasifikovať negatívne vzorky (v našom neškodné), teda vyjadruje koľko vzoriek bolo správne klasifikovaných ako negatívne z celkového počtu negatívnych vzoriek. Je vyjadrená ako [Hossin(2015)]:

$$\text{špecifickosť} = \frac{TN}{TN + FP} \quad (3)$$

2) **Extrémne náhodné stromy**: Tabuľka VI zobrazuje 5 najlepších výsledkov, ktoré boli dosiahnuté použitím algoritmu ET na dátach statickej analýzy. Klasifikačná presnosť v najlepšom prípade dosahuje v priemere 91,786% so senzitivitou v priemere 97,176% a špecifickosťou v priemere 74,067%. O niečo lepšie výsledky boli dosiahnuté použitím tohto algoritmu na dátach dynamickej analýzy, ktoré zobrazuje tabuľka VII. Klasifikačná presnosť v najlepšom prípade dosahuje pri týchto dátach v priemere 96,387%, so senzitivou v priemere 98,165% a špecifickosťou v priemere 90,281%. V prípade dát statickej analýzy možno pozorovať nižšie hodnoty pri metrike špecifickosti, s čím súvisí väčšie množstvo falošne pozitívnych prípadov a tým aj nižšia klasifikačná presnosť.

TABLE VI: Najlepšie výsledky pri algoritme ET (statická analýza)

Parametre	Klasifikačná presnosť (%)	Senzitivita (%)	Špecifickosť (%)	ROC AUC
ET_S_1	91,786 ± 0,697	97,176 ± 0,890	74,067 ± 3,107	0,856 ± 0,014
ET_S_2	91,775 ± 0,623	97,322 ± 0,817	73,541 ± 3,326	0,854 ± 0,014
ET_S_3	91,774 ± 0,625	97,489 ± 0,845	72,990 ± 3,239	0,852 ± 0,014
ET_S_4	91,769 ± 0,650	97,445 ± 0,853	73,110 ± 3,295	0,853 ± 0,014
ET_S_5	91,769 ± 0,676	97,198 ± 0,838	73,923 ± 3,403	0,856 ± 0,015

TABLE VII: Najlepšie výsledky pri algoritme ET (dynamická analýza)

Parametre	Klasifikačná presnosť (%)	Senzitivita (%)	Špecifickosť (%)	ROC AUC
ET_D_1	96,387 ± 0,685	98,165 ± 0,497	90,281 ± 2,302	0,942 ± 0,012
ET_D_2	96,375 ± 0,530	98,039 ± 0,525	90,663 ± 1,753	0,944 ± 0,009
ET_D_3	96,375 ± 0,540	98,202 ± 0,568	90,102 ± 1,530	0,942 ± 0,008
ET_D_4	96,375 ± 0,705	98,016 ± 0,495	90,740 ± 2,229	0,944 ± 0,012
ET_D_5	96,375 ± 0,605	97,875 ± 0,510	91,224 ± 1,832	0,945 ± 0,010

3) *XGBoost*: Algoritmus *XGBoost* dosahoval o niečo lepšie výsledky aj v prípade dát statickej analýzy, aj v prípade dát dynamickej analýzy. V prípade statickej analýzy, zobrazenej v tabuľke VIII, dosahuje klasifikačná presnosť v najlepšom prípade v priemere 91,920%, senzitivita v priemere 98,253% a špecifickosť v priemere 71,100 %. V prípade dynamickej analýzy, zobrazenej v tabuľke IX, dosahuje klasifikačná presnosť v najlepšom prípade v priemere 96,467%, so senzitivitou v priemere 98,514% a špecifickosťou 89,439%.

TABLE VIII: Najlepšie výsledky pri algoritme *XGBoost* (statická analýza)

Parametre	Klasifikačná presnosť (%)	Senzitivita (%)	Špecifickosť (%)	ROC AUC
XGB_S_1	91,920 ± 0,658	98,253 ± 0,472	71,100 ± 3,051	0,847 ± 0,014
XGB_S_2	91,914 ± 0,617	98,508 ± 0,380	70,239 ± 2,762	0,844 ± 0,013
XGB_S_3	91,897 ± 0,609	98,231 ± 0,426	71,077 ± 2,980	0,847 ± 0,014
XGB_S_4	91,858 ± 0,635	98,464 ± 0,452	70,143 ± 3,048	0,843 ± 0,014
XGB_S_5	91,853 ± 0,717	97,584 ± 0,492	73,014 ± 3,338	0,853 ± 0,016

TABLE IX: Najlepšie výsledky pri algoritme *XGBoost* (dynamická analýza)

Parametre	Klasifikačná presnosť (%)	Senzitivita (%)	Špecifickosť (%)	ROC AUC
XGB_D_1	96,467 ± 0,560	98,514 ± 0,602	89,439 ± 2,484	0,940 ± 0,011
XGB_D_2	96,438 ± 0,522	98,670 ± 0,780	88,776 ± 2,499	0,937 ± 0,011
XGB_D_3	96,387 ± 0,517	98,789 ± 0,722	88,138 ± 2,466	0,935 ± 0,011
XGB_D_4	96,381 ± 0,521	98,782 ± 0,563	88,138 ± 2,482	0,935 ± 0,011
XGB_D_5	96,341 ± 0,529	98,841 ± 0,559	87,755 ± 2,382	0,933 ± 0,011

F. Výber najlepšieho modelu

Najlepšie dosiahnuté výsledky sme využitím niekoľkých metrik porovnali s našimi predchádzajúcimi prácami [Sutorcik(2021)][Spakovsky(2021)], v ktorých sme sa venovali porovnaniu algoritmov RF, DT, SVM a NB na rovnakej dátovej sade.

Tabuľka X zobrazuje porovnanie najlepších výsledkov pri jednotlivých algoritmoch implementovaných v tomto článku a našom predchádzajúcom článku, a to v prípade dátovej sady získanej statickou analýzou vzoriek programom *Dependency Walker*. Výsledky sú zoradené podľa metriky klasifikačná presnosť od najlepšieho po najhorší.

TABLE X: Porovnanie najlepších výsledkov pri jednotlivých algoritmoch (statická analýza)

Algoritmus	Klasifikačná presnosť (%)	Senzitivita (%)	Špecifickosť (%)	ROC AUC
XGB	91,92 ± 0,66	98,25 ± 0,47	71,10 ± 3,05	0,85 ± 0,01
ET	91,79 ± 0,70	97,18 ± 0,89	74,07 ± 3,11	0,86 ± 0,01
RF	91,32 ± 0,92	96,94 ± 0,60	72,82 ± 2,93	0,85 ± 0,02
DT	89,74 ± 0,98	95,76 ± 1,08	69,97 ± 2,56	0,83 ± 0,01
SVM	88,48 ± 0,78	95,46 ± 0,66	65,55 ± 2,54	0,81 ± 0,01
NB	42,27 ± 1,28	26,61 ± 1,66	93,72 ± 1,12	0,60 ± 0,01

Podobne aj tabuľka XI zobrazuje porovnanie najlepších výsledkov pri jednotlivých algoritmoch implementovaných v tomto článku a našom predchádzajúcom článku, no v prípade dátovej sady získanej dynamickou analýzou vzoriek nástrojom

Cuckoo Sandbox. Výsledky sú opäť zoradené podľa metriky klasifikačná presnosť od najlepšieho po najhorší.

TABLE XI: Porovnanie najlepších výsledkov pri jednotlivých algoritmoch (dynamická analýza)

Algoritmus	Klasifikačná presnosť (%)	Senzitivita (%)	Špecifickosť (%)	ROC AUC
XGB	96,47 ± 0,56	98,51 ± 0,60	89,44 ± 2,48	0,94 ± 0,01
ET	96,39 ± 0,69	98,17 ± 0,50	90,28 ± 2,30	0,94 ± 0,01
RF	95,95 ± 0,58	98,08 ± 0,49	88,66 ± 2,34	0,93 ± 0,01
DT	94,53 ± 0,74	96,37 ± 0,73	88,20 ± 2,53	0,92 ± 0,01
SVM	92,38 ± 0,83	97,82 ± 0,62	73,69 ± 2,88	0,86 ± 0,02
NB	59,53 ± 1,76	48,58 ± 2,42	97,14 ± 1,31	0,73 ± 0,01

Keďže našim hlavným cieľom pri navrhovanom programe je správne detegovať čo najväčšie množstvo škodlivých vzoriek, hlavnou porovnanou metriku je metrika senzitivity. Algoritmus *XGBoost* dosahuje v oboch prípadoch lepšie výsledky ako ostatné algoritmy. Modely s kombináciou hyperparametrov dosahujúcimi najlepšie výsledky (v tabuľke VIII označené ako *XGB_S_1* a v tabuľke IX ako *XGB_D_1*) tak sú exportované využitím knižnice *joblib*, uložené a pripravené na ďalšie použitie. Hodnoty hyperparametrov pri týchto modeloch zobrazuje tabuľka XII.

TABLE XII: Hodnoty testovaných hyperparametrov najlepších modelov

Model	n_estimators	max_depth	learning_rate	colsample_bytree	subsample	scale_pos_weight	gamma
XGB_S_1	300	11	0.01	0.7	1	0.3	1
XGB_D_1	300	11	0.01	0.3	1	0.3	0

G. Program a používateľské rozhranie

Pre použitie natrénovaných modelov v praxi na predikovanie nových vzoriek používateľom bol navrhnutý program, ktorý umožňuje skenovanie nových spustiteľných súborov využitím používateľského grafického rozhrania. Okrem toho ponúka aj ďalšiu základnú funkcionálnu, a to:

- skenovanie väčšieho množstva súborov z priečinka ako aj celého systému,
- umiestňovanie škodlivých spustiteľných súborov do karantény,
- plánované skenovanie podľa podmienok zadaných používateľom.

Doplňujúcou funkciou je taktiež skenovanie a vyhodnotenie škodlivosti webovej stránky využitím nástroja *Cuckoo Sandbox*, ktorý ponúka aj takúto funkcionálnu.

Na vytvorenie grafického rozhrania bola využitá Python knižnica *PyQt5*. Program je zložený z niekoľkých základných obrazoviek, ktoré sú doplnené doplňujúcimi modálnymi obrazovkami. Základné obrazovky programu sú:

- obrazovka s možnosťou výberu súboru, zadania URL adresy na skenovanie alebo skenovanie celého systému,
- obrazovka s prehľadom naplánovaných jednorazových alebo opakovaných skenovaní súborov a priečinkov,
- obrazovka s prehľadom už dokončených skenovaní,
- obrazovka pre zobrazenie súborov uložených v karanténe,
- obrazovka pre zobrazenie a zmenu nastavení programu.

Hlavnou obrazovkou programu je obrazovka s možnosťou výberu súboru, zadania URL adresy na skenovanie alebo skenovanie celého systému, zobrazená na obrázku 5, ktorá je spustená po zapnutí programu.

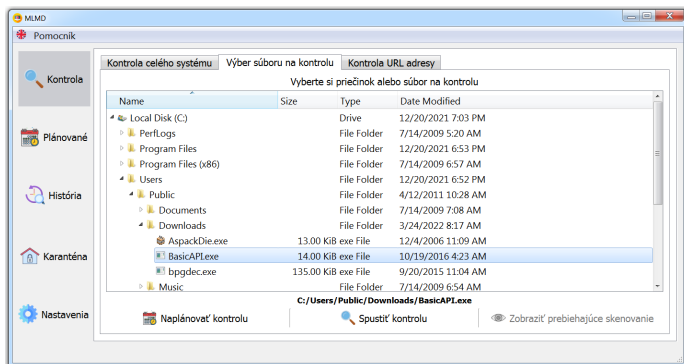


Fig. 5: Hlavná obrazovka programu

1) *Skenovanie súboru*: Po spustení skenovania testovacej vzorky nasleduje táto postupnosť krokov:

- 1) vykonanie statickej analýzy pomocou programu *Dependency Walker* a dynamickej analýzy pomocou programu *Cuckoo Sandbox*, výsledkom čoho sú 2 výstupy analýzy, z ktorých sú extrahované príznaky a vytvorená dátová sada (rovnakým spôsobom ako bolo opísané v prípade dát na tréningu),
- 2) načítanie 2 klasifikátorov strojového učenia natrénovaných na tréningových dátach, a to zvlášť pre dátovú sadu získanú statickou analýzou a zvlášť pre dátovú sadu získanú dynamickou analýzou,
- 3) vykonanie klasifikácie,
- 4) skombinovanie získaných výsledkov (hlasovanie) pre rozhodnutie o škodlivosti testovacej vzorky.

Postupnosť týchto krokov zobrazuje obrázok 6.

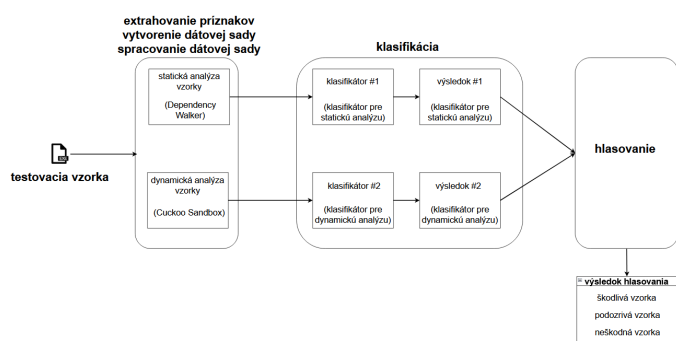


Fig. 6: Navrhnutá metóda detekcie

a) *Získanie výsledku o škodlivosti*: Klasifikátor priradí hodnotu 1 v prípade škodlivej vzorky a 0 v prípade neškodnej vzorky. Keďže systém detekcie tvoria 2 klasifikátori, pre určenie výsledku je nutné skombinovanie ich výsledkov, ako to zobrazuje tabuľka XIII.

TABLE XIII: Získanie konečného výsledku predikcie súboru

Výsledok (statická analýza)	Výsledok (dynamická analýza)	Konečný výsledok
škodlivá (1)	škodlivá (1)	škodlivá
škodlivá (1)	neškodná (0)	podozrivá
neškodná (0)	škodlivá (1)	podozrivá
neškodná (0)	neškodná (0)	neškodná

Po vyhodnotení škodlivosti vzorky je v prípade nájdenie škodlivej alebo podozrivej vzorky používateľovi vypísaný výsledok s možnosťou výberu akcie, ktorá sa má nad skenovaným súborom vykonať, ako to zobrazuje obrázok 7.

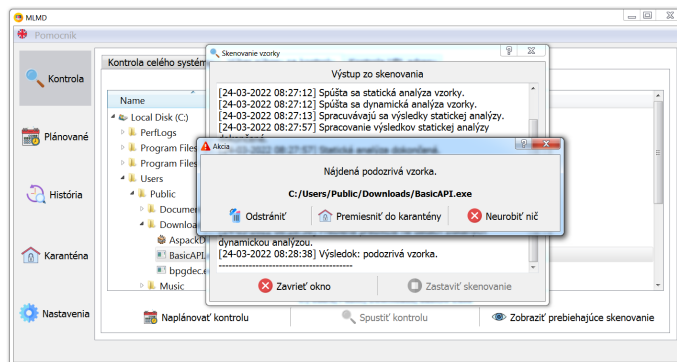


Fig. 7: Akcia nad skenovaným súborom

2) *Skenovanie webovej adresy*: Skenovanie webovej adresy je doplnujúcou funkciou programu, ktorá využíva a sprostredkováva funkcionality nástroja *Cuckoo Sandbox*. Ten okrem analýzy súborov rôzneho formátu umožňuje aj zadanie URL adresy, ktorá je následne navštívená v prehliadači a analyzovaná v izolovanom prostredí virtuálneho systému. Po ukončení analýzy je vrátený výsledok v podobe súboru vo formáte *JSON*, ktorý obsahuje výsledné skóre v rozmedzí 0-10 (objekt *info*) spolu s popisom nájdených hrozieb a závažnosťou (pole *signatures*), ako to zobrazuje obrázok 8. 0 znamená najmenšie nebezpečenstvo a 10 znamená najvyššie nebezpečenstvo.

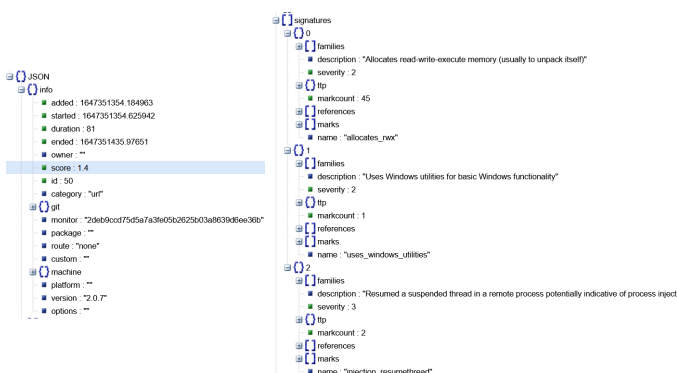


Fig. 8: Výstup po analýze URL adresy nástrojom Cuckoo Sandbox

Výsledné skóre spolu so slovným ohodnotením je vrátený používateľovi podľa tabuľky XIV.

TABLE XIV: Získanie konečného výsledku po ukončení URL adresy

Skóre škodlivosti (0 - min., 10 - max.)	Konečný výsledok
>= 0 a < 4	bezpečná
>= 4 a < 7	podozrivá
>= 7 a <= 10	veľmi podozrivá

3) *Karanténa*: V prípade detekcie podozrivého alebo potenciálne škodlivého súboru umožňuje množstvo antivírusových riešení umiestnenie tohto súboru do karantény s cieľom zabrániť jeho spusteniu a tým aj rozšíreniu do iných častí systému. Súbory umiestnené v karanténe nie sú zmazané, no sú presunuté do skrytého priečinka pred používateľmi a prípadne upravené tak, aby bolo zabránené ich spusteniu. Používateľ následne môže rozhodnúť o tom, či súbor v karanténe bude ponechaný, úplne odstránený z karantény a z celého systému alebo obnovený na pôvodné umiestnenie v systéme, a to napríklad využitím ďalších informácií o tomto súbore.

Náš program využíva priečinok *Quarantine*, ktorý je umiestnený v adresári projektu a ktorý slúži na uchovanie premiestnených súborov programom alebo používateľom. Súbory v tomto priečinku sú kódované využitím *base64* kódovania a ukladané do formátu *.b64*. Umiestnenie súboru do karantény opisujú nasledujúce kroky (podobným spôsobom je zabezpečené aj obnovenie súboru z karantény - dekódovaním):

- 1) otvorenie pôvodného súboru v binárnom formáte na čítanie (mód *read binary*),
- 2) otvorenie nového súboru na zápis (mód *write binary*) v priečinku *Quarantine*,
- 3) čítanie pôvodného súboru riadok po riadku a transformácia (kódovanie) pomocou knižnice *base64io*,
- 4) zápis transformovaného riadku do nového súboru,
- 5) zmazanie pôvodného súboru.

Program obsahuje obrazovku, ktorá poskytuje prehľad súborov v karanténe a poskytuje používateľovi možnosť obnovenia na pôvodné umiestnenie a trvalého odstránenia z karantény, ako možno vidieť na obrázku 9.

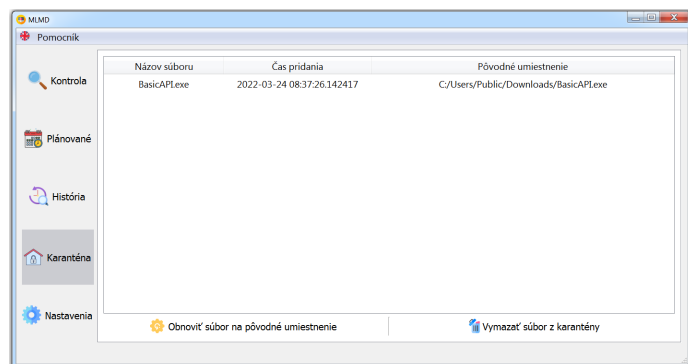


Fig. 9: Obrazovka karantény

IV. VYHODNOTENIE

Pre vyhodnotenie úspešnosti sme nami dosiahnuté výsledky porovnali:

- s existujúcimi prácami opísanými na začiatku článku,
- s voľne dostupným nástrojom *VirusTotal*.

A. Porovnanie s existujúcimi prácami

Porovnanie s existujúcimi prácami opísanými na začiatku článku zobrazuje tabuľka XV, ktorá zobrazuje hodnoty presnosti a senzitivity v najlepších prípadoch, a to pri použití uvedených algoritmov. Udávané hodnoty presnosti a senzitivity pri našej práci sú vyjadrené aritmetickým priemerom bez udávania štandardnej odchýlky a boli získané pri testovaní počas jednotlivých iterácií a hľadani najlepších parametrov na dátovej sade o veľkosti 25% pôvodnej dátovej sady.

TABLE XV: Porovnanie s existujúcimi prácami

práca	typ analýzy	veľkosť dátovej sady (škodlivé / vše)	najlepší algoritmus	presnosť (%)	senzitivita (%)
Schultz a kol.	statická	3 265 / 1 001	RF	97,11	97,43
Bai a kol.	statická	10 521 / 8 992	148	95,1-99,1	91,3-99,4
Kumar a kol.	statická	2 722 / 2 488	RF	98,78	99,9
Bragen	statická	992 / 771	RF	95,58	96,77
Chowdhury a kol.	statická	41 265 / 10 920	ANN	97,7	91
Shijo a Salim	statická	997 / 490	SVM	95,88	95,9
naša práca	statická	2747 / 837	XGBoost	91,92	98,25
naša práca	dynamická	2897 / 838	XGBoost	96,48	98,81
Shijo a Salim	dynamická	997 / 490	SVM	97,16	97,2
Shijo a Salim	kombinovaná	997 / 490	SVM	98,71	98,7
Pridant a kol.	dynamická	220 / 250	148	96,8	95,9
Moshi a kol.	dynamická	3 130 / 1 157	RF	91,4	91,1
Kumar Goetha	Embed dataset	300K / 300K	Gaussian NB, KNN, Linear SVC, DT, AdaBoost, RF, Extra Trees, GB, XGBoost	96,5	0,99-0,99
Dhamija, Dhamija	Open data source	4060 / 2709	NB, DT, RF, GB, XGBoost	99,95	99,95
Shhadat et al.	Open data source	984 / 172	KNN, SVM, Bernoulli NB, RF, DT, LR, HV	98,2	92

Dobré výsledky sme dosiahli najmä pri metrike senzitivity a sme tak schopní dobre klasifikovať škodlivé vzorky so zachovaním nízkej miery falošne negatívnych vzoriek. V prípade dátovej sady získanej statickou analýzou vzoriek mierne zaostávame pri metrike klasifikačnej presnosti, s čím súvisí hlavne počet falošne pozitívnych prípadov, čo pripisujeme nevyváženosti dát. V prípade dátovej sady získanej dynamickou analýzou vzoriek sú výsledky o niečo lepšie ako v prípade statickej analýzy, čoho dôvodom môže byť lepší výber príznakov na rozlíšenie medzi škodlivou a neškodnou vzorkou v prípade dát získaných dynamickou analýzou.

B. Porovnanie programu MLMD s nástrojom VirusTotal

VirusTotal je jedným z najpopulárnejších online nástrojov na skenovanie podozrivých súborov a web stránok. Je to akýsi agregátor informácií, ktorého výstup je kombinovaným výstupom rôznych antivírusových produktov, pričom na skenovanie ich v súčasnosti využíva viac ako 70. Nasledujúce porovnanie je porovnaním výstupov nášho programu *MLMD* s výstupmi nástroja *VirusTotal*.

1) *Testovacie vzorky*: Testovaciu sadu tvorilo 105 vzoriek, pričom ich zloženie z pohľadu škodlivosti možno vidieť v tabuľke XVI.

TABLE XVI: Zloženie testovacích vzoriek

Trieda	Počet
škodlivé	70
neškodné	35
spolu	105

Škodlivé vzorky tvorili spustiteľné súbory, ktoré boli prebraté z repozitára škodlivých vzoriek *VirusShare*, a to z balíka *VirusShare_00164.zip*. Ten bol využitý aj na získanie našich tréningových dát. Keďže spomenutý balík obsahuje celkovo 65 536 škodlivých vzoriek a na tréningovanie bolo využitých prvých 3 000 vzoriek, vybrali sme vzorky také, ktoré neboli využité pri tréningu našich modelov strojového učenia. Neškodné vzorky tvorili taktiež spustiteľné súbory, ktoré boli vybrané náhodne z voľne dostupnej stránky *PortableApps*.

Testovacie vzorky boli postupne otestované našim programom a taktiež nástrojom *VirusTotal*. Po otestovaní každej vzorky našim programom bol pre každú vzorku zaznamenaný jeden z nasledujúcich výsledkov:

- **škodlivá** - obidva klasifikačné modely predikovali vzorku ako škodlivú,
- **podozrivá** - jeden klasifikačný model predikoval vzorku ako škodlivú, druhý ako neškodnú,
- **neškodná** - obidva klasifikačné modely predikovali vzorku ako neškodnú.

V prípade nástroja *VirusTotal* bolo pre každú vzorku zaznamenané skóre pozostávajúce z nasledujúcich údajov:

- **počet systémov, ktoré detegovali vzorku ako škodlivú,**
- **počet systémov, ktoré detegovali vzorku ako neškodnú.**

2) *Porovnanie predikcií:* Tabuľka XVII a grafy zobrazené na obrázku 10 zobrazujú percentuálne rozloženie predikcií nášho programu a systému *VirusTotal* na škodlivých vzorkách.

Ako možno vidieť, náš program *MLMD* z celkového počtu vzoriek predikoval 94,29% vzoriek ako škodlivé, čo bolo správnou predikciou. Zvyšných 5,71% tvorili predikcie vzoriek ako podozrivé, čo taktiež možno v prípade škodlivých vzoriek považovať za čiastočne správnou odpoveď.

TABLE XVII: Porovnanie predikcií pri škodlivých vzorkách

	MLMD	VirusTotal
celkový počet vzoriek	70	70
škodlivé vzorky (%)	94,29	71,17
podozrivé vzorky (%)	5,71	-
neškodné vzorky (%)	0	26,83

Pri systéme *VirusTotal* celkovo 71,17% systémov označilo vzorky ako škodlivé. Množstve systémov však nesprávne označilo vzorku ako neškodnú, čo tvorilo celkovo 26,83% prípadov.

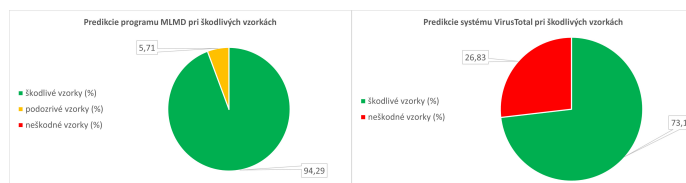


Fig. 10: Porovnanie predikcií pri škodlivých vzorkách

Tabuľka XVIII a grafy zobrazené na obrázku 11 zobrazujú percentuálne rozloženie predikcií nášho programu *MLMD* a systému *VirusTotal* na neškodných vzorkách.

TABLE XVIII: Porovnanie predikcií pri neškodných vzorkách

	MLMD	VirusTotal
celkový počet vzoriek	35	35
škodlivé vzorky (%)	0	1,94
podozrivé vzorky (%)	17,14	-
neškodné vzorky (%)	82,86	98,06

V prípade neškodných vzoriek možno za správnou predikciu považovať len predikciu vzorky ako neškodnej. Náš program *MLMD* z celkového počtu vzoriek predikoval 82,86% vzoriek ako neškodné, čo bolo správnou predikciou. Nesprávnu predikciu vykonal pri 17,14% vzoriek, pri ktorých vzorky predikoval ako podozrivé, a teda 1 z klasifikačných modelov vykonal nesprávnu predikciu. Ani v jednom prípade však program neoznačil vzorku za škodlivú.



Fig. 11: Porovnanie predikcií pri škodlivých vzorkách

Lepšie výsledky v prípade neškodných vzoriek dosahoval systém *VirusTotal*, respektíve systémy, ktoré využíva. V prípade neškodných vzoriek celkovo 98,06% systémov správne označilo vzorky za neškodné a len 1,94% ich nesprávne označilo ako škodlivé.

Tabuľka XIX zobrazuje počet vzoriek s nesprávnou predikciou pri škodlivých vzorkách. Za hranicu úspešnosti sme si definovali hodnotu 35, čo je približne polovica dostupných detekčných systémov v nástroji *VirusTotal*. Ide teda o počet vzoriek, pri ktorých menej ako 35 systémov správne označilo vzorku ako škodlivú a väčšia časť ju označila nesprávne ako neškodnú. Takýchto vzoriek bolo celkovo 8. Náš program však pri týchto vzorkách nesprávne predikoval len 2 vzorky, ktoré označil len ako podozrivé a nie ako škodlivé.

TABLE XIX: Počet vzoriek s nesprávnou predikciou pri škodlivých vzorkách

Nástroj	Počet
VirusTotal	8
MLMD	2

Podobné porovnanie možno vidieť v tabuľke XX, ktoré však zobrazuje počty pri neškodných vzorkách. Podobne ako v predchádzajúcom prípade, hranicu úspešnosti sme si definovali na hodnotu 35. Tabuľka tak zobrazuje počet vzoriek, pri ktorých menej ako 35 systémov správne označilo vzorku ako neškodnú a väčšia časť ju označila nesprávne ako škodlivú. Takéto vzorky však pri systéme *VirusTotal* neboli žiadne a pri väčšine vzoriek väčšina systémov správne detegovala vzorky za neškodné. Náš program označil 6 vzoriek ako podozrivé, čo bolo nesprávnou predikciou.

TABLE XX: Počet vzoriek s nesprávnou predikciou pri neškodných vzorkách

Nástroj	Počet
VirusTotal	0
MLMD	6

Vo všetkých uvedených porovnaníach možno pozorovať dobrú úspešnosť algoritmu *XGBoost*, a to najmä pri predikovaní vybranej skupiny škodlivých vzoriek, čo potvrdzuje úspešnosť na testovacích dátach. Vyzdvihnúť je však nutné všetky algoritmy využívajúce rozhodovacie stromy, keďže práve tieto algoritmy sa ukázali ako najlepšie pri probléme detekcie škodlivého softvéru a s náskokom prekonalí iné algoritmy.

V. ZÁVER

V tomto článku sme sa zaoberali detekciou škodlivého softvéru pomocou algoritmov strojového učenia s učiteľom. Modely strojového učenia boli natréňované na reálnych škodlivých a neškodných vzorkách, na ktorých bola vykonaná statická analýza využitím programu *Dependency Walker* a dynamická analýza využitím programu *Cuckoo Sandbox*. Statické príznaky tvorili počty importovaní vybraných funkcií a dynamické príznaky tvorili počty volaní funkcií z knižnice Windows API. Porovnávali sme úspešnosti dvoch algoritmov strojového učenia, ktoré využívajú rozhodovacie stromy - algoritmus ET a *XGBoost*, pri ktorých sme hľadali aj najlepšiu kombináciu hyperparametrov. Najlepšie výsledky boli dosiahnuté využitím algoritmu *XGBoost* aj v prípade dátovej sady po statickej analýze, aj v prípade dátovej sady po dynamickej analýze, a tak boli natréňované modely uložené na ďalšie použitie v programe MLMD.

Program MLMD bol navrhnutý tak, aby umožnil využiť natréňované modely aj na nových vzorkách, a to využitím grafického rozhrania vo forme desktopovej aplikácie. Okrem tejto funkcionality ponúka aj základnú funkcionality iných bezpečnostných programov - karanténu súborov a plánované spustenie kontroly súborov. Pri vyhodnotení sme úspešnosť získanú počas tréňovania modelov strojového učenia porovnali s opísanými podobnými prácami venujúcimi sa tejto problematike a taktiež sme program porovnali s nástrojom *VirusTotal* využitím testovacích vzoriek. Dosiahnuté výsledky naznačujú výbornú úspešnosť pri detekcii škodlivých vzoriek, pri neškodných vzorkách sa však vyskytli prípady falošne pozitívnych vzoriek.

REFERENCES

[Monnappa(2018)] Monnappa, K.; *Learning Malware Analysis*, 1st ed.; Publisher: Packt Publishing, 2018; Chapter 1, ISBN 978-178-839-250-1.

[Malwarebytes(2020)] 2020 State of Malware Report. Available online: https://www.malwarebytes.com/resources/files/2020/02/2020_state-of-malware-report.pdf (accessed on 28 March 2022).

[Elisan(2018)] Elisan, Ch.; *Malware, Rootkits & Botnets A Beginner's Guide*, 1st ed.; Publisher: McGraw-Hill Education, 2012; Chapter 1, ISBN 978-007-179-206-6.

[Lawrynowicz(2014)] Lawrynowicz, A.; Tresp, V. Introducing Machine Learning. *Perspectives on Ontology Learning* **2014**, pp. 35–50.

[Mohanta(2020)] Mohanta, A.; Saldanha, A. *Malware Analysis and Detection Engineering: A Comprehensive Approach to Detect and Analyze Modern Malware*, 1st ed.; Apress, 2020; ISBN 978-148-426-192-7.

[Fedak(2020)] Fedak, A.; Stulrajter, J. Fundamentals of static malware analysis: principles, methods, and tools. *Science & Military* **2014**, *15*, pp. 45–53.

[Hisham(2015)] Hisham, S.G. Behavior-based features model for malware detection. *Journal of Computer Virology and Hacking Techniques* **2015**, *12*, pp. 59–67.

[Damodaran(2020)] Damodaran, A.; Troia, F.D.; Visaggio, C.A. et al. A comparison of static, dynamic, and hybrid analysis for malware detection. *Journal of Computer Virology and Hacking Techniques* **2017**, *13*, pp. 1–12.

[Cisar(2020)] Cisar, P.; Joksimovic, D. Heuristic scanning and sandbox approach in malware detection. *Thematic Conference proceedings of international significance* **2020**, *9*.

[Seker(2021)] Advanced Heuristics to Detect Zero-Day Attacks. Available online: <https://hackernoon.com/advanced-heuristics-to-detect-zero-day-attacks-8e3335lt> (accessed on 28 March 2022).

[Schultz(2001)] Schultz, G.M.; Eskin, E.; Zadok, F.; Stolfo, J.S. Data Mining Methods for Detection of New Malicious Executables. In Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, 1 February 2001; pp. 38–49.

[Bai(2014)] Bai, J.; Wang, J.; Guozhong, Z. A Malware Detection Scheme Based on Mining Format Information. *The Scientific World Journal* **2014**.

[Kumar(2019)] Kumar, A.; Kuppasamy, K.S.; Aghila, G. A learning model to detect maliciousness of portable executable using integrated feature set. *Journal of King Saud University - Computer and Information Sciences* **2019**, *31*, pp. 252–265.

[Bragen(2015)] Bragen, R.S. Malware detection through opcode sequence analysis using machine learning. Master thesis, Gjøvik University College, Gjøvik, 2015.

[Moser(2007)] Moser, A.; Kruegel, C.; Kirida, E. Limits of Static Analysis for Malware Detection. In Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007), 2007; pp. 421–430.

[Chowdhury(2018)] Chowdhury, M.; Rahman, A.; Islam, M. Protecting data from malware threats using machine learning technique. In 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2017; pp. 1691-1694.

[Shijo(2015)] Shijo, P.V.; Salim, A. Integrated Static and Dynamic Analysis for Malware Detection. *Procedia Computer Science* **2015**, *46*, pp. 804–811.

[Firdausi(2010)] Firdausi, I.; Lim, C.; Erwin, A.; Nugroho, A.S. Analysis of machine learning techniques used in behavior-based malware detection. In 2010 Second International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2010; pp. 201-203.

[Mosli(2017)] Mosli, R.; Yuan, B.; Li, R.; Pan, Y. A Behavior-Based Approach for Malware Detection. *13th IFIP International Conference on Digital Forensics (DigitalForensics)* **2017**, pp.187–201

[VirusShare(2021)] VirusShare malware repository. Available online: <https://virusshare.com/> (accessed on 29 March 2022).

[PortableFreeware(2021)] The Portable freeware collection. Available online: <https://www.portablefreeware.com/> (accessed on 29 March 2022).

[PortableApps(2021)] Portable software repository. Available online: <https://portableapps.com/> (accessed on 29 March 2022).

[DependencyWalker(2022)] Dependency Walker website. Available online: <https://www.dependencywalker.com/> (accessed on 29 March 2022).

[Cuckoo(2022)] Cuckoo Sandbox website. Available online: <https://cuckoosandbox.org/> (accessed on 29 March 2022).

[Hossin(2015)] Hossin, M.; Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process* **2015**, *5(2)*, pp. 1–11.

[Sutorčík(2021)] Sutorčík, K. Využitie algoritmov strojového učenia na detekciu malvérových vzoriek pomocou metód dynamickej analýzy. Master thesis, Technická univerzita v Košiciach, Košice, 2021.

[Spakovský(2021)] Špakovský, E. Využitie algoritmov strojového učenia na detekciu malvérových vzoriek pomocou metód statickej analýzy. Master thesis, Technická univerzita v Košiciach, Košice, 2021.

Aplikácia Rozličných Foriem Strojového Učenia pre Pohyblivú RAN Sieť

Marek Ružička

Fakulta Elektrotechniky a Informatiky
Katedra Počítačov a Informatiky
Technická univerzita v Košiciach
Košice, Slovensko
marek.ruzicka@tuke.sk

Juraj Gazda

Fakulta Elektrotechniky a Informatiky
Katedra Počítačov a Informatiky
Technická univerzita v Košiciach
Košice, Slovensko
juraj.gazda@tuke.sk

Abstrakt—V článku rozoberáme existujúce riešenia predikcie pohybu koncových užívateľov a optimálneho rozmiestňovania pohyblivých bázových staníc pomocou viacerých algoritmov. Dôležitým aspektom je rovnako šetrenie energie drona, ktorá je obmedzená. Nosným prínosom je návrh spôsobu rozšírenia existujúceho algoritmu pre maximalizáciu súčtovej kapacity koncových užívateľov pokrytých pomocou bázovej stanice umiestnenej na drone. Rozšírenie berie do úvahy vplyv vetra na spotrebu energie a značne znižuje energetickú náročnosť systému. Ďalej riešime možnosť optimalizácie rozmiestnenia koncových bázových staníc na základe rozmiestnenia užívateľov v priestore. Ponúkame prístup, ktorý umožní niekoľkonásobné časové vylepšenie existujúceho algoritmu. V závere ponúkneme ďalšie možnosti smerovania výskumu v danej oblasti.

Abstract—In the article, we discuss existing solutions for predicting the movement of end users and optimal placement of mobile base stations using multiple algorithms. An important aspect is also conserving the drone's energy, which is limited. The main contribution is the proposed method for extending the existing algorithm to maximize the total capacity of end users covered by a base station located on a drone. The extension takes into account the impact of wind on energy consumption and significantly reduces the energy demand of the system. We further address the possibility of optimizing the placement of end base stations based on the location of users in space. We offer an approach that allows for multiple time improvements to the existing algorithm. In conclusion, we offer further possibilities for directing research in this area.

I. ÚVOD

Bezdrôtová komunikácia sa stala pilierom modernej doby. Podľa výročnej správy Cisco o internete [1], sa medzi rokmi 2018 a 2023 očakával nárast počtu mobilných zariadení o 1,5-2 krát. V závislosti od kontinentu je 25-50% zariadení pripojených k sieti RAN. V strednej Európe to znamená 1 miliardu zariadení pripojených k sieti RAN, zatiaľ čo v roku 2018 to bolo 0,6 miliardy. Tento masívny nárast pripojení povedie k obrovskému dátovému prenosu.

Stará infraštruktúra 4G sietí sa postupne nahrádza novou 5G sieťou. Avšak, 5G výrazne spolieha na malé bunky fungujúce na vysokých frekvenciách [2]. Signál vysokých frekvencií je náchylnejší na problémy spojené s priamou linkou (LOS) a má nižší dosah; zatiaľ čo signál v 4G môže dosahovať

stovky metrov až kilometrov, v prípade 5G malých buniek je to zvyčajne desiatky až stovky metrov. Mimo budov sa ľudia zvyčajne pohybujú a v prípade, že sú v kanceláriách alebo domovoch, zvyčajne sa pripájajú k Wi-Fi sieťam. Avšak, ľudské zvyky sa môžu rýchlo meniť. Správa o trendoch mobility Google ukazuje, že uzatvorenie a obmedzenia na celom svete spôsobili pokles mobility v mestských oblastiach o 20-30%, zatiaľ čo mobilita v obytných lokalitách sa zvýšila o 15-20% a v parkoch dosiahla mobilita na svojom vrchole 12. augusta 2020 rast o 109% [3]. Takéto udalosti naznačujú, že v niektorých situáciách môže byť dopyt po lepšom pripojení, ktoré nemôže pokryť aktuálne používaná statická infraštruktúra. Dynamická sieťová architektúra sa javí ako dobré riešenie.

Možným riešením pre scenáre, kedy sa koncové zariadenie (UE) presúva na územie so zlým pokrytím signálu bez statickej infraštruktúry, je využitie dynamickej siete pohyblivých bázových staníc (BS) - buď na zemi alebo vo forme bezpilotných lietadiel (UAV), často označovaných ako drony. Avšak v takomto scenári je potrebné efektívne algoritmy pre presun UAV na najoptimálnejšiu polohu. Nevyhnutné vlastnosti, na ktoré musíme dávať pozor, sú životnosť batérie UAV, počet UAV potrebných na pokrytie oblasti, spokojnosť koncových užívateľov a šírka pásma alebo kapacita.

II. PREHĽAD PROBLEMATIKY

V prípade UAV bázových staníc, ktoré vzhľadom na jednoduchosť a flexibilitu pri nasadení uvažujeme v tomto článku, poznáme nasledujúce typy:

- **Multirotory:** Toto sú UAV, ktoré majú viacero rotorov poskytujúcich vztlak a ťah. Môžu lietať na mieste, pohybovať sa v akomkoľvek smere a ľahko meniť výšku. Sú vhodné pre mestské prostredie a komunikáciu na krátke vzdialenosti. Často sa označujú ako UAV s rotujúcimi krídlami. Viac informácií možno nájsť napríklad v [4].
- **S pevnými krídlami:** Toto sú UAV, ktoré majú krídla ako konvenčné lietadlá. Môžu lietať rýchlejšie a dlhšie ako multirotory, ale potrebujú viac priestoru na vzlet a pristátie. Sú vhodné pre vidiecke prostredie a komunikáciu na dlhé vzdialenosti [5].

- Balóny: Používajú hélium alebo horúci vzduch na vznášanie. Môžu zostávať vo vzduchu dlhé obdobia a pokrývať veľké oblasti. Sú vhodné pre núdzové situácie a pomoc pri katastrofách [6].
- Vzducholode: Používajú pevnú alebo polopevnú štruktúru naplnenú plynom na vznášanie sa. Môžu niesť ťažké náklady a fungovať vo veľkých výškach. Sú vhodné na pokrytie širokých oblastí a konštantné sledovanie [7].

V našej práci budeme pre aplikáciu v mestskom prostredí uvažovať o multirotorových UAV, ktoré sú na túto aplikáciu najvhodnejšie. Bázové stanice založené na UAV s rotujúcimi krídlami čelia niekoľkým problémom, ako sú i) krátke časy letu, ii) optimalizácia pozície UAV a plánovanie optimálnej trasy, iii) výkonnosť komunikácie a iiiii) počet potrebných uzlov.

Optimalizácia komunikačnej výkonnosti bázových staníc založených na UAV s rotujúcimi krídlami je stále výzvou. Skoršie štúdie o umiestnení UAV a optimalizácii trajektórie zvažovali najmä model kanála založený na LoS v rôznych komunikačných systémoch umožňujúcich UAV, a to na základe skutočnosti, že A2G (vzduch-zem) linky majú oproti pozemným linkám vysokú pravdepodobnosť LoS [8]. UAV môže lietať blízko každého pozemného uzlu, aby sa znížila vzdialenosť linky.

Jedným z najväčších problémov UAV, ktoré sa efektívne využívajú v mnohých oblastiach, sú nepochybne časy letu. UAV s rotujúcimi krídlami sa môžu vznášať nad konkrétnym miestom pri nízkej rýchlosti, avšak obvykle je doba letu menej ako 1 hodina. UAV s pevnými krídlami môžu lietať niekoľko hodín, ale aby zostali vo vzduchu, musia sa pohybovať vpred.

Plánovanie trasy je jedným z najdôležitejších krokov pri navigácii a kontrole UAV. Zaisťuje optimálnu cestu bez kolízií medzi dvoma miestami od východiskového bodu k cieľu pre autonómne UAV a zároveň splňuje požiadavky týkajúce sa charakteristík UAV a obsluhovanej oblasti. Schopnosť autonómneho riadenia letu a vyhýbania sa prekážkam je predpokladom pre zabezpečenie bezpečného letu UAV. Rovnako dôležité je aj umiestnenie UAV, aby sa zaisťovalo čo najlepšie pokrytie s čo najmenším počtom potrebných UAV, čo ďalej znižuje náklady na sieť.

V tomto článku sa budeme venovať optimalizácii dvoch parametrov - energie potrebnej pre let drona a umiestneniu dronov v priestore za účelom kvalitného pokrytia koncových užívateľov.

1) *Optimalizácia založená na signálnych charakteristikách viazaných na pozíciu drona:* S rozšíreným využívaním bezpečných lietadiel (UAV) v sieťach 5G a 6G sa zvyšuje potreba optimalizovať ich pozíciu, aby sa dosiahlo lepšie pokrytie a vyššie dátové rýchlosti. Optimalizáciou pozície UAV je možné vylepšiť výkon siete a znížiť interferenciu medzi rôznymi zariadeniami. To môže viesť k efektívnejšiemu využitiu sieťových zdrojov a lepšiemu celkovému používateľskému zážitku. Pri optimalizácii pozícií UAV sa môžu využiť tri hlavné charakteristiky signálu: priepustnosť, pokrytie a kapacita. Priepustnosť sa týka množstva dát, ktoré je možné preniesť v danom období. Pokrytie sa týka oblasti siete, ktorú je možné dosiahnuť daným

zariadením. Kapacita sa vzťahuje na maximálne množstvo dát, ktoré je možné preniesť za dané obdobie. Optimalizácia pozície pomáha zlepšovať výkon siete a redukovať interferenciu.

Z hľadiska riešení umiestňovania delíme riešenia na analytické a na riešenia s použitím strojového učenia. Zatiaľ čo analytické riešenia využívajú exaktný matematický prístup, strojové učenie sa učí ako drony rozmiestňovať. Preto analytické riešenia majú zvyčajne vyššiu presnosť, no za cenu vyššej časovej náročnosti.

Pokrytie koncových zariadení pomocou dronov sa vo všeobecnosti považuje za problém k -centier. Problém k -centier je klasickým optimalizačným problémom v informatike a operačnom výskume. Vzhľadom na množinu bodov v metrickom priestore je cieľom problému k -centier nájsť k centier, ktoré minimalizujú maximálnu vzdialenosť medzi akýmkoľvek klientom a jeho najbližším centrom. Matematicky je definícia nasledovná: "Vzhľadom na množinu n bodov $P \in \mathbb{R}^d$ a kladné celé číslo k je potrebné pokryť P k kongruentnými guľkami čo najmenšej možnej veľkosti. d tu označuje počet rozmerov." [9]. Skutočné aplikácie zahŕňajú zhlukovanie dát pre ťažbu dát, segmentáciu alebo vyhľadávanie informácií a problémy s umiestnením zariadení (napr. umiestnenie skladov a tovární na minimalizáciu času cestovania medzi miestami).

Keďže problém k -centier je podproblémom k -zhlukovania, najjednoduchšie riešenie tohto problému je použitie algoritmu k -means [10] na nájdenie optimálnych centier zhlukov prítomných v vybranej oblasti. Algoritmus k -means je definovaný ako: daná je množina n pozorovaní $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ d -rozmerných reálnych vektorov, k -means zhlukovanie rozdelí n pozorovaní do $k \leq n$ zhlukov $\mathbf{S} = S_1, S_2, \dots, S_k$ s cieľom minimalizovať súčet štvorcov v rámci zhlukov, ako je napríklad variancia. Cieľová funkcia k -means zhlukovania je formálne definovaná ako:

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_{\mathbf{S}} \sum_{i=1}^k |S_i|^2 \text{Var } S_i. \quad (1)$$

μ_i je priemer bodov v zhluku S_i , zatiaľ čo Var reprezentuje varianciu. Algoritmus používa techniku iteratívneho vylepšovania a pracuje v dvoch krokoch. V priradzovacom kroku je každému pozorovaniu priradený najbližší priemer z náhodne inicializovanej sady k priemerov $m_1^{(1)}, m_2^{(1)}, \dots, m_k^{(1)}$ podľa najnižšej euklidovskej vzdialenosti medzi pozorovaním a každým z priemerov. Druhý krok sa nazýva aktualizácia, kde sa priemery aktualizujú na nové hodnoty podľa priemeru medzi priradenými pozorovaniami. Proces sa opakuje, kým sa fáza priradenia už nezmení. Algoritmus je považovaný za NP-ťažký a ak je počet k centier a d dimenzií pevný, komplexnosť je $\mathcal{O}(n_{dk+1})$ [11]. K-means je veľmi jednoduchý algoritmus z hľadiska riešenia, ale jeho časová zložitosť je vysoká. K-means nie je tiež optimalizovaný pre pozicionovanie UAV. Z toho dôvodu boli navrhnuté sofistikovanejšie algoritmy.

Základný algoritmus riešiaci problém k -centier je založený na jadrách (core-sets). Jadro je malá množina bodov, ktorá

aproximuje pôvodnú väčšiu množinu bodov. Predpokladá sa, že v redukovanej množine (jadrá) je výpočet stredy približne rovnaký ako v pôvodnej množine. Mnoho prirodzených geometrických optimalizačných problémov má jadrá, ktoré aproximujú optimálne riešenie v lineárnom čase alebo takmer lineárnom čase, a ktoré majú veľkosť obmedzenú funkciou $\frac{1}{\epsilon}$ nezávislou od veľkosti vstupu, kde ϵ je ľubovoľné kladné číslo [12]. Problém riešenia algoritmu hrubou silou [13] je však považovaný za INP-ťažký, čo znamená, že jeho časová zložitosť pre optimálne riešenie je $\mathcal{O}(k^n)$, kde n je počet bodov. Ďalšie zjednodušenie základných jadier zahŕňa dynamické programovanie s hornou hranicou časovej zložitosti $\mathcal{O}(n \log k) + (k/\epsilon)^{\mathcal{O}(k^{1-1/d})}$ [14]. V súčasnosti najoptimálnejší prístup k umiestneniu UAV podľa počtu UAV a ich polohy, kde autori založili svoj prístup na zmiešanej stratégii hĺbkového vyhľadávania pre vetvený a hranový strom euklidovského problému k -centier, používa zmiešané prehľadávanie do hĺbky a do šírky [9].

Na základe tohto algoritmu Lyu et al. [15] predstavili metódu metódu pri ktorej sa hľadajú centrá v špirále, čo značne znižuje časovú náročnosť. Komplexnosť tohto prístupu je $\mathcal{O}(K[K \log K + KC(K)]) = \mathcal{O}(K^2[\log K + C(K)])$, kde $C(K)$ je čas jedného behu podprogramu centra. Lokálny algoritmus pokrytia berie do úvahy aktuálne centrum kruhu u , pokrytých používateľov ako množinu P_{prio} a nepokrytých používateľov P_{sec} . Algoritmus prebieha cez celú množinu P_{sec} a v prvom kroku aktualizuje P_{sec} tak, aby obsahovala iba používateľov, ktorí sú vo vzdialenosti menšej ako $2r$ od u , a P_{prio} tak, aby obsahovala iba používateľov, ktorí sú vo vzdialenosti $< r$ od centra. Druhá časť skontroluje najbližšiu pozíciu koncového používateľa v sekundárnej množine používateľov k aktuálnej pozícii centra u . Ak môže byť tento používateľ pokrytý presunom pozície centra bez ovplyvnenia už pokrytých bodov, potom sa aktualizuje centrum a tento používateľ sa presunie zo sekundárnej sady do prioritnej sady. Ak je množina P_{sec} prázdna, algoritmus sa zastaví a presunie sa k ďalšiemu nepokrytému používateľovi v špirále [15]. $C(K)$ predstavuje druhú časť algoritmu. Keďže tento prístup je pod $\mathcal{O}(K^3)$, je výrazne rýchlejší ako jadrá. Pri veľkom počte používateľov v oblasti sa tento prístup stáva pomalým.

Sharma a kol. [16] dosiahli vysokú priepustnosť a zlepšenie pravdepodobnosť správneho umiestnenia pomocou prístupu založeného na farbení matíc s využitím backtrackingu a použitím samohojacej neurónovej siete. Samohojacej neurónovej sieti sa podarilo autonómne regulovať jej mieru učenia pomocou koncepcie umelých spojení, ktoré pomáhajú dosiahnuť stabilizovaný stav [17]. Farbenie matíc je prístup inšpirovaný farbením grafov [18]. Algoritmus je založený na rozdelení oblasti na bunky $n \times n$ a ich farebnom kódovaní podľa počtu koncových používateľov v bunke. V tomto prístupe je každému UAV dovolené obsluhovať maximálne dve susediace bunky tak, aby celkový počet používateľov v susediacich bunkách neprekročil limit podporovaný UAV. Ak bunky obsahujú viac UEs, než dokáže obsluhovať jedno UAV, sú nasadené ďalšie UAV.

Bolo urobených niekoľko pokusov o riešenie problému

optimalizácie umiestnenia UAV pomocou techník strojového učenia alebo hlbokého učenia. Techniky posilňovacieho učenia sa úspešne použili na vytvorenie autonómneho modelu pre navigáciu UAV [19]. Tento prístup sa však nedá použiť na plánovanie celej topológie, ale iba na reálne plánovanie lietacieho času. Relevantnejší a úspešnejší prístup bol použitý v [20], kde autori použili kombinovanú sieť MLP-LSTM. Presnosť pre správne umiestnenie UAV dosiahla vo väčšine prípadov nad 95% s časovou zložitou $\mathcal{O}(1)$. Tento prístup bol však testovaný len na obmedzenom území a s veľmi malým počtom koncových používateľov a UAV, konkrétne 35 a 3. Toto obmedzené testovacie prostredie by viedlo k dobrým výsledkom aj v predchádzajúcich algoritmoch, a preto je spochybniteľné.

[21] predstavuje prístup na maximalizáciu priepustnosti pre pozemné siete asistované UAV zmenšením oneskorenia a straty paketov prostredníctvom optimalizácie trajektórie UAV a posilnením preťažených uzlov a prenosových kanálov. Prístup charakterizuje uzly, linky a celkovú topológiu prostredníctvom oneskorenia, straty, priepustnosti a vzdialenosti. Na charakterizáciu, predikciu a dynamické zlepšenie trajektórie UAV sa používa grafická neurónová sieť (GNN) citlivá na polohu. Autori uvádzajú, že výrazne prekonáva klasický prístup tým, že ukazuje významné zisky v priepustnosti a pomere doručených paketov s významným poklesom oneskorenia a straty paketov.

Hlboká metóda posilňovania, konkrétne hlboké Q-učenie (DQN), bola použitá na spoločnú optimalizáciu plánovania trajektórie UAV a alokácie časových zdrojov na maximalizáciu minimálnej priepustnosti všetkých UEs v multi-UAV-ovom WPCN v [22]. Navrhnutá metóda používa hlbokú neurónovú sieť na aproximáciu optimálnej politiky pre plánovanie trajektórie UAV a alokáciu časových zdrojov. Motiváciou na použitie DQN bola skutočnosť, že vyvinutý problém optimalizácie priepustnosti, ktorý zahŕňa spoločnú optimalizáciu návrhu 3D trasy a priradenia zdrojov kanálov s obmedzením rýchlosti letu UAV a vysielacieho výkonu IoT zariadení, nie je konvexný a preto je jeho priame riešenie mimoriadne ťažké. Učenie posilňovaním je najbežnejším prístupom k riešeniu rozmiestňovania, viď napr. aj [23].

Poskytli sme stručný prehľad niekoľkých prác, ktoré optimalizujú dynamické RAN siete. Vidíme, že sa používa široké spektrum rôznych algoritmov. Vynechali sme veľa prác, ktoré využívajú podobné prístupy a vybrali sme len niekoľko reprezentatívnych prác. Vidíme, že sa používajú hlboké neurónové siete, ako aj prístupy založené na posilňovanom učení. V niektorých prípadoch sa používajú modely strojového učenia, ako je Random Forest, zatiaľ čo v iných prípadoch sa používajú rôzne typy neurónových sietí, ako sú LSTM alebo CNN. V praxi sme si všimli, že využitie GAN sietí bolo doposiaľ veľmi málo preskúmané.

2) *Optimalizácia energie drona*: Optimálne pokrytie UAV v oblasti znamená kompromis s ich životnosťou batérie. Kapacita batérie UAV je obmedzená a vyžaduje si nabíjanie alebo výmenu, keď klesne pod určitý prah. Spotreba energie UAV závisí od palubného obvodu, komunikácie (vysielací (Tx)

výkon) a lietania (pohonový výkon). Energetická účinnosť UAV pôsobiacich ako základňové stanice je prominentnou témou výskumu. Niekoľko prác navrhlo metódy na predĺženie životnosti batérie úpravou vysielacieho výkonu. Rovnako ako v predchádzajúcej časti, existujú ako analytické prístupy, tak aj prístupy využívajúce strojové učenie.

V [24] autori vyjadrujú optimálne umiestnenie UAV na základe problému umiestnenia kruhu, ktorý maximalizuje počet pokrytých UE pomocou minimálnej vysielacej energie. Táto práca dekapluje problém nasadenia UAV-BS vo vertikálnom a horizontálnom rozmere bez straty optimality. Ďalej modeluje nasadenie UAV-BS v horizontálnom rozmere ako problém umiestnenia kruhu a problém najmenšieho obalujúceho kruhu. Simulácie boli vykonané na hodnotenie výkonnosti navrhovanej metódy pre rôzne priestorové rozloženie používateľov.

Článok s názvom "Energy-aware coverage path planning of UAVs" [25] navrhuje algoritmus plánovania trasy s ohľadom na energetickú efektívnosť, ktorý minimalizuje spotrebu energie pri dodržiavaní súboru ďalších požiadaviek ako pokrytie a rozlíšenie. Algoritmus je založený na energetickom modeli odvodenom z reálnych meraní. Autori hodnotia navrhovaný algoritmus pomocou simulácií aj skutočných experimentov. Výsledky ukazujú, že navrhovaný algoritmus môže efektívne znížiť spotrebu energie pri zachovaní pokrytia a rozlíšenia. Tento článok zdôrazňuje, že vysielacia energia je oveľa nižšia ako pohonová energia kvôli vysokej energetickej náročnosti rotorov UAV s otáčavými krídlami [25]. Preto hlavná výzva spočíva v optimalizácii pohonovej energie.

Bežným predpokladom pre UAV s otáčavými krídlami je, že sa vznášajú na statickej pozícii, ktorá optimalizuje kapacitu UE. Tento predpoklad však spochybňuje [26], ktorý rozširuje prácu [27] a ukazuje, že vznášanie sa nie je najenergetickejšou scénarom pre UAV s otáčavými krídlami. Zeng et al. [27] predstavujú energeticky efektívny prístup k umiestneniu UAV. Autori zdôrazňujú kompromis medzi pokrytím a energetickou efektívnosťou. Hlavným cieľom tohto výskumu je optimalizovať trajektóriu UAV s ohľadom na spotrebu pohonovej energie. Teoretický model založený na lineárnej aproximácii stavového priestoru a sekvenčnej konvexnej optimalizácii sa aplikuje na optimalizáciu trajektórií. Bez akýchkoľvek obmedzení na optimalizáciu trajektórií viedu návrhy s maximalizáciou prenosovej rýchlosti a minimalizáciou energie k vymiznutiu energetickej efektívnosti. Optimálna letová trajektória UAV je kruhová trajektória s optimalizovaným polomerom letu a rýchlosťou. Optimalizované kruhové trajektórie preukázali vyššiu energetickú efektívnosť. Autori ďalej rozšírili výskum v [28]. Tento článok navrhuje perspektívu dizajnu riadenia na optimalizáciu 3D trajektórií UAV pre energeticky efektívnu komunikáciu. Autori hodnotia navrhovaný algoritmus pomocou simulácií aj skutočných experimentov. Výsledky ukazujú, že navrhovaný algoritmus môže efektívne znížiť spotrebu energie pri zachovaní kvality komunikácie.

Článok "Optimizing Transmission and Propulsion Powers for Flying Base Stations" [29] navrhuje metódu

na optimalizáciu vysielacej a pohonovej energie pre lietajúce základňové stanice (FlyBSs), ktoré sa považujú za efektívny nástroj na zlepšenie kapacity mobilných sietí a na zabezpečenie komunikácie v núdzových prípadoch. Článok navrhuje dva algoritmy: 1) algoritmus pre pridelenie vysielacej energie, ktorý maximalizuje minimálny pomer signálu k interferencii plus šum (SINR) všetkých používateľov za dodržania celkového obmedzenia vysielacej energie FlyBSs a 2) algoritmus pre pridelenie pohonovej energie, ktorý minimalizuje celkovú spotrebu pohonovej energie FlyBSs za dodržania minimálneho požiadavku na SINR všetkých používateľov. Dosiahli až 26% úsporu energie v režime nízkej rýchlosti a až 60% v režime vysokých rýchlostí pomocou číselného riešenia pre optimálnu polohu FlyBSs.

Rozsiahly výskum aplikácií ML/AI pre optimalizáciu výkonu lietajúcej základňovej stanice vykonali autori z laboratória pre výskum mobilnej 6G na Českej technickej univerzite v Prahe. Autori Becvar a Nikooro publikovali viacero článkov, ktoré navrhujú niekoľko metód na optimalizáciu vysielacej energie s pomocou ML/AI. V [30] sa používa zhukovanie na zoskupenie používateľov tak, aby sa vysielacia energia FlyBS mohla efektívne alokovať. Navrhované riešenie zoskupuje používateľov a efektívne pridelí vysielaciu energiu FlyBS používateľom tak, že komunikačné pokrytie poskytované FlyBS sa rozšíri o 67% - 270% v porovnaní s existujúcimi riešeniami, zatiaľ čo pohonová energia sa nezvyšuje.

Prístup pomocou RL na optimalizáciu pozície lietajúcich základňových staníc namontovaných na palube UAV na podporu makrozakladňovej stanice (MBS) a maximalizáciu celkovej dátovej prenosovej rýchlosti inteligentným nasadením UAV BS v downlinku celulárneho systému bol navrhnutý v [31]. Autori navrhli algoritmus, ktorý používa RL na optimalizáciu pozície lietajúcich základňových staníc namontovaných na palube UAV na podporu MBS. Podobne ako v predchádzajúcom článku, aj tu sa používa Q-learning.

V [32] sú definované pravidlá pre pozicionovanie a priradenie pre transparentné lietajúce reléové stanice (FlyRS), reprezentované transparentnými relémi namontovanými na UAV, ktoré majú potenciál zlepšiť kapacitu a pokrytie celulárnej siete pri minimálnych dodatočných komplikáciách a energetických nákladoch, najmä v porovnaní s netransparentnými relémi. Článok navrhuje riešenie umožňujúce efektívne priradenie používateľov k FlyRS a určenie vhodných pozícií FlyRS prostredníctvom aglomeratívneho hierarchického zhukovania, ktoré zlepšuje súčet kapacity používateľov vo vzťahu k existujúcim riešeniam, ktoré sa spoliehajú na znalosť pozícií používateľov. Aglomeratívne zhukovanie je typ hierarchického algoritmu zhukovania, ktorý začína s každým dátovým bodom ako individuálnym zhukom (nazývaným aj list) a potom spája každý jednotlivý zhuk s procesom, ktorý sa opakuje, až kým sa všetky dáta nestanú jedným zhukom. Algoritmus aglomeratívneho hierarchického zhukovania použitý v tomto článku je založený na stratégii úplnej väzby. V tejto práci sú použité dve prepojené hlboké neuronové siete, kde prvá predpovedá x a y súradnice UAV podľa informácií o zisku kanála od okolitých BS a druhá

predpovedá výšku.

Prehľad, ktorý sme poskytli, vytvára jasný obraz o tom, že AI a ML majú široké využitie v 5G. Mnoho prístupov sa snaží zlepšiť charakteristiky siete, ako je priepustnosť, path loss a zisk kanála, kapacita alebo celková topológia siete. V nasledujúcich kapitolách poskytujeme pohľad na naše výskumy o nových spôsoboch využitia AI/ML pre dynamické RAN siete, konkrétne drony.

III. OPTIMALIZÁCIA SPOTREBY ENERGIE

V tejto kapitole opíšeme nami navrhnutý spôsob optimalizácie spotreby energie vzhľadom na poveternostné podmienky.

A. Model systému

Zvažujeme 3D mestskú oblasť $\mathcal{A} \subset \mathbb{R}^3$ s budovami. Budovy zaberajú oblasť definovanú súradnicami \mathcal{A}' . Okrem toho je na miestach $\mathcal{U} = \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N$ umiestnených N UE, kde $\mathbf{u}_n = [x_n, y_n, z_n] \in \mathcal{A}^o, \forall n \in \langle 1, N \rangle$ predstavuje oblasť, kde sa môžu pohybovať používatelia a dronové bázové stanice (UAV-BS). V oblasti sa nachádzajú aj M statických bázových staníc (SBS) umiestnených na strechách budov na súradniciach $\mathcal{S} = \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_M$, kde $\mathbf{s}_m = [x_m, y_m, z_m] \in \mathcal{A}', \forall m \in \langle 1, M \rangle$. V oblasti je ďalej K dronových základňových staníc (UAV-BS), ktoré prenášajú dáta zo statických základňových staníc (SBS) k používateľom na súradniciach $\mathcal{D}(t) = \mathbf{d}_1(t), \mathbf{d}_2(t), \dots, \mathbf{d}_K(t)$, kde $\mathbf{d}_k(t) = [x_k(t), y_k(t), z_k(t)] \in \mathcal{A}^o, \forall k \in \langle 1, K \rangle$ v čase t .

Realistické UAV-BS sú obmedzené v manévrování a ich pohyb by mal byť hladký vzhľadom na zmeny smeru, aby sa zabránilo vysokej spotrebe energie [33]. Preto navrhujeme kruhovú trajektóriu, ktorá dostatočne aproximuje optimálnu trajektóriu s len malým vplyvom na výkon, ako je demonštrované v [34]. Navyše, menšia odchýlka v pozícii UAV-BS oproti optimálnemu všeobecnému tvaru trajektórie má len nepatrný vplyv na kapacitu UEs [7]. Zvažujeme kruhovú trajektóriu s polomerom $\mathbf{r} = [r_1, r_2, \dots, r_K]$ v strede kruhu $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K] \in \mathcal{R}^{K \times 3}$, kde $\mathbf{f}_k = [f_{k,x}, f_{k,y}, f_{k,z}]$. Pozícia k -tej UAV-BS v čase t je potom definovaná ako $\mathbf{d}_k(t) = [f_{k,x} - r_k \cos \alpha(t), f_{k,y} - r_k \sin \alpha(t), f_{k,z}]$, kde $\alpha(t)$ je relatívny uhol UAV-BS k x-ovej osi.

1) *komunikačný model*: Zvažujeme ortogonálny downlinkový prenos k UEs. Kapacitu prístupového kanála (a) medzi k -tou UAV-BS a priradeným n -tým UE vyjadrujeme ako

$$C_{m,k}^b = B_n \log_2 \left(1 + \frac{P_m^{\text{Tx}} g_{m,k}^b \theta_{m,k}^b}{B_n \sigma^2 + I} \right), \quad (2)$$

kde B_n je šírka pásma priradená k n -tému UE. Dostupná šírka pásma k -tej UAV-BS je rovnomerne rozdelená medzi jej priradené UE. Parameter $P_{k,n}^{\text{Tx}}$ je vysielač výkon k -tej UAV-BS priradenej k n -tému UE, $g_{k,n}^a$ predstavuje zisk prístupového kanála medzi k -tou UAV-BS a n -tým UE, $\theta_{k,n}^a$ označuje skreslenie, σ^2 predstavuje šumovú výkonovú hustotu a I je interferencia so susednými bunkami.

Analogicky je kapacita backhaul kanála (b) medzi m -tou SBS a k -tou UAV-BS, ktorá je potrebná na prenos dát n -tého UE, definovaná ako

$$C_{m,k}^b = B_n \log_2 \left(1 + \frac{P_m^{\text{Tx}} g_{m,k}^b \theta_{m,k}^b}{B_n \sigma^2 + I} \right), \quad (3)$$

kde P_m^{Tx} je vysielač výkon m -tej SBS, $g_{m,k}^b$ je zisk backhaul kanála medzi m -tou SBS a k -tou UAV-BS. Keďže alokácia šírky pásma neovplyvňuje spotrebu energie na pohon UAV-BS, predpokladáme rovnaké B_n pridelené pre všetky UEs obsluhované rovnakou UAV-BS.

Pre komunikáciu UAV-BS sa používa prenos s polovičnou dvojcestnosťou typu decode-and-forward. Teda kapacita kanála medzi m -tou SBS a n -tým UE cez k -tu UAV-BS je určená ako

$$C_{m,k,n} = \min(T_n^b C_{m,k}^b, (1 - T_n^b) C_{k,n}^a), \quad (4)$$

kde $T_n^b \in [0, 1]$ je normalizovaný čas naplánovaný pre prenos n -tého UE cez backhaul a $(1 - T_n^b)$ je normalizovaný čas pre prenos cez prístupový kanál. Aby sa zabránilo preťaženiu na niektorom z týchto kanálov, pre k -tu UAV, ktorá preposiela dáta z m -tej SBS, predpokladáme [35]

$$\sum_{n=1}^N T_n^b C_{m,k}^b = \sum_{n=1}^N (1 - T_n^b) C_{k,n}^a. \quad (5)$$

2) *Dynamický model vetra*: Dynamické účinky vetra modelujeme pomocou všeobecne uznávaného k - ε modelu založeného na Reynoldsových priemerovaných Navier-Stokesových (RANS) rovniciach [36]. K - ε model zohľadňuje turbulentný tok vetra, ktorý je prítomný v mestskom prostredí s množstvom prekážok v ceste vetra a používa časovo priemerované rovnice pohybu pre vietor. V praxi je model prostredia popísaný prostredníctvom metódy konečných objemov (FVM) pomocou mriežky, ktorá reprezentuje geometriu prostredia. Presnosť modelu je určená citlivosťou mriežky a počtom iterácií [37]. Mriežka by mala byť dostatočne citlivá, aby zachytila geometriu modelu. Príliš citlivá mriežka však vedie k vysokej výpočtovej zložitosti. Počet iterácií určuje, koľkokrát sa model spustí, kým dôjde ku konvergencii k správnej predpovedi vetra. Mriežku vytvárame buď ručne, alebo pomocou nástroja, napríklad na základe skenovania vybranej oblasti pomocou UAV. Mriežka sa potom premieta do virtuálneho veterného tunela a súčasne sa vyberie vstup a výstup reprezentujúci stranu, z ktorej vietor prúdi dovnútra a von z tunela. Rýchlosť vetra na vstupe $\vec{I} = [i_x, i_y, i_z]$, a turbulentná intenzita a viskozita sa merajú v danom referenčnom bode (RP), napríklad na meteorologickej stanici na budove, a vložia sa do modelu.

K - ε model poskytuje informácie o rýchlosti vetra $\vec{w}_a = [w_x, w_y, w_z]$ pre každú diskretnú pozíciu $a \in \mathcal{A}^o$. Časový vývoj vektorov rýchlosti vetra je vyjadrený ako časový vývoj poľa priemerných vektorov rýchlosti (rýchlosť prúdenia) z ľavej strany konvektívnej formy RANS rovníc, ako je definované v [36]. Dĺžka prevádzkovej doby UAV-BS (\sim minúty) je

dostatočne veľká v porovnaní s časovou škálou turbulencií vo veternom prúde (\sim milisekundy). Preto môžeme využiť priemerné hodnoty veterného prúdenia z RANS pre návrh trajektórie UAV-BS namiesto okamžitých hodnôt vetra, ktoré sú v praxi ovplyvnené ťažko odhadnuteľnou malomierkovou turbulenciou [36].

B. Model spotreby energie pri zohľadnení vetra

V tejto časti predstavujeme nový model spotreby pohonnej energie pre UAV-BS s prihliadnutím na vietor.

Aby sme mohli odvodiť model vhodný pre podmienky vetra, začneme s modelom pre spotrebu pohonnej energie pri rýchlosti \tilde{V} bez vetra, ako je uvedené v [26].

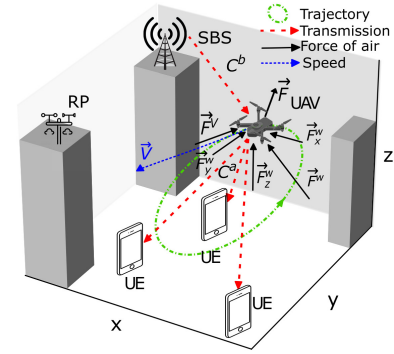
$$P(\tilde{V}) = \underbrace{P_0 \left(1 + \frac{3\tilde{V}^2}{V_{\text{tip}}^2}\right)}_{\text{blade profile}} + \underbrace{P_i \left(\sqrt{1 + \frac{\tilde{V}^4}{4v_0^4}} - \frac{\tilde{V}^2}{2v_0^2}\right)}_{\text{induced}} + \underbrace{\frac{1}{2}d_0\rho s A\tilde{V}^3}_{\text{parasite}} \quad (6)$$

kde P_0 , P_i , d_0 a s sú konštanty špecifické pre hardware UAV (pozri [26]), ρ definuje hustotu vzduchu, $V_{\text{tip}} = R\omega$ je rýchlosť konca lopatky rotora pre uhlovú rýchlosť lopatky ω a polomer rotora R , v_0 reprezentuje rýchlosť rotora indukovanú preletom, a A je plocha rotorového disku [26].

Poznamenajme, že tento model je výrazne rozšírený o zohľadnenie dynamiky UAV pre praktické realizácie v [28]. Modely v oboch prácach [26], [28] vedú k podobným trendom v spotrebe energie vzhľadom na rýchlosť UAV, pozri [28]. Navyše, ani jeden z týchto dvoch modelov nezohľadňuje vietor a jeho vplyv na energiu, avšak všeobecný princíp rozšírenia o zahrnutie vetra je analogický pre oba. Keďže model z [26] je oveľa menej zložitý ako model v [28] a integrácia vetra je ľahšie ilustrovateľná na modeli v [26], zameriavame sa na tento model a do tohto modelu následovne zahrnieme faktory súvisiace s vetrom.

Vo všeobecnosti sa sila vetra F_w pôsobiaca na efektívnu plochu A_w UAV vypočíta ako $F_w = \frac{1}{2}\rho w^2 A_w$. Efektívna plocha A_w je celková plocha zasiahnutá príslušným prúdením vzduchu. Okrem toho sa odčíta sila vzdušného odporu pôsobiaca v opačnom smere letu vypočítaná ako $F_v = \frac{1}{2}\rho V^2 A_v$ pôsobiaca na efektívnu plochu A_v .

Uvažujeme rýchlosť vetra $\vec{w} = w_x\vec{e}_x + w_y\vec{e}_y + w_z\vec{e}_z$ vyjadrenú pomocou ortogonálnych jednotkových vektorov $\vec{e}_x, \vec{e}_y, \vec{e}_z$ reprezentujúcich zložky rýchlosti v každej osi. Analogicky predpokladáme, že rýchlosť UAV v systéme vzťahujúcom sa k pozemnému pozorovateľovi je $\vec{V} = V_x\vec{e}_x + V_y\vec{e}_y + V_z\vec{e}_z$. Využitím zložiek rýchlosti vetra môžeme vypočítať x zložku sily vetra F_x^w ako $F_x^w = \frac{1}{2}\rho w_x w A_w \vec{e}_x$. F_y^w a F_z^w sú vypočítané analogicky. Sila odporu vzduchu F^v pozostávajúca z F_x^v, F_y^v a F_z^v sa vypočíta analogicky z \vec{V} a A_v . Konečná sila pôsobiaca na UAV je súčet zložiek F^v a F^w , takže $\vec{F} = \vec{F}^v + \vec{F}^w$. Považovaný model je znázornený na obr. 1.



Obr. 1: Model znázorňujúci sily pôsobiace na UAV, vektor rýchlosti, trajektóriu a backhaul/access kanály [38]

Celková hnacia energia k -teho UAV-BS k -th UAV-BS $E_k = E_k^p + E_k^{\text{air}}$ je súčtom: *i*) energie E_k^p zodpovedajúcej výkonu k -teho UAV potrebnému na udržanie pohybu rýchlosťou V_k po celkový čas letu τ , a *ii*) disipovanej energie E_k^{air} v dôsledku sily \vec{F}_k súvisiacej s prúdením vzduchu pôsobiacej na povrchy k -teho UAV.

Obidve zložky celkovej spotreby hnacej energie sa vypočítavajú časovou integráciou príslušných premenných, t.j., $E_k^p = \int_0^\tau dt P(V_k)$, $E_k^{\text{air}} = \int_0^\tau dt \vec{F}_k(t) \cdot \left(\frac{d\vec{s}_k(t)}{dt}\right)$, kde $\vec{s}_k(t)$ reprezentuje projekciu dráhy k -teho UAV v priestore.

C. Formulácia problému

Cieľom je minimalizovať spotrebu hnacej energie UAV-BS prostredníctvom určenia trajektórií UAV-BS. Kvôli obmedzeniam, ktoré sú vysvetlené v sekcii III-A, je naším cieľom určiť kruhové trajektórie, vrátane ich centier; $\mathbf{F}^* = [f_1^*, f_2^*, \dots, f_K^*]$ pre všetky UAV-BS a príslušné polomery $\mathbf{r}^* = [r_1^*, r_2^*, \dots, r_K^*]$. Problém optimalizácie úspory energie je teda definovaný ako

$$\mathbf{F}^*, \mathbf{r}^* = \underset{\mathbf{F}, \mathbf{r}}{\operatorname{argmin}} \sum_{k=1}^K E_k, \quad (7)$$

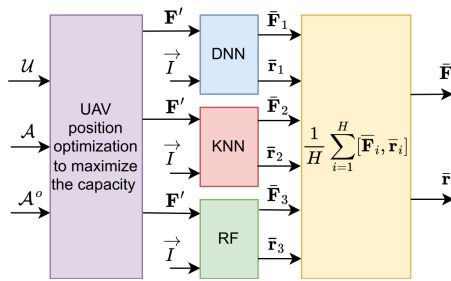
$$\text{s.t. } (f_k^*) \in \mathcal{A}^o, \quad \forall k \in \langle 1, K \rangle \quad (a)$$

$$f_{k,z}^* \in \langle z_{\min}, z_{\max} \rangle, \quad \forall k \in \langle 1, K \rangle \quad (b)$$

$$r_k^* \in \langle r_{\min}, r_{\max} \rangle, \quad \forall k \in \langle 1, K \rangle \quad (c)$$

$$C_{m,k,n}(d_k(t)) \geq (1 - \gamma)C'_{m,k,n}, \quad \forall k \in \langle 1, K \rangle, \\ \forall m \in \langle 1, M \rangle, \forall n \in \langle 1, N \rangle. \quad (d)$$

Obmedzenie (a) definuje množinu všetkých možných polôh \mathcal{A}^o v oblasti s vylúčením budov a prekážok, kde nie je prítomnosť UAV-BS povolená. Obmedzenie (b) definuje rozsah možných výšok UAV-BS a obmedzenie (c) definuje rozsah možných letových rádiusov. Obmedzenie (d) zabezpečuje, aby kapacita $C_{m,k,n}$ n -tej UE spojennej s m -tým SBS prostredníctvom k -tej UAV na pozícii d_k v čase t neklesla pod $(1 - \gamma)C'_{m,k,n}$, kde γ predstavuje maximálny relatívny povolený pokles kapacity UE vo vzťahu k maximalizovanej



Obr. 2: Metóda súborového učenia na determináciu energeticky efektívnej trajektórie UAV-BS. Ensemble learning approach to determine energy-efficient UAV-BS trajectory. Trom základným algoritmom (DNN, KNN, RF) sú predložené maximálne kapacitné pozície UAV-BS F' a informácie o vetre \vec{I} v referenčnom bode [38].

kapacite $C'_{m,k,n}$ n -tej UE bez ohľadu na spotrebu energie. Teda obmedzenie (d) zabezpečuje, že degradácia kapacity je zanedbateľná.

Optimalizačný problém v (7) je NP-ťažký, nekonvexný nelineárny programovací problém s vnoreným numerickým hodnotením parciálnych diferenciálnych rovníc reprezentovaných RANS s lineárnymi obmedzeniami na premenné pomocou FVM. Časová zložitosť FVM je $\mathcal{O}(\eta \log(\eta))$, kde η je citlivosť mriežky charakterizujúca prostredie FVM. FVM aproximuje hodnoty časovým hodnotením v ψ diskretných časových krokoch, takže zložitosť sa stáva $\mathcal{O}(\eta \log(\eta)\psi)$ [37]. Nákladová funkcia v (7) je teda mimoriadne výpočetne náročná. Preto uvažujeme strojové učenie, aby bolo riešenie uskutočniteľné pre praktické aplikácie.

D. Navrhnuté riešenie

V tejto časti najprv určíme teoretickú minimálnu spotrebu energie. Potom navrhne riešenie problému (7). Riešenie je založené na využití učenia súboru pre predpovedanie sub-optimálnych energeticky účinných pozícií UAV-BS \bar{F} a príslušných rádiusov \bar{r} . Využívame rozloženie vetra v oblasti a hľadáme trajektóriu, kde je protivietor čo najnižší, zatiaľ čo zadný vietor je čo najvyšší, aby sa znížila spotreba energie.

Na identifikáciu teoretickej minimálnej energetickej spotreby UAV-BS definovanej v (7) používame vyhľadávanie, ktoré odvodzuje optimálne F^* a r^* testovaním všetkých možných možností nasadenia UAV-BS (s prihliadnutím na obmedzenia a , b a c) a výberom toho, ktorý vedie k minimálnej energetickej spotrebe pri splnení kapacitného obmedzenia (d). Pre praktické aplikácie nie je takéto vyhľadávanie uskutočniteľné kvôli obrovskému výpočtovému zložitosti. Preto tiež riešime (7) ako predikčný problém prostredníctvom súborového učenia, kde sa F^* a r^* používajú iba ako ciele v tréningovej fáze a ich výpočet online nie je požadovaný.

Súborové učenie kombinuje niekoľko základných modelov na zlepšenie výkonnosti predikcie v regresných problémoch oproti samostatným prediktorom. Prijímame heterogénnu sadu

učiteľov $\mathcal{F} = \{f_i, i = 1, 2, \dots, H\}$, kde $H = 3$, ako je zobrazené na obr. 2. V navrhovanom riešení používame hlbokú neuronovú sieť (DNN) v kombinácii s tradičnými náhodnými lesmi (RF) a metódou najbližších susedov (KNN) na vytvorenie heterogénneho súboru, ktorý využíva výhody nižších výpočtových nákladov (plytká DNN a relatívne nízke požiadavky na výpočet RF a KNN) a vyššej diverzity, ktorá potenciálne vedie k zlepšeniu výkonu [39]. Okrem toho sú jednotlivé základné modely charakterizované vysokou citlivosťou na vzorky datasetu a dokonca aj malé zmeny v tréningových vzorkách by mohli v našom probléme viesť k veľkým zmenám v predpovedanom výstupe. Keď sú však kombinované do súborového učenia, výsledná chyba je nižšia ako u jednotlivých klasifikátoroch [39].

Vektor príznačkov pre súborové učenie je identický pre všetky tri základné modely a pozostáva z F' , určeného pomocou akéhokoľvek existujúceho algoritmu na pozicionovanie UAV-BSs na maximalizáciu kapacity, a informácií o vetre na RP $\vec{I} = [i_x, i_y, i_z]$. V štádiu nasadenia je výstup súboru implementovaný ako priemer základných učiteľov, t.j. $[\bar{F}, \bar{r}] = \frac{1}{H} \sum_{i \in \mathcal{F}} \hat{f}_i[F', \vec{I}]$.

Hyperparametre základných modelov sú identifikované pomocou optimalizácie grid-search. Prediktor DNN pozostáva z troch skrytých vrstiev s 30, 40 a 40 neurónmi, ktoré využívajú aktivačnú funkciu ReLU. RF dosahuje najvyššiu výkonnosť, keď je počet odhadovateľov nastavený na 160 a maximálna hĺbka je rovná 13. Pre KNN je počet susedov nastavený na počet UE N delený požadovaným počtom UAV-BSs K .

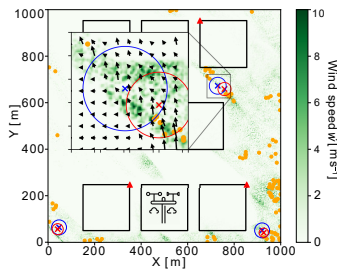
E. Analýza výkonu

V našej simulácii zvažujeme obdĺžnikovú mestskú oblasť o veľkosti 1×1 km s 8 budovami rôznych výšok. Zvažujeme päť SBS umiestnených na náhodných pozíciách na budovách. UE sú umiestnené náhodne podľa binomiálneho bodového rozdelenia. Zvažujeme $B = 20$ MHz a spektrálnu hustotu šumu -174 dBm/Hz. Vysielacie výkony SBS a UAV-BS sú rovné 46 dBm a 30 dBm, respektíve. Model strát cesty pre kanály s priamou viditeľnosťou (LoS) je v súlade s [32]. Pre kanály bez priamej viditeľnosti (non-LoS) sa na vrch LoS útlmu pridáva útlm stien/prekážok, ako je to v [32]. Rýchle zložky zblednutia sa generujú ako exponenciálne rozdelené náhodné premenné s jednotkovým priemerom. Simulačné výstupy sú priemerované z 10 000 behov.

Polohu pre minimálnu spotrebu energie hľadáme v diskretizovanom priestore pomocou absolútneho prehľadávania priestoru s krokom 1m. Parameter zhoršenia kapacity uvedený v (7d) je nastavený na $\gamma = 0.03$.

Zvažujeme UAV-BS reprezentované modelom DJI Spreading Wings S900 s tvarom a hardvérovými charakteristikami uvedenými v technickom manuáli¹. Vizualizáciu oblasti, ktorú sme zvažovali v našom simulačnom nastavení, nájdete na Obr. 3.

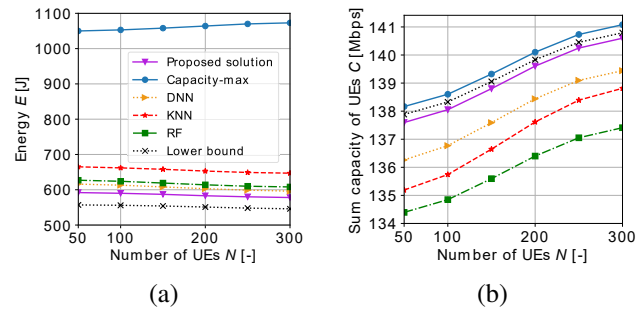
Porovnáваме výkon nášho návrhu s dvoma referenčnými schémami: *i*) dôkladné vyhľadávanie (označované ako *lower-bound*), ktoré poskytuje F a r , a *ii*) najmodernejší algoritmus (state-of-the-art), ktorý maximalizuje kapacitu pomocou DNN,



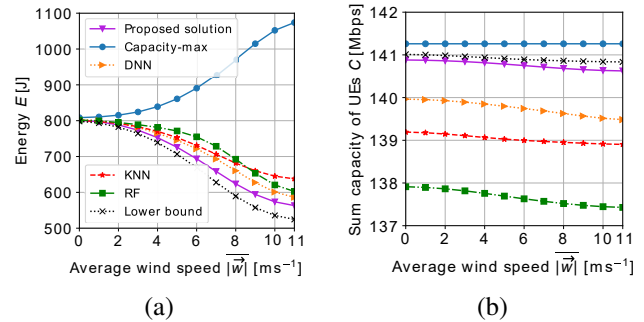
ako je navrhnuté v [32] (označované ako *capacity-max*). Treba poznamenať, že algoritmus *capacity-max* nezohľadňuje efekt vetra pri návrhu trajektórie.

Akometriku výkonnosti sme použili priemernú spotrebu pohonu $E = \frac{1}{\tau} E_k^p$ a priemernú súčtovú kapacitu definovanú ako priemer súčtu všetkých kapacít UE premierený cez čas, t.j. $C = \frac{1}{\tau} \int_0^{\tau} dt \sum_{m=1}^M \sum_{k=1}^K \sum_{n=1}^N (C_{m,k,n}(\mathbf{d}_k(t)))$. Obrázok 4a ukazuje úsporu energie vo výške 47%, ktorú prináša navrhovaný prístup v porovnaní s riešením *capacity-max* pre priemernú rýchlosť vetra $|\bar{w}| = 10 \text{ ms}^{-1}$ bez ohľadu na počet nasadených UEs v systéme. Takéto úspory sa očakávajú, pretože prístup *capacity-max* nezohľadňuje vietor pri návrhu trajektórie a UAV-BS sa teda môže stretávať s potenciálne silným vetrom počas letu. Samostatné základné modely preukazujú mierny pokles výkonnosti v porovnaní s navrhovaným prístupom. Preto rozmanitosť medzi základnými modelmi vedie k vynikajúcej výkonnosti navrhovaného sýborového učenia. Na obrázku 4b sme vykreslili priemerný súčet kapacity v závislosti od počtu nasadených UEs v systéme pre priemernú rýchlosť vetra $|\bar{w}| = 10 \text{ ms}^{-1}$. Navrhované riešenie takmer zodpovedá výkonnosti algoritmu *capacity-max* (strata pod 0,3%) bez ohľadu na počet UEs. Samostatné základné modely preukazujú o niečo väčší pokles výkonnosti približne o 1,2%, 1,4% a 2% pre DNN, KNN a RF, v poradí.

Na obrázku 5a zobrazujeme priemernú spotrebovanú pohonnú energiu v závislosti od rôznych rýchlostí vetra. Algoritmus *capacity-max* bez kompenzácie vetra je charakterizovaný narastajúcou spotrebou energie s rýchlosťou vetra. Naopak, náš navrhovaný prístup a základné modely využívajú turbulentný prúd vetra tým, že prispôbujú trajektórie, čo vedie k poklesu spotreby energie s narastajúcou rýchlosťou vetra. Nižšia spotreba energie pri silnejšom vetre je výsledkom optimalizovanej trajektórie UAV-BS, ktorá zohľadňuje silu vetra na



Obr. 4: Average energy consumption (subplot a) and sum-capacity (subplot b) over varying number of UEs.



Obr. 5: Average energy consumption (subplot a) and sum capacity (subplot b) as a function of the average wind speed [38].

rôznych miestach, takže UAV-BS sleduje trajektóriu s silným závetrím a slabým protivetrím. Úspora energie nášho návrhu v porovnaní so štandardným algoritmom *capacity-max* dosahuje 47% úsporu energie pre rýchlosť vetra $|\bar{w}| = 10 \text{ ms}^{-1}$. Okrem toho, návrh poskytuje takmer identický výkon ako výpočtovo náročný *lower-bound* s rozdielom vždy pod 2%. Spotreba energie u základných modelov je až o 8% horšia v porovnaní so súborovým učením. Na obrázku 5b pozorujeme, že súčty kapacity podľa navrhovaného algoritmu sú takmer zhodné (rozdiel pod 0,1%) vzhľadom na súčet kapacity *lower-bound* a *capacity-max* pre všetky skúmané charakteristiky vetra. Základné modely preukazujú väčšie zhoršenie medzi 1,37 a 2,75%.

IV. OPTIMALIZÁCIA POLOHY DRONOV VZHLÁDOM NA SIGNÁLNE CHARAKTERISTIKY

V tejto časti navrhujeme nové riešenie pre optimálny návrh pokrytia založené na podmienených generatívnych súperiaciach sieťach (cGAN). cGAN patria do triedy generatívnych hlbokých učiacich sa modelov, ktoré sú široko používané pre rôzne výzvy v systémoch piatych generácií (5G) [40], [41]. Základnou motiváciou pre použitie cGAN je schopnosť naučiť sa mapovanie z vstupných údajov na cieľové rozdelenie údajov. Hoci väčšina výskumu cGAN sa zameriava na generovanie obrazov alebo videa, existuje obrovský potenciál pre technológiu cGAN v iných oblastiach informatiky, ako je zhlukovanie [42] a detekcia narušenia [43]. V našom výskume sú vstupné údaje reprezentované

¹http://dl.djicdn.com/downloads/s900/en/S900_User_Manual_v1.2_en.pdf

rozdelením UEs v priestore a cieľové údaje sú reprezentované rozdelením UAVs poskytujúcich pokrytie pre UEs. Výpočet optimálneho rozdelenia UAVs je vo všeobecnosti NP-ťažký [15] a preto, vzhľadom na extrémne časové obmedzenia, nie je prakticky uskutočniteľný. Veríme, že prostredníctvom rýchlejšej a výpočtovo efektívnej aplikácie cGAN dokážeme poskytnúť dostatočne presnú aproximáciu rozdelenia UAVs so zníženou časovou zložitou. Ako základ pre porovnanie sme si vybrali dva algoritmy, ktoré poskytujú buď optimálne riešenie pre pokrytie založené na UAV, nazývané algoritmus jadrových množín [13], [9], alebo kvázi optimálne riešenie, nazývané špirálový algoritmus [15]. Upozorňujeme, že v článku sa nezameriavame na špecifickú mobilitu UAVs, ale sústredíme svoju pozornosť na výpočet polôh UAVs pre daný momentálny záber polôh UEs. Napriek tomu je dôležitá vlastnosť, že náš algoritmus je schopný spolupracovať so špecifickými algoritmi, ktoré riešia pohyb dronov pre získanie lepšej výkonnosti aj vo veľmi komplexných a dynamických scenároch.

A. Model systému

Pre optimalizáciu pokrytia rozdelíme skúmanú oblasť na neprekývajúce sa segmenty veľkosti $n \times n$. Takto reprezentujeme scenár v maticovej forme, kde každý prvok matice reprezentuje počet UEs v segmente a ich príslušné polohy v oblasti. Rovnaký postup sa uplatňuje na základové stanice založené na UAV. Segmentácia skúmaného regiónu umožňuje použitie techník hlbokého učenia a rýchle spracovanie matíc na určenie kvázi optimálneho pokrytia pre UEs. Predpokladáme, že UE je pokryté, ak je umiestnené v rámci vyhradeného perimetra UAV. Náš navrhovaný prístup je zameraný na určenie minimálneho potrebného počtu UAV a ich príslušných polôh na dosiahnutie kvázi optimálneho pokrytia cieľovej oblasti pomocou cGAN.

V tomto výskume používame nasledujúce označenia. Matica $\mathbf{X} = (x_{ij}) \in \mathcal{N}^{n \times n}$ reprezentuje rozdelenie UEs v oblasti, pričom x_{ij} označuje počet UEs v (i, j) segmente. Podobne, $\mathbf{K} = (k_{ij}) \in \mathcal{N}^{n \times n}$ a $\mathbf{Y} = (y_{ij}) \in \mathcal{N}^{n \times n}$ označujú umiestnenie UAVs v segmentovanom regióne vypočítané pomocou analytickej stratégie jadrových množín a heuristicky založeného cGAN.

B. Navrhovaný prístup

Navrhovaný prístup pozostáva z dvoch fáz - tréningovej fázy cGAN a fázy nasadenia. Tréningová fáza zahŕňa tréningovanie oboch architektúr cGAN: generátora G a diskriminátora D . Fáza nasadenia zahŕňa spoločné použitie generátora G a unikátneho korekčného postupu, ktorý zabraňuje rozmazaniu polôh UAVs. Tréningovanie sa vykonáva iba raz a to offline; teda neovplyvňuje časovú zložitú navrhovaného prístupu pri praktických nasadeniach.

1) *tréningová fáza*: cGAN [44] súčasne používa dve hlboké neurónové siete, teda generátor (G) a diskriminátor (D). Tréningový proces cGAN prebieha striedavou aktualizáciou parametrov G a D . V konvenčných modeloch GAN [45] je úlohou generátora generovať realistické vzorky a úlohou diskriminátora rozlišovať syntetické (generované) vzorky od

skutočných vzoriek. Pre našu úlohu používame cGAN, ktorý umožňuje prídanie externej podmienky priestorového rozdelenia UEs \mathbf{X} v pokrývanej oblasti pre generátor aj diskriminátor. Teda cieľom generátora G je vypočítať požadovaný minimálny počet UAV a ich kvázi optimálne polohy \mathbf{Y} pre daný vstup cGAN. Úlohou diskriminátora D je rozhodnúť, či sú generované polohy UAV vytvorené optimálnym algoritmom jadrových množín alebo heuristicky založeným generátorom. V tomto prípade, optimálne polohy UAV predstavujú polohy, pri ktorých sú všetky UEs pokryté minimálnym počtom UAV. Optimálne riešenie \mathbf{K} vypočítame na základe euklidovského k -centrového problému pomocou *zmiešanej stratégie prechádzania štrkou-hĺbkou*. Vo všeobecnosti sa riešenie nazývané algoritmus jadrových množín zakladá na existencii malého podmnožstva bodov nazývaného *jadrové množiny*, ktoré vedie k schéme polynomiálnej časovej aproximácie pre pevný počet UAV.

Problém optimalizácie pokrytia reprezentujeme v maticovej forme. Formálne, generátor G sa učí mapovanie z vstupných údajov $\mathbf{X}^{n \times n}$ na generovanie výstupných údajov $\mathbf{Y}^{n \times n}$, $G : \mathbf{X} \rightarrow \mathbf{Y}$. Skutočné vzorky údajov (optimálne polohy UAV) vypočítané algoritmom core-sets reprezentuje matica $\mathbf{K}^{n \times n}$.

Objektívna funkcia generátora G minimalizuje pravdepodobnosť, že generované vzorky $G(\mathbf{X}) \rightarrow \mathbf{Y}$ budú diskriminátorom klasifikované ako umelé. Objektívna funkcia G je definovaná ako:

$$\min_G \mathcal{L}_{cGAN}(G) = \mathbb{E}_{\mathbf{Y}}[\log(1 - D(\mathbf{X}, \mathbf{Y}))]. \quad (8)$$

Diskriminátor D produkuje skalárnu hodnotu, ktorá reprezentuje pravdepodobnosť že \mathbf{Y} je vzorka produkovaná generátorom G a nie vzorka z rozloženia \mathbf{K} . Objektívna funkcia D je teda definovaná ako:

$$\max_D \mathcal{L}_{cGAN}(D) = \mathbb{E}_{\mathbf{X}, \mathbf{K}}[\log D(\mathbf{X}, \mathbf{K})] + \mathbb{E}_{\mathbf{Y}}[\log(1 - D(\mathbf{X}, \mathbf{Y}))]. \quad (9)$$

D aj G sú tréňované paralelne, za účelom minimalizácie straty $\mathcal{L}_{cGAN}(G)$ a súčasne maximalizácie straty $\mathcal{L}_{cGAN}(D)$. Autori článku [46] ďalej rozšírili tréningový proces o L1 normu (najmenšia absolútna chyba (MAE)). L1 norma je definovaná ako MAE medzi cieľovou distribúciou \mathbf{K} a generovanou vzorkou \mathbf{Y} . Vzhľadom na túto definíciu môžeme formulovať optimalizačný problém ako min-max problém cez $\mathcal{L}_{cGAN}(G, D)$ nasledovne:

$$\arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G), \quad (10)$$

$\lambda = 100$ je váhový koeficient [46].

C. Generátor a diskriminátor

Generátor G a diskriminátor D sú implementované podľa schémy vyobrazenej na Obr. 6 a), ktorá zobrazuje tréningový proces navrhovaného riešenia pomocou siete cGAN. Presný opis generátora a diskriminátora sa dajú nájsť v práci [46]. Zjednodušene, generátor G je implementovaný ako U-Net sieť so vstupom o veľkosti 256×256 , ktorý reprezentuje segmentáciu posudzovanej oblasti. Rozmer každej ďalšej vrstvy sa

zmenšuje 2-krát v oboch smeroch (riadky a stĺpce), s úzkym profilom o veľkosti 2×2 na siedmej vrstve. Discriminátor D je implementovaný ako bežná konvolučná neurónová sieť s veľkosťou vstupu 256×256 .

Keďže UAV zaujímajú iba niekoľko segmentov v skúmanej oblasti, cieľová matica \mathbf{Y} je mimoriadne riedka. Preto aplikácia konvenčnej stratovej funkcie pre cGAN, napr. norma L1, ako je uvedené v rovnici 10, vedie k slabému výkonu navrhovanej metódy.

Aby sme prekonalí tento problém, navrhujeme vlastnú stratovú funkciu viacvrstvového sum-poolingu, ktorá sa vypočíta podľa Alg. 1. Najprv algoritmus aplikuje pooling s filtrom veľkosti 1 a rovnakým paddingom, čo zodpovedá konvenčnému výpočtu strednej kvadratickej chyby (MSE) medzi cieľovou šablónou \mathbf{K} a výstupom generátora \mathbf{Y} . Cieľová šablóna \mathbf{K} je vzorka z rozsiahlej databázy, ktorá obsahuje UEs a zodpovedajúce optimálne polohy UAV vytvorené algoritmom jadrových množín. Potom zdvojnásobením veľkosti pooling-u počas každej iterácie poskytujeme generátoru dôležitú spätnú väzbu o priestorovej dislokácii generovaných polôh UAV \mathbf{Y} vo vzťahu k ich optimálnym pozíciám \mathbf{K} .

Algorithm 1 Viacvrstvová sum-pooling stratová funkcia

- 1: Vstupy: $\mathbf{Y}^{n \times n}$, $\mathbf{K}^{n \times n}$
 - 2: Inicializácie stratovej matice $\mathcal{L}_{pool} = \emptyset$
 - 3: Nastavenie sum-pooling filtra na $f = 1$, $stride = 1$, a $padding = 1$
 - 4: **while** $f \leq 64$ **do**
 - 5: Výpočet poolovanej verzie matice s optimálnymi pozíciami pomocou filtra o veľkosti f : $P_1 = pool(\mathbf{K}, f, padding)$
 - 6: Výpočet poolovanej verzie matice s generovanými pozíciami pomocou filtra o veľkosti f : $P_2 = pool(\mathbf{Y}, f, padding)$
 - 7: Výpočet štvorcovej chyby medzi vypočítanými dvoma maticami $Loss = (P_1 - P_2)^2$
 - 8: Prirátanie straty k celkovej strate $\mathcal{L}_{pool} = \mathcal{L}_{pool} + Loss$
 - 9: $f = 2 \times f$, $padding = 2 \times padding$
 - 10: Výstup: \mathcal{L}_{pool}
-

Nakoniec môžeme formulovať optimalizačnú funkciu cGAN s navrhnutou stratovou funkciou \mathcal{L}_{pool} nasledovne:

$$\arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{pool}(G). \quad (11)$$

D. Korekčný algoritmus na zamedzenie rozostrenia výstupu

Fáza nasadenia sa vykonáva online na určenie polohy UAV v reálnom čase, s rešpektovaním polohy UEs v priestore. Tu sa spoliehame na aplikáciu redukovanej architektúry cGAN, konkrétne na jej natrénovaný generátor G (pozri obr. 6 b)). Podľa svojej povahy generátor G generuje rozmazané obrázky, čo vedie k prítomnosti väčšieho počtu UAV, než je nevyhnutné. Aby sme tento problém riešili, v rámci fázy nasadenia je vyvinutý a zahrnutý algoritmus na korekciu.

Konvenčná aplikácia cGAN v oblasti spracovania obrazu nevyžaduje vysokú presnosť a často trpí rozostrením obrazu. V pôvodnej doméne spracovania obrazu to nie je veľkým problémom a môže byť potlačená pokročilými technikami spracovania obrazu. Avšak v našom prípade vedie rozostrenie výstupu ako vedľajší efekt k zvýšeniu počtu UAV mierne posunutých od seba v pomerne riedkej matici \mathbf{Y} .

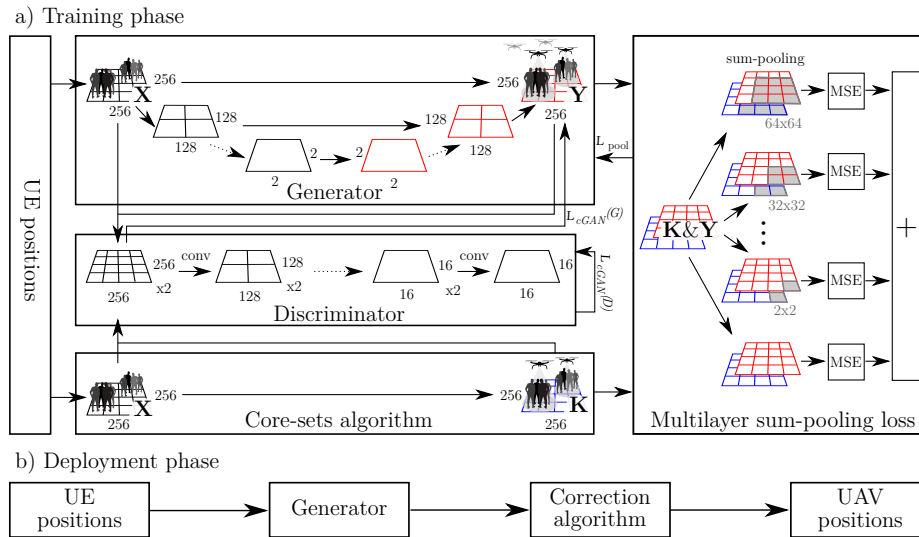
Na riešenie tohto problému sme použili korekčný mechanizmus, ktorý je navrhnutý nasledovne. Navrhovaný korekčný algoritmus berie riedku maticu \mathbf{Y} a na urýchlenie výpočtov ju transformuje na škálovateľnú schému ukladania súradníc, reprezentovanú množinou Y [48], obsahujúcej iba priestorové súradnice $[x, y]$, kde sú prítomné UAV. Poznamenajme, že transformácia f riedkej matice na komprimovanú sadu $f : \mathbf{Y} \rightarrow Y$ umožňuje rýchlejšie výpočty než s obvyklými riedkymi maticami. V prvom kroku algoritmus náhodne vezme UAV a jeho polohu $[x, y]$ z Y . Potom určí všetky ostatné UAV umiestnené v Y v perimetri veľkosti ϵ a vypočíta vzájomnú vzdialenosť. Vybrané UAV a susedné UAV v perimetri ϵ sú stiahnuté z Y . Skupina venovaných UAV je nahradená novým konkrétnym UAV uloženým v novej sade L so súradnicami $[\Delta x, \Delta y]$ vypočítanými ako priemer všetkých súradníc ktoré sa nachádzajú v relatívnej blízkosti. Výstup algoritmu je reprezentovaný sadou L , ktorá obsahuje výsledný počet UAV s ich koordinátami v priestore. V našom prípade, hyperparameter ϵ bol získaný na základe veľkého množstva výpočtových experimentov s cieľom maximalizovať stratovú funkciu 8. Detailný opis krokov algoritmu je sumarizovaný v 7. Obr. 7 graficky ilustruje problém rozostrenia UAV (červené kružnice) a korekčný mechanizmus Alg. 2 použitý na rozostrený výstup generátora (čierné kružnice).

Algorithm 2 korekčný algoritmus

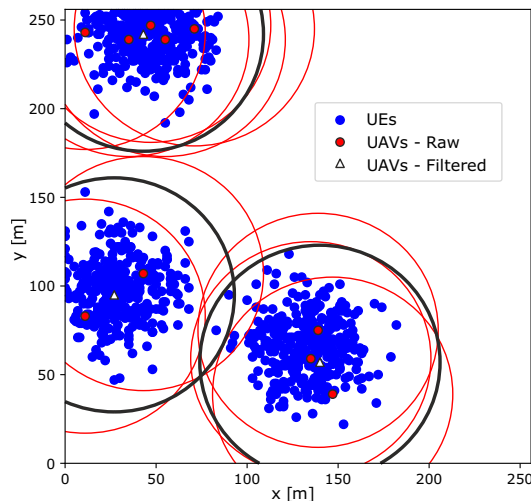
- 1: Vstup: surový výstup generátora G s pozíciami UAV $\mathbf{Y}^{n \times n}$
 - 2: Transformácia vstupnej matice: $f : \mathbf{Y} \rightarrow Y$
 - 3: Inicializácia prázdnej sady $L = \emptyset$ a prahovej hodnoty ϵ
 - 4: **while** Y nie je prázdna **do**
 - 5: Výber prvého prvku z Y , $y = pop(Y)$
 - 6: Výpočet vzdialenosti k všetkým zostávajúcim UAV v Y : $D_i = dist(y, Y_i), \forall i \in \langle 1, size(Y) \rangle$
 - 7: Výber m UAVs pre ktoré je vzájomná vzdialenosť nižšia ako prahová hodnota ϵ , $P = pop(Y_i, D_i < \epsilon)$
 - 8: Výpočet priemerných súradníc $[\Delta x, \Delta y]$ $\Delta x = \frac{\sum_{i=1}^m x_i}{m}, \forall x_i \in P, i \in \langle 1, size(P) \rangle$, $\Delta y = \frac{\sum_{i=1}^m y_i}{m}, \forall y_i \in P, i \in \langle 1, size(P) \rangle$
 - 9: Vloženie nových vypočítaných súradníc $[\Delta x, \Delta y]$ do L , $L = L \cup [\Delta x, \Delta y]$
 - 10: Výstup: L
-

E. Výsledky

Simulačný model je definovaný vo forme geometrického problému pokrytia diskom [49]. Cieľom problému je pokryť súbor UE v cieľovej oblasti minimálnym počtom UAV s daným polomerom pokrytia.



Obr. 6: a) Detailná schéma tréningovej fázy navrhnutého algoritmu. pozície UAVs generované pomocou core-set algoritmu slúžia ako tréningové vzorky a sú využívané iba v offline fáze. Na dôvažok zobrazujeme sum-pooling stratovú funkciu pomocou grafickej reprezentácie. b) Bloková schéma fázy nasadenia ktorá je spúšťaná online. Táto fáza pozostáva z generátora G a korekčného algoritmu pre predchádzanie rozostrenia výstupu [47].



Obr. 7: Fáza nasadenia: priamo predikované z generátora G (červené kruhy) a výstup korekčného mechanizmu (biele trojuholníky). $R = 4$, $p = 1000$ [47].

Výkonnosť sa hodnotí z dvoch hľadísk: kvality pokrytia a výpočtovej efektívnosti. Ako referenciu pre kvalitu pokrytia používame algoritmus core-set, ktorý poskytuje optimálne riešenie, ale nie je výpočtovo efektívny. Výpočtová efektívnosť sa hodnotí voči špirálovému algoritmu, ktorý je v súčasnosti najlepšou praxou medzi kvázi optimálnymi riešeniami z hľadiska výpočtovej efektívnosti [15].

1) *Hodnotenie zložitosti:* Sústredíme sa na analýzu zložitosti generátora G a navrhovaného postupu na filtrovanie počtu UAV na výstupe generátora G , čím z nášho preskúmania vynechávame zložitost tréningového pos-

Tabuľka I: Výpočtové parametre algoritmov

Alg.	Časová zložitosť	Tréning	Riešenie
Proposed	$\mathcal{O}(k^2)$	Áno	Heuristické
Core-sets	$\mathcal{O}(p^k)$	Nie	Deterministické
K-means	$\mathcal{O}(p^{dk+1})$	Nie	Heuristické
Spiral	$\mathcal{O}(k^3)$	Nie	Deterministické

tupu. Poznamenajme, že tréningový postup sa realizuje pred nasadením UAV (offline) a preto neovplyvňuje negatívne výpočtové požiadavky navrhovanej metódy v praktických realizáciách. V ďalšom texte teda používame termín *navrhovaný prístup* na označenie spoločnej aplikácie funkcie generátora G a navrhovaného korekčného algoritmu (t.j. fáza nasadenia). Ako základný scenár predstavujeme zložitost algoritmov core-set a spiral a konvenčného zhlukovacieho algoritmu K-means.

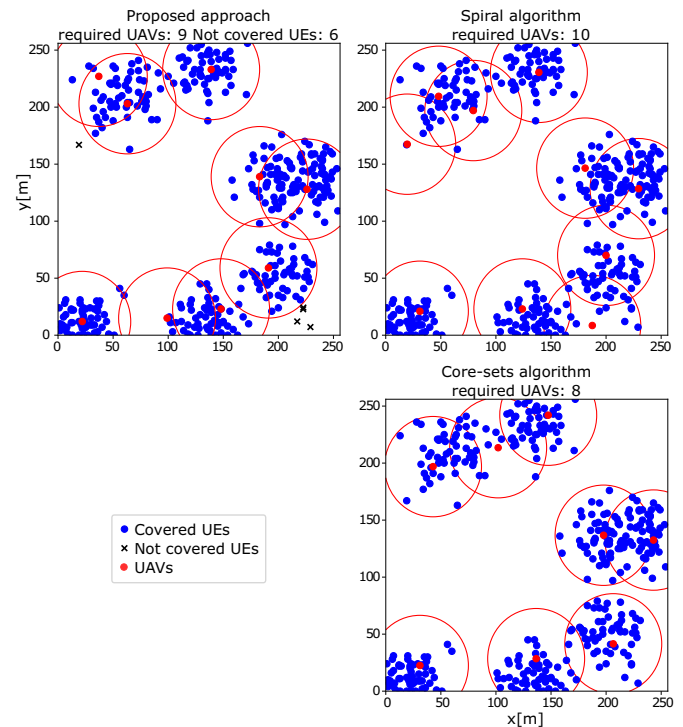
Označme počet UAV na výstupe generátora G ako y a počet filtrovaných UAV na výstupe korekčného mechanizmu ako k . Na základe mnohých numerických experimentov môžeme tvrdiť, že vo všeobecnosti platí $y \sim 2.4k$. Ako ukazuje pseudokód v Alg. 2, rutina pripomína typický postup triedenia poľa so zložitostou $\mathcal{O}(y^2)$. Keďže $y \sim 2.4k$, môžeme zložitost aproximovať ako $\mathcal{O}(k^2)$. Poznamenajme, že ide o najhorší prípad, pretože zmenšenie triediacej množiny vo všeobecnosti vedie k zmenšeniu počtu vykonaných vnútorných krokov. Časová zložitost analytickej riešenia poskytnutého algoritmom core-set je podľa definície $\mathcal{O}(p^k)$, kde p je počet UEs v oblasti. Časová zložitost špirálového algoritmu je $\mathcal{O}(k^3)$, čo je výrazne viac než u navrhovaného prístupu. Medzi konvenčnými, dobre známymi technikami strojového učenia má algoritmus K-means časovú zložitost $\mathcal{O}(p^{dk+1})$, kde d označuje počet dimenzií stavového priestoru (v našom prípade $d = 2$) [50]. Kým časová zložitost algoritmov core-set a K-

means závisí na počte UEs v cieľovej oblasti, pre špirálový a navrhovaný prístup je časová zložitosť nezávislá na počte UEs. Časová zložitosť špirálového algoritmu je kubická, zatiaľ čo náš prístup má v najhoršom prípade kvadratický čas. Ako nevýhodu nášho prístupu treba spomenúť, že tréningový postup musí byť vykonaný pred nasadením optimalizácie UAV, čo nie je prípad ostatných skúmaných algoritmov. Porovnanie časových zložítostí daných algoritmov je zobrazené v tabuľke I. Pre experimenty sme použili pracovnú stanicu vybavenú AMD Ryzen Threadripper 1950X s 64 GB RAM a simulácie boli vykonávané na GPU NVIDIA GeForce GTX 1080Ti 11 GB GPU. Simulátor bol vytvorený v Python ver. 3.6, na MS Windows 10 Pro.

F. Analýza pokrytia

Definujeme metriku pomeru pokrytia $R = D/r$, kde D je dĺžka strany cieľovej štvorcovej pokrývanej oblasti a r je polomer pokrytia jedného UAV. Testovali sme tri pomery pokrytia $R = 2, 4, 6$ pre dva rôzne počty UEs: 400 a 1000. cGAN bol tréningovaný iba raz pre každý R a rovnaké váhy boli prenesené na predpovedanie pozícií UAV pre rôzny počet UEs. Okrem toho bol počas simulácií upravený korekčný parameter ϵ na odstránenie efektu rozmazania UAV.

V tabuľke II vidíme priemerné výsledky simulácie navrhovaného algoritmu v porovnaní so špirálovým a algoritmom core-set. Štatistiky boli spriemerované cez dataset s veľkosťou 1000 vzoriek (snímky rôznych okamžitých pozícií UEs). Môžeme oceniť redukciu časovej zložítosti navrhovaného prístupu vo všetkých skúmaných scenároch. Navyše, počet potrebných UAV pre navrhovaný prístup je vo všeobecnosti bližšie k optimálnemu riešeniu core-set než k riešeniu špirálového algoritmu. Grafické reprezentácie pokrývajúcich riešení navrhovaného prístupu, špirálového algoritmu a algoritmu core-set sú zobrazené na obrázku 8. Ako nevýhodu navrhovaného prístupu musíme poznamenať, že pokrytie nie je optimálne a existuje veľmi málo UEs, ktoré nie sú pokryté (takisto je uvedené v tabuľke II). Tieto sú väčšinou umiestnené na okrajoch zhlukov UEs alebo sú odľahlé body, ktoré nezachytil generátor G . Pokiaľ je našim záujmom pokryť zhluky UEs s vysokou hustotou, nepovažujeme túto nevýhodu za veľký problém. Samozrejme, tieto UEs môžu byť ľahko pokryté tradičnou pozemnou infraštruktúrou, t.j. sieťovými prvky eNodeB. Vzhľadom na rastúci dopyt po mobilných aplikáciách a exponenciálny rast pripojených zariadení v inteligentných mestách môžeme predpokladať, že typický počet UEs v dohľadnej budúcnosti bude o rád vyšší než tie, ktoré boli simulované v tejto práci. Preto predpokladáme, že navrhovaný prístup s asymptotickou zložitosťou $\mathcal{O}(k^2)$ môže byť prijatý pre nasadenie leteckých bezdrôtových komunikačných systémov, ako aj pre iné aplikácie UAV, ktoré čelia problému pokrytia diskom. *Napokon, vzhľadom na to, že cGAN môže byť tréningovaný na akýkoľvek typ dát (t.j., zložitejšie datasety), navrhovaný algoritmus môže byť ľahko rozšírený na zložitú optimalizáciu pokrytia na základe viacerých kritérií výkonnosti bezdrôtovej siete, ako je dosiahnuteľný prietok a spektrálna účinnosť oblasti.* Takže riešenie založené na aplikácii cGAN



Obr. 8: Porovnanie výstupov navrhovaného riešenia, core-set algoritmu a špirálového algoritmu, $R = 6$, počet UE 1000. Okrem faktu, že navrhované riešenie je rýchlejšie, vyžaduje aj menej UAV ako špirálový algoritmus [47].

navrhnuté v tejto práci sa zdá byť uskutočniteľné, otvára ďalšie sľubné výskumné smerovania v tejto oblasti.

V. ZÁVER A PRÍNOSY PRE VEDNÝ ODBOR

Statická topológia siete sa môže stať nevhodnou pre potreby dynamických zmien v modernom svete. Postupné zvyšovanie počtu mobilných koncových zariadení a zvyšovanie mobility koncových používateľov spolu so zvyšovaním počtu zariadení IoT vyzýva na potrebu dynamickejšej topológie. Pandémia Covid-19 v roku 2020 dokázala potrebu topológie, ktorá sa dokáže rýchlo prispôbiť zmenám správania koncových používateľov. Veľmi sľubným prístupom k vytvoreniu dynamickej topológie siete je využitie malých buniek namontovaných na vozidlách, a to buď na pozemných vozidlách, ale aj na lietadlách. Pozemné vozidlá sú obmedzené pravidlami cestnej premávky a prekážkami. V prípade vzdušných uzlov buniek je obmedzení oveľa menej, pretože sa ľahko dostanú cez prekážky ako sú stromy, rieky a budovy a môžu dosiahnuť pešie zóny. Použitie dynamickej topológie je však nákladnejšie, pretože zariadenia sú buď drahšie ako statické a prevádzkové náklady sú vyššie. Aby sa minimalizovali náklady na sieť, je potrebné nasadiť najoptimálnejší počet buniek namontovaných na UAV na najoptimálnejšie pozície. Problém optimálneho pokrytia koncových používateľov s minimálnym počtom UAV je známy ako k -stredový problém a nemôže byť riešený v krátkom čase pomocou aktuálne známych algoritmov podľa potrieb 5G s minimálnou latenciou a najvyššou možnou šírkou pásma.

Tabuľka II: Časové charakteristiky každého algoritmu podľa R a počtu UEs (vľavo), príslušné charakteristiky pokrytia a priemerný počet potrebných UAV (vpravo). Pre špirálový a core-set algoritmus je pokrytie podľa definície 100%. Na porovnanie, navrhovaný prístup poskytuje obrovské zlepšenie časovej zložitosti doplnené len o miernu degradáciu výkonnosti v oblasti pokrytia (pozri tiež obr. 8). Poznamenajme, že požiadavky na časovú zložitost' pre špirálový a core-set algoritmus výrazne rastú s rastúcim počtom UEs, zatiaľ čo pre navrhovaný prístup zostávajú takmer konštantné. Najlepší dosiahnutý čas pre dané nastavenie je zvýraznený tučným písmom.

Nastavenia			Čas [s]			Priem. pokrytie [%] (Navrhnuté riešenie)	Priemerný počet UAV		
R	UEs	ϵ	Navrhnuté riešenie	Špirál	Core-Sets		Navrhnuté riešenie	Špirál	Core-sets
2	400	160	0,07191	0,83652	0,16217	98,9	2	1,8	1,7
	1000	160	0,08767	5,49683	0,71645	99,8	1,9	2	1,9
4	400	60	0,05183	0,45599	5,12239	98,9	4,1	4,3	3,7
	1000	60	0,05295	2,80836	38,92758	99,7	3,9	4,4	3,7
6	400	40	0,0804	0,32621	264,4874	97,6	7,1	7,6	6,5
	1000	40	0,0967	1,41399	8493,39611	99,1	7,9	8,5	6,8

Náklady môžu byť tiež znížené optimalizáciou spotreby energie UAV-BS, čo je v tejto práci podrobne skúmané.

Bol poskytnutý realistický scenár 3D umiestnenia UAV v mestských oblastiach s vetrom. Existujúci model spotreby energie pohonu UAV bol v tejto práci rozšírený o pokrytie spotreby energie súvisiacej s vetrom, ktorá nebola predtým zohľadnená. Charakteristiky vetra boli modelované pomocou softvéru CFD využívajúceho model $k - \epsilon$ založený na rovniciach RANS. Použitím modelu učenia v súbore pozostávajúceho z KNN, RF a DNN bolo možné predpovedať pozície pre kruhový let, ktorý bol zvolený namiesto jednoduchého vznášania na minimalizáciu spotreby energie zvažovaného dronu. Simulácie ukázali, že tento prístup je schopný ušetriť takmer 50% energie v porovnaní s vznášaním sa na optimálnych pozíciách kapacity poskytnutých konvenčnými algoritmi. Navrhnutá metóda bola publikovaná v IEEE Communication Letters ako *Energy Consumption Optimization for Rotary-wing UAV Base Stations with Wind Compensation* [38].

Okrem toho bola vykonaná analýza, aby sa opísali najnovšie algoritmy a možné vylepšenia pri optimálnom umiestňovaní UAV. Sieť GAN bola použitá s cieľom efektívne riešiť problém optimálneho umiestnenia UAV v dynamickom RAN. V tomto prípade základné vedomosti o neurónových sieťach a široko používaných paradigmaty nestačili kvôli povaha dát poskytnutých neurónovej sieti. Bola navrhnutá vlastná stratová funkcia na riešenie problému veľmi riedkych matíc. Použitie tejto funkcie v sieti GAN vyriešilo problém, keď GAN generovala len nulové matice. Štandardná MSE strata nebola schopná poskytnúť správnu spätnú väzbu. Sieť potom bola schopná generovať pozície UAV s nízkou varianciou voči skutočným pozíciám UAV poskytnutým špirálovým algoritmom alebo core-sets. Problém nesprávneho (vyššieho) počtu generovaných pozícií bol vyriešený použitím algoritmu podobného tomu, ktorý sa používa v špirálovom algoritme. Počet iterácií bol nižší ako v prípade špirálovom algoritmu a navrhnutý algoritmus sa ukázal byť časovo efektívnejší. Avšak na plne náhodne rozmiestnených pozíciách v priestore tento prístup nebol schopný dobre sa učiť a počet UAV a ich pozície neodzrkadľovali ideálny prípad. Tento výskum bol publiko-

vaný v International Journal of Distributed Sensor Networks ako *Fast and computationally efficient generative adversarial network algorithm for unmanned aerial vehicle-based network coverage optimization* [47]. Počas vykonávania tohto výskumu sme tiež publikovali viaceré ďalšie články, ktoré sa snažili buď generovať novú databázu trajektórií UEs [51], [52] a riešiť zhlukovanie [53].

Objavy opísané v tejto práci otvárajú priestor pre ďalší výskum možností použitia generatívnych modelov v organizácii topológie siete 5G. Budúca práca môže mať prospech z použitia skutočných dát, namiesto spoliehania sa na simulované dáta a kolekcie z databáz, ktoré nie sú úplne vhodné pre takéto prípady použitia. Rôzne typy GAN, ako je Wasserstein GAN [54], by mohli zlepšiť výsledky optimálneho umiestnenia UAV. Výskum GAN by mohol mať prospech z použitia 3D Chamferovej stratovej funkcie [55].

V oblasti výskumu spotreby energie súvisiacej s vetrom môžu byť použité relevantnejšie neurónové siete pre predikciu charakteristik vetra, ako sú fyzikálne informované neurónové siete (PINNs) [56]. Fyzickú topológiu možno vytvoriť pokročilejšími spôsobmi, ako napr. cez NeRF siete [57] alebo LIDAR [58], ktoré môžu zlepšiť celkovú použiteľnosť navrhovaného prístupu. Do budúcnosti môžeme navrhnuť lacný, realistický a časovo efektívny prístup, ktorý by sa mal zbaviť použitia softvéru CFD a CAD. Namiesto použitia CAD na definovanie geometrie máme v úmysle použiť NeRF siete, ktoré sú schopné generovať 3D bodové mračná alebo mriežky z fotografií. Všetko, čo treba urobiť, je odfotiť miesto z rôznych uhlov a natrénovať sieť NeRF. Mriežku potom možno použiť v softvéri CFD na predpovedanie vetra, avšak efektívnejšie je využiť PINN. Ak je stratová funkcia PINN dobre nastavená, môže sa učiť a predpovedať charakteristiky vetra bez tréningových dát z CFD. Použitie takéhoto prístupu by viedlo k technike, ktorá by mohla úplne odstrániť dodatočný softvér, ktorý zvyčajne vyžaduje ľudskú interakciu od vysoko skúseného odborníka. Navyše, PINN je tiež schopná predpovedať šírenie signálu v priestore, čo by mohlo byť použité na ešte väčšie vylepšenie prístupu.

Hlavnými prínosmi pre vednú disciplínu sú teda:

- 1) Generovanie umelej dátovej sady obsahujúcej trajektórie

- fiktívnych užívateľov. Táto metóda umožňuje vytvoriť dátovú sadu pre budúce spracovanie inými algoritmi.
- 2) Použitie GAN siete pre detekciu a odstránenie anomálií v jednotlivých vzorkách užívateľských trajektórií. Práca je návaznosťou na predchádzajúcu aplikáciu, do značnej miery dokáže odstrániť nedokonalosti generovaných trajektórií.
 - 3) Použitie GAN siete pre optimalizáciu algoritmu na rozloženie pohyblivých bázových staníc. Vďaka aplikácii GAN siete s vlastnou stratovou funkciou bolo možné značne urýchliť proces rozloženia dronov v priestore s minimálnou zmenou pokrytia oproti optimu.
 - 4) Použitie neurónovej siete v spojení s algoritmi náhodného lesa a k najbližších susedov pre optimálne umiestnenie bázevej stanice na lietajúcom drone vo vetre. Aplikácia je schopná ušetriť vo veternom prostredí značnú časť energie, zatiaľ čo jej náročnosť je statická k meniacemu sa počtu koncových zariadení. Experimentálnou formou bola preukázaná úspora až 47

POĎAKOVANIE

Táto práca bola financovaná grantami Slovenskej agentúry pre vedu a výskum (čísla projektov APVV-18-0214 a APVV SK-CZ-RD-21-0028) a Vedeckej grantovej agentúre Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky pod číslom: 1/0268/19.

REFERENCIE

- [1] U. Cisco, "Cisco annual internet report (2018–2023) white paper," 2020.
- [2] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5g wireless communication networks," *IEEE communications magazine*, vol. 52, no. 2, pp. 122–130, 2014.
- [3] H. Ritchie, "Google mobility trends: How has the pandemic changed the movement of people around the world?" <https://ourworldindata.org/covid-mobility-trends>, accessed: 2021-01-19.
- [4] X. JIANG, M. SHENG, N. ZHAO, C. XING, W. LU, and X. WANG, "Green uav communications for 6g: A survey," *Chinese Journal of Aeronautics*, vol. 35, no. 9, pp. 19–34, 2022.
- [5] M. Strada, "Uavs as mobile base stations," *arXiv preprint arXiv:2005.06523*, 2020.
- [6] A. Merwaday, A. Tuncer, A. Kumbhar, and I. Guvenc, "Improved throughput coverage in natural disasters: Unmanned aerial base stations for public-safety communications," *IEEE Vehicular Technology Magazine*, vol. 11, no. 4, pp. 53–60, 2016.
- [7] Z. Becvar, M. Nikooroo, and P. Mach, "On energy consumption of airship-based flying base stations serving mobile users," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 7006–7022, 2022.
- [8] J. Won, D.-Y. Kim, Y.-I. Park, and J.-W. Lee, "A survey on uav placement and trajectory optimization in communication networks: From the perspective of air-to-ground channel models," *ICT Express*, 2022.
- [9] H. A. Fayed and A. F. Atiya, "A mixed breadth-depth first strategy for the branch and bound tree of euclidean k-center problems," *Computational Optimization and Applications*, vol. 54, no. 3, pp. 675–703, 2013.
- [10] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.
- [11] M. Inaba, N. Katoh, and H. Imai, "Applications of weighted voronoi diagrams and randomization to variance-based k-clustering," in *Proceedings of the tenth annual symposium on Computational geometry*, 1994, pp. 332–339.
- [12] P. K. Agarwal, S. Har-Peled, K. R. Varadarajan *et al.*, "Geometric approximation via coresets," *Combinatorial and computational geometry*, vol. 52, no. 1, 2005.

- [13] M. Bădoiu, S. Har-Peled, and P. Indyk, "Approximate clustering via core-sets," in *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, 2002, pp. 250–257.
- [14] P. K. Agarwal and C. M. Procopiuc, "Exact and approximation algorithms for clustering," *Algorithmica*, vol. 33, no. 2, pp. 201–226, 2002.
- [15] J. Lyu, Y. Zeng, R. Zhang, and T. J. Lim, "Placement optimization of uav-mounted mobile base stations," *IEEE Communications Letters*, vol. 21, no. 3, pp. 604–607, 2016.
- [16] V. Sharma, D. N. K. Jayakody, and K. Srinivasan, "On the positioning likelihood of uavs in 5g networks," *Physical Communication*, vol. 31, pp. 1–9, 2018.
- [17] V. Sharma, R. Kumar, and P. S. Rana, "Self-healing neural model for stabilization against failures over networked uavs," *IEEE Communications Letters*, vol. 19, no. 11, pp. 2013–2016, 2015.
- [18] T. R. Jensen and B. Toft, *Graph coloring problems*. John Wiley & Sons, 2011, vol. 39.
- [19] H. X. Pham, H. M. La, D. Feil-Seifer, and L. V. Nguyen, "Autonomous uav navigation using reinforcement learning," *arXiv preprint arXiv:1801.05086*, 2018.
- [20] Y. Y. Munaye, H.-P. Lin, A. B. Adege, and G. B. Tarekgn, "Uav positioning for throughput maximization using deep learning approaches," *Sensors*, vol. 19, no. 12, p. 2775, 2019.
- [21] M. A. Sayeed, R. Kumar, V. Sharma, and M. A. Sayeed, "Efficient deployment with throughput maximization for uavs communication networks," *Sensors*, vol. 20, no. 22, p. 6680, 2020.
- [22] J. Tang, J. Song, J. Ou, J. Luo, X. Zhang, and K.-K. Wong, "Minimum throughput maximization for multi-uav enabled wpcn: A deep reinforcement learning method," *IEEE Access*, vol. 8, pp. 9124–9132, 2020.
- [23] X. Zhang, H. Zhao, J. Wei, C. Yan, J. Xiong, and X. Liu, "Cooperative trajectory design of multiple uav base stations with heterogeneous graph neural networks," *IEEE Transactions on Wireless Communications*, 2022.
- [24] M. Alzenad, A. El-Keyi, F. Lagum, and H. Yanikomeroglu, "3-D placement of an unmanned aerial vehicle base station (UAV-BS) for energy-efficient maximal coverage," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 434–437, 2017.
- [25] C. Di Franco and G. Buttazzo, "Energy-aware coverage path planning of UAVs," in *2015 IEEE international conference on autonomous robot systems and competitions*. IEEE, 2015, pp. 111–117.
- [26] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, 2019.
- [27] Y. Zeng and R. Zhang, "Energy-efficient uav communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, 2017.
- [28] B. Li, Q. Li, Y. Zeng, Y. Rong, and R. Zhang, "3D trajectory optimization for energy-efficient UAV communication: A control design perspective," *IEEE Transactions on Wireless Communications*, 2021.
- [29] M. Nikooroo and Z. Becvar, "Optimizing transmission and propulsion powers for flying base stations," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–8.
- [30] —, "Optimal positioning of flying base stations and transmission power allocation in noma networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1319–1334, 2022.
- [31] S. P. Gopi and M. Magarini, "Reinforcement learning aided uav base station location optimization for rate maximization," *Electronics*, vol. 10, no. 23, p. 2953, 2021.
- [32] M. Najla, Z. Becvar, P. Mach, and D. Gesbert, "Positioning and association rules for transparent flying relay stations," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1276–1280, 2021.
- [33] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, 2017.
- [34] M. Yang, S.-W. Jeon, and D. K. Kim, "Optimal trajectory for curvature-constrained UAV mobile base stations," *IEEE Wireless Communications Letters*, vol. 9, no. 7, pp. 1056–1059, 2020.
- [35] M. K. Shehzad, A. Ahmad, S. A. Hassan, and H. Jung, "Backhaul-aware intelligent positioning of UAVs and association of terrestrial base stations for fronthaul connectivity," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 4, pp. 2742–2755, 2021.
- [36] O. Reynolds, "IV. On the dynamical theory of incompressible viscous fluids and the determination of the criterion," *Philosophical transactions of the royal society of London.(a.)*, no. 186, pp. 123–164, 1895.

- [37] J. Yang, T. Michael, S. Bhushan, A. Hanaoka, Z. Wang, and F. Stern, "Motion prediction using wall-resolved and wall-modeled approaches on a cartesian grid," in *Proc. of the 28th Symposium on Naval Hydrodynamics (USA, Pasadena)*. Citeseer, 2010.
- [38] M. Ruzicka, Z. Becvar, and J. Gazda, "Energy consumption optimization for uav base stations with wind compensation," *IEEE Communications Letters*, 2023.
- [39] M. A. Ganaie, M. Hu *et al.*, "Ensemble deep learning: A review," *arXiv preprint arXiv:2104.02395*, 2021.
- [40] Y. Qu, J. Zhang, R. Li, X. Zhang, X. Zhai, and S. Yu, "Generative adversarial networks enhanced location privacy in 5g networks," *Science China Information Sciences*, vol. 63, no. 12, pp. 1–12, 2020.
- [41] B. Hughes, S. Bothe, H. Farooq, and A. Imran, "Generative adversarial learning for machine learning empowered self organizing 5g networks," in *2019 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2019, pp. 282–286.
- [42] B. Wu, L. Liu, Y. Yang, K. Zheng, and X. Wang, "Using improved conditional generative adversarial networks to detect social bots on twitter," *IEEE Access*, vol. 8, pp. 36 664–36 680, 2020.
- [43] M. H. Haghghat and J. Li, "Intrusion detection system using voting-based neural network," *Tsinghua Science and Technology*, vol. 26, no. 4, pp. 484–495, 2021.
- [44] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, pp. 2672–2680, 2014.
- [46] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [47] M. Ružička, M. Vološin, J. Gazda, T. Maksymyuk, L. Han, and M. Dohler, "Fast and computationally efficient generative adversarial network algorithm for unmanned aerial vehicle-based network coverage optimization," *International Journal of Distributed Sensor Networks*, vol. 18, no. 3, p. 15501477221075544, 2022.
- [48] S. Hossain and T. Steihaug, "Sparse matrix computations with application to solve system of nonlinear equations," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 5, pp. 372–386, 2013.
- [49] A. Srinivas, G. Zussman, and E. Modiano, "Construction and maintenance of wireless mobile backbone networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 239–252, 2009.
- [50] J. R. Shewchuk, "Delaunay refinement algorithms for triangular mesh generation," *Computational geometry*, vol. 22, no. 1-3, pp. 21–74, 2002.
- [51] M. Ružička, M. Vološin, J. Gazda, and T. Maksymyuk, "The extension of existing end-user mobility dataset based on generative adversarial networks," in *2020 30th International Conference Radioelektronika (RADIOELEKTRONIKA)*. IEEE, 2020, pp. 1–4.
- [52] M. Ružička, M. Vološin, J. Gazda, T. Maksymyuk, and Y. Bobalo, "Correcting defective trajectories using conditional gan," in *2021 IEEE 4th International Conference on Advanced Information and Communication Technologies (AICT)*. IEEE, 2021, pp. 206–211.
- [53] M. Ružička and M. Dopiriak, "Clustering using conditional generative adversarial networks (cgans)," in *2023 33rd International Conference Radioelektronika (RADIOELEKTRONIKA)*. IEEE, 2023, pp. 1–6.
- [54] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.
- [55] J. Lu, Z. Li, J. Bai, and Q. Yu, "Oriented and directional chamfer distance losses for 3d object reconstruction from a single image," *IEEE Access*, vol. 10, pp. 61 631–61 638, 2022.
- [56] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.
- [57] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [58] B. Douillard, J. Underwood, N. Kuntz, V. Vlaskine, A. Quadros, P. Morton, and A. Frenkel, "On the segmentation of 3d lidar point clouds," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 2798–2805.

Výskum a diagnostika štandardných izolačných materiálov pre izolátory

Luboš Šárpataky

Katedra elektroenergetiky
FEI TUKE
Košice, Slovenská republika
lubos.sarpataky@tuke.sk

Bystrík Dolník

Katedra elektroenergetiky
FEI TUKE
Košice, Slovenská republika
bystrik.dolnik@tuke.sk

Abstrakt— Izolátory sú dôležitou súčasťou elektroenergetického systému a ich správne fungovanie je jedným z hlavných pilierov spoľahlivosti dodávky elektrickej energie. Na to, aby izolátory vydržali zaťaženie, na ktoré sú konštruované, je dôležité dohliadať na ich bezchybný stav. Problémy s chybnými kusmi sa eliminujú už po ich výrobe typovou skúškou a preto je najväčším problémom vplyv prostredia počas ich prevádzky. Prostredie, v ktorom sa izolátory nachádzajú ovplyvňuje ich životnosť, elektrické straty a taktiež spoľahlivosť, kde môže nastať vplyvom znečistenia a vysokej vlhkosti rapidný pokles preskakového napätia na kritické hodnoty. Táto práca sa venuje opisu a poznaniu izolátorov, environmentálnym vplyvom na izolátory ako aj poruchám a poškodeniam, ktoré môžu vďaka týmto vplyvom nastať. Na elimináciu rizika porúch a poškodení sa vyvíjajú rôzne metódy, ktoré dokážu odhadnúť stav povrchu izolátorov ako aj vnútorné poruchy. Táto práca opisuje nové metódy, ktoré využívajú stratový činiteľ a zvodový prúd na diagnostiku stavu povrchu izolátorov. Na povrch izolátora sú aplikované vplyvy v podobe umelého znečistenia a zmeny vlhkosti. Stratový činiteľ je vyhodnocovaný a skúmaný ako diagnostická veličina na určenie závažnosti znečistenia povrchu izolátora. Z meraní je vyhodnotená vhodnosť použitia stratového činiteľa pri nízkom aj vysokom napätí a pri rôznych frekvenciách. Výsledky meraní poukazujú na výhody použitia stratového činiteľa.

Kľúčové slová— *Izolátory, poruchy na izolátoroch, stratový činiteľ, vlhkosť, znečistenie*

I. ÚVOD

Vývojom izolácií sa zaoberá mnoho výskumných pracovísk a vynakladá sa naň mnoho finančných prostriedkov. Izolátory, ktoré slúžia na ochranu živých častí pred dotykom s neživými časťami konštrukcií a prístrojov, sú jedným z najdôležitejších izolačných prvkov najmä pri prenose a rozvoде elektrickej energie na vysokonapäťových vedeniach.

Prvé izolátory boli používané už v deväťdesiatych rokoch osemnásteho storočia na telegrafných linkách s nízkym napätím. Tieto izolátory boli vyrábané z dreva, cementu a neskôr zo skla a mali nízku spoľahlivosť. Výroba spoľahlivejších izolátorov začala koncom devätnásteho storočia pri rozvoji priemyslu, kedy rástla potreba vysokého napätia a taktiež izolátorov. Do sedemdesiatych rokov dvadsiateho storočia dominovali porcelánové a sklenené izolátory. V tejto dobe však začala éra nekeramických izolátorov. Keramické izolátory sa sice

používajú dodnes, no ich používanie sa značne obmedzilo a nekeramické kompozitné izolátory sa začali používať v oveľa väčšej miere [1]-[4].

V oblasti vývoja izolačných materiálov pokračuje výskum používaných materiálov, kde sa odstraňujú ich nedostatky, ktoré sú spojené najmä so životnosťou materiálov. Ich životnosť je ovplyvnená rôznymi faktormi ako je počasie, teplota, znečistenie, či UV žiarenie. V súčasnosti sa taktiež kladie dôraz na environmentálnu stopu, ktorú izolátor zanechá. To sa týka použitých surovín na výrobu, výrobný proces a likvidácie po vyradení z prevádzky. V budúcnosti budú preferované izolátory vyrábané z recyklovaných materiálov, no ich výskum musí prejsť ešte dlhým procesom.

Výskum a diagnostika izolátorov sa vyvíja rôznymi smermi, no najnovšie trendy kladú dôraz na meranie a diagnostiku, ktorá môže byť realizovaná online. Štandardné metódy ako napríklad vizuálna kontrola sú naďalej používané, no pre prevádzkovateľa prenosových a distribučných vedení predstavujú nemalé náklady. Výbava, cestovanie a kvalifikovaný personál sú nutné pre diagnostiku izolátorov, ktorá sa vykonáva priamo v teréne. Aj na obsluhu prístrojov ako je kamera na nočné videnie, prístroj na infračervenú termografiu alebo na akustickú emisiu, je potrebné mať školený personál, ktorý spĺňa požiadavky na elektrotechnickú spôsobilosť. Vďaka modernizácii technológií je možný zber a analýza veľkého množstva dát. Preto sa výskum sústreďuje na online meranie, kde je zber dát nepretržitý a systém dokáže vyhodnocovať merané veličiny okamžite, čím sa dá zabrániť zlyhaniam izolátorov a tým sa zvýši spoľahlivosť prenosu a rozvođu elektrickej energie.

Článok sa venuje diagnostike znečistených izolátorov v laboratórnom prostredí. Na diagnostiku sa využíva stratový činiteľ, ktorý je bežne používaný na vyhodnotenie stavu izolácií káblov, transformátorov a iných vysokonapäťových zariadení. Stratový činiteľ sa na izolátoroch používal len na vyhodnotenie starnutia povrchu, no táto práca poukazuje na využiteľnosť tejto veličiny aj na vyhodnotenie znečistenia izolátorov. Na porovnanie a potvrdenie výsledkov merania stratového činiteľa sa súčasne meral aj zvodový prúd, ktorý sa na určenie kontaminácie izolátorov už bežne používa.

II. IZOLÁTORY

Izolácia vonkajších vedení sa rieši najmä pomocou izolátorov. Izolátory vysokého napätia sú zariadenia určené na flexibilnú a pevnú podporu elektrických vodičov alebo zariadení a na izolovanie vodičov a zariadení od zeme alebo od iných vodičov a zariadení.

Izolátor je prispôbený svojej funkcii, na ktorú bol vyrobený a to svojím tvarom, druhom a akosťou izolantu. Čo sa týka elektrických vlastností, medzi najdôležitejšie patria elektrická pevnosť, elektromechanická pevnosť a vodivosť (najmä povrchová).

Izolátory zvyčajne pozostávajú z vodivých a izolačných materiálov. Existuje mnoho spôsobov ako ich klasifikovať [5], [6]:

- hlavný účel použitia (energetické, trolejové, oznamovacie),
- umiestnenie izolátora (vonkajšie a vnútorné),
- spôsob upevnenia (závesné, podperné),
- konštrukčné riešenie izolátora (tanierové, tyčové),
- materiál izolátora (porcelánový, sklenený, polymérny),
- dĺžka prieraznej dráhy (neprierazné, prierazné).

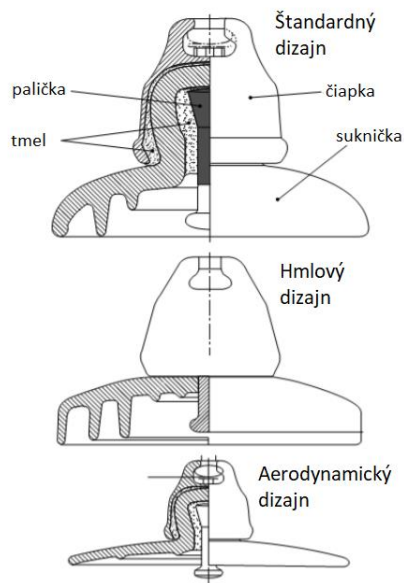
A. Klasifikácia podľa konštrukčného riešenia – tanierové izolátory

Izolačná časť tanierových izolátorov má tvar taniera alebo zvonu, ktorá je nazývaná aj strieška. Závesná (upevňovacia) časť pozostáva z čiapky (vrchná časť upevňovacieho systému) a paličky (spodná časť upevňovacieho systému), ktoré sú ku sebe spojené osovo. Tanierové izolátory sa v minulosti používali hlavne na vonkajšie vedenia všetkých napätových úrovní. Tieto izolátory sú (klasifikované ako prierazné kvôli malej vzdialenosti elektród, čo umožňuje vznik elektrického prierazu).

Na vyššie napätia sa tanierové izolátory skladajú do reťazcov. Pre napätovú hladinu 22 kV sa používa reťazec z dvoch až troch izolátorov, pre napätovú hladinu 110 kV to je šesť až sedem izolátorov, 220 kV vedenie potrebuje trinásť izolátorov a 400 kV vedenie dvadsať až dvadsaťtri izolátorov.

Tanierové izolátory sa vyznačujú svojou pevnosťou v ťahu, čo nadobudli vďaka kónickému tvaru paličiek vložených do tmelu, ktorý ich pevne spojil s čiapkami. Kovové časti sú slabinou týchto izolátorov. Majú schopnosť obmedziť šírenie oblúku, no výrazne zvyšujú hmotnosť izolátora. Taktiež sú náchylné na koróziu a zvyšujú elektrické namáhanie izolačného materiálu.

Tvary striešok sú na hornej časti povrchu podobné, pričom horná časť má hladký tvrdý povrch, aby sa zabránilo hromadeniu nečistôt a vlhkosti a sklon sukničky je stále väčší ako 5°. To slúži na podporu samočistenia a stekania nahromadenej vody. Spodné časti striešok sa líšia podľa použitia. Pre nadmerne znečistené miesta a pobrežné oblasti sa využívajú izolátory s predĺženou povrchovou cestou a nazývajú sa hmlové izolátory. Ďalší typ tanierového izolátora môže mať aerodynamický dizajn, čo sa využíva v púštnych oblastiach.



Obr. 1 Typy tanierových izolátorov

B. Klasifikácia podľa konštrukčného riešenia – tyčové izolátory

Reťazce tvorené z tanierových izolátorov sú postupne nahradzované tyčovými izolátormi. Ich výhodou je neprieraznosť a rovnomernejšie rozloženie napätie. Sú konštruované na báze keramiky alebo kompozitných materiálov. Častejšie sa používajú kompozitné materiály, ktoré sú ľahšie a lepšie odolávajú vonkajšiemu prostrediu. Tyčové izolátory majú iba dve kovové upevňovacie armatúry [7].

C. Klasifikácia podľa materiálu izolátora – porcelánové izolátory

Porcelán je najviac používaný materiál na výrobu izolátorov. Tento porcelán je zvyčajne tvorený približne z 50 % ílom, z 25 % kremeňom alebo oxidom hliníovým a z 25 % živcom. Základom porcelánových materiálov je silná chemická väzba medzi kremíkom a kyslíkom. To zabezpečuje stabilné vlastnosti voči chemickému pôsobeniu na porcelán a taktiež odolnosť voči vysokým teplotám. To však spôsobuje vysokú hodnotu povrchovej energie, čo má za následok nedostatočné odpudzovanie vody, v dôsledku čoho sa taktiež ľahko znečisťujú vo vonkajšom prostredí. Tradičné izolátory pre prenosové a distribučné vedenia sa spájajú do izolátorových reťazcov, ktoré sú tvorené jednotlivými článkami. Základné typy porcelánových izolátorov sú tanierové a tyčové. Najčastejšie používané typy armatúr sú vidlicové a čiapkové, kde vidlicové majú vyššiu hmotnosť ako čiapkové pri rovnakej mechanickej odolnosti [5], [6], [8].

Okrem porcelánovej striešky sa izolátor skladá aj z dvoch elektród (čiapky a paličky), ktoré sú pripevnené na izolátor cementovou kašou. Porcelánová časť izolátora sa vyrába z vysoko rafinovanej hliny, ktorá sa lisuje do foriem a tie nám vytvárajú požadovaný tvar. To sa následne natrie glazúrou a vypáli v peci. Glazúra vyhladí povrch izolátora, čo napomáha

ku zníženiu zmáčania a taktiež znižuje usadzovanie nečistôt. Rebrá sú tvarované na spodnej časti porcelánovej sukničky, aby sa predĺžila zvodová vzdialenosť a aby sa znížil gradient napätia pri danom napätí. Pri rebrách treba dbať na pomer hĺbky a hrúbky. Pri zlom pomere sa rebrá stávajú krehkými a sú náchyľnejšie na poškodenie napríklad zlou manipuláciou [9], [10].

D. Klasifikácia podľa materiálu izolátora – sklenené izolátory

Aj napriek tomu, že začiatkom 21. storočia sa začali kompozitné izolátory využívať vo väčšom množstve, dnes je situácia medzi sklenenými, porcelánovými a kompozitnými izolátormi vyrovnaná. Sklenené izolátory sú stále používané a budujú sa nové fabriky na ich výrobu. Najväčšou výhodou oproti porcelánovým izolátorom je ich nižšia cena. Ďalšou výhodou sklenených izolátorov je ich využitie v rôznych poveternostných podmienkach od arktických až po tropické [11]-[13].

Sklenené izolátory sa zvyčajne vyrábajú pri vysokých teplotách zmiešaním rôznych materiálov vrátane vápna a kremenného prášku. Následne sa náhle ochladia vo forme, ktorá je požadovaná pre tvar izolátora. Aby boli splnené mechanické požiadavky vysokonapäťových izolátorov, veľká väčšina sklenených jednotiek je „tvrdeného“ typu na rozdiel od „žilhaného“ typu. Proces vytvrdzovania zahŕňa zrýchlené ochladzovanie povrchu izolátora, zatiaľ čo vnútro vyrábaného skla sa ochladzuje pomalšie. Rozdielna rýchlosť tuhnutia vytvára tlakové predpínanie vonkajších vrstiev, aby sa účinne zabránilo tvorbe povrchových mikrotrhlín a zadržalo šírenie trhlín [14]-[16].

E. Klasifikácia podľa materiálu izolátora – kompozitné izolátory

Polymérne materiály začali byť vo veľkom využívané najmä kvôli ich výbornej spracovateľnosti a taktiež cene. Prvý polymér, ktorý bol použitý na výrobu izolátorov bola cykloalifatická epoxidová živica. Tá sa dodnes používa na výrobu elektrických izolátorov pre vnútorné a vonkajšie aplikácie. Vonkajšie polymérne izolátory sa pre prenosové vedenia rozšírili začiatkom roku 1964 v Nemecku. Ďalší výrobcovia boli v Spojenom kráľovstve, Francúzsku, Taliansku a Spojených štátoch. V Nemecku sa v roku 1967 uskutočnili prvé skúšky v teréne, kde sa skúmala prvá generácia kompozitných (polymérnych) izolátorov. Podľa údajov o degradácii, niektorí výrobcovia prestali vyrábať vysokonapäťové jednotky a iní začali s intenzívnym výskumom vytvárania kompozitných izolátorov prenosových vedení druhej generácie. V sedemdesiatych rokoch minulého storočia sa začali používať izolátory, ktoré mali jadro tvorené zo sklenených vlákien a ako vonkajšia vrstva sa používali polyméry, čo začalo byť úspešnou alternatívou k porcelánu a sklu. Kovové koncovky sú pripevnené na obidva konce sklenenej tyče a utesnené proti vniknutiu vlhkosti. Medzi najúspešnejšie typy izolátorov patria tyčové izolátory a podperné izolátory s plným jadrom, ktoré sa používajú ako izolátory nízkeho napätia a taktiež ako izolátory pre prenosové vedenia. V roku 2009 mali polymérne izolátory početnosť poškodení 0,5 z 10 000 za rok, čo je nižšie ako u porcelánových

izolátoroch, ktoré mali početnosť poškodení 0,5-1 z 10 000 za rok [5], [14], [17], [18].

Kompozitné izolátory pozostávajú z najmenej dvoch izolačných materiálov, kde jeden má zabezpečiť elektrickú pevnosť a druhý zabezpečuje mechanické vlastnosti. Kompozitné izolátory sú založené na použití kompozitného jadra, ktorého funkciou je odolávať mechanickému zaťaženiu spôsobenému vodičom a prenášať túto ťahovú silu na stožiar. Jadro kompozitného izolátora pozostáva z kompozitnej tyče pozostávajúcej z dvoch hlavných zložiek matrice a výstužných vlákien. Matrica je vyrobená z epoxidovej živice a výstužné vlákno, ktoré je vyrobené zo skleneného vlákna, je paralelne a v rovnakom smere ako matrica po celej dĺžke jadra.

III. NAMÁHANIE IZOLÁTOROV VONKAJŠÍCH VEDENÍ

Prevádzkové podmienky, v ktorých izolátory pracujú sa líšia na základe konštrukčného riešenia uchytenia izolátorov (mechanické namáhanie), prostredia a napäťovej hladiny, na ktorú sú inštalované (elektrické namáhanie). Na izolátory neustále pôsobí kombinácia týchto namáhaní a to najmä elektrické namáhanie v kombinácii s nevhodným prostredím alebo počasím. Námraza alebo sneh okrem elektrického namáhania zvyšuje a náhle mení mechanické namáhanie izolátora. Tieto namáhania spôsobujú na izolačných systémoch degradáciu a starnutie. Izolátory budú pracovať spoľahlivo celú ich životnosť iba v prípade, že ich konštrukcia a použitie zohľadňuje všetky príslušné namáhanie. Izolátory používané vo vonkajšom prostredí sú neustále vystavované trom druhom namáhania [5], [19]:

- elektrické namáhanie,
- mechanické namáhanie,
- vonkajšie prostredie.

A. Elektrické namáhanie

Mechanizmus starnutia elektrickým namáhaním elektrických izolačných systémov je veľmi zložitý jav. Zatiaľ neboli definované presné metódy, ktoré by popisovali zrýchlené starnutie na určenie životnosti izolácie ako aj analytické metódy pre tepelné starnutie. Existujú však dobre vypracované pracovné postupy založené na niektorých empirických poznatkoch. Je dôležité si uvedomiť, že negatívne účinky elektrického starnutia sú silne závislé od povahy a druhu elektrického izolačného systému. Veľkosť elektród, hrúbka izolácie, miera namáhania, maximálna intenzita a frekvencia aplikovaného napätia a miera pravdepodobnosti poruchy sú faktory a premenné, ktoré ovplyvňujú veľkosť elektrického namáhania. Elektrické namáhanie je spôsobené striedavým, jednosmerným alebo impulzným napätím. Tieto napätia spôsobujú [20]-[23]:

- čiastkové výboje vznikajúce na povrchu alebo v izolátore,
- praskliny rozširujúce sa z miest, kde vznikajú čiastkové výboje,
- rozvoj elektrických stromčekov,
- elektrolyza,

- zvýšenie teploty spôsobené vysokými dielektrickými stratami.

B. Vonkajšie prostredie

Namáhanie izolátorov vplyvom vonkajšieho prostredia zahŕňa účinky:

- vlhkosti,
- kyslíka,
- chemické, biologické a klimatické zmeny,
- nečistoty,
- zmena tlaku,
- vibrácie,
- žiarenia.

Tieto namáhania ovplyvňujú úroveň degradácie izolácie a to najmä v kombinácii s elektrickým alebo mechanickým namáhaním. Ďalším dôležitým aspektom je zmena rozloženia napätia na izolácií, ktorá je spôsobená vplyvom prostredia ako napríklad vplyv nečistôt na zmeny elektrických pomerov na povrchu izolácie. Ak nie sú dostupné dáta o prostredí, kde bude izolácia umiestnená, musia sa brať do úvahy najnáročnejšie podmienky aké môžu na izoláciu pôsobiť [23], [24].

IV. FAKTORY OVPLYVŇUJÚCE PRESKOK PO POVRCHU IZOLÁTORA

Preskok po povrchu izolátora je ovplyvnený najmä znečistením a atmosférickými podmienkami. Tieto dva faktory sú hlavným aspektom zníženia povrchového odporu izolátora, čo môže viesť k preskoku po povrchu izolátora a následne k jeho možnému poškodeniu alebo narušeniu spoľahlivosti dodávky elektrickej energie.

A. Vplyv znečistenia

Znečistenie na izolátoroch vrátane dlhodobého hromadenia častíc prachu alebo piesku, prívalu vlhkosti z dažďa, okolitej vlhkosti a rosy môže vytvoriť vodivú vrstvu. To môže následne viesť k povrchovým výbojom a vzniku ciest pre zvodový prúd, ktorý môže negatívne ovplyvniť životnosť izolátora alebo viesť k preskoku, čím sa preruší dodávka elektrickej energie. Znečistenie izolátora prebieha tak, že vďaka vetru sa privádzajú znečisťujúce častice na izolátor, kde sa pri dopade častice zníži jej rýchlosť a to spôsobí ukladanie znečistenia na izolátore. Neustálym ukladáním častíc na izolátor sa na jeho povrchu vytvára stále hrubšia vrstva znečistenia a po určitom čase vznikne súvislá vrstva znečistenia. Vlhkosť, dážď, hmla a rosa navlhčujú vrstvu znečistenia, rozpúšťajú látky v nahromadenom znečistení a vytvárajú vodivú vrstvu, ktorá znižuje preskokové napätie. Znečistenie môže znížiť preskokové napätie izolátora aj o 20 – 25 % [25], [26].

Existujú dva hlavné typy ukladania a pôsobenia znečistenia. Vopred nahromadené znečistenie, ktoré sa ukladá na izolátore priebežne a potrebuje vlhkosť na to, aby sa stal vodivým elektrolytom a okamžité znečistenie, ktoré je vodivým elektrolytom hneď po usadení na izolátor [27].

Vopred nahromadené znečistenie sa rozdeľuje do dvoch kategórií. Takzvané aktívne znečistenie, ktoré tvorí vodivú vrstvu a inertné (nerozpustné) znečistenie, ktoré vytvára väzobnú vrstvu aktívnej zložke a vhodné podmienky pre šírenie zvodového prúdu. Aktívne znečistenie sa vyhodnocuje na základe vodivosti a inertné na základe hmotnosti znečistenia [27].

Okamžité znečistenie sa dá definovať ako kontaminant s vysokou vodivosťou, ktorý sa rýchlo usadzuje na povrchu izolátora, čo vedie k tomu, že sa stav izolátora v krátkom čase (< 1 hodina) zmení z čistého stavu s nízkou vodivosťou do stavu, kde je nebezpečenstvo preskoku po povrchu izolátora, ktoré trvá tak dlho, kým tento stav nepominie. Oblasti s vysokou koncentráciou soli alebo oblasti, kde sa vyskytuje vodivá priemyselná hmla sú dobrými príkladmi výskytu okamžitého znečistenia. Tento druh znečistenia sa dá analyzovať monitorovaním povrchovej vodivosti alebo zvodového prúdu [27].

B. Vlhkosť a hmla ako faktor ovplyvňujúci preskok po povrchu izolátora

Ako už bolo spomenuté, izolačné materiály používané na izoláciu vysokého napätia sú počas ich dlhoročnej prevádzky namáhané rôznymi faktormi. Jedným z dôležitých faktorov, ktoré prispievajú k starnutiu izolácie je vlhkosť, ktorá v kombinácii s ďalšími faktormi môže meniť stav izolačného povrchu. Tieto zmeny na povrchu môžu spôsobiť čiastkové výboje a prípadne postupnú degradáciu izolačného materiálu. Výskumy jednoznačne dokázali, že vlhkosť môže spôsobovať výboje, ktoré poškodzujú plášť zo silikónovej gúmy, ktorý pokrýva jadro izolátora. Vlhkosť takisto zvyšuje dielektrické straty a znižuje priernu pevnosť izolátora, čo sa dá považovať za nežiaduce [28], [29].

Kritická hodnota navlhčenia izolátora sa najčastejšie vyskytuje v hmle, keď voda kondenzuje na izolátore takou rýchlosťou, že dokáže zvlhčiť celý povrch izolátora a zároveň nie takou rýchlosťou, aby z izolátora vytekala alebo odvádzala nečistoty z povrchu izolátora. Kritické navlhčenie vedie k najnižšej elektrickej pevnosti v štandardizovaných testoch pre čisté aj slané hmly. Na odstránenie 50 % nečistôt je z pravidla potreba akumulácia 2mm vrstvy dažďa na hornom povrchu izolátora. Táto vrstva dažďa dokáže znížiť aj spodný povrch znečistenia o 15 % [5].

V. POUŽÍVANÉ DIAGNOSTICKÉ METÓDY

Značné úsilie bolo venované hľadaniu spoľahlivých metód na hodnotenie stavu izolátorov v prevádzke. Pre konvenčné sklenené a porcelánové izolátory sú k dispozícii metódy, ktoré sa osvedčili a majú vysokú úroveň spoľahlivosti, a taktiež sa vyvíjajú nové metódy. Vývoj spoľahlivých metód na hodnotenie stavu kompozitných izolátorov však stále prebieha. Okrem vizuálneho vyšetrenia sa v súčasnosti používa niekoľko techník, ale získané informácie ešte nie sú úplne uspokojivé. Zvyčajne ide o meranie alebo detekciu [30]:

- vizuálna detekcia,
- meranie čiastkových výbojov,

- metóda akustickej emisie,
- infračervená termografia,
- rozloženie elektrického poľa,
- sledovanie korónových javov.

Medzi progresívne metódy patrí [27]:

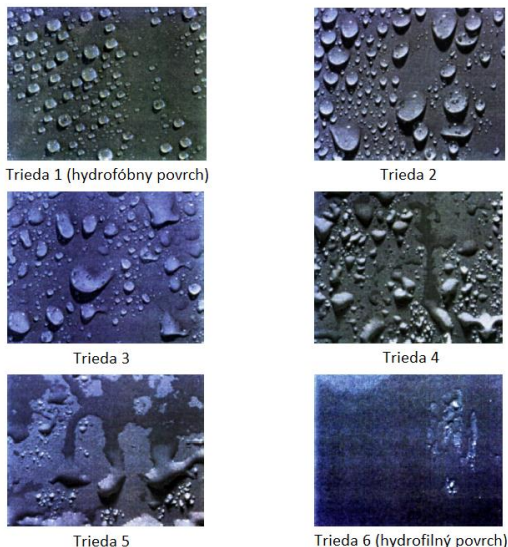
- harmonická analýza zvodového prúdu,
- numerické výpočtové metódy.

Diagnostické metódy sa dajú rozdeliť do dvoch hlavných kategórií. Tie, ktoré vyžadujú kontakt s izolantom a na tie, ktoré sa vykonávajú v určitej vzdialenosti od izolátora (meranie na diaľku).

V nasledujúcich podkapitolách budú stručne opísané metódy, ktoré sú dôležité pre experimentálne merania.

A. Hodnotenie hydrofóbnosti

Táto diagnostická metóda sa dá zaradiť aj medzi vizuálne kontroly. Kontrola hydrofóbnosti materiálu vie zhodnotiť úroveň starnutia povrchu izolátora. Hydrofóbnosť povrchu izolátorov (najmä silikónovej gumeny) hrá dôležitú úlohu pri obmedzovaní toku zvodového prúdu a pri predchádzaní preskoku po povrchu znečisteného izolátora. Úroveň hydrofóbnosti sa však môže časom meniť, napríklad v dôsledku účinkov životného prostredia, čiastkových výbojov alebo koróny. Hydrofóbnosť môže byť definovaná z hľadiska kontaktného uhla vytvoreného medzi kvapkou vody a povrchom.



Obr. 2 Porovnanie jednotlivých tried hydrofóbnosti [31]

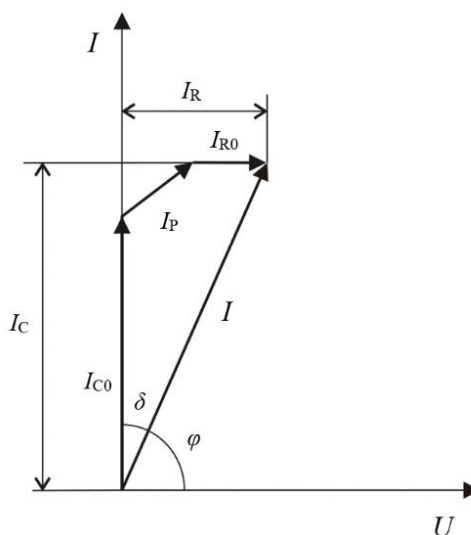
Tento uhol je však v teréne ťažko merateľný a zjednodušený spôsob klasifikácie hydrofóbnosti je zahrnutý do normy IEC 62073 (Príručka na meranie hydrofóbnosti povrchu izolátorov). Triedy hydrofóbnosti z normy sú na Obr. 2.

B. stratový činiteľ

Diagnostika pomocou merania stratového činiteľa sa bežne využíva a to najmä na diagnostiku transformátorov a káblov. Merania sú realizované v širokom frekvenčnom spektre, ako aj pri stabilnej frekvencii. Zmena stratového činiteľa indikuje starnutie izolačného materiálu, chemické zmeny v materiály a taktiež čiastočné poškodenie. Stratový činiteľ je závislý najmä od vodivosti, polarizácie dielektrika a teploty.

Podľa výsledkov merania a naštudovanej literatúry, je stratový činiteľ vhodný parameter na diagnostiku stavu povrchu izolátorov a vyhodnotenie stavu znečistenia.

Izolácia vysokonapäťových zariadení sa dá opísať ako kondenzátor s vysokonapäťovou a nízkonapäťovou (uzemnenou) časťou. Keď sa na kondenzátor priloží napätie, cez dielektrikum (izoláciu) preteká elektrický prúd. Ak je dielektrikum dokonalé a nemá straty, vektor nabíjacieho prúdu dielektrika má kapacitný charakter a je posunutý o 90° voči napätiu.



Obr. 3 Fázorový diagram reálneho dielektrika

Ak dielektrikum nie je dokonalé, je prítomná nielen jalová zložka prúdu, ale vďaka vodivosti a polarizácii dielektrika je prítomná aj činná zložka. Na Obr. 3 je znázornený príklad fázorového diagramu reálneho dielektrika, kde IC reprezentuje jalovú zložku, IR je činná zložka, IP reprezentuje polarizáciu dielektrika a IR0 predstavuje vodivosť dielektrika. δ je uhol medzi dokonalým dielektrikom, ktorý má kapacitný charakter, a reálnym dielektrikom, v ktorom prebiehajú polarizačné deje, ktoré majú okrem jalovej zložky aj zložku činnú a má určitú vodivosť keďže odpor izolácie nie je nekonečný.

V elektroenergetike sa stratový činiteľ využíva najmä na diagnostiku transformátorov, káblov, vinutí a kondenzátorov. Spektroskopia sa používa pri transformátoroch na analýzu degradácie izolačného systému transformátora. Zmeny stratového činiteľa v rôznych častiach frekvenčného pásma indikujú starnutie izolácie [32]-[34]. Diagnostika káblov sa môže vykonávať pomocou nízko-frekvenčných meraní stratového činiteľa, kde sa sledujú hodnoty vodivostných

a polarizačných strát [35], [36]. Čo sa týka izolátorov, stratový činiteľ bol použitý ako veličina na sledovanie starnutia izolátorov. Výsledky ukazujú, že starnutie koreluje s hodnotami stratového činiteľa. Výskumy poukázali na využiteľnosť stratového činiteľa na starnutie a degradáciu povrchu izolátorov [37]-[40].

VI. PRÍPRAVA VZORIEK A KLASIFIKÁCIA ZNEČISTENIA

Na meranie stratového činiteľa sa využilo viacero izolátorov a taktiež viacero úrovní znečistenia. Podľa typu konštrukcie sa použili tanierové (čiapkové) a tyčové izolátory. Merania prebiehali na sklenených aj porcelánových tanierových izolátoroch. Využitie izolačné materiály tyčových izolátorov boli silikónová guma a EVA (etylén-vinyl-acetát). Využitý bol aj obmedzovač prepätia so silikónovým plášťom pre merania pri nízkom napätí, kde sa využíval iba povrch obmedzovača. Informácie ku použitým izolátorom sú v tabuľke 1.

TABUĽKA 1 SÚPIS POUŽITÝCH IZOLÁTOROV A ICH ZÁKLADNÉ PARAMETRE

Mat. izolátora	Konštrukčné riešenie	Typ	Parametre
Sklo	Tanierový	U120 B	U_p (50 Hz, mokrý): 40 kV *PoV: 330 mm
Porcelán	Tanierový	U120 B	U_p (50 Hz, mokrý): 45 kV *PoV: 330 mm
Porcelán	Tanierový	U70 BL	U_p (50 Hz, mokrý): 45 kV *PoV: 292 mm
Silikónová guma	Obmedzovač prepätia	POLIM-S 07 N	U_p (50 Hz, mokrý): 68 kV *PoV: 983 mm
EVA	Tyčový - závesný	RST-DS28R-CT	U_p (50 Hz, mokrý): 98 kV *PoV: 690 mm

** PoV – povrchová vzdialenosť

Na znečistenie izolátorov sa použila metóda predkontaminácie. Pri tejto metóde sa na rozdiel od metódy znečistenia slanou hmlou dajú vyhodnotiť parametre aj za sucha,

kedy je vrstva znečistenia zaschnutá na povrchu izolátora. Nevýhodou je, že povrch izolátora nie je znečistený tak rovnomerne ako v prípade merania v slanej hmle.

Ako prvé sa pripravili roztoky, ktoré budú reprezentovať rôzne úrovne kontaminácie (štyri úrovne umelého znečistenia (L1, L2, L3 a L4)). Jednotlivé roztoky boli namiešané použitím určeného množstva soli, kaolínu a destilovanej vody. Kaolín pridávaný do destilovanej vody napomáha ku vytvoreniu rovnomernej rozloženej vrstvy znečistenia na povrchu izolátora a reprezentuje nerozpustnú zložku znečistenia. Na každý liter destilovanej vody sa pridalo 40 g kaolínu. V tabuľke 2 sú uvedené jednotlivé úrovne znečistenia s množstvom pridanej soli na liter destilovanej vody a príslušné hodnoty vodivosti roztokov podľa IEC/TS 60815-1:2008.

TABUĽKA 2 ZÁKLADNÉ PARAMETRE JEDNOTLIVÝCH ROZTOKOV ZNEČISTENIA

Úroveň znečistenia	Množstvo soli (NaCl) (g/l)	Vodivosť roztoku (mS/cm)
L1	1	2.88
L2	6	12.7
L3	10	21.7
L4	16	27.7

Okrem vodivosti roztokov sa podľa normy IEC/TS 60815-1:2008 vyhodnotil aj pomer rozpustných a nerozpustných zložiek znečistení získaných z povrchov jednotlivých izolátorov. Tieto parametre sa vyhodnocujú pomocou ekvivalentnej hustoty nánosov soli (ESDD) a hustotou nánosov nerozpustnej zložky (NSDD). Tieto parametre udávajú množstvo znečistenia, ktoré môžu izolačné materiály absorbovať. Keďže na izolačné materiály sa aplikovali rovnaké roztoky, ktoré predstavujú rôzne úrovne kontaminácie, dá sa tým poukázať na schopnosť použitých materiálov odpudzovať znečistenie obsiahnuté v roztokoch. Meranie a výpočty boli vykonané podľa spomínanej normy IEC/TS 60815-1:2008.

Prvým krokom bola predkontaminácia. Izolátor (alebo jeho časť) sa ponoril do roztoku a po vynorení sa zavesil aby prešiel 24 hodinovým procesom sušenia. Po vytvorení suchej vrstvy znečistenia na povrchu izolačnej časti izolátora sa začal proces zmývania. Zvoleným množstvom vody bol povrch izolátora dôkladne umytý. Pri výpočtoch sa zohľadňuje množstvo vody, ako aj veľkosť povrchu, z ktorého bolo znečistenie zmyté. Vodivosť roztoku zozbieraného z povrchu izolátora sa merala pomocou prístroja na meranie konduktivity Toledo S230. Zachytená voda sa miešala 5 minút a vodivosť sa merala desaťkrát, z čoho sa vypočítala priemerná hodnota. Prístroj nezohľadňoval teplotu roztoku. Teplota roztoku sa zaznamenávala a následne sa nameraná vodivosť prepočítala pomocou faktora závislého od teploty b . Po výpočte faktora b sa dá vypočítať vodivosť roztoku pri 20°C a následne sa vypočítala salinita S_a . Po výpočte salinity sa vypočítala ESDD podľa (1), kde sa berie do úvahy množstvo použitej destilovanej vody V a plocha povrchu S , ktorá bola umytá:

$$ESDD = S_a \cdot \frac{V}{S} \quad (1)$$

Aby sa dali jednotlivé materiály medzi sebou porovnať, bolo potrebné aby bol pomer V/S rovnaký pre všetky typy izolátorov. V prípade, ak by tieto pomery neboli dodržané, materiály nie je možné porovnať, keďže vodivosť nerastie lineárne so spomenutými veličinami.

Po meraní vodivosti sa roztok prefiltraval cez filter. Filter bol pred filtráciou 24 hodín sušený a vážený pred procesom filtrovania. Po procese filtrovania sa filter opäť sušil 24 hodín. NSDD sa následne vypočíta podľa (2) z váhy filtra pred filtrovaním W_i , váhy filtra po filtrovaní W_f a plochy S , z ktorej bola vzorka odobraná:

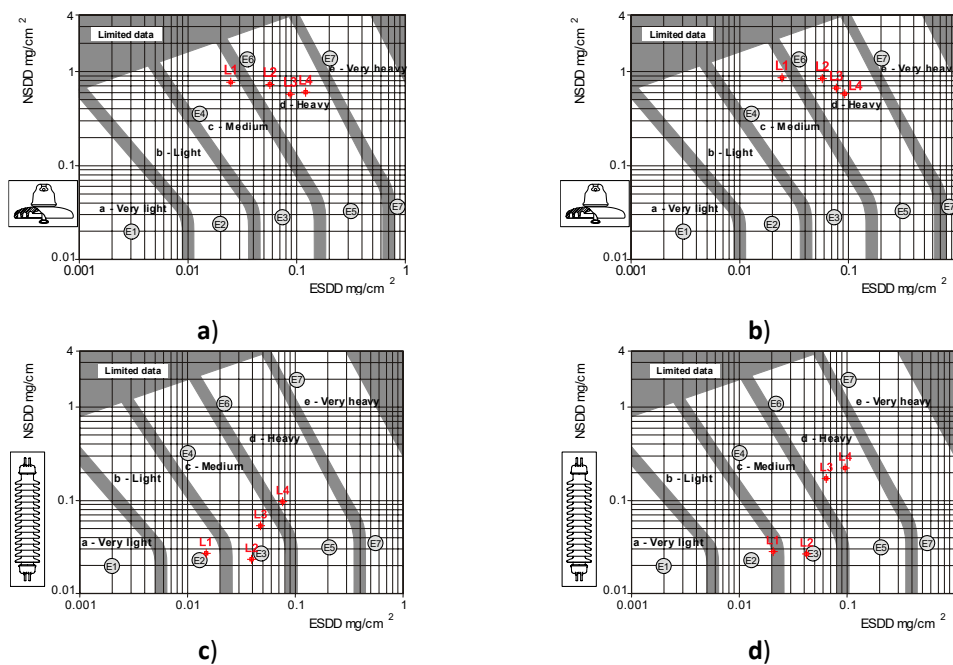
$$NSDD = \frac{1000(W_i - W_f)}{S} \quad (2)$$

Namerané hodnoty ESDD a NSDD sú v tabuľke 3.

V norme IEC/TS 60815-1:2008 sú zobrazené grafy, kde sa na základe hodnôt ESDD a NSDD dá graficky učiť stupeň závažnosti znečistenia. Tieto grafy boli empiricky vytvorené na základe predošlých meraní.

TABUĽKA 3 VYPOČÍTANÉ HODNOTY ESDD A NSDD PRE RÔZNE TYPY IZOLÁTOROV A ZNEČISTENÍ

		L1	L2	L3	L4
Glass	ESDD (mg/cm ²)	0.0244	0.0564	0.0882	0.1095
	NSDD (mg/cm ²)	0.7585	0.7113	0.5729	0.6025
Porcelain	ESDD (mg/cm ²)	0.0206	0.0573	0.0773	0.0908
	NSDD (mg/cm ²)	0.8319	0.8209	0.6619	0.5778
Silicone rubber	ESDD (mg/cm ²)	0.0147	0.0385	0.0466	0.0743
	NSDD (mg/cm ²)	0.0257	0.0217	0.0515	0.0961
EVA	ESDD (mg/cm ²)	0.0204	0.0413	0.0633	0.0945
	NSDD (mg/cm ²)	0.2679	0.2554	0.1698	0.2117



Obr. 4 Stupeň znečistenia vyhodnotený na základe hodnôt ESDD a NSDD pre izolátor zo: a) skla; b) porcelánu; c) silikónovej gummy; d) EVA materiálu

Na Obr. 4 sú znázornené jednotlivé stupne znečistenia merané na rôznych typoch izolátorov. Podľa nameraných a vypočítaných hodnôt ESDD a NSDD sa graficky určila závažnosť znečistenia. Z výsledkov je jasné, že sklenený a porcelánový izolátor dosahujú vyššiu závažnosť znečistenia ako je v prípade izolátorov zo silikónovej gummy a EVA materiálu.

Najmenej závažné znečistenie metódou predkontaminácie bolo vytvorené na izolátore so silikónovej gummy, kde znečistenie L1 dosiahlo len ľahkú úroveň znečistenia, L2 a L3 boli v poli stredne závažného znečistenia a L4 zodpovedalo ťažkému znečisteniu. EVA izolátor dosahoval celkovo vyššiu závažnosť

znečistenia ako silikónová guma. Znečistenia L3 a L4 sa približovali hodnotám nameraným na sklenenom a porcelánovom izolátore.

Sklenené a porcelánové izolátory mali nameranú vyššiu závažnosť znečistenia ako kompozitné izolátory a najvyššie hodnoty sa dosiahli na sklenenom izolátore. Hodnoty na porcelánovom izolátore však boli len mierne nižšie ako v prípade skla. Už znečistenie L1 na keramických izolátoroch patrilo do oblasti strednej závažnosti znečistenia. Ostatné znečistenia sa nachádzali v oblasti ťažkého znečistenia.

Najväčšie rozdiely sa dosiahli pri prvých dvoch stupňoch znečistenia, kde sa na keramických izolátoroch vytvorila o dosť ťažšia vrstva znečistenia ako v prípade kompozitných materiálov. Okrem materiálu izolátora na to môže mať vplyv aj konštrukcia izolátora, kde sa používali v prípade skla a porcelánu tanierové izolátory a izolátory zo silikónovej gumy a EVA materiálu boli tyčového typu. Mierka v týchto grafoch je pre obidve osi logaritmická, čo v prípade vyšších znečistení spôsobuje zdanlivo menšie rozdiely.

VII. MERANIE STRATOVÉHO ČINITELĽA PRI NÍZKOM NAPÄTÍ V ŠIROKOM FREKVENČNOM SPEKTRE

Meranie stratového činiteľa a na sklenenom izolátore a porcelánovom izolátore sa uskutočnilo vo vysokonapäťovom laboratóriu Technickej univerzity v Košiciach.

A. Postup merania

Meranie prebiehalo v uzavretej komore pri relatívnej vlhkosti od 40 % do 90 %. Použité boli všetky štyri úrovne znečistenia. Vzorky boli znečistené predkontamináciou a sušené minimálne 24 hodín. Ako elektródový systém boli použité vodivé medené pásky. Vzdialenosť medzi meracími elektródami kruhového tvaru bola 4 cm (Obr. 5). Použitý sklenený izolátor aj porcelánový izolátor sú typu U120 B.



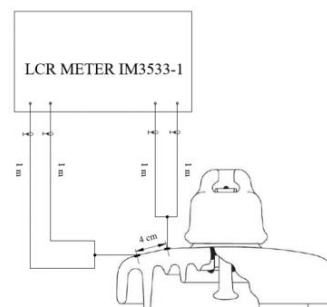
Obr. 5 Sklenený izolátor s elektródovým systémom

Čisté a kontaminované vzorky rôznymi úrovňami znečistenia, boli merané individuálne. Meranie sa začalo pri izbovej vlhkosti v rozmedzí 30-40 %. Po meraní za sucha sa relatívna vlhkosť v komore zvýšila na 60 %. Táto vlhkosť je ekvivalentná ku štandardnej absolútnej vlhkosti, ktorá má hodnotu 11 g/m³. Štandardná absolútna vlhkosť je špecifikovaná v medzinárodnej norme IEC 60060-1:2010. Po dosiahnutí požadovanej relatívnej vlhkosti sa určila perióda saturácie znečistenej vrstvy na 10 minút. Táto doba by mala byť dostatočná aby zabezpečila absorbovanie vlhkosti znečistenou vrstvou. Po odmeraní stratového činiteľa sa relatívna vlhkosť zvýšila o 10 % a proces sa opakoval až po relatívnu vlhkosť 90 %.

Na meranie stratového činiteľa ($\tan \delta$) sa použil analyzátor impedancie HIOKI LCR meter IM3533-1 s frekvenčným rozsahom od 1 mHz do 200 kHz. Prepojenie izolátora a LCR metra bolo realizované pomocou štyroch samostatných tienených koaxiálnych káblov s dĺžkou 1 m. Podľa odporúčaní výrobcu, bola na začiatku merania každej vzorky vykonaná

kalibrácia obvodu naprázdno a nakrátko. Koaxiálne káble boli po kalibrácii pripojené k elektródam na povrchu vzoriek.

LCR meter meral $\tan \delta$ meracím napätím 5 V vo frekvenčnom pásme od 1 Hz do 200 kHz. Frekvenčné pásmo bolo rozdelené do 100 meracích bodov, pričom v každom bode bolo spriemerovaných 20 meraní. Schéma zapojenia analyzátoru impedancie pripojeného k meracím elektródam pomocou koaxiálnych káblov je na Obr. 6. LCR meter vyhodnocuje stratový činiteľ z fázového posunu medzi prúdom a napätím nameraným medzi elektródami. Teplota okolia sa počas merania pohybovala od 23 °C do 25 °C. Táto zmena teploty nemôže výrazne ovplyvniť výsledky merania.



Obr. 6 Schéma merania stratového činiteľa LCR metrom IM3533-1

B. Meranie stratového činiteľa v širokom frekvenčnom pásme

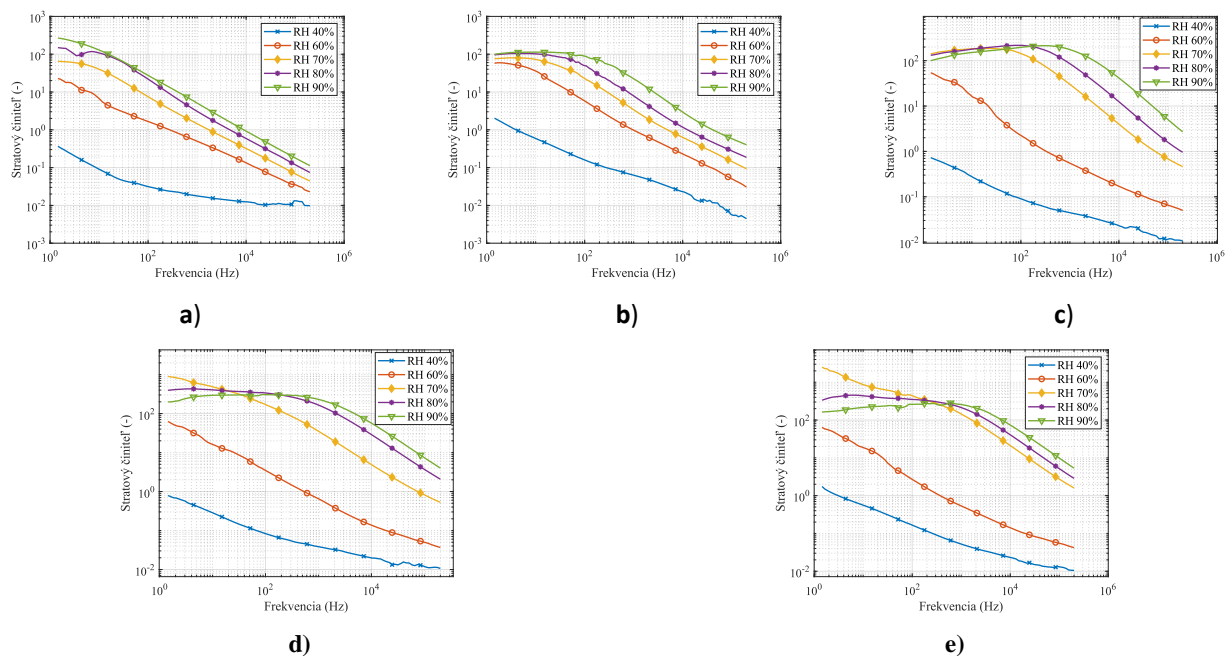
Z merania stratového činiteľa v širokom frekvenčnom pásme sa zistilo, že krivka závislosti stratového činiteľa od frekvencie nemá lineárny charakter ale dosahuje lokálne maximum pri rôznych frekvenciách. Ako prvé boli vyhodnotené merania v závislosti od relatívnej vlhkosti (RH). Z grafov na Obr. 7 a Obr. 8 je jasne vidieť stúpajúci charakter stratového činiteľa v závislosti od relatívnej vlhkosti.

Výsledky meraní na porcelánovom izolátore na Obr. 34 poukazujú na zvyšovanie stratového činiteľa so zvyšujúcou sa relatívnu vlhkosťou. Z nameraných hodnôt sa dá vyčítať, že stratový činiteľ taktiež rastie so zvyšujúcou sa úrovňou znečistenia. Frekvenčná odozva stratového činiteľa na Obr. 34 a) a Obr. 34 b) pri relatívnej vlhkosti 40 % je nižšia ako pri vyšších hodnotách relatívnej vlhkosti. Merania s relatívnu vlhkosťou 60 % a vyššou sú viac „zospúpené“ a rozdiel medzi nimi nie je až tak výrazný. Pre druhú, tretiu a štvrtú úroveň znečistenia na Obr. 34 c), d) a e) je viditeľný rýchly nárast stratového činiteľa medzi relatívnu vlhkosťou 60 % a vyššou relatívnu vlhkosťou. Najvyšší nárast dielektrického stratového činiteľa je teda medzi 60 % a 70 % relatívnej vlhkosti. Nárast Stratového činiteľa medzi týmito dvoma relatívnymi vlhkosťami je pri frekvencii 1 kHz, o 51 až 263 násobne vyšší pre rôzne úrovne znečistenia. Pre štvrtú úroveň znečistenia má stratový činiteľ hodnotu 0,48 (pri relatívnej vlhkosti 60 %), ktorá pri relatívnej vlhkosti 70 % narástla na 126 pri meranej frekvencii 1 kHz.

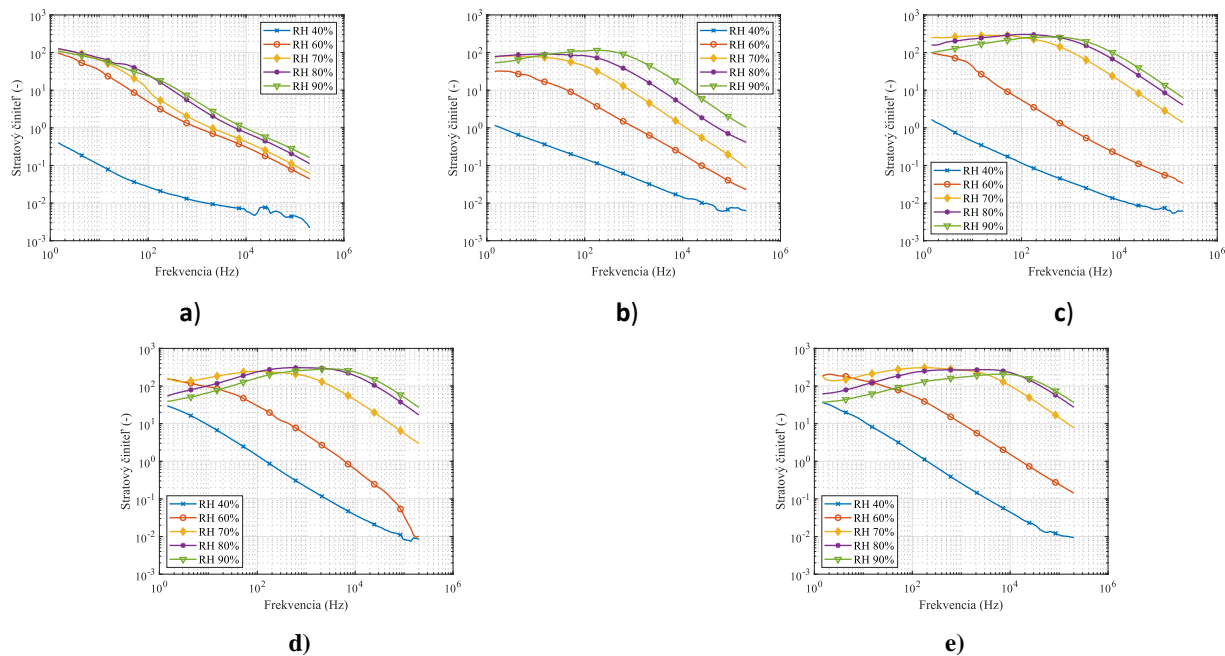
Pri porovnaní Obr. 35 a Obr. 34 sa dá zhodnotiť, že tvar kriviek sa vo väčšine prípadov podobá s výsledkami nameranými na porcelánovom izolátore. Krivky pre vlhkosť

40 % a 60 % sú súběžné a krivky od relatívnej vlhkosti 70 % sú zoskupené a majú podobný tvar, s výnimkou merania na čistom izolátore, kde sú okrem merania za sucha všetky krivky blízko seba a hodnota stratového činiteľa sa mierne zvyšuje vplyvom

narastajúcej vlhkosti. Jednotlivé grafy zobrazujú frekvenčnú odozvu stratového činiteľa pre čistý izolátor a pre štyri rôzne úrovne znečistenia. Stratový činiteľ zvyšuje svoju hodnotu aj v dôsledku zvyšovania úrovne umelého znečistenia.



Obr. 7 Výsledky merania Stratového činiteľa vo frekvenčnom spektre pri rôznej vlhkosti na porcelánovom izolátore: a) čistý izolátor; b) L1; c) L2; d) L3; e) L4



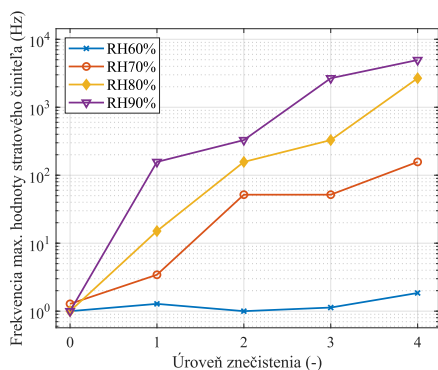
Obr. 8 Výsledky merania Stratového činiteľa vo frekvenčnom spektre pri rôznej vlhkosti na porcelánovom izolátore: a) čistý izolátor; b) L1; c) L2; d) L3; e) L4

Porcelánový izolátor má maximá v nižších frekvenciách ako zvyšné dva materiály. Podobnosť v krivkách nastáva až vo vyšších frekvenciách (od 10 kHz), kde je pokles kriviek súbežný. V tabuľke 4 je porovnanie medzi materiálmi pri frekvencii 10 kHz, kedy sa všetky krivky stabilizovali a začali klesať po dosiahnutí lokálnych maxim. Rozdiel v hodnotách stratového činiteľa medzi skleneným izolátorom a porcelánovým je značný. Najväčší rozdiel je medzi prvým, druhým a tretím stupňom znečistenia pre obidva izolátory, kedy stratový činiteľ stúpal niekoľkonásobne. Rozdiel medzi meraniami na treťom a štvrtom stupni znečistenia nie sú tak výrazné.

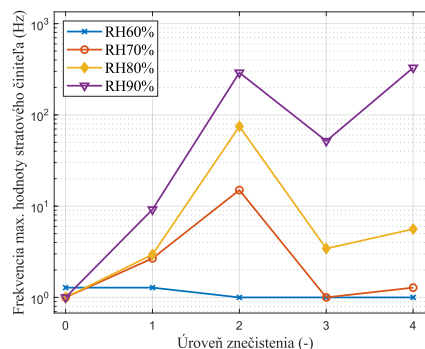
TABUĽKA 4 VÝSLEDKY MERANIA STRATOVÉHO ČINITEĽA PRI RELATÍVNEJ VĽHKOSTI 80 % A FREKVENCII 10 KHz

Stratový činiteľ (-)	Relatívna vlhkosť 80 %			Frekvencia 10 kHz
	L1	L2	L3	L4
Porcelán	0,99	10,10	23,56	33,33
Sklo	3.34	43.8	166.22	209.48

Ďalšia hodnota ovplyvnená zmenou relatívnej vlhkosti a zmenou znečistenia je poloha maxima stratového činiteľa na frekvenčnej osi pre sklenený izolátor a porcelánový izolátor. Veľkosť najvyššej hodnoty stratového činiteľa nezodpovedá zvýšeniu úrovne znečistenia ale jeho posun vo frekvenčnej oblasti áno. To znamená, že so zvyšujúcou sa relatívnou vlhkosťou a zvyšujúcou sa úrovňou kontaminácie sa maximá stratového činiteľa posúvajú k vyšším frekvenciám. Táto skutočnosť je znázornená na . Zvýšenie frekvencie maximálnej hodnoty stratového činiteľa je výraznejšie so zvyšujúcou sa vlhkosťou. Závisí to aj od použitého izolačného materiálu. Pri meraní na sklenenom izolátore sa lokálne maximum posúva do vyšších frekvencií viditeľne, a to najmä pri relatívnych vlhkosťach 80 % a 90 %.



a)



b)

Obr. 9 Závislosť polohy maxima stratového činiteľa od stupňa znečistenia, relatívnej vlhkosti: a) sklenený izolátor; b) porcelánový izolátor

Meranie na porcelánovom izolátore nepreukázalo spomenuté skutočnosti v prípade závislosti polohy maxima od znečistenia v celom rozsahu. Dá sa však podľa polohy maxima rozoznať či bol izolátor meraný za sucha a čistý alebo bol znečistený a v prostredí s vysokou relatívnou vlhkosťou. Krivky v prípade porcelánového izolátora vo väčšine prípadov dosiahli maximum pravdepodobne v oblasti nižších frekvencií ako bol rozsah merania alebo pri frekvenciách, ktoré boli namerané na začiatku meranej frekvenčnej oblasti.

C. Diskusia

Po meraniach stratového činiteľa pri nízkom napätí 5 V v rôznych podmienkach sa zistilo, že so zvyšujúcou sa relatívnou vlhkosťou a zvyšujúcou sa úrovňou znečistenia sa zvyšuje aj stratový činiteľ. Po podrobnejšej analýze je potvrdené aj to, že sa lokálne maximá stratového činiteľa posúvajú k vyšším frekvenciám, ako na Obr. 7 a Obr. 8. Výnimkou bolo meranie na porcelánovom izolátore pre druhý stupeň znečistenia, kde bola pozícia lokálneho maxima výrazne vyššia ako pre tretí a štvrtý stupeň znečistenia. Poloha lokálnych maxim sa líšila aj v závislosti od použitého izolačného materiálu. Hodnoty frekvencie lokálnych maxim stratového činiteľa pre sklenený izolátor boli vyššie ako u porcelánu. Tento trend sa zachoval aj pri porovnaní hodnôt na rôznych materiáloch pri frekvencii 10 kHz, kedy stratový činiteľ boli vo fáze poklesu za lokálnym maximum. Oblasť vo vyšších frekvenciách, za bodom dosiahnutia lokálnych maxim bola najvhodnejšia na porovnanie jednotlivých znečistení a materiálov. Po dosiahnutí maxima začal vo všetkých prípadoch stratový činiteľ konzistentne klesať.

Tvar kriviek stratového činiteľa sa mení v závislosti od vlhkosti a znečistenia. Zmeny v stratovom činiteľi pri kontaminácii rovnakou úrovňou znečistenia súvisia s hydrofóbnosťou materiálu a schopnosťou odpudzovať kontamináciu. Polarizačné procesy ovplyvňujúce stratový činiteľ patria k pomaly sa formujúcej polarizácii, nazývanej aj relaxačný typ polarizácie.

Ak porovnáme čisté izolačné materiály so znečistenými, nárast stratového činiteľa v závislosti od relatívnej vlhkosti nie je u vzoriek s čistým povrchom taký výrazný. To je spôsobené schopnosťou znečistených vrstvy absorbovať vodu zo vzduchu.

Experimentálne merania publikované v [41] ukazujú, že nasýtenie kontaminovanej vrstvy závisí od zloženia tejto vrstvy znečistenia na povrchu izolátora. V prípade použitia roztoku s iným zložením možno očakávať prudký nárast meraných veličín pri inej relatívnej vlhkosti. Ďalšie štúdie zaoberajúce sa znečistením v rôznych oblastiach [42] ukázali, že znečisťujúca vrstva môže pozostávať z rôznych zlúčenín s rôznymi fyzikálnymi vlastnosťami (vodivosť, hydrofilnosť). Preto by sa mali sledovať namerané hodnoty v celom spektre relatívnych vlhkostí.

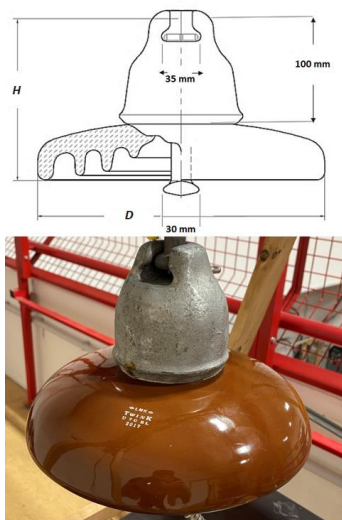
Monitorovanie relatívnej vlhkosti sa javí ako veľmi užitočné pre zvýšenie spoľahlivosti. Podľa normy IEC 60071: 2010 je referenčná hodnota relatívnej vlhkosti 60 % (11 g/m³). Výrazná zmena nárastu stratového činiteľa bola medzi vlhkosťami 60 % a 70 %. V súčasnej literatúre sa za najnebezpečnejšiu situáciu pre vznik preskoku po povrchu izolátora považuje hmlisté počasie. Vo všeobecnosti je pri hmle relatívnu vlhkosť vyššia ako 70 % a hodnota sa pohybuje okolo 99 %. Z výsledkov našich meraní sa dá usúdiť, že nebezpečenstvo poruchy začína pri relatívnej vlhkosti 70 % dosiahnutej v miernych poveternostných podmienkach a závislej od druhu kontaminácie.

VIII. MERANIE STRATOVÉHO ČINITELĽA PRI VYSOKOM NAPÄTÍ V ŠIROKOM FREKVENČNOM SPEKTRE

Výskum a merania prebiehali vo vysokonapäťovom laboratóriu inštitútu techniky vysokého napätia Technickej univerzity v Grazi. Merania boli vykonané na novom porcelánovom izolátore U 70 BL spomínanom v tabuľke 1.

A. Postup merania

Použitý porcelánový izolátor je vhodný pre prenosové vedenia ako súčasť izolátorového reťazca a jeho povrch je glazúrovaný. Na sú zobrazené rozmery izolátora a použitý izolátor, ktorý spĺňa štandardy podľa normy IEC 60305:2021. Priemer izolátora (D) je 255 mm a výška (H) je 146 mm.



Obr. 10 Porcelánový izolátor U 70 BL

Ako prvé bolo potrebné pripraviť roztoky reprezentujúce znečistenia na predkontamináciu izolátora. Boli vytvorené dva roztoky reprezentujúce dve úrovne znečistenia (L1 a L2) pomocou vody, kaolínu a soli ako v predchádzajúcom výskume. Množstvo kaolínu bolo rovnaké pre obe úrovne znečistenia, do ktorých sme pridali 40 g kaolínu na každý liter vody.

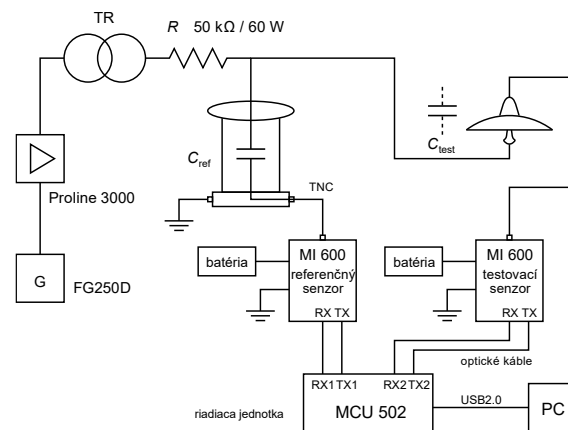
Množstvo soli na liter vody pre jednotlivé úrovne znečistenia a príslušné hodnoty vodivosti zmiešaných roztokov podľa IEC/TS 60815-1:2008 sú v tabuľke 5. Tieto dve úrovne znečistenia, označené v tejto časti ako L1 a L2, majú rovnaké parametre ako roztoky znečistenia L2 a L4 z tabuľky 2.

TABUĽKA 5 ZÁKLADNÉ PARAMETRE JEDNOTLIVÝCH ROZTOKOV ZNEČISTENIA POUŽITÝCH PRI MERANÍ

Úroveň znečistenia	Množstvo soli (NaCl) (g/l)	Vodivosť roztoku (mS/cm)
L1	6	12.7
L2	16	27.7

Prvé meranie prebiehalo na čistom izolátore. Po meraní na čistom izolátore sa vzorka namočila do vody a meranie sa opakovalo. Proces namáčania trval približne minútu. Vzduch zo spodnej časti izolátora medzi jeho rebrami sa odstránil vhodným otáčaním a nakláňaním vo vode, aby sa zabezpečil proces namáčania celého povrchu izolátora. Následne boli realizované merania na kontaminovanom izolátore. Meranie na kontaminovanom izolátore prebiehalo za sucha, kedy bol izolátor sušený minimálne 24 hodín a za mokra, kedy meranie prebiehalo 10 minút po ponorení porcelánového izolátora do roztoku umelého znečistenia.

Meral sa stratový činiteľ. Vysoké napätie bolo pripojené na paličku izolátora a čiapka izolátora bola uzemnená.



Obr. 11 Schéma zapojenia meracieho pracoviska

Generátor napätia FG250D (G) a zosilňovač Proline 3000 (►) boli pripojené ku paličke izolátora cez transformátor (TR). Pre meranie stratového činiteľa pri vysokom napätí boli použité napäťové úrovne 5 kV, 7,5 kV a 10 kV a na izolátor bolo aplikované napätie v širokom frekvenčnom pásme. Frekvenčný rozsah merania začínal na úrovni 10 Hz a frekvencia sa postupne zvyšovala až na hladinu 1 kHz. Pre nižšie frekvencie bol použitý

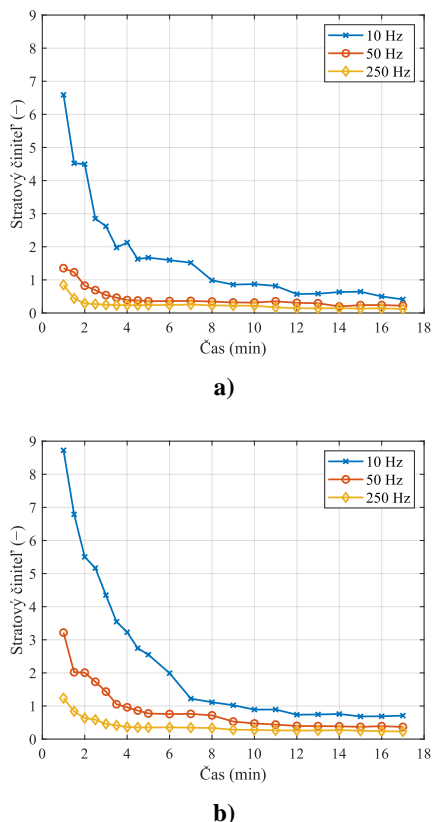
krok 10 Hz do bodu, kedy sa dosiahla úroveň frekvencie 100 Hz. Po dosiahnutí 100 Hz, sa krok zvyšovania frekvencie navýšil na 50 Hz. Počítač (PC) pripojený k meracím zariadeniam cez USB port riadil meranie a zaznamenával hodnoty stratového činiteľa každých 0,9 s. Namerané dáta boli uložené na pamäťové médium pomocou PC.

Pomocou meracieho softvéru OMICRON pre použité zariadenie OMICRON MI 600 boli vypočítané priemerné hodnoty zo štatistickej množiny 100 hodnôt. Referenčný kondenzátor, manuálne nastavený v softvéri OMICRON, bol pripojený k referenčnému senzoru MI 600. Referenčný kondenzátor mal kapacitu 37,86 pF. Porcelánový izolátor sa pripojil k testovaciemu senzoru MI 600 a obidva senzory boli optickými káblami pripojené k riadiacej jednotke MCU 502. Schéma zapojenia meracieho obvodu na meranie stratového činiteľa je na Obr. 11.

Merací systém priamo vyhodnotil hodnotu stratového činiteľa. OMICRON MI 600 vypočítava stratový činiteľ na základe fázového posunu medzi nameraným prúdom a aplikovaným napätím na vzorku. Následne porovnáva nameranú vzorku s referenčnou vzorkou.

B. Výsledky merania stratového činiteľa

Ako prvé sa uskutočnili testovacie merania hneď po ponorení do roztoku, aby sa zistila zmena v hodnotách meraných veličín na znečistenej vrstve.

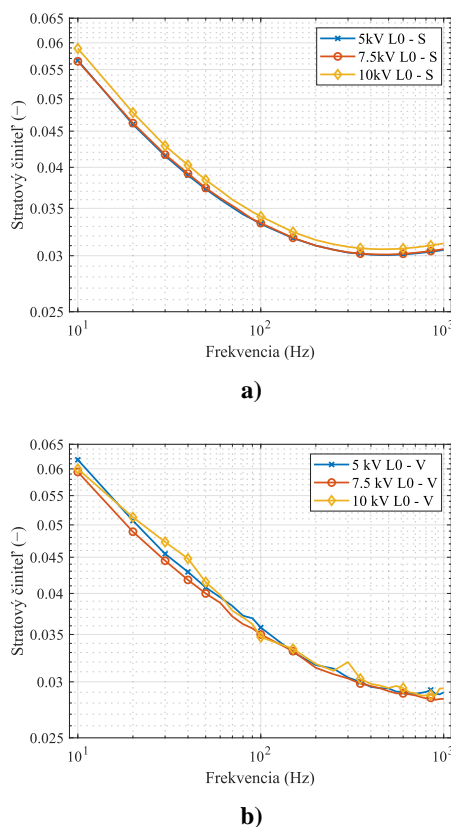


Obr. 12 Zmena veličín v závislosti od času: a) stratový činiteľ (L1); b) stratový činiteľ (L2)

Bezprostredne po namočení izolátora do roztoku došlo na jeho povrchu k čiastočným výbojom, ktoré ovplyvnili stratový činiteľ. Preto sa na mokrej vrstve znečistenia meral až 10 minút po namočení.

Obr. 12 zobrazuje závislosť stratového činiteľa od času, počas ktorého bolo priložené napätie 5 kV pre prvú aj druhú úroveň znečistenia. Na porovnanie boli zvolené tri rôzne meracie frekvencie. Stratový činiteľ sa postupne znižuje vplyvom zvyšovania frekvencie. Najvyššie hodnoty stratového činiteľa pri porovnávaní meraní medzi frekvenciami boli namerané pri frekvencii 10 Hz. Ak sa porovná meranie pri rôznom znečistení, hodnota stratového činiteľa narastá so zvyšujúcou sa úrovňou znečistenia, no tvar kriviek je podobný. Porovnaním výsledkov meraní v 1., 10. a 15. minúte sme zistili, že pokles stratového činiteľa medzi 10. a 15. minútou bol v priemere 5,8 %.

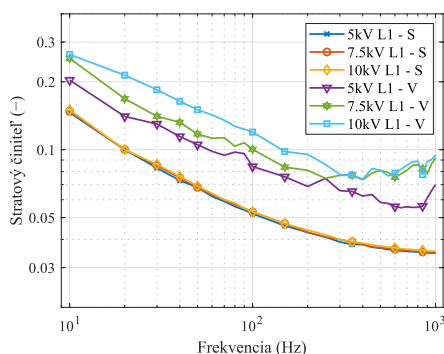
Najprv sa porovnávali merania pri rôznych napätiach. Na Obr. 13 a Obr. 14 sú frekvenčné závislosti stratového činiteľa pre rôzne napäťové hladiny. Aby boli výsledky prehľadnejšie, merania na čistom izolátore sa rozdelili do dvoch grafov na meranie za sucha (S) a za mokra (V).



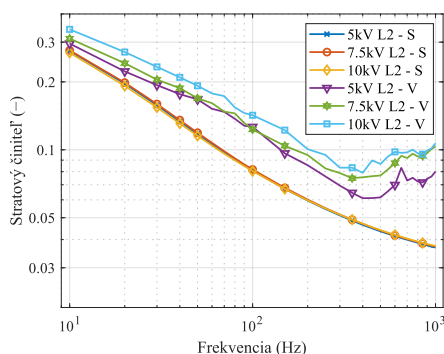
Obr. 13 Frekvenčná závislosť stratového činiteľa pre rôzne napäťové hladiny: a) čistý izolátor (suchý); b) čistý izolátor (mokrý)

Z nameraných hodnôt a grafických zobrazení sa dá vyhodnotiť zmena stratového činiteľa v závislosti od napäťovej hladiny. V Tabuľke 6 sú údaje predstavujúce percentuálny rozdiel medzi meraniami pri rôznych napätiach. Tabuľka tiež

poukazuje na to, ako napätie ovplyvňuje stratový činiteľ v závislosti od stavu povrchu izolátora (čistý-znečistený, suchý-mokrý). Grafy na Obr. 13 ukazujú, že zmena hladiny napätia významne neovplyvňuje stratový činiteľ. Výsledky merania na čistom izolátore ukázali, že stratový činiteľ sa pri zmene napätia líši len minimálne a krivky sú tvarovo takmer totožné. Pri meraní za sucha, na Obr. 13 a), je stratový činiteľ na napät'ovej hladine 10 kV o niečo vyšší ako na napät'ových hladinách 5 kV a 7,5 kV. Priemerný rozdiel v stratovom činiteli medzi napätiami v percentách bol 0,26 % – 2,3 %. Maximálny rozdiel medzi jednotlivými hodnotami bol 4,1 % pri frekvencii 10 Hz pri porovnaní napät'ových hladín 7,5 kV a 10 kV. Obr. 13 b) predstavuje merania na čistom a mokrom izolátore. Priemerný rozdiel je vyšší ako u suchého izolátora a pohybuje sa od 1,52 % do 2,1 %. Pri frekvencii 40 Hz dosahuje najvyšší rozdiel jednotlivých hodnôt 6,7 % medzi napätiami 7,5 kV a 10 kV.



a)



b)

Obr. 14 Frekvenčná závislosť stratového činiteľa pre rôzne napät'ové hladiny: a) prvá úroveň znečistenia; b) druhá úroveň znečistenia

Merania na porcelánovom izolátore znečistenom prvou úrovňou umelého znečistenia (L1) ukazujú, že výsledky merania na suchom izolátore sú takmer identické pre všetky napät'ové úrovne. Priemerný rozdiel v stratovom činiteli pri rôznych napät'och je 1,05 % – 2,22 %. Maximálny rozdiel medzi jednotlivými hodnotami bol 3,94 % pri frekvencii 40 Hz medzi napätiami 5 kV a 10 kV. Rozdiel v stratovom činiteli pri rôznych hladinách napätia sa prejavil pri meraní na mokrom a znečistenom izolátore. Od frekvencie 10 Hz do 300 Hz stratový

činiteľ narastal so zvyšujúcou sa hladinou napätia. Od úrovne frekvencia 350 Hz boli namerané hodnoty pre napät'ové úrovne 7,5 kV a 10 kV opäť blízko seba. Krivky nie sú hladké ako pri meraní na suchej vrstve znečistenia, ale trend kriviek je podobný. Pri meraní na mokrom a prvom stupňom znečistenom izolátore bol priemerný rozdiel medzi hodnotami stratového činiteľa od 9,76 % do 26,8 %. Maximálny rozdiel medzi jednotlivými hodnotami dosiahol 37,38 % pri frekvencii 800 Hz medzi napät'ovými hladinami 5 kV a 10 kV.

Meranie na izolátore kontaminovanom druhým stupňom umelého znečistenia (L2) poukazuje na rovnaký trend zmeny stratového činiteľa ako pri prvom stupni znečistenia (L1), pri porovnaní meraní za sucha aj za mokra. Izolátor so suchou vrstvou znečistenia vykazoval takmer identické výsledky pre všetky napät'ové hladiny. Meranie na druhom stupni znečistenia je zobrazené na Obr. 48 d). Priemerný rozdiel v hodnotách stratového činiteľa na všetkých aplikovaných úrovniach napätia bol v rozmedzí 0,93 % – 1,31 %. Maximálny rozdiel dosiahol hodnotu 3,58 % medzi napätiami 7,5 kV a 10 kV pri frekvencii 30 Hz. Pri porovnaní meraní na stupňoch znečistenia L1 a L2 za mokra dochádza k zvýšeniu hodnôt stratového činiteľa vplyvom napätia. Priemerný rozdiel medzi napät'ovými hladinami sa pohybuje od 8,83 % do 20,26 %. Pri frekvencii 450 Hz medzi 5 kV a 10 kV napätím sa dosiahol maximálny rozdiel medzi jednotlivými hodnotami a to 31,92 %. Všetky priemerné rozdiely stratového činiteľa sú spísané v tabuľke 6. Táto tabuľka porovnáva, ako úroveň napätia ovplyvňuje stratový činiteľ.

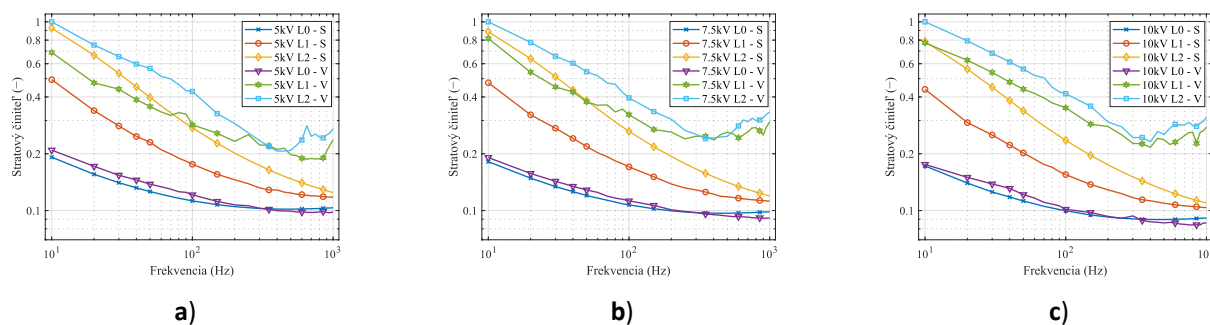
TABUĽKA 6 POROVNANIE ROZDIELOV STRATOVÉHO ČINITEA MEDZI RÔZNYMI ÚROVNAMI NAPÄTIA

Stupeň znečistenia	Priemerný rozdiel (%)		
	5 – 7,5 kV	5 – 10 kV	7,5 – 10 kV
L0 (suchý)	0,26	2,03	2,08
L1 (suchý)	1,05	2,22	1,18
L2 (suchý)	0,93	1,29	1,31
L0 (mokrý)	1,69	1,52	2,1
L1 (mokrý)	18,9	26,8	9,76
L2 (mokrý)	12,96	20,26	8,83

Ako druhé boli porovnané merania na čistom izolátore a na obidvoch úrovniach znečistenia pri rovnakej napät'ovej hladine. Obr. 15 zobrazuje závislosti stratového činiteľa od frekvencii pri rôznych úrovniach znečistenia a rovnakej úrovni napätia.

Na vyhodnotenie rozdielov stratového činiteľa medzi úrovňami znečistenia sa použili vz'ážné hodnoty k maximálnej hodnote stratového činiteľa, ktorá bola nameraná pri frekvencii 10 Hz, pri druhom stupni znečistenia a za mokra. Najvyššia hodnota stratového činiteľa bola pri rovnakej frekvencii a rovnakom znečistení pre všetky použité napätia.

Z jednotlivých grafov na Obr. 15 je zrejmé, že stratový činiteľ má klesajúcu tendenciu takmer v celom frekvenčnom rozsahu. Medzi meraniami na čistom izolátore je minimálny rozdiel a vplyv namáčania je zanedbateľný.



Obr. 15 Závislosti stratového činiteľa na frekvencii pre izolátor v rôznych podmienkach merané pri napätí: a) 5 kV; b) 7,5 kV; c) 10 kV

Pri meraní na znečistených izolátoroch je rozdiel medzi meraním na suchej a mokrej vrstve znečistenia zjavný. Aj napriek tomu, že sa meranie uskutočnilo 10 minút po namáčaní do roztoku a po pripojení napätia, rozdiel medzi meraniami suchej a mokrej vrstvy bol významný, percentuálne v rozmedzí 33,8 % – 54,2%. Čím vyššie bolo napätie, tým väčší bol rozdiel medzi suchými a mokrými podmienkami.

Ak sa porovnajú merania za sucha, najvýraznejší rozdiel bol dosiahnutý pri nízkych frekvenciách. krivky majú rovnaký tvar a so zvyšujúcou sa frekvenciou sa k sebe približujú. Priemerný rozdiel medzi meraniami na suchej znečistenej vrstve bol v rozmedzí 23,7 % – 25,1 %.

Ak sa porovnajú merania na mokrej vrstve znečistenia, krivky sú rovnobežné a nárast stratového činiteľa so zvyšujúcou sa úrovňou znečistenia je zreteľný do dosiahnutia frekvencie v rozmedzí 200 Hz až 300 Hz. Pri vyšších frekvenciách sa tvar kriviek mení a taktiež ich hodnotový rozdiel už nie je tak výrazný. V niektorých častiach frekvenčného spektra sú hodnoty stratového činiteľa rovnaké pre obidve znečistenia. Od frekvencie 400 Hz sa začína hodnota stratového činiteľa pri meraní za mokra zvyšovať (k zvýšeniu dochádza o niečo skôr na druhej úrovni znečistenia), čo by mohlo byť spôsobené polarizáciou. Polarizácia sa prejavuje len v konkrétnych podmienkach v danom frekvenčnom rozsahu. Priemerný rozdiel medzi meraniami za mokra, ktorý sa dosiahol hlavne pri nízkych frekvenciách, je 21,3 % – 14,5 %. Všetky priemerné rozdiely stratového činiteľa sú uvedené v tabuľke 7.

TABUĽKA 7 PRIEMERNÝ ROZDIEL V STRATOVOM ČINITELI PRE RÔZNE ÚROVNE ZNEČISTENIA A RÔZNE PODMIENKY

porovnanie	Priemerný rozdiel (%)		
	5 kV	7,5 kV	10 kV
L1 (S)-L1 (V)	38,7	49,8	54,2
L2 (S)-L2 (V)	33,8	41,1	47,0
L1 (S)-L2 (S)	25,1	25,0	23,7
L1 (V)-L2 (V)	21,3	15,8	14,5

C. Diskusia

Meranie stratového činiteľa vykazuje nízku citlivosť na zmeny napätia. Napriek významnej zmene napätia bola zmena stratového činiteľa minimálna. Boli vypočítané priemerné rozdiely v stratovom činiteli pre účely porovnania meraní na rôznych úrovniach napätia (Tabuľka 6). Iba meranie stratového činiteľa za mokra na znečistenej vrstve poukazujú na jeho

zvýšenie v závislosti od úrovne napätia. Na suchej vrstve znečistenia boli zmeny stratového činiteľa minimálne, čo znamená, že zmena napätia minimálne ovplyvňuje výsledky (rozdiel do 2,5 %). Tento poznatok je potvrdený v štúdiu [93], kde bol účinok zmeny napätia taktiež minimálny. Porovnanie stratového činiteľa na rovnakej hladine napätia ukázalo, že stratový činiteľ je výrazne závislý od stupňa znečistenia izolátora. Stratový činiteľ taktiež závisí od vlhkosti kontaminovanej vrstvy. Účinok namáčania čistého izolátora do vody bol viditeľný, ale rozdiely medzi meraniami na znečistených izolátoroch boli podstatne väčšie. Najmä merania na suchej znečistenej vrstve pre obidve úrovne znečistenia boli porovnateľné a výrazné pri akejkoľvek frekvencii. Najvyššie rozdiely v stratovom činiteli sa dosiahli pri nižších frekvenciách. Ak sa porovnajú merania na mokrých znečistených izolátoroch, krivky boli súbežné s jasným rozdielom do frekvencie okolo 100–200 Hz. Stratový činiteľ sa používa hlavne na diagnostiku starnutia izolácie. Starnutie spôsobuje zhoršenie izolačných schopností čím sa zvyšuje stratový činiteľ, čo je potvrdené v štúdiách na izolátoroch [37]-[40]. Náš výskum poukazuje, že tak ako starnutie aj znečistenie zhoršuje izolačné vlastnosti povrchu izolátora a zvyšuje stratový činiteľ. Ďalej dochádza k zväčšeniu stratového činiteľa vplyvom vlhkosti a zrejme aj polarizácie. Stratový činiteľ bol väčší pri meraní na druhom stupni znečistenia v celom frekvenčnom spektre v porovnaní s nižšou úrovňou znečistenia.

IX. MERANIE STRATOVÉHO ČINITELA PRI VYSOKOM NAPÄTÍ A PRIEMYSELNEJ FREKVENCII 50 HZ

Po vyhodnotení merania na Technickej univerzite v Grazi, výskum a merania pokračovali vo vysokonapäťovom laboratóriu Technickej univerzity v Košiciach. Merania boli realizované na štyroch rôznych izolátoroch. Keramické izolátory boli reprezentované izolátorom U 120 B zo skla a novým porcelánovým izolátorom U 70 BL, ktorý bol použitý pri meraní v predchádzajúcej kapitole VIII. Kompozitné izolátory použité pri meraní boli zo silikónovej gúmy (SMS 25/0/4/M) a etylén-vinyl-acetátu (RST-DS28R-CT). Bližšie informácie o izolátoroch sú v tabuľke 1.

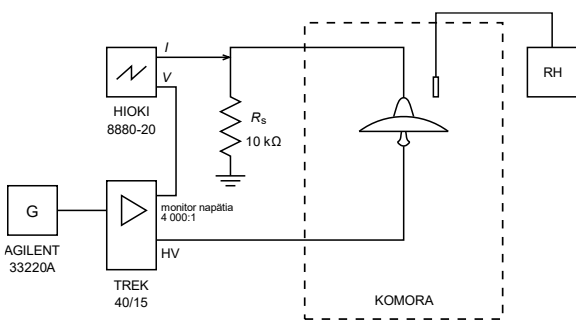
A. Postup merania

Všetky štyri izolátory sú vhodné pre aplikácie na vedeniach s vysokým napätím. Minimálne výdržné napätia za mokra sa pohybuje v rozmedzí 40 kV až 98 kV. Vzorky sa líšia aj v povrchovej vzdialenosti, ktorá je v rozmedzí 330 mm až 760

mm. Hydrofóbnosť kompozitných izolátorov je vyššia ako u keramických. Hydrofóbnosť kompozitných izolátorov je triedy 1 až 2 a izolátor zo silikónovej gumeny mal o niečo väčšiu hydrofóbnosť ako izolátor z etylén-vinyl-acetátu. Porcelánový izolátor s glazúrou by mohol patriť do triedy hydrofóbnosti 4 až 5 a sklenený izolátor mal najhoršie hydrofóbné vlastnosti a dá sa považovať za hydrofílny (trieda 6). Triedy hydrofóbnosti sú na Obr. 2.

Merania boli realizované na čistých izolátoroch a pri všetkých štyroch úrovniach znečistenia spomenutých v tabuľke 2. Po namočení do roztokov znečistenia sa urobila analýza rozpustnej zložky (ESDD) a nerozpustnej zložky (NSDD) znečistenia. Výsledky meraní týchto parametrov sú v tabuľke 3. Nanesená vrstva znečistenia prešla minimálne 24 hodinovým procesom sušenia. Jednotlivé izolátory sa postupne umiestnili do izolovanej komory, kde sa sledovala vlhkosť prostredia. Meranie stratového činiteľa začínalo na suchej vrstve znečistenia. Použité napäťové hladiny boli 8 kV, 13,856 kV, 19 kV a 24 kV. Zvolená hodnota 13,856 kV je fázovou hodnotou ku združenému napätiu 24 kV. Po odmeraní izolátorov za sucha sa metódou naparovania začala zvyšovať vlhkosť v komore. Meranie prebiehalo pri vlhkostiach od 50 % do 90 %. Zvyšovanie vlhkosti bolo postupné s krokom 10 %. Merat' sa začalo po 10 minútach od dosiahnutia jednotlivých úrovní vlhkosti, aby prebehla potrebná saturácia znečistenej vrstvy. Po odmeraní stratového činiteľa na všetkých úrovniach vlhkosti sa meranie ukončilo a meraný izolátor sa dôkladne umyl, vysušil a namočil do roztoku vyššej úrovne znečistenia. Proces sa opakoval pri všetkých typoch izolátorov a úrovniach znečistenia rovnako, aby bola zachovaná konzistentnosť meraní.

Na generovanie napätia sa použil generátor Agilent 33210A (G), na ktorom sa menili hodnoty napätia a následne sa nízke napätie generátora zosilnilo na vysoké napätie pomocou zosilňovača Trek 40/15, ktorý má pomer 4000:1. Pomocou využitia nezávislého zdroja sa zabránilo rušeniu zo siete. Vysoké napätie bolo privedené na spodnú časť izolátora. Prúd z hornej časti izolátora bol meraný na snímacom rezistore s hodnotou odporu 40 kΩ pomocou analyzátora siete Hioki 8880 20. Hodnota vlhkosti sa sledovala pomocou snímača vlhkosti, ktorý mal senzor umiestnený v blízkosti izolátora v komore. Schéma zapojenia je na Obr. 16.

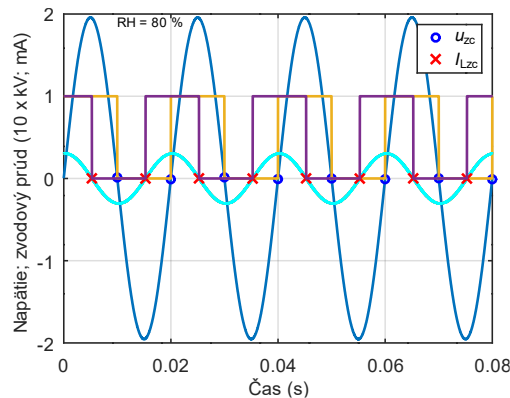


Obr. 16 Schéma zapojenia meracích prístrojov pre meranie pri frekvencii 50 Hz a vysokom napätí

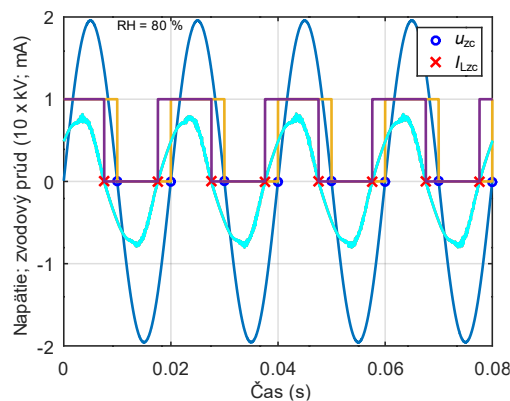
Stratový činiteľ bol počítaný ako:

$$\tan\delta = \tan(90 - \phi) \tag{3}$$

kde ϕ reprezentuje fázový posun medzi použitým napätím a pretekajúcim zvodovým prúdom po povrchu izolátora. Bola použitá metóda The zero crossing method (ZCM), ktorá určuje fázový posun medzi napätím a zvodovým prúdom rozpoznávaním časových bodov, v ktorých napätie a zvodový prúd menia svoju polaritu z kladnej na zápornú a opačne. Na Obr. 17 je zobrazený tvar vln štyroch periód aplikovaného napätia (fázového) a zvodového prúdu pretekajúceho cez izolátor pre dve rôzne úrovne znečistenia.



a)



b)

Obr. 17 Fázový posun medzi aplikovaným fázovým napätím a zvodovým prúdom pri relatívnej vlhkosti 80 % na: a) čistom izolátore; b) znečistenom izolátore stupňom znečistenia L4

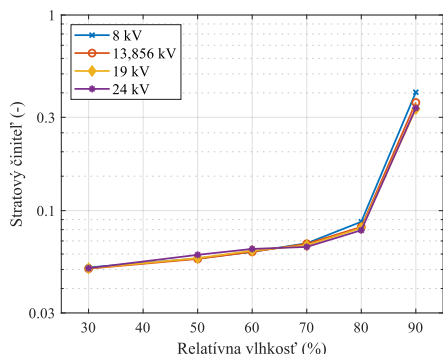
ZCM teda možno použiť na určenie časového rozdielu medzi aplikovaným napätím a prúdom pre každú polperiódu a na výpočet zodpovedajúceho fázového posunu medzi napätím a zvodovým prúdom podľa (19):

$$\phi = (t_{uzc} - t_{izc}) \cdot 18^3 = \Delta t_{zc} \cdot 18^3 \tag{4}$$

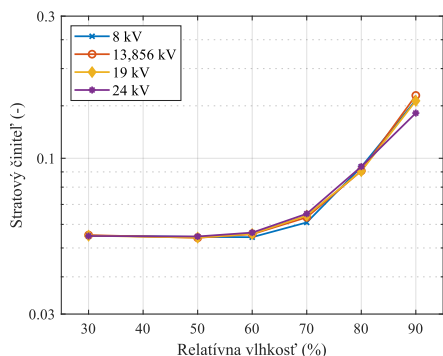
Napätie a zvodový prúd boli nepretržite merané počas jednej sekundy, výsledkom čoho bolo päťdesiat meraných periód. Vypočítal sa zodpovedajúci stratový činiteľ pre každú polperiódu a určila sa jeho priemerná hodnota so smerodajnou odchýlkou. Smerodajná odchýlka pre stratový činiteľ bola priemerné 0,004. Pre lepšiu prehľadnosť sú časové priebehy prechodov nulou napätia (u_{zc}) a zvodového prúdu (I_{zsc}) znázornené aj vo farebne odlišených pravouhlých priebehoch na Obr. 17.

B. Vplyv napätia na stratový činiteľ

Ako prvé bolo vyhodnotený vplyv napätia na stratový činiteľ. Kvôli podobnosti vplyvu napätia pri meraní jednotlivých stupňov znečistenia sa graficky vyhodnotia iba výsledky merania na čistom izolátore a pri štvrtom stupni znečistenia.



a)



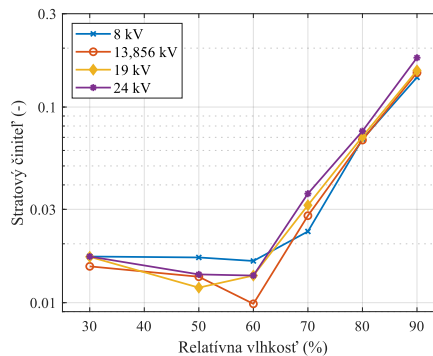
b)

Obr. 18 Porovnanie merania stratového činiteľa pri rôznom napätí na čistých izolátoroch: a) sklo; b) porcelán

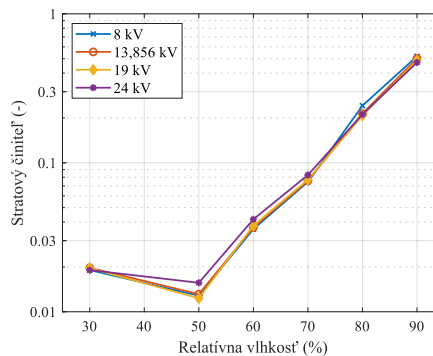
Z grafov merania na čistom izolátore na Obr. 18 a Obr. 19 je zjavné, že vplyv zmeny napätia pri meraní na čistých izolátoroch nie je výrazný. Na sklennom izolátore bol rozdiel najvýraznejší pri vyšších relatívnych vlhkosťach a dosiahol 20,33 % medzi napätiami 8 kV a 24 kV. Pri nízkych relatívnych vlhkosťach to bolo do 5 %. Na porcelánovom izolátore dosahovali percentuálne rozdiely najnižšie hodnoty. Najvyšší bol nameraný pri relatívnej vlhkosťi 90 % a to 10,51 %.

V prípade merania na izolátore zo silikónovej gummy zobrazenom na Obr. 19 boli rozdiely pri relatívnych vlhkosťach 50 % až 70 % výraznejšie. Stratový činiteľ na izolátore zo silikónovej gummy však dosahoval trikrát nižšie hodnoty ako na keramických izolátoroch. Najväčší rozdiel bol dosiahnutý pri vlhkosťi 60 % medzi napätovými hladinami 8 kV a 13,856 kV a to 65,6 %. V týchto prípadoch mohlo dôjsť ku chybe merania vplyvom citlivosti meracích prístrojov. Zvyšné najvyššie rozdiely sa pohybovali okolo 18 % až 35 % a v prípade merania za sucha bol rozdiel zanedbateľný. Vplyv napätia na EVA

izolátor nebol výrazný a najvyšší dosiahnutý rozdiel pri relatívnej vlhkosťi 50 % bol 18 %.



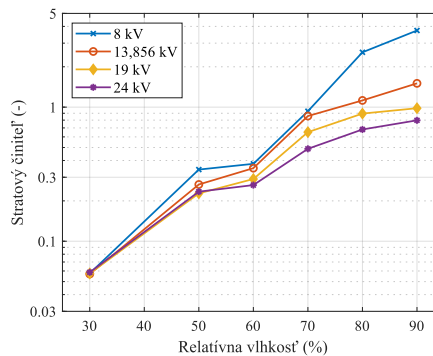
a)



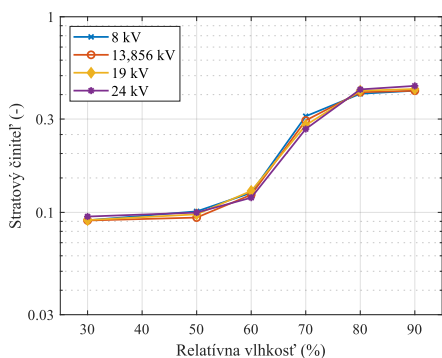
b)

Obr. 19 Porovnanie merania stratového činiteľa pri rôznom napätí na čistých izolátoroch: a) silikónová guma; b) EVA

Na Obr. 20 a Obr. 21 sú zobrazené merania na rôznych napätových hladinách pre štvrtý stupeň znečistenia. Najväčší vplyv mala zmena napätia na sklenný izolátor. V suchom stave bol rozdiel len 1,83 %, no pri zvyšovaní vlhkosťi tento rozdiel narástol na 44 % pri relatívnej vlhkosťi 60 %, na 90 % pri relatívnej vlhkosťi 70 % a až na 367 % pri relatívnej vlhkosťi 90 %.



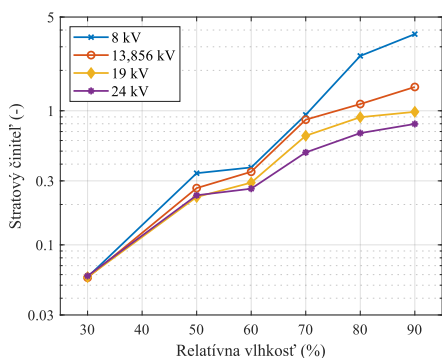
a)



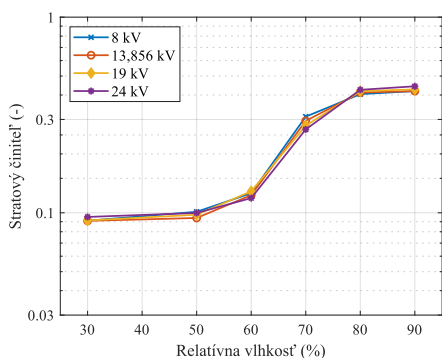
b)

Obr. 20 Porovnanie merania stratového činiteľa pri rôznom napätí na izolátoroch znečistených štvrtým stupňom znečistenia: a) sklo; b) porcelán

Vplyv napätia nebol tak výrazný pri meraní na zvyšných izolátoroch. V prípade porcelánového izolátora to bolo 16 % pri relatívnej vlhkosti 70 % medzi napäťovými hladinami 8 kV a 24 kV.



a)



b)

Obr. 21 Porovnanie merania stratového činiteľa pri rôznom napätí na izolátoroch znečistených štvrtým stupňom znečistenia: a) sil. guma; b) EVA

Na silikónovom izolátore sa najvyššie rozdiely pohybovali v rozmedzí 8 % až 17,8 %. Najvyššia hodnota bola dosiahnutá pri relatívnej vlhkosti 60 % medzi napäťovými hladinami 8 kV a 24 kV. Vplyv zmeny napätia na EVA izolátor bol vo väčšine

prípadoch nízky do 6 % až na meranie pri relatívnej vlhkosti 70 %, kde bol rozdiel 23,67 %.

Vo väčšine prípadov bol najväčší rozdiel pri porovnaní meraní na napäťovej hladine 8 kV a 24 kV.

Ak sa porovnajú merania na jednotlivých znečistených izolátoroch, potvrdzuje sa, že najväčší vplyv mala zmena napätia na sklenený izolátor. Výrazný nárast stratového činiteľa bol zaznamenaný už pri treťom stupni znečistenia. V ostatných prípadoch sa rozdiely pohybovali v rozmedzí 0,11 % až 45 %, kde nie je pozorované stabilné zvyšovanie alebo znižovanie stratového činiteľa v závislosti od relatívnej vlhkosti alebo stupňa znečistenia. Nepravidelnosť zmeny v stratovom činiteli vplyvom zmeny napätia platí aj pre ostatné izolátory. V prípade porcelánového izolátora bol najvyšší rozdiel nameraný na druhom stupni znečistenia pri relatívnej vlhkosti 90 % a to 41,5 %. Odhliadnuc od merania na čistom izolátore, izolátor zo silikónovej gummy dosiahol najväčší rozdiel 31,5 % na treťom stupni znečistenia pri najvyššej relatívnej vlhkosti. EVA izolátor dosahoval podobne ako silikónový izolátor nízke hodnoty rozdielu medzi napäťovými hladinami. Jediný prípad kedy bol dosiahnutý vyšší rozdiel než hodnoty v tabuľkách, bol medzi napäťovými hladinami 8 kV a 14 kV na prvom stupni znečistenia pri relatívnej vlhkosti 50 % a to 95 %.

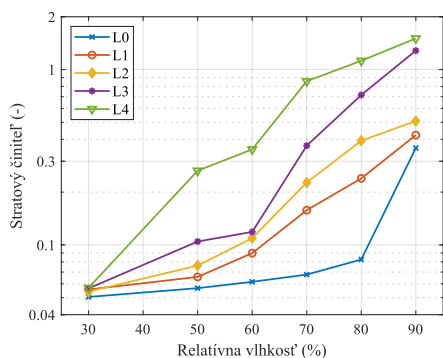
C. Vplyv znečistenia na stratový činiteľ

Hlavný zámer merania bolo vyhodnotenie vplyvu znečistenia na stratový činiteľ. Stratový činiteľ bol vyhodnotený pre všetky hladiny napätia, no na porovnanie meraní pri rôznom znečistení bola zvolená napäťová hladina 13,856 kV ako fázová hodnota používaných 24 kV vedení. Na Obr. 22 a Obr. 23 sú zobrazené výsledky merania na rôznych izolátoroch na napäťovej hladine 13,856 kV. Obrázky poukazujú na to, že znečistenie povrchu izolátorov vplyva na hodnotu stratového činiteľa.

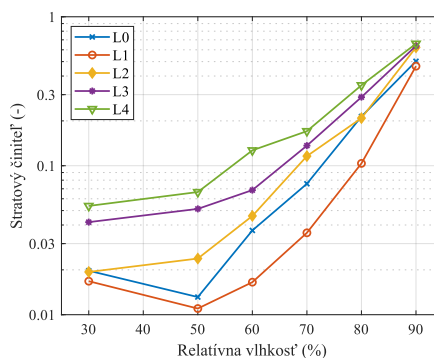
V prípade merania na sklenenom izolátore, stratový činiteľ zvyšoval vplyvom zväčšujúcej sa úrovne znečistenia pri všetkých vlhkostiach okrem merania za sucha. Na suchej vrstve znečistenia bol rozdiel v stratovom činiteli minimálny, no vplyv znečistenia v porovnaní s čistým izolátorom je badaťelný.

Na porcelánovom izolátore (Obr. 22 b)) je taktiež vidieť jasný rozdiel medzi meraním na čistom a znečistenom izolátore. V prípade tohto izolátora boli rozdiely medzi znečisteniami zreteľné pri relatívnej vlhkosti 70 % a 80 %. Ku priblíženiu alebo zoskupeniu hodnôt stratového činiteľa pre jednotlivé znečistenia dochádzalo pri nižších relatívnych vlhkostiach.

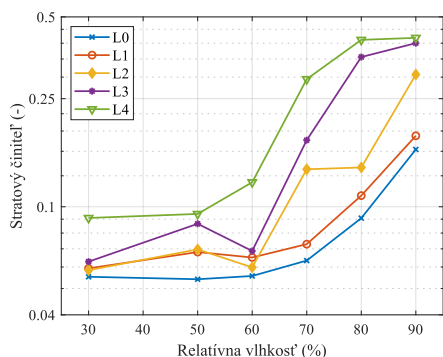
Silikónový izolátor na Obr. 23 a) preukázal najväčší rozdiel medzi meraniami na čistom a znečistenom izolátore. Znečistenia majú však podobné hodnoty stratového činiteľa a stratový činiteľ sa nezvyšuje vplyvom znečistenia pri nízkych hodnotách relatívnej vlhkosti. Až štvrtý stupeň znečistenia pri relatívnej vlhkosti 60 % začal narastať a odčlenil sa od ostatných znečistení. Tretí stupeň znečistenia sa odčlenil až pri relatívnej vlhkosti 90 %. To môže byť spôsobené hydrofóbnosťou materiálu, s čím súvisí aj menšie množstvo znečistenia, ktoré sa dokáže usadiť na izolátore.



a)



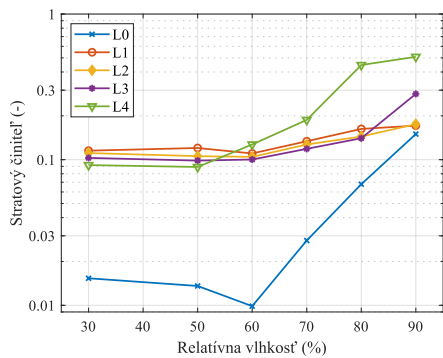
b)



b)

Obr. 22 Porovnanie merania stratového činiteľa pri rôznom znečistení na izolátoroch na napät'ovej hladine 13,856 kV: a) sklo; b) porcelán

Eva izolátor na Obr. 23 b) bol jediný, pri ktorom mal stratový činiteľ vyššie hodnoty na čistom izolátore ako na prvom stupni znečistenia. Pri relatívnej vlhkosti 80 % dokonca dosiahol vyššie hodnoty ako znečistenie L2. Pri porovnaní merania na čistom izolátore s meraniami na treťom a štvrtom stupni znečistenia však boli hodnoty stratového činiteľa oveľa nižšie.



a)

Obr. 23 Porovnanie merania stratového činiteľa pri rôznom znečistení na izolátoroch na napät'ovej hladine 13,856 kV: a) silikónová guma; b) EVA

Celkovo najvýraznejšie rozdiely boli namerané na sklenenom izolátore. Rozsah stratového činiteľa bol od 0,05 po 1,505. Pri ostatných meraných materiáloch stratový činiteľ nedosiahol vyššiu hodnotu ako 0,659, ktorá bola nameraná na EVA izolátore pri relatívnej vlhkosti 90 % a štvrtom stupni znečistenia.

Ak medzi sebou porovnáme sklenený a porcelánový izolátor, sklenený izolátor dosahoval nižšie hodnoty stratového činiteľa len v suchom stave. Pri všetkých ostatných vlhkostiach bol stratový činiteľ na sklenenom izolátore vyšší. To sa zhoduje s triedou hydrofóbnosti materiálu. Glazúra na porcelánovom izolátore zvyšuje hydrofóbnosť a tým lepšie odoláva vlhkosti a znečisteniu.

Silikónový izolátor dosahoval najnižšie hodnoty stratového činiteľa v čistom stave. No ak ho porovnáme s EVA izolátorom, v znečistenom stave mal stratový činiteľ vyššie hodnoty. To platí vo väčšine prípadov do relatívnej vlhkosti 70 %. Od relatívnej vlhkosti 80 % mal EVA izolátor vyššie hodnoty stratového činiteľa s výnimkou merania pri relatívnej vlhkosti 80 % a štvrtým stupňom znečistenia. Izolátor zo silikónovej gummy teda lepšie odolával najťažším podmienkam pri relatívnej vlhkosti 80 % a 90 % a štvrtom stupni znečistenia.

Ak sa porovnajú všetky materiály, tak v najťažších podmienkach pri relatívnej vlhkosti 90 % a štvrtom stupni znečistenia najlepšie obstál porcelánový izolátor. No s výnimkou štvrtého stupňa znečistenia, dosiahol silikónový izolátor najnižšie hodnoty stratového činiteľa pre vysoké hodnoty relatívnej vlhkosti.

D. Diskusia

Po meraní stratového činiteľa pri vysokom napätí v kapitole VIII, bolo meranie v tejto kapitole modifikované tak, aby zohľadňovalo pozitíva predošlého merania a zameralo sa na použiteľnosť v praxi. Keďže výsledky merania v kapitole VIII poukazovali na to, že stratový činiteľ má najrelevantnejšie výsledky pri nižších frekvenciách, zvolila sa pre nové meranie priemyselná frekvencia 50 Hz, s využitím zdroja, ktorý eliminoval rušenie zo siete, aby výsledky meraní neboli skreslené. Využila sa metodika zvyšovania vlhkosti z kapitoly VII a taktiež všetky pripravené znečistenia.

Meranie potvrdilo to, že stratový činiteľ je len mierne závislý od zmeny napätia aj napriek tomu, že napätie bolo trikrát vyššie pri porovnaní najnižšieho a najvyššieho použitého napätia. Významnejší vplyv napätia sa prejavil len na sklenenom izolátore a to v najťažších podmienkach pri najvyššej relatívnej vlhkosti a najvyššom stupni znečistenia.

Stratový činiteľ preukázal závislosť na zmene relatívnej vlhkosti a úrovne kontaminácie. Stratový činiteľ vykazoval málo prípadov podobnosti výsledkov medzi znečisteniami a rozdiely medzi hodnotami boli výrazné.

V prípade kompozitných materiálov boli výsledky menej zreteľné. To je spôsobené vynikajúcimi hydrofóbnymi vlastnosťami. Hydrofóbné vlastnosti spôsobujú nerovnomernosť kontaminácie na povrchu izolátora, čo spôsobuje to, že usadená vrstva kontaminácie po namočení do roztoku môže byť rôzna. Taktiež výsledky merania na čistých izolátoroch dokazujú, že ich izolačné vlastnosti sú lepšie ako v prípade sklenených a porcelánových izolátorov. Rozdiely v stratovom činiteľi medzi jednotlivými materiálmi neboli tak výrazné. To znamená, že stratový činiteľ nie je tak výrazne závislý od použitého izolačného materiálu a jeho meranie môže byť štandardizované pre väčší rozsah používaných izolačných materiálov pre izolátory.

X. ZÁVER

Izolátory sú nevyhnutnou súčasťou elektrizačnej sústavy. Ich spoľahlivosť je dôležitým faktorom pre neprerušovanú dodávku elektrickej energie. Ekonomické dopady pri ich zlyhaní sú niekoľkonásobne vyššie ako ich cena. Aby izolátor plnil svoju úlohu počas celej jeho životnosti, je nutné, aby bol pre daný účel zvolený taký typ izolátora, ktorý zvládne všetky druhy namáhania a je čo najviac prispôbený okolitým podmienkam. Izolátor vhodný na použitie v prímorských oblastiach, ktorý musí zvládať vysokú úroveň znečistenia s vysokou vlhkosťou vzduchu, nemusí vyhovovať v oblastiach, kde je väčšie sucho a teplo ako napríklad v púštnych oblastiach. Preto treba pri výbere izolátorov zhodnotiť všetky namáhania a faktory, ktoré budú ovplyvňovať jeho funkčnosť a životnosť.

Znečistenie a vonkajšie podmienky sú dlhodobo najväčšími problémami spoľahlivosti vonkajších izolátorov. S týmito faktormi súvisia najčastejšie poruchy spojené so zlyhaním izolátora, ktoré sú spomenuté v práci spomenutej. V práci sú taktiež spomenuté diagnostické metódy, ktorými sa izolátory v prevádzke kontrolujú.

Experimentálna časť publikácie je venovaná výskumu a diagnostike štandardných materiálov, ktoré sa používajú na prenosových vedeniach. Na výskum sa použili porcelánové, sklenené a kompozitné izolátory. Na meranie bol použitý stratový činiteľ, ktorý sa v elektroenergetike využíva najmä na diagnostiku transformátorov a káblov.

Súčasťou meraní bolo aj vyhodnotenie závažnosti znečistenia pomocou štandardizovanej metódy merania ESDD a NSDD, ktorá je opísaná v norme IEC/TS 60815-1:2008. Vďaka týmto meraniam sa dalo predpokladať, aké množstvo znečistenia sa na izolátoroch nachádza. Množstvo znečistenia, ktoré sa usadilo na izolátoroch bolo ovplyvnené hydrofóbnymi vlastnosťami izolátorov. Zatiaľ čo sa na porcelánovom a

sklenenom izolátore vytvárala rovnomerná vrstva znečistenia, kompozitné izolátory boli voči vytvoreniu tejto vrstvy odolnejšie. Ak by sa znečistenie na izolátoroch diagnostikovalo vizuálnou kontrolou, tak na kompozitných izolátoroch bolo znečistenia oveľa menej. Merania však ukázali, že rozdiely v hodnotách rozpustnej zložky nie sú až tak výrazné a zhluky znečistenia na kompozitných izolátoroch dokážu zachytiť vysoké množstvo znečistenia. To taktiež vytvára väčšie nehomogenity, čo môže viesť ku poškodeniam povrchu v týchto miestach. Nedostatkom normy bolo taktiež nejasné určenie množstva destilovanej vody a veľkosť povrchu, z ktorého sa znečistenie má zmývať. Pre porovnanie rôznych druhov izolátorov je vhodné dodržať rovnaký pomer zmývanej oblasti a množstva použitej destilovanej vody.

Diagnostika povrchu znečistených izolátorov pomocou merania stratového činiteľa sa vykonala tromi rôznymi metódami pri nízkom aj vysokom napätí. Všetky merania potvrdili použiteľnosť stratového činiteľa za účelom určenia závažnosti znečistenia.

Ako prvé sa merali izolátory pri nízkom napätí 5 V a v prípade stratového činiteľa vo frekvenčnom pásme 1 Hz až 200 kHz. Výsledky merania ukázali, že táto metodika je vhodná na určenie znečistenia izolátorov. Štyri rôzne stupne znečistenia zvyšovali hodnoty stratového činiteľa pri rôznych úrovniach relatívnej vlhkosti. Stratový činiteľ, ktorý bol meraný v širokom frekvenčnom pásme poukázal na vytváranie lokálnych maxím, ktoré sa vplyvom vlhkosti a znečistenia posúvajú do vyšších frekvencií. Vyhodnotením hodnôt stratového činiteľa sa prišlo na to, že najvhodnejšou oblasťou frekvencií pre vyhodnotenie vplyvu znečistenia bola oblasť vysokých frekvencií nad 10 kHz, kde vo všetkých prípadoch dochádza ku poklesu stratového činiteľa po dosiahnutí spomínaných lokálnych maxím. Výhodou je, že táto frekvenčná oblasť je vyššia ako priemyselná frekvencia, čo zabraňuje rušeniu zo siete.

Po úspešnom meraní pri nízkom napätí sa výskum posunul na overenie metodiky merania stratového činiteľa pri vysokom napätí. Na Technickej univerzite v Grazi sa realizovalo meranie stratového činiteľa na novom porcelánovom izolátore pri rôznych úrovniach vysokého napätia. Na diagnostiku sa použili tri rôzne úrovne od 5 kV po 10 kV vo frekvenčnom rozsahu od 1 Hz po 1 kHz. Merania na čistom izolátore a dvoch rôznych stupňoch znečistenia za sucha a po namočení do roztoku znečistenia ukázali, že hodnoty stratového činiteľa sa zvyšujú vplyvom znečistenia aj vlhkosti vrstvy. Dôležitým poznatkom bol vplyv úrovne napätia na stratový činiteľ. V prípade stratového činiteľa bol vplyv zmeny napätia minimálny. Na rozdiel od merania pri nízkom napätí boli výsledky stratového činiteľa najzreteľnejšie pri nižších frekvenciách do 100 Hz. Ukázalo sa, že stratový činiteľ vďaka spomenutým záverom, môže byť vhodná veličina na určenie úrovne kontaminácie izolátorov.

Na základe zistení z predchádzajúceho merania bolo ďalším krokom meranie pri priemyselnej frekvencii 50 Hz pri napätíach, ktoré sa používajú v praxi. Použili sa napätia v rozsahu 8 kV až 24 kV, čo je hodnota združeného napätia používaného pre vedenia vysokého napätia. Toto meranie bolo realizované na štyroch rôznych druhoch izolátorov, štyroch úrovniach znečistenia a relatívnej vlhkosti od 30 % po 90 %.

Meranie potvrdilo to, že stratový činiteľ nie je výrazne závislý od zmeny napätia. Jediná významná závislosť sa prejavila na sklenenom izolátore a pri najvyššej úrovni znečistenia. Aj tretie rozsiahle meranie potvrdilo použiteľnosť stratového činiteľa na diagnostiku znečistenia izolátorov a vysokú citlivosť na zmenu stavu povrchu. Nízka citlivosť stratového činiteľa na zmenu napätia môže zabrániť znehodnoteniu nameraných výsledkov vplyvom výkyvov napätia v sieti.

REFERENCES

- [1] A. E. Schwalm, "Insulators", v *IEEE PES T&D 2010*, New Orleans, LA, USA, 2010, s. 1–2. doi: [10.1109/TDC.2010.5484355](https://doi.org/10.1109/TDC.2010.5484355).
- [2] E. A. Cherney, "50 years in the development of polymer suspension-type insulators", *IEEE Electr. Insul. Mag.*, roč. 29, č. 3, s. 18–26, máj. 2013, doi: [10.1109/MEL.2013.6507410](https://doi.org/10.1109/MEL.2013.6507410).
- [3] J. F. Hall, "History and bibliography of polymeric insulators for outdoor applications", *IEEE Trans. Power Delivery*, roč. 8, č. 1, s. 376–385, jan. 1993, doi: [10.1109/61.180359](https://doi.org/10.1109/61.180359).
- [4] M. T. Gençoğlu, "THE COMPARISON OF CERAMIC AND NON-CERAMIC INSULATORS", *Engineering Sciences*, vol. 2, no. 4, pp. 274–294, May. 2007, doi: [10.1109/61.180359](https://doi.org/10.1109/61.180359).
- [5] M. Farzaneh a W. A. Chisholm, *Insulators for icing and polluted environments*. Piscataway, NJ : Hoboken, NJ: IEEE Press ; J. Wiley, 2009.
- [6] K. MARTON, *Technika vysokých napätí I*. Bratislava : ALFA, 1984.
- [7] V. Mach, *Technika vysokého napětí*. Ostrava: Vysoká škola báňská - Technická univerzita Ostrava, Fakulta elektrotechniky a informatiky, 2001.
- [8] K. Marimuthu, S. Vynatheya, N. Vasudev, a P. Raja, "Quality Analysis of Ceramic Insulators Under Electro Thermal Stresses", v *2019 International Conference on High Voltage Engineering and Technology (ICHVET)*, Hyderabad, India, feb. 2019, s. 1–6. doi: [10.1109/ICHVET.2019.8724303](https://doi.org/10.1109/ICHVET.2019.8724303).
- [9] M. Rabelo, S. Sanyal, T. Kim, J.-A. Son, I.-H. Choi, a J. Yi, "Influence of electrolytic and crevice corrosion on mechanical resistance of porcelain insulators", *Engineering Failure Analysis*, roč. 124, s. 105317, jún. 2021, doi: [10.1016/j.engfailanal.2021.105317](https://doi.org/10.1016/j.engfailanal.2021.105317).
- [10] L. Yang *et al.*, "Experimental investigation on flashover characteristics of hollow porcelain insulator under extreme rainfall", *Electric Power Systems Research*, roč. 203, s. 107659, feb. 2022, doi: [10.1016/j.epsr.2021.107659](https://doi.org/10.1016/j.epsr.2021.107659).
- [11] INMR, "Manufacturing Glass Insulators", *INMR*, apr. 13, 2019. <https://www.inmr.com/manufacturing-glass-insulators/> (cit nov. 04, 2021).
- [12] H. de Santos a M. Á. Sanz-Bobi, "Research on the pollution performance and degradation of superhydrophobic nano-coatings for toughened glass insulators", *Electric Power Systems Research*, roč. 191, s. 106863, feb. 2021, doi: [10.1016/j.epsr.2020.106863](https://doi.org/10.1016/j.epsr.2020.106863).
- [13] N. A. Othman, M. A. M. Piah, a Z. Adzis, "Contamination effects on charge distribution measurement of high voltage glass insulator string", *Measurement*, roč. 105, s. 34–40, júl. 2017, doi: [10.1016/j.measurement.2017.03.044](https://doi.org/10.1016/j.measurement.2017.03.044).
- [14] R. Taherian, "Application of Polymer-Based Composites", v *Electrical Conductivity in Polymer-Based Composites: Experiments, Modelling and Applications*, Elsevier, 2019, s. 131–181. doi: [10.1016/B978-0-12-812541-0.00006-9](https://doi.org/10.1016/B978-0-12-812541-0.00006-9).
- [15] N. A. Othman *et al.*, "Characterization of charge distribution on the high voltage glass insulator string", *Journal of Electrostatics*, roč. 72, č. 4, s. 315–321, aug. 2014, doi: [10.1016/j.elstat.2014.05.003](https://doi.org/10.1016/j.elstat.2014.05.003).
- [16] INMR, "Manufacturing & Quality of Toughened Glass Insulators", *INMR*, máj. 08, 2020. <https://www.inmr.com/manufacturing-quality-toughened-glass-insulators/> (cit nov. 05, 2021).
- [17] "Insulators", v *Transmission and Distribution Electrical Engineering*, Elsevier, 2007, s. 163–180. doi: [10.1016/B978-075066673-2/50010-0](https://doi.org/10.1016/B978-075066673-2/50010-0).
- [18] S. M. Gubanski, "Ageing of composite insulators", v *Ageing of Composites*, Elsevier, 2008, s. 421–447. doi: [10.1533/9781845694937.3.421](https://doi.org/10.1533/9781845694937.3.421).
- [19] E. Kuffel, W. S. Zaengl, a J. Kuffel, *High voltage engineering: fundamentals*, 2nd ed. Oxford ; Boston: Butterworth-Heinemann, 2000.
- [20] C. Mayoux, "Degradation of insulating materials under electrical stress", *IEEE Trans. Dielect. Electr. Insul.*, roč. 7, č. 5, s. 590–601, okt. 2000, doi: [10.1109/TDEI.2000.879355](https://doi.org/10.1109/TDEI.2000.879355).
- [21] . Al-Gheilani, W. Rowe, Y. Li, a K. L. Wong, "Stress Control Methods on a High Voltage Insulator: A Review", *Energy Procedia*, roč. 110, s. 95–100, mar. 2017, doi: [10.1016/j.egypro.2017.03.112](https://doi.org/10.1016/j.egypro.2017.03.112).
- [22] G. Carlo Montanari, P. Morshuis, P. Seri, a R. Ghosh, "Ageing and reliability of electrical insulation: the risk of hybrid AC/DC grids", *High Voltage*, roč. 5, č. 5, s. 620–627, okt. 2020, doi: [10.1049/hve.2019.0371](https://doi.org/10.1049/hve.2019.0371).
- [23] M. Váry a E. Firický, "Electrical insulation system ageing caused by electrical, multifactorial and environmental stresses", *Posterus*, jún. 12, 2013. <http://www.posterus.sk/?p=15765> (cit nov. 09, 2021).
- [24] Ali. A. Salem, R. A. Rahman, M. S. Kamarudin, a N. A. Othman, "Factors and models of pollution flashover on high voltage outdoor insulators: Review", v *2017 IEEE Conference on Energy Conversion (CENCON)*, Kuala Lumpur, Malaysia, okt. 2017, s. 241–246. doi: [10.1109/CENCON.2017.8262491](https://doi.org/10.1109/CENCON.2017.8262491).
- [25] N. A. Othman, M. A. M. Piah, a Z. Adzis, "Contamination effects on charge distribution measurement of high voltage glass insulator string", *Measurement*, roč. 105, s. 34–40, júl. 2017, doi: [10.1016/j.measurement.2017.03.044](https://doi.org/10.1016/j.measurement.2017.03.044).
- [26] R. Castillo-Sierra, O. Oviedo-Trespalcacios, J. E. Candelo-Becerra, J. D. Soto, a M. Calle, "A novel method for prediction of washing cycles of electrical insulators in high pollution environments", *International Journal of Electrical Power & Energy Systems*, roč. 130, s. 107026, sep. 2021, doi: [10.1016/j.ijepes.2021.107026](https://doi.org/10.1016/j.ijepes.2021.107026).
- [27] W. L. Vosloo, R. E. Macey, a C. de Tourreil, *The practical guide to outdoor high voltage insulators*. Johannesburg: Crown Publications, 2004.
- [28] C. Charalambous, M. Danikas, Y. Yin, N. Vordos, J. W. Nolan, a A. Mitropoulos, "Study of the Behavior of Water Droplets Under the Influence of a Uniform Electric Field on Conventional Polyethylene and on Crosslinked Polyethylene (XLPE) with MgO Nanoparticles Samples", *Eng. Technol. Appl. Sci. Res.*, roč. 7, č. 1, s. 1323–1328, feb. 2017, doi: [10.48084/etasr.813](https://doi.org/10.48084/etasr.813).
- [29] A. S. Paramane a K. S. Kumar, "A Review on Nanocomposite Based Electrical Insulations", *Transactions on Electrical and Electronic Materials*, roč. 17, č. 5, s. 239–251, okt. 2016, doi: [10.4313/TEEM.2016.17.5.239](https://doi.org/10.4313/TEEM.2016.17.5.239).
- [30] Selection and dimensioning of high-voltage insulators intended for use in polluted conditions — Part I: Definitions, information and general principles, DD IEC/TS 60815-1:2008, 2008.
- [31] Guidance of the measurement of hydrophobicity of insulator surface, IEC 62073:2016, 2016
- [32] M. Zhang *et al.*, "Research on nonlinear characteristics for frequency domain dielectric response of transformer oil-paper insulation", *Measurement*, roč. 204, s. 112103, nov. 2022, doi: [10.1016/j.measurement.2022.112103](https://doi.org/10.1016/j.measurement.2022.112103).
- [33] V. Valentic, S. Grzinic, a D. Dobrec, "Testing the electrical insulation system of power transformer based on measuring factor of dielectric losses", v *IEEE EUROCON 2017 -17th International Conference on Smart Technologies*, Ohrid, Macedonia, júl. 2017, s. 423–427. doi: [10.1109/EUROCON.2017.8011146](https://doi.org/10.1109/EUROCON.2017.8011146).
- [34] Y. Xian, M. Zhiqin, X. Xiang, L. Chunyao, R. Zhangquan, a Z. Yuan, "Research of Influence of Garton Effect on the Measurement Values of Dielectric Loss of Capacitor Voltage Transformer", v *2018 International Conference on Power System Technology (POWERCON)*, Guangzhou, nov. 2018, s. 4826–4829. doi: [10.1109/POWERCON.2018.8601602](https://doi.org/10.1109/POWERCON.2018.8601602).
- [35] X. Cheng, G. Ye, H. Sun, T. Li, a C. Sun, "Analysis of low-frequency dielectric loss of XLPE cable insulation based on extended Debye model", *AIP Advances*, roč. 11, č. 8, s. 085103, aug. 2021, doi: [10.1063/5.0060939](https://doi.org/10.1063/5.0060939).

- [36] S. Qin *et al.*, “Study on temperature rise characteristics of 110 kV XLPE cable under different service years considering dielectric loss”, *Energy Reports*, roč. 8, s. 493–501, nov. 2022, doi: [10.1016/j.egy.2022.09.153](https://doi.org/10.1016/j.egy.2022.09.153).
- [37] S. Morsalin a B. T. Phung, “Modeling of dielectric dissipation factor measurement for XLPE cable based on Davidson-Cole model”, *IEEE Trans. Dielect. Electr. Insul.*, roč. 26, č. 3, s. 1018–1026, jún. 2019, doi: [10.1109/TDEI.2019.007784](https://doi.org/10.1109/TDEI.2019.007784).
- [38] N. Bashir, H. Ahmad, a M. Shaffuan Suddin, “Ageing studies on transmission line glass insulators using dielectric dissipation factor test”, v *2010 Conference Proceedings IPEC*, Singapore, Singapore, okt. 2010, s. 1062–1066. doi: [10.1109/IPECON.2010.5696973](https://doi.org/10.1109/IPECON.2010.5696973).
- [39] F. Gerdinand, M. Budde, a M. Kurrat, “Electrical and mechanical strength of mineral filled epoxy insulators in correlation to power loss factor”, v *Proceedings of the 2004 IEEE International Conference on Solid Dielectrics, 2004. ICSD 2004.*, Toulouse, France, 2004, s. 320–323. doi: [10.1109/ICSD.2004.1350355](https://doi.org/10.1109/ICSD.2004.1350355).
- [40] M. Ehsani, H. Borsi, E. Gockenbach, J. Morshedian, G. R. Bakhshandeh, a A. A. Shayegani, “Effect of aging on dielectric behavior of outdoor polymeric insulators”, v *Proceedings of the 2004 IEEE International Conference on Solid Dielectrics, 2004. ICSD 2004.*, Toulouse, France, 2004, s. 312–315. doi: [10.1109/ICSD.2004.1350353](https://doi.org/10.1109/ICSD.2004.1350353).
- [41] K. Chrzan, Z. Pohl, a T. Kowalak, “Hygroscopic properties of pollutants on HV insulators”, *IEEE Trans. Electr. Insul.*, roč. 24, č. 1, s. 107–112, feb. 1989, doi: [10.1109/14.19874](https://doi.org/10.1109/14.19874).
- [42] K. Chrzan, “Conductivity of Aqueous Dust Solutions”, *IEEE Trans. Electr. Insul.*, roč. EI-22, č. 3, s. 241–244, jún. 1987, doi: [10.1109/TEI.1987.298984](https://doi.org/10.1109/TEI.1987.298984).

Výskum alternatívnych kvapalných dielektrík pre vysokonapäťové aplikácie

Miloš Šárpataky

Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach
Košice, Slovensko
milos.sarpataky@tuke.sk

Juraj Kurimský

Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach
Košice, Slovensko
juraj.kurimsky@tuke.sk

Abstract— Tento článok sa zaoberá výskumom alternatívnych kvapalných dielektrík pre vysokonapäťové aplikácie, konkrétne biodegradovateľnými nanokvapalinami. Cieľom tohto článku je priblížiť výskum dielektrických vlastností nanokvapalín na báze syntetických a prírodných esterov s nanočasticami fullerénu a magnetitu. Dielektrické vlastnosti boli skúmané pomocou merania striedavého preskokového napätia, čiastkových výbojov a veličín ako sú stratový činiteľ, objemová rezistivita a relatívna permitivita pri teplotách medzi 25°C a 140°C. Všetky merania boli vyhodnotené aj pomocou bodového systému, aby bolo možné určiť kvalitu jednotlivých vzoriek komplexne, s uvažovaním všetkých meraní. Výsledky ukázali, že vzorka 3_PE (prírodný ester s fullerénovými nanočasticami s koncentráciou 0,03 %w/V) dosiahla celkovo najlepšie výsledky v porovnaní s ostatnými meranými vzorkami.

Keywords— nanokvapaliny; dielektrické vlastnosti; preskokové napätie; čiastkové výboje

I. ÚVOD

V posledných desaťročiach je možné pozorovať nárast množstva aktivít a projektov zameraných na inováciu súčasných technológií a materiálov. Je to vyvolané požiadavkou na trvalo udržateľný rozvoj priemyslu, čo je zjavné aj v elektroenergetike. Táto dôležitá časť priemyslu čelí veľkému tlaku na zdokonaľovanie technológií, aby bola zabezpečená dodávka elektrickej energie s ohľadom na všetky prísne predpisy týkajúce sa ochrany životného prostredia, kvality a spoľahlivosti dodávky elektrickej energie. Veľká pozornosť sa venuje výrobe a spotrebe elektrickej energie, ale mnohé ďalšie časti tohto odvetvia, ako je prenos a distribúcia elektrickej energie, sú kľúčové pre splnenie cieľov a potrieb týkajúcich sa trvalo udržateľného rozvoja [1].

Vysokonapäťové zariadenia s kvapalnou izoláciou, najmä transformátory, káble, kondenzátory, ističe, priechodky atď., sú v elektrizačnej sústave často využívané. V týchto zariadeniach sa väčšinou používajú rovnaké druhy izolačných olejov. Tento článok bude zameraný na kvapalné dielektriká pre výkonové transformátory, ktoré sa ale používajú aj v iných spomínaných zariadeniach. Výkonové transformátory sa využívajú na transformáciu napätia z jednej úrovne na druhú, alebo na galvanické oddelenie rôznych elektrických systémov na rovnakej úrovni napätia. Kvapalná izolácia tohto zariadenia zabezpečuje prenos tepla a elektrické oddelenie jeho živých častí od konštrukcie. Väčšinou sa vo výkonových

transformátoroch využívajú oleje v kombinácii s pevnou izoláciou. Medzi dôležité vlastnosti takéhoto izolačného systému patrí preskokové napätie, stratový činiteľ, merný odpor kvapaliny, viskozita, tepelná vodivosť, bod vzplanutia, bod vznietenia a biodegradovateľnosť [2]. Často používaným izolačným olejom vo svete je minerálny olej. V súčasnosti sa však minerálne oleje považujú za neuspokojivé, pretože majú nízku biodegradovateľnosť, čo znamená, že sa v prírode vedia rozložiť len sčasti. V globálnom meradle je možné pozorovať snahu nahradiť minerálne oleje ekologickejšou alternatívou, napr. prírodnými (rastlinné oleje), alebo syntetickými esterami, poprípade biodegradovateľnými olejmi, ktoré sú vyrábané pomocou metódy skvapalňovania plynov [3]. Prírodné alebo syntetické estery, ako oleje šetrné k životnému prostrediu, majú biodegradovateľnosť viac ako 80 %, zatiaľ čo minerálne oleje dosahujú hodnoty menej ako 30 % [4]. Dielektrické vlastnosti minerálnych olejov (MO) sú porovnateľné alebo lepšie ako v esteroch, čo je jedným z hlavných dôvodov, prečo sa minerálne oleje doteraz používajú najčastejšie. V [5] je predpoklad, že sa dopyt v energetike zvýši do roku 2040 o 30 % a taktiež sa skončí životnosť veľkému množstvu výkonových transformátorov, ktoré budú musieť byť nevyhnutne nahradené. Predpoveď udáva investíciu vo výške 30 miliárd dolárov. Aktuálne je dopyt po izolačných kvapalinách do transformátorov na úrovni približne 1,5 miliónov ton a z toho sú 95 % minerálne oleje. Vzhľadom na bezpečnostné problémy spojené s minerálnymi olejmi, hlavne ich horľavosťou, ktorá súvisí s nebezpečnými haváriami a stratou 339 miliónov dolárov v rokoch 2008-2013 (prieskum FM Global), je aj výskum izolačných kvapalín zameraný čoraz viac na estery, ktoré sú ekologickejšou a bezpečnejšou alternatívou. Aktuálny výskum taktiež poukazuje na mnohé možnosti zlepšenia dielektrických a tepelných vlastností úpravou vyššie spomenutých kvapalín pridávaním nanočastíc rôznych tuhých materiálov. Táto zmes oleja, ako základnej kvapaliny a nanočastíc, tvorí koloidný roztok a vo všeobecnosti sa nazýva nanokvapalina. Špecificky pripravené nanokvapaliny sa považujú za potenciálnu náhradu súčasnej generácie izolačných kvapalín používaných vo vysokonapäťových zariadeniach [6]. Zmes nanočastíc a biodegradovateľných kvapalín sa javí ako jedno z možných riešení zlepšenia úžitkových vlastností izolačných olejov s ohľadom na trvalo udržateľný rozvoj. Zlepšenie odvodu tepla vo výkonovom transformátore pridaním nanočastíc sa preukázalo v mnohých publikáciách. Ak sa zvýši množstvo vyprodukovaného tepla alebo sa zníži prenos tepla, môže to spôsobiť zvýšenie teploty

vinutia a skrátit' životnosť izolácie. Distribučné transformátory majú zvyčajne plánovanú životnosť 30 rokov. Platí bežné pravidlo, že každé zvýšenie teploty o 10 °C znižuje životnosť izolačného materiálu na polovicu. Použitie nanočastíc v izolačných olejoch môže predĺžiť životnosť distribučného transformátora, zvýšiť prenášaný výkon a umožniť zmenšenie konštrukcie stroja vďaka zlepšenej schopnosti prenosu tepla chladiaceho média a vďaka lepším dielektrickým parametrom. Tento článok skúma dielektrické vlastnosti nanokvapalín, pomocou ktorých sa dá určiť, či sú testované vzorky vhodné pre vysokonapäťové aplikácie.

II. NANOKVAPALINY

Nanokvapaliny sú koloidné roztoky vytvorené pridaním nanočastíc do základnej kvapaliny. Počas histórie výskumu nanokvapalín bolo testovaných mnoho rôznych roztokov s rôznymi izolačnými kvapalinami, alebo s rôznymi nanočasticami. Spôsob prípravy nanokvapalín sa všeobecne rozdeľuje na jednodušný a dvojitý proces. Podľa [6] je pôvod slova „nanofluid“ spojený s prácou Choia a kol. [7] z roku 1995, kde bol prezentovaný výskum zvýšenia tepelnej vodivosti kvapalín zmiešaných s nanočasticami a diskusiou o ich potenciálnom prínose v aplikáčnej sfére [7]. Od tejto prelomovej publikácie sa počet vedeckých prác, ktoré sa zaoberajú nanokvapalinami, exponenciálne zvyšuje. Počet publikácií v tejto problematike nepresiahol 200 ročne až do roku 2008. Od roku 2008 dochádza k výraznejšiemu nárastu vedeckých prác a z roka na rok sa ich počet zvyšuje [8]. Všeobecne sa príprava nanokvapalín delí na jednodušnú a dvojitú metódu. Vzhľadom na náročnosť prípravy nanokvapalín jednodušnou metódou sa vo väčšine prípadov využíva dvojitá príprava (Fig. 1)

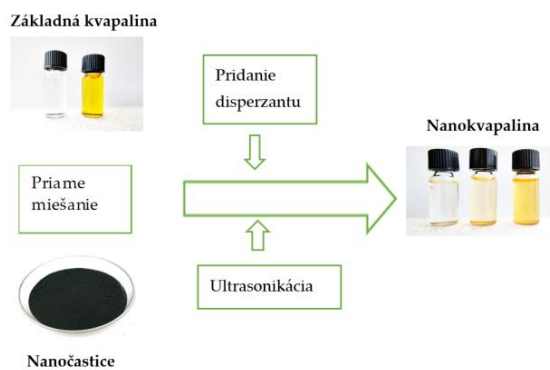


Fig. 1. Jednoduchá ilustrácia procesu výroby nanokvapaliny pomocou dvojitú metódu

III. SÚČASNÝ STAV VO VÝSKUME DIELEKTRICKÝCH NANOKVAPALÍN

Tento výskum bol zameraný na nanokvapaliny s nanočasticami fullerénu (C_{60}) a magnetitu (Fe_3O_4). Táto kapitola popisuje podobné výskumy ako budú prezentované v tomto článku, aby bolo možné porovnanie výsledkov so zahraničnými publikáciami.

Khaled a kol. [9] porovnávali výsledky striedavého preskokového napätia (AC-PN) syntetického estera (SE) MIDEL 7131 s nanočasticami Fe_3O_4 . Nanočastice Fe_3O_4 s priemerom 50 nm zvýšili AC-PN až o 47,78 % pri koncentrácii 0,4 g/l, čo bolo najvyššie zlepšenie spomedzi všetkých skúmaných vzoriek. Zvýšenie AC-PN 17,83 % bolo zaznamenané pri koncentrácii 0,3 g/l, a vzorky s nižšou koncentraciou (0,05 g/l a 0,2 g/l) vykazovali zníženie AC-PN o 6,06 % a 0,05 %.

Mendez a kol. v [10] porovnávali nanokvapaliny s nanočasticami TiO_2 a nanočasticami Fe_2O_3 s veľkosťou priemeru od 10 nm do 20 nm. Ako izolačný olej bol zvolený prírodný ester (PE) a koncentrácie nanočastíc sa pohybovali v rozmedzí 0,1 kg/m³ a 0,5 kg/m³. Výsledky tepelnej vodivosti boli takmer rovnaké pre obe nanočastice a výsledky v porovnaní s PE nevykazovali významné zmeny. Najväčšie zvýšenie AC-PN spomedzi všetkých vzoriek bolo 15,1 % pri optimálnej koncentrácii 0,2 kg/m³. Koncentrácie 0,1 kg/m³ a 0,3 kg/m³ zvýšili hodnotu AC-PN a vyššie koncentrácie 0,4 kg/m³ a 0,5 kg/m³ hodnotu AC-PN znížili. V porovnaní s nanočasticami TiO_2 so zlepšením o 33,2 % je zmes nanočastíc PE a Fe_3O_4 podľa týchto výsledkov menej vhodná na výrobu nanokvapalín.

Hussain a kol. [11] pripravili nanokvapaliny vyrobené z PE a SE s nanočasticami Fe_3O_4 s veľkosťou od 50 nm do 100 nm. AC-PN bolo merané pomocou dvoch elektródových systémov (sférický a semiférický). Rozdiel medzi výsledkami týchto systémov bol približne 30 kV. Semiférický elektródový systém s vyššími hodnotami AC-PN vykazoval rozdielne výsledky pre nanokvapaliny na báze PE a SE. Najvyššie zlepšenie 20,7 % bolo namerané pri nanokvapaline na báze SE s koncentraciou 0,0022 w/w. Nanokvapalina na báze PE vykazovala najvyššie zlepšenie hodnoty AC-PN 12,1 % pri koncentrácii 0,004 w/w. Sférický elektródový systém potvrdil optimálne koncentrácie nanokvapalín a najlepšie výsledky vykazovali vzorky nanokvapalín SE a PE 30,7 % (0,0022 w/w) a 33,4 % (0,004 w/w). Výsledky potvrdzujú, že pri každej kombinácii základnej kvapaliny a nanočastíc je možné nájsť optimálnu koncentráciu, ktorá bola v tomto experimente potvrdená pri rôznych druhoch elektródových systémov.

Olmo a kol. [12] použili nanočastice Fe_2O_3 s priemerom medzi 10 nm a 20 nm zmiešané s PE. Rôzne fyzikálne a dielektrické vlastnosti, ako je viskozita, tepelná vodivosť, alebo dielektrická pevnosť boli skúmané na vzorkách so šiestimi rôznymi koncentraciami od 0,1 kg/m³ do 0,5 kg/m³. Nanočastice nemali takmer žiaden vplyv na viskozitu a tepelnú vodivosť nanokvapaliny, čo bolo pravdepodobne spôsobené nízkou koncentraciou nanočastíc. Optimálna koncentrácia nanočastíc Fe_2O_3 bola 0,2 kg/m³ pri ktorej zlepšenie AC-PN dosahovalo hodnotu 16 %. Nanokvapalina s koncentraciou nanočastíc 0,1 kg/m³ a 0,3 kg/m³ taktiež vykazovala zlepšenie hodnôt, avšak zvyšné koncentrácie 0,4 kg/m³ a 0,5 kg/m³ hodnotu AC-PN znížili.

Beroual a kol. [13] skúmali impulzné prekovkové napätie (LI-PN) so zápornou polaritou pre nanokvapaliny na báze PE (MIDEL 1204) a SE (MIDEL 7131) s rôznymi nanočasticami. Pri nanokvapalinách na báze SE zvýšenie hodnôt klesalo s koncentraciou z 25,57 % (0,05 g/l) na 6,37 % (0,4 g/l). PE

ako čistý olej, mal vyššiu hodnotu LI-PN ako SE a rozdiel bol 2,6 %. Po pridaní nanočastíc PE nedosahoval také vysoké hodnoty ako SE. Najvyššie zlepšenie pri PE s nanočasticami Fe_3O_4 bolo 7,51 % pri koncentrácii nanočastíc 0,2 g/l a zvyšné koncentrácie 0,05 g/l, 0,3 g/l a 0,4 g/l vykazovali zníženie hodnoty o 2,62 %, 4,45 % a 8,29 %. Tento experiment ukázal, že SE MIDEL 7131 je vhodnejší ako PE MIDEL 1204 ako izolačný olej pre výrobu nanokvapalín s Fe_3O_4 .

Kurimský a kol. [14] vykonali testy zrýchleného tepelného starnutia na nanokvapalinách so superparamagnetickými nanočasticami oxidu železa. Bolo zistené, že transformátorový olej je odolnejší voči tepelnému starnutiu ako nanokvapaliny. Tepelne ovplyvnený rozklad nanokvapalín zapríčinil zhoršené preskokové napätie so zvyšujúcou sa koncentráciou nanočastíc. Taktiež Rajnák a kol. [15] dokázali, že prítomnosť magnetitových (Fe_3O_4) nanočastíc v transformátorovom oleji nemusí nutne viesť k zvýšeniu AC-PN. V spomínanom zdroji uviedli hypotézu, že modifikácia povrchu nanočastíc môže hrať rozhodujúcu úlohu v mechanizme rozvoja preskoku v nanokvapaline. Ďalšia štúdia taktiež naznačila, že aj malé množstvo nanočastíc magnetitu (0,001 hm.%) môže výrazne zvýšiť aktivitu čiastkových výbojov v naftenicom oleji [16]. Okrem toho boli v prototypoch transformátorov aplikované nanokvapaliny na báze transformátorového oleja obsahujúce magnetické nanočastice Fe_3O_4 . Z experimentov v [17] a [18] vyplýva, že prototypové transformátory naplnené magnetickými kvapalinami vykazujú sľubné prevádzkové vlastnosti. Zvýšenie účinnosti chladenia bolo čiastočne dosiahnuté zvýšenou termomagnetickou konvekciou. Aplikácia vonkajšieho magnetického poľa na magnetické nanokvapaliny môže spôsobovať ich dielektrickú odozvu, čo môže zvýšiť viskozitu a tepelnú vodivosť, ale to závisí od konkrétnej konfigurácie magnetického poľa a meracieho senzora [19].

Szcześniak a kol. [20] skúmali nanočastice C_{60} s priemerom približne 0,7 nm zmiešané s PE FR3 pred a po 164 hodinách zrýchleného starnutia pri teplote 150°C . Čo sa týka termofyzikálnych vlastností, viskozita nanokvapaliny sa takmer nemenila počas celej doby starnutia. Hodnoty AC-PN skúmaných nanokvapalín pred tepelným starnutím boli nižšie o približne 10 % a 5 % pri koncentráciách 500 mg/l a 250 mg/l. Po tepelnom starnutí došlo ešte k výraznejšiemu poklesu približne o 23 % (rozdiel medzi izolačným olejom a nanokvapalinou s koncentráciou 500 mg/l po tepelnom starnutí). Stratový činiteľ sa zvyšoval s koncentráciou nanočastíc v kvapaline pred, aj po zrýchlenom starnutí.

Huang a kol. [21] zmiešali PE získaný z repkového oleja a taktiež MO s nanočasticami C_{60} s priemerom od 4 nm do 6 nm. Stratový činiteľ sa znížil o 20,1 % pre PE s nanočasticami C_{60} s koncentráciou 100 mg/l a pokles v MO bol okolo 50 % pri koncentrácii 50 mg/l. Hodnoty AC-PN v PE sa zvýšili pri koncentráciách 50-150 mg/l, no pri vyšších koncentráciách došlo k poklesu hodnôt. Zlepšenie sa pohybovalo od približne 2 % do približne 8 %. MO zmiešaný s nanočasticami C_{60} vykazoval zníženie hodnoty len pri koncentrácii 50 mg/l. Optimálna koncentrácia fullerénu pre MO je 200 mg/l, keď zlepšenie dosiahlo približne 21 %.

Beroual a kol. [22] skúmali AC-PN a LI-PN kladnej polarite na vzorkách vyrobených z PE MIDEL eN 1204 a nanočastíc fullerénu o priemernej veľkosti 21 nm. Skúmaných bolo 5 koncentrácií a to 0,05 g/L, 0,1 g/L, 0,2 g/L, 0,3 g/L a 0,4 g/L. Výsledky AC-PN ukázali, že koncentrácie od 0,05 g/L do 0,2 g/L vykazovali zníženie hodnoty preskokového napätia od 5 % do 14 %. Zlepšenie nastalo pri 2 najvyšších koncentráciách a to o 5 % pri koncentrácii 0,03 g/L a o 7,8 % pri najvyššej koncentrácii 0,04 g/L. Rozdielny charakter vplyvu nanočastíc bol nameraný pri impulznom napätí, kde zvýšenie hodnoty bolo pozorované pri koncentráciách od 0,05 g/L do 0,3 g/L. Hodnoty zlepšenia sa pohybovali od 3 % pri koncentrácii 0,3 g/L do 8 % pri koncentrácii 0,1 g/L. Hodnota LI-PN vzorky s koncentráciou nanočastíc 0,4 g/L sa znížila o 12,5 %. Vzorka s koncentráciou fullerénu na úrovni 0,3 g/L, ako jediná dosiahla vyššie hodnoty ako čistý PE pri oboch experimentoch.

Khelifa a kol. [23] preverovali na vzorkách SE MIDEL 7131 s nanočasticami fullerénu o veľkosti 4-8 nm zmenu AC-PN a parametrov čiastkových výbojov. Koncentrácie nanočastíc sa pohybovali od 0,1 g/L do 0,5 g/L s krokom 0,1 g/L. Pri všetkých koncentráciách sa zvýšila hodnota PN od 5,49 % (0,1 g/L) do 12,67 % (0,4 g/L). Meranie čiastkových výbojov vykonali na vzorke s najlepším výsledkom z merania PN, čiže na vzorke s koncentráciou 0,4 g/L. Došlo k zlepšeniu hodnoty počiatočného napätia o 5,1 %, ale zhašacie napätie čiastkových výbojov sa znížilo o 14,98 %. Hodnoty veličín priemerný náboj, maximálny náboj a početnosť výbojov boli zlepšené o 18,18 %, 34,63 % a 42 %.

IV. EXPERIMENTÁLNA ANALÝZA ZMIEN DIELEKTRICKÝCH VELIČÍN V NANOKVAPALINÁCH

A. Materiály a príprava nanokvapalín

Základné kvapaliny použité na prípravu nanokvapalín v tejto experimentálnej štúdiu boli prírodný izolačný olej na báze esteru PE MIDEL eN 1204 (vyrobený z repkového oleja) a syntetický izolačný olej na báze esteru SE MIDEL 7131. Fyzikálno-chemické vlastnosti týchto kvapalín sú uvedené v TABLE I.

TABLE I. FYZIKÁLNO-CHEMICKÉ VLASTNOSTI ZÁKLADNÝCH KVAPALÍN [24]

Vlastnosti	MIDEL 7131	MIDEL eN 1204
Hustota pri 20°C (g/cm^3)	0,97	0,92
Viskozita pri 40°C (mm^2/s)	29	8
Bod tuhnutia ($^\circ\text{C}$)	-56	-31
Bod vzplanutia ($^\circ\text{C}$)	260	>315
Bod vznietenia ($^\circ\text{C}$)	316	>350
Acidita (mg KOH/g)	<0,03 mg	0,04

Príprava nanokvapalín prebiehala v spolupráci s ústavom experimentálnej fyziky SAV v Košiciach, kde boli do vzoriek esterov primiešané nanočastice fullerénu C_{60} a magnetitu Fe_3O_4 . Pre experimentálnu fázu výskumu bolo vytvorených 10 vzoriek a to 3 vzorky pre kombináciu PE a nanočastic fullerénu, rovnako 3 vzorky pre SE s nanočasticami fullerénu a 2 vzorky pre obe estery s nanočasticami Fe_3O_4 . Koncentrácie nanočastic a označenie vzoriek je v TABLE II.

Nanočastice oxidu železa (magnetit) boli syntetizované chemickou metódou koprecipitácie, po ktorej nasledovala stérická stabilizácia nanočastic jednou vrstvou kyseliny olejovej adsorbovanej na povrchoch častíc (surfaktant). Stabilizované nanočastice boli dispergované v nosných olejoch mechanickým miešaním a ultrazvukom pri teplote $60^\circ C$. Na dvoch základných kvapalinách boli pripravené dve vzorky magnetických nanokvapalín s koncentráciou nanočastic 0,05 %w/V a 0,1 %w/V. Vyššie koncentrácie nanočastic v oboch esteroch vykazovali koloidnú nestabilitu vedúcu k sedimentácii nanočastic.

Na získanie nanokvapalín s nanočasticami fullerénu C_{60} bol použitý práškový fullerén s čistotou 99,5 % (Merck) tak, ako bol zakúpený bez akejkoľvek ďalšej úpravy. Nanočastice boli homogénne dispergované v základných olejoch pomocou ultrazvuku pri teplote $60^\circ C$ počas 4 hodín. Pre každý olej boli pripravené tri vzorky s koncentráciou od 0,01 %w/V do 0,03 %w/V.

TABLE II. SKÚMANÉ VZORKY. KONCENTRÁCIA NANOČASTÍC C_{60} A Fe_3O_4 V NANOKVAPALINÁCH JE VYJADRENÁ AKO PEVNÝ OBJEMOVÝ PODIEL V % HMOTNOSTI/OBJEM

Základná kvapalina	Nanočastice	Koncentrácia	Označenie vzorky
PE	-	-	PE
PE	C_{60}	0,01 %w/V	1_PE
PE	C_{60}	0,02 %w/V	2_PE
PE	C_{60}	0,03 %w/V	3_PE
PE	Fe_3O_4	0,05 %w/V	1_PE_FE
PE	Fe_3O_4	0,1 %w/V	2_PE_FE
SE	-	-	SE
SE	C_{60}	0,01 %w/V	1_SE
SE	C_{60}	0,02 %w/V	2_SE
SE	C_{60}	0,03 %w/V	3_SE
SE	Fe_3O_4	0,05 %w/V	1_SE_FE
SE	Fe_3O_4	0,1 %w/V	2_SE_FE

B. Metodika merania striedavého preskakového napätia

Meranie AC-PN bolo vykonávané podľa normy IEC 60156 pomocou zariadenia Haefely Hipotronics používaného na testy prieraznej pevnosti materiálov. Schéma zapojenia meracieho pracoviska je uvedená na Fig. 2. Na elimináciu degradácie vzoriek v dôsledku vysokých prúdov počas preskoku, bol pripojený 5 M Ω rezistor, ktorý plnil funkciu obmedzovača prúdu. Použité boli elektródy hubovitého tvaru s priemerom 12,5 mm, ktoré boli vodorovne umiestnené v skúšobnej nádobe

o objeme 350 ml. Elektródy boli od seba vzdialené $2,50 \pm 0,05$ mm. Rýchlosť nárastu napätia bola 2 kV/s, pokiaľ nedošlo k preskoku. Po výmene vzorky a príprave meracej nádoby podľa štandardného protokolu (čistenie, sušenie a naliatie kvapaliny po stene nádoby) nasledovala 10 minútová pauza, aby sa eliminovali akékoľvek bubliny, ktoré by mohli ovplyvniť výsledky merania. Čas medzi dvomi po sebe idúcimi meraniami bol 2 minúty, počas ktorých bola kvapalina miešaná pomocou magnetického miešadla (pri vzorkách s nanočasticami magnetitu bolo miešanie mechanické) a meranie bolo opakované 30 krát.

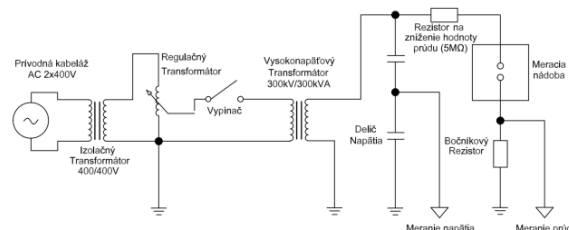


Fig. 2. Jednoduchá ilustrácia procesu výroby nanokvapaliny pomocou dvojstupňovej metódy

Merania boli realizované v laboratóriu CVVOZE PowerLab na technickej univerzite v Brne. Vzhľadom na degradáciu vzoriek počas skúšobných meraní bol prúd obmedzený pomocou rezistora na hodnoty pohybujúce sa okolo 10 mA. Odpojenie obvodu od napájania bolo následne vykonávané pomocou zariadeného prístroja na detekciu nízkych prúdov, ktoré automaticky odpojilo meranú vzorku od napájania a výška napätia sa uložila do systému Haefely Hipotronics.

C. Metodika merania teplotnej závislosti stratového činiteľa, relatívnej permitivity a objemovej rezistivity

Pre zjednodušenie bude ďalej v texte súbor meraní stratového činiteľa, relatívnej permitivity a objemovej rezistivity označovaný aj ako dielektrická výkonnosť. Meranie dielektrickej výkonnosti bolo vykonané podľa normy IEC 60247 2004. Testovacia nádoba HAEFELY TEST AG 2903 H s objemom 40 ml bola napojená na zariadenie Tettex 2830 Precision Oil and Solid Dielectric Analyzer, z ktorého boli odčítavané hodnoty. Zostava meracích prístrojov je znázornená na Fig. 3.



Fig. 3. Pracovisko na meranie teplotnej závislosti stratového činiteľa, relatívnej permitivity a objemovej rezistivity

Na určenie teplotnej závislosti dielektrickej výkonnosti boli vzorky testované pri teplotách medzi 25°C a 140°C s krokom 10°C. Pred každým meraním (po výmene jednotlivých vzoriek) bolo testovacie prostredie v nádobe upravené pomocou vákuovej pumpy. Merania boli realizované v laboratóriu CVVOZE PowerLab na technickej univerzite v Brne.

D. Metodika merania čiastkových výbojov

Meranie čiastkových výbojov bolo realizované podľa normy IEC 60270 a špecializovanej normy na meranie čiastkových výbojov v izolračných olejoch IEC 61294 [25]. Je nutné podotknúť, že podľa normy IEC 61294 by sa do štatistiky mali zapisovať výboje vyššie ako 100 pC, ale v prípade tohto merania, sa vzhľadom na skúsenosti, nastavenú citlivosť a pri najodolnejších vzorkách nízku početnosť výbojov, brali do úvahy výboje so zdanlivým nábojom nad 3 pC. Rušenie, ktoré by mohlo ovplyvniť výsledky merania sa pohybovalo okolo hranice 2 pC. Priebeh aplikovaného napätia, s maximom na úrovni 20 kV (frekvencia 50 Hz) a rýchlosťou nárastu 1 kV/s. Táto napät'ová rampa bola nastavená pomocou lineárneho napájacieho zdroja GWINSTEK APS-7050E. Čas aplikácie napätia na elektródový systém bol 5 minút (amplitúda napätia) a 40 sekúnd (stúpanie a klesanie napätia na požadovanú hodnotu). Pri vzorkách, ktoré mali nízku početnosť výbojov sa vykonalo aj meranie s amplitúdou napätia 24 kV, aby sa presnejšie určil rozdiel medzi kvalitou týchto vzoriek.

Koróna a strímre, ktoré generujú náboj čiastkových výbojov, boli indukované napätím aplikovaným na konfiguráciu ihlová elektróda a uzemnená doska (Fig. 4), pričom ihlová elektróda bola ponorená do testovanej kvapaliny v nádobe s objemom 30 ml. Vzdialenosť medzi ihlovou elektródou a doskou bola $20 \pm 0,05$ mm. Polomer zakrivenia ihlovej elektródy bol medzi 20-30 μ m. Čiastkové výboje boli merané pomocou MPD600 PD meracieho systému a softvéru mtronix v1.0 od firmy Omicron Electronics. Kalibrácia veľkosti indukovaného náboja bola vykonaná prístrojom Tettex

PD typ 9216. Za napät'ový transformátor bol sériovo pridaný odpor na zníženie veľkosti prúdov, ktoré by mohli spôsobiť degradáciu meranej vzorky, alebo znehodnotenie ihly zväčšením polomeru zakrivenia. Vplyv polomeru zakrivenia na parametre čiastkových výbojov je popísaný v [25].



Fig. 4. Elektródový systém na meranie čiastkových výbojov

Nanokvapaliny s nanočasticami magnetitu boli merané aj pri externom statickom magnetickom poli, ktoré bolo vytvorené dvoma permanentnými magnetmi pripevnenými k nádobe. Hodnota magnetickej indukcie v mieste hrotovej elektródy bola 60 mT.

E. Metodika spracovania výsledkov meraní

Základné vyhodnotenie dát pre jednotlivé merania bolo vykonané výpočtom percentuálnej zmeny priemeru meranej veličiny skúmanej vzorky, vo vzťahu k priemeru skúmanej veličiny čistej základnej kvapaliny (v texte označované ako *Zmena*) a štandardnej odchýlky.

Napríklad pre meranie AC-PN bola percentuálna zmena vypočítaná ako:

$$Zmena = \frac{\bar{U}_{NF} - \bar{U}_{PF}}{\bar{U}_{PF}} \cdot 100\% \quad (1)$$

Kde \bar{U}_{NF} a \bar{U}_{PF} značia priemerné hodnoty striedavého preskokového napätia vzorky nanokvapaliny a čistej základnej kvapaliny. Kladná percentuálna zmena značí zlepšenie (zvýšenie) AC-PN (alebo inej veličiny), zatiaľ čo záporná percentuálna zmena znamená jeho zhoršenie (zníženie). Priemerná hodnota \bar{U} a štandardná odchýlka σ sú určené všeobecnými vzorcami:

$$\bar{U} = \frac{1}{n} \sum_{i=1}^n U_i \quad (2)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (U_i - \bar{U})^2} \quad (3)$$

Kde n je celkový počet meraní, i je index konkrétneho merania a U_i je hodnota konkrétneho meranie preskokového napätia.

F. Experimentálne výsledky a štatistická analýza meraní striedavého preskokového napätia

Priemerné hodnoty preskokového napätia, Zmena a štandardná odchýlka, ktoré boli namerané a vyhodnotené na všetkých skúmaných vzorkách základných kvapalín SE a PE, sú zapísané v TABLE III.

TABLE III. PRIEMERNÁ HODNOTA AC-PN, ŠTANDARDNÁ ODCHÝLKA A ZMENA MERANÝCH VZORIEK [26]

Vzorka	Priemerná hodnota \bar{U} (kV)	Štandardná odchýlka σ (kV)	Zmena (%)
SE	57,31	2,728	-
PE	55,64	3,311	-
1_SE	53,22	5,270	-7,69
2_SE	55,06	6,125	-4,09
3_SE	56,71	4,804	-1,06
1_PE	34,61	6,232	-60,76
2_PE	56,36	4,763	1,29
3_PE	58,57	5,025	5,27
1_SE_FE	63,79	4,899	11,31
2_SE_FE	65,06	4,508	13,52
1_PE_FE	63,81	4,714	14,68
2_PE_FE	57,18	4,030	2,77

Na Fig. 5 a Fig. 6 je možné vidieť grafy, ktoré obsahujú krabicové grafy definované prvým a tretím kvartilom. V krabicových grafoch je priemerná hodnota znázornená ako štvorec a medián znázornený ako hrubá čiara. Cez rámček prechádza zvislá čiara, ktorá predstavuje štandardnú odchýlku σ . Vpravo sú namerané hodnoty rozdelenia zobrazené vo farebnej škále, ktorá zobrazuje poradie jednotlivých meraní. Hodnota preskokového napätia v červenej farbe patrí prvým meraniam AC-PN a modrá naopak zobrazuje posledné merania vzoriek. Z rámcových grafov je zrejmé, že neexistuje jasný vzťah medzi poradím jednotlivých meraní preskoku a nameranou hodnotou preskokového napätia. To znamená, že vzorky sa počas meraní nemenili a predradený odpor zamedzil degradácii vzoriek. Vlnitá krivka predstavuje hustotu výskytu hodnôt, pri jednotlivých úrovniach napätia.

Vo všeobecnosti je možné konštatovať, že nanokvapaliny s nanočasticami magnetitu okrem vzorky 2_PE_FE vykazujú najvyššie hodnoty AC-PN. Fulerénové nanočastice v kvapalinách znížili hodnotu preskokového napätia pri všetkých vzorkách so základnou kvapalinou SE a pri najnižšej koncentrácii v PE.

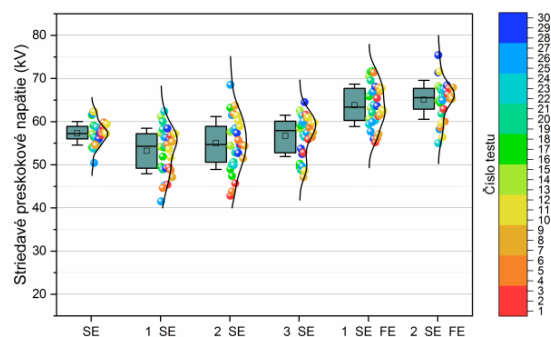


Fig. 5. Grafické zhodnotenie striedavého preskokového napätia vzorky SE a nanokvapalín z tejto základnej kvapaliny [26]

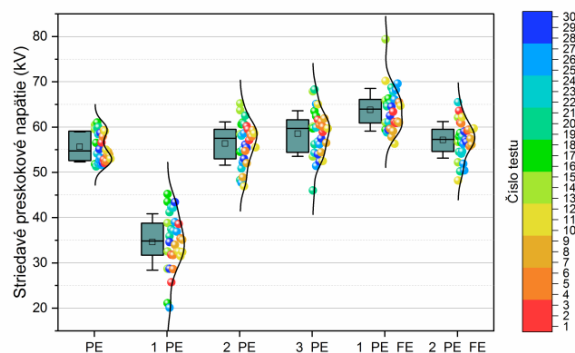


Fig. 6. Grafické zhodnotenie striedavého preskokového napätia vzorky PE a nanokvapalín z tejto základnej kvapaliny [26]

Z TABLE III si možno všimnúť, že fullerénové nanočastice v SE pri všetkých skúmaných vzorkách (1_SE, 2_SE a 3_SE) spôsobili zhoršenie hodnoty AC_PN (zodpovedajúce hodnoty zmeny sú 7,69 %, 4,09 %, 1,06 %). Naopak nanokvapalina obsahujúca magnetitové nanočastice v syntetickom esteri (vzorka 2_SE_FE) má najvyššiu hodnotu preskokového napätia dosahujúcu 65,06 kV. V porovnaní s AC-PN v čistom syntetickom esteri, táto vzorka zvýšila svoju hodnotu o 13,52 %. Vzorky na báze prírodných esterov obsahujúce nanočastice Fe₃O₄ vykazovali zlepšenie AC-PN v rozsahu od 2,77 % (2_PE_FE) do 14,68 % (1_PE_FE), kde najvyššia priemerná hodnota preskokového napätia bola 63,81 kV. Mierne zlepšenie je možné pozorovať pri vzorkách 2_PE a 3_PE (obe s vyššou koncentráciou fullerénu), ktoré dosiahli zlepšenie 1,29 % a 5,27 %.

Z týchto výsledkov vyplýva, že určenie optimálnej koncentrácie magnetitových nanočastíc závisí od použitej základnej kvapaliny. Syntetický ester vykazoval vyššie hodnoty AC-PN pri vyššej koncentrácii (0,1 % w/V) nanočastíc, zatiaľ čo v prípade magnetickej kvapaliny na báze prírodného esteru je najvyššie preskokové napätie pri koncentrácii 0,05 % w/V (nižšia koncentrácia). Je možné konštatovať, že nanočastice magnetitu sú vhodné na zlepšenie AC-PN kvapalných dielektrík na báze prírodných aj

syntetických esterov. Aplikácia fullerénových nanočastíc v skúmaných olejoch na báze esterov, nepreukázala výrazné zlepšenie. Experimentálne výsledky ukázali, že nanočastice C_{60} pri skúmaných nízkych koncentráciách cca. 0,01 %w/V zhoršili hodnoty preskovového napätia.

Hodnoty smerodajnej odchýlky σ ovplyvňujú rozloženie nameraných údajov. Čím nižšia je štandardná odchýlka, tým vyšší je vrchol početnosti javu okolo priemernej hodnoty, inými slovami, pri vysokej hodnote štandardnej odchýlky sú hodnoty viac rozptýlené, čo spôsobí, že má amplitúda nižšiu hodnotu. Spôhlivosť transformátora môže byť spojená s hodnotou smerodajnej odchýlky, pretože zmena AC-PN pri nízkej kumulatívnej pravdepodobnosti preskoku môže byť spojená s nízkou hodnotou σ [23].

Možná interpretácia zvýšeného AC-PN v dôsledku prítomnosti nanočastíc Fe_3O_4 v kvapaline, môže byť založená na modeli zachytávania elektrónov opísaného v [27]. Podľa tohto modelu, splňajú nanočastice v dielektrických kvapalinách úlohu zachytávačov elektrického náboja. Mechanizmus je účinný, keď je v nanokvapaline „dielektrický kontrast“, čo znamená zvýšený rozdiel hodnôt permitivity a elektrickej vodivosti základnej kvapaliny a dispergovaných nanočastíc. Potom pôsobením vonkajšieho elektrického poľa, vzniká efektívna dielektrická polarizácia a rýchla relaxačná dynamika nanočastíc v základnej kvapaline. Ak je relaxačný čas kratší ako vývoj strímra, potom polarizované nanočastice menia rozloženie elektrického poľa, vedúce k voľným nábojom (iniciačným elektrónom) na povrchu nanočastíc. Následne nabité nanočastice migrujú smerom ku gradientu elektrického poľa (k elektróde) oveľa pomalším spôsobom, ako voľné náboje. To má za následok zmenu podmienok pre iniciovanú ionizáciu spôsobenú kolíziou, ktorá priamo ovplyvňuje vývoj strímra a následne zvyšuje hodnotu preskovového napätia. Keďže hodnoty permitivity a elektrickej vodivosti prírodných a syntetických esterov sú porovnateľné s MO, je možné porovnať spomenuté zákonitosti s dejmi v tejto experimentálnej štúdiu a konštatovať, že zvýšenie AC-PN oboch esterov s nanočasticami Fe_3O_4 , možno vysvetliť modelom zachytávania elektrónov. Skutočnosť, že prírodný ester s nanočasticami magnetitu má vyššiu hodnotu preskovového napätia pri nižšej koncentrácii (1_PE_FE) ako pri vyššej (2_PE_FE) a syntetický ester s Fe_3O_4 nanočasticami vykazuje vyššie AC-PN pri vyššej koncentrácii nanočastíc, môže súvisieť s viskozitou základných kvapalín. Dá sa predpokladať, že nižšia kinematická viskozita prírodného esteru (8 mm²/s) umožňuje rýchlu migráciu a interakciu nanočastíc s elektrickým poľom, čo vedie k tvorbe vodivých dráh. Potom je pri vyššej koncentrácii nanočastíc tvorba vodivých dráh pravdepodobnejšia a mechanizmus zachytávania náboja obmedzený, čo vedie k zníženiu hodnoty AC-PN. Na druhej strane, vyššia viskozita syntetického esteru (29 mm²/s) poskytuje horšie podmienky pre tvorbu vodivých dráh a priaznivé podmienky pre zachytávanie náboja a pomalý vývoj strímra aj pri vyššej koncentrácii.

G. Experimentálne výsledky meraní teplotných závislostí stratového činiteľa, relatívnej permitivity a objemovej rezistivity

Výsledky budú prezentované postupne podľa meranej veličiny. Pri vzorkách PE s nanočasticami magnetitu, nebolo možné namerať hodnoty pri vyšších teplotách, kvôli limitom používaného prístroja, ktorý hlásil chybu „nadprúd“ („overcurrent“).

Dielektrické straty primárne spojené s oxidačnými derivátmi, obsahom vody v kvapaline a prítomnosťou nečistôt reprezentuje veličina nazývaná stratový činiteľ (tg δ). Stratový činiteľ sa používa na určenie kvality a úrovne degradácie izolačných olejov. Grafické znázornenie výsledkov je na Fig. 7.

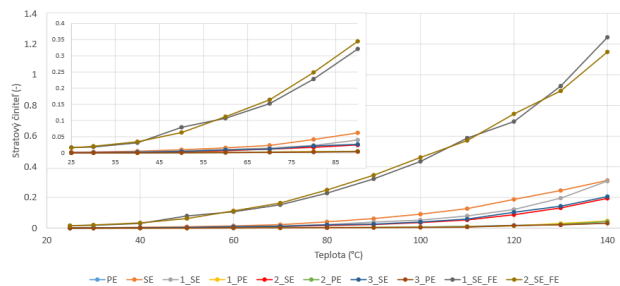


Fig. 7. Teplotná závislosť stratového činiteľa [28]

Stratový činiteľ izolačných olejov vo všeobecnosti stúpa so zvyšujúcou sa teplotou. S vyššou teplotou rozdiely medzi jednotlivými vzorkami rastú a zvyrazňujú medzi nimi úroveň kvality. Čistý PE má v porovnaní s čistým SE pri všetkých teplotách približne 10-krát nižšie hodnoty tg δ . Pri 90°C mal SE približne 12-krát vyššiu hodnotu ako PE.

Vzorky s magnetickými nanočasticami z hľadiska stratového činiteľa vykazovali slabé dielektrické vlastnosti. Tg δ magnetických kvapalín vyrobených zo SE sa pohyboval od 0,01583 pri 25°C do 1,244 pri 140 °C. Rozdiely medzi vyššou a nižšou koncentráciou nanočastíc neboli výrazné. Výsledky PE s magnetickými nanočasticami nie sú znázornené na Fig. 7 z dôvodu vysokých hodnôt. Hodnota tg δ začínala na 4,693 (25°C) a najvyššia nameraná hodnota bola 35,8. Vzorky 1_PE_FE a 2_PE_FE neboli merané pri vysokých teplotách z dôvodu meracieho rozsahu zariadenia (nadprúdy).

Stratový činiteľ esterov s fullerénovými nanočasticami sa zvyšoval pri rôznych úrovniach teploty. Krivky na Fig. 7 možno rozdeliť do 2 skupín, okolo čistého SE a PE. Je zrejme, že všetky vzorky PE zmiešané s fullerénom majú nižšie hodnoty tg δ ako zmes nanočastíc SE a nanočastíc C_{60} . Vplyv fullerénu na pokles hodnôt tg δ v SE je viditeľný pri všetkých teplotách. Pokles hodnoty pri 90°C (v porovnaní s čistým SE) bol pre vzorky 1_SE, 2_SE a 3_SE približne 54 %, 152 % a 133 %. Podľa priemerných hodnôt je optimálna koncentrácia fullerénových nanočastíc v SE 0,02 %w/V s približne 135 % poklesom hodnôt tg δ .

Fullerénové nanočastice s PE vykazovali odlišné výsledky. Len vzorka 3_PE mala nižšie hodnoty stratového činiteľa, ako

čistý PE pri všetkých skúmaných teplotách. Preto môžeme koncentráciu 0,03 %w/V z pohľadu výsledkov merania tgδ považovať za optimálnu koncentráciu. Hodnota pri 90°C sa znížila pri meraní vzorky 3_PE (26,6 %) a naopak zvýšila pri meraní vzoriek 1_PE a 2_PE o 6,5 %, resp. 0,8 %.

Relatívna permitivita určuje stupeň polarizácie, ako odozvu na aplikované elektrické pole a uvádza sa aj ako dielektrická konštanta [29]. Výsledky merania sú graficky znázornené na Fig. 8.

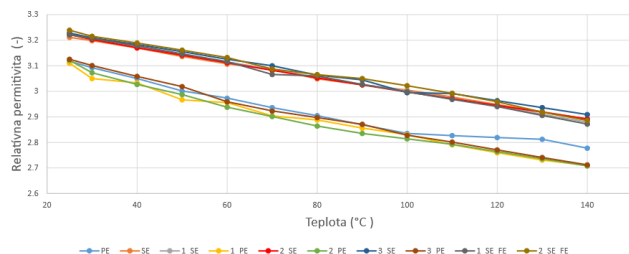


Fig. 8. Teplotná závislosť relatívnej permitivity [28]

Teplotná závislosť relatívnej permitivity skúmaných izolačných olejov vytvára 2 skupiny okolo čistého SE a PE. Hodnoty relatívnej permitivity v PE sú pri všetkých skúmaných teplotách nižšie v porovnaní so SE a rozdiely sa pohybovali od 3 % do 6 %.

Nanokvapaliny na báze PE a fullerénových nanočastíc znížili hodnotu relatívnej permitivity v porovnaní s čistým PE. Vzorky 1_PE a 2_PE vykazovali nižšie hodnoty pri všetkých meraných teplotách. Rozdiely medzi čistým PE a vzorkou 1_PE sa pohybovali od 0,2 % pri teplote 100°C do približne 3 % pri 130°C a medzi PE a vzorkou 2_PE sa pohybovali od 0 % do 2,8 %. Kombinácie PE a magnetických nanočastíc boli merané súbežne so stratovým činiteľom, takže boli mimo rozsah meracieho prístroja. Hodnoty 1_PE_FE a 2_PE_FE začali pri 25°C s hodnotami 3,26 a 5,24.

Rozdiel medzi nanokvapalinami vyrobenými zo SE (fullerén a magnetické nanočastice) a čistým SE sa pohyboval do 1 %. Pri vyšších teplotách, sa rozdiely takmer nemenia a priemerný percentuálny rozdiel medzi čistým olejom a nanokvapalinami je medzi -0,48 % a 0,06 %. Možno konštatovať, že relatívnu permitivitu pri 50 Hz výrazne neovplyvňujú použité nanočastice a ich koncentrácia.

Objemová rezistivita materiálu je odpor toku elektrického prúdu cez materiál skúmanej vzorky. Elektrická vodivosť materiálu klesá so zvyšujúcou sa hodnotou objemového odporu, čo značí vyššiu priernu pevnosť materiálu [30]. Grafické znázornenie priebehov meranej veličiny na skúmaných vzorkách je na Fig. 9.

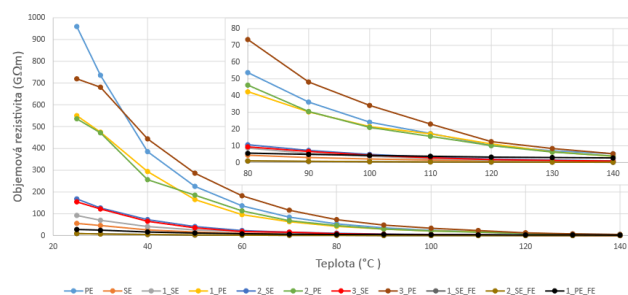


Fig. 9. Teplotná závislosť objemovej rezistivity [28]

Objemová rezistivita vo všeobecnosti klesala so zvyšujúcou sa teplotou. Hodnota rezistivity čistého prírodného esteru je vyššia ako u SE a rozdiel je výrazný hlavne pri nižších teplotách s maximálnym rozdielom okolo 900 GΩm pri 25°C. Pri teplote 90°C je odpor čistého PE asi 11,8-krát vyšší ako odpor SE. Nanokvapaliny esterov s nanočasticami C₆₀ majú podobný priebeh, ako ich základné kvapaliny a tvoria 2 skupiny.

Zvýšenie hodnoty objemovej rezistivity pomocou fullerénových nanočastíc v PE bolo pozorované len pri najvyššej skúmanej koncentrácii 0,03 %w/V a pri teplotách vyšších ako 40°C. Zvýšenie rezistivity vzorky 3_PE v porovnaní s čistým PE (teploty medzi 40-140 °C) sa pohybovalo približne od 1,1 % pri 140°C do približne 42 % pri 100°C. Vzorky 1_PE a 2_PE dosahovali približne rovnaké hodnoty v celom teplotnom priebehu a to pod hodnotami čistého PE a vzorky 3_NE. Rezistivita PE s magnetickými nanočasticami vzorky 1_PE_FE klesala s teplotou z 28,26 GΩm na 2,71 GΩm a vzorku 2_PE_FE rovnako ako pri predošlých meraniach nebolo možné zaznamenať.

Fullerénové a magnetitové nanočastice mali rozdielny vplyv na objemovú rezistivitu v SE. Zmes SE a nanočastíc Fe₃O₄ mala stabilnú, ale relatívne nízku hodnotu objemového odporu. Magnetické nanokvapaliny vykazovali nižšie hodnoty rezistivity pri všetkých teplotách. Hodnota rezistivity bola približne 3-krát nižšia pre vzorku 2_SE_FE pri 140°C a viac ako 6-krát nižšia pre vzorku 1_SE_FE pri 25°C. Fullerénové nanočastice pozitívne ovplyvnili objemovú rezistivitu SE s výnimkou hodnoty pri 140°C (vzorka 1_SE), kde nanokvapalina vykázala pokles odporu približne o 18 %. Zvýšenie objemového odporu sa pohybovalo približne od 4 % (vzorka 1_SE pri 25°C) do 201 % (vzorka 2_SE pri 25°C). Koncentráciu 0,02 %w/V možno považovať za optimálnu koncentráciu podľa priemerného zvýšenia objemového odporu s 2,33-krát vyššími hodnotami v porovnaní s čistým SE.

Podľa IEC 60247 sa meranie veličín spomínaných v predošlých kapitolách vykonáva pri teplote 90°C, ale v praxi vysokonapäťové zariadenia pracujú pri rôznych teplotách v závislosti od okolitého prostredia, úrovne zaťaženia a zmien spôsobených rôznymi poruchami, ktoré môžu ovplyvniť izolačné vlastnosti.

Stratový činiteľ sa na rozdiel od relatívnej permitivity a objemovej rezistivity zvyšuje s teplotou, čo súvisí so zhoršením dielektrickej výkonnosti. Najnižší vplyv zvýšenia

teploty bol pozorovaný u PE a jeho zmesi s fullerénovými nanočasticami (vzorky 1_PE, 2_PE a 3_PE). Optimálna koncentrácia 0,03 %w/V mala najnižšie priemerné hodnoty tgδ, vďaka čomu je táto vzorka vhodná na zvýšenie dielektrickej výkonnosti. Všetky vzorky SE s fullerénovými nanočasticami mali vyššie hodnoty ako vzorky PE. Nanočastice C₆₀ v SE zlepšili dielektrickú výkonnosť z pohľadu stratového činiteľa pri všetkých koncentráciách a teplotách. Optimálna koncentrácia fullerénových nanočastíc v SE je 0,02 %w/V s najnižšími hodnotami tgδ spomedzi vzoriek na báze SE. Magnetické nanočastice (Fe₃O₄) výrazne zvýšili hodnoty stratového činiteľa, takže možno konštatovať, že nie sú vhodné na zvýšenie dielektrickej výkonnosti.

Relatívnu permitivitu nanokvapalín podľa týchto meraní možno rozdeliť do 2 skupín, okolo krivky čistého PE a okolo krivky čistého SE. Skupina čistého PE a zmesi PE a fullerénových nanočastíc dosiahla nižšie hodnoty permitivity pri všetkých skúmaných teplotách. Vzorky SE mali hodnotu o približne 0,1 vyššiu, čo nepredstavuje výrazný rozdiel. Možno konštatovať, že relatívna permitivita nie je významne ovplyvnená koncentráciou a druhom nanočastíc, ale typom izolačnej kvapaliny.

Objemová rezistivita je výrazne ovplyvnená stúpajúcou teplotou. Nanokvapaliny s PE a fullerénovými nanočasticami a čistý PE, ukázali vyššie hodnoty rezistivity v porovnaní so zvyšnými vzorkami, čo je možné pozorovať hlavne pri nižších teplotách. Optimálna koncentrácia (podľa objemovej rezistivity) fullerénových nanočastíc v PE je 0,03 %w/V s najvyššími hodnotami rezistivity pri teplotách od 40°C do 140°C. Fullerénové nanočastice zvýšili hodnoty rezistivity pri všetkých koncentráciách do 100°C a pri vyšších teplotách sa rozdiely medzi jednotlivými vzorkami vyrovnali. Magnetické nanočastice vykazovali nízke, ale stabilné hodnoty a pri teplotách nad 100°C vzorka 1_PE_FE dosahovala vyššie hodnoty ako zmesi SE a fullerénových nanočastíc.

Podľa meraní v tejto kapitole, nanokvapaliny PE s fullerénovými nanočasticami majú lepšiu dielektrickú výkonnosť, ako nanokvapaliny na báze SE. Podľa dielektrickej výkonnosti vzorka 3_PE dosiahla najlepšie výsledky spomedzi všetkých meraných vzoriek, čiže je možné ju považovať za optimálnu koncentráciu z pohľadu dielektrickej výkonnosti. Optimálna koncentrácia fullerénových nanočastíc v SE je 0,02 %w/V. Magnetické nanočastice pri všetkých meraniach znížili dielektrickú výkonnosť ich základných olejov.

H. Experimentálne výsledky merania čiastkových výbojov

Pri hodnotení parametrov čiastkových výbojov budú výsledky rozdelené do 2 hlavných skupín, výboje v kladnej polperióde a výboje v zápornej polperióde priloženého striedavého napätia s frekvenciou 50 Hz. V prípade týchto meraní výboje v kladnej polperióde nad hodnotou 50 pC vytvárajú strímre a ich hodnota zdanlivého náboja sa pohybuje v porovnaní s hodnotami v zápornej polperióde rádovo vyššie (väčšinou nad hodnotou 100 pC). V zápornej polperióde sa vyskytujú výboje koróny, ktoré vznikajú okolo hrotovej elektródy. Tieto poznatky boli publikované v práci od Kurimský a spol. [31] a z porovnania hodnôt zdanlivého

výboja aplikované v tomto článku (Fig. 10). Vzhľadom na charakter týchto meraní, nebolo určené počiatkové ani zhášacie napätie čiastkových výbojov, pretože pri experimentálne určenej hladine napätia, čiastkové výboje vznikali postupne počas 5 minútového intervalu a nie počas nárastu napätia.

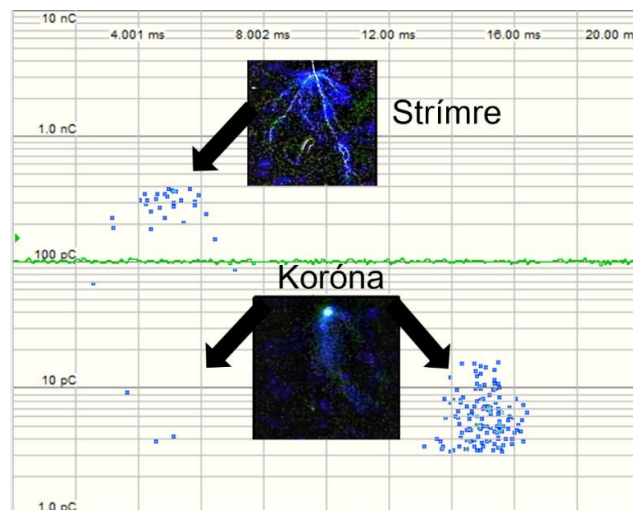


Fig. 10. Analýza druhu výbojov z merania vzorky 3_SE [31]

Hlavným kritériom pre určenie kvality izolačnej kvapaliny je v tomto prípade počet výbojov, ktoré vznikli v kladnej, alebo zápornej polperióde. Korónové výboje v kladnej polperióde vznikali len v určitých prípadoch a v malom počte, takže výboje v tejto polperióde môžeme klasifikovať ako strímre. Na Fig. 11 a Fig. 12 sú graficky znázornené počty výbojov v jednotlivých vzorkách.

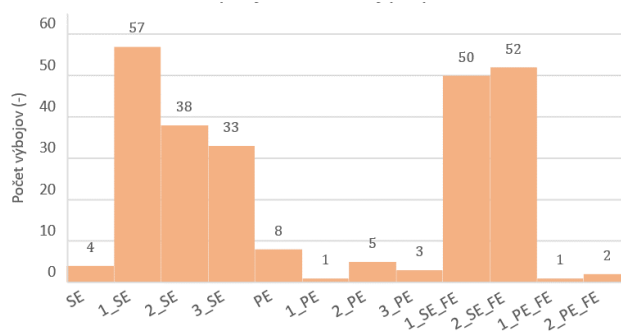


Fig. 11. Počet čiastkových výbojov v kladnej polperióde napätia 20 kV

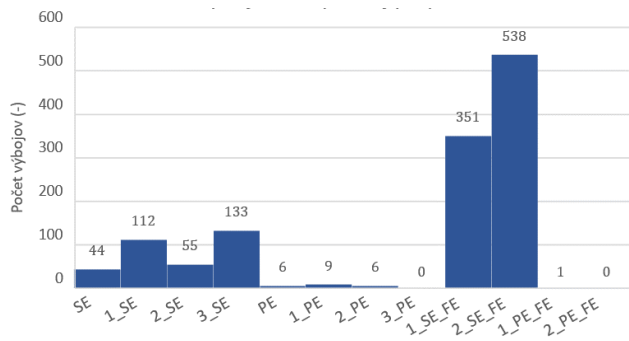


Fig. 12. Počet čiastkových výbojov v zápornej polperióde napätia 20 kV

Počet čiastkových výbojov v SE a jeho nanokvapalinách je vyšší, ako pri čistom PE a nanokvapalinách na jeho báze. Čistý SE vykazoval málo strímrových výbojov (kladná polperióda) v porovnaní v jeho nanokvapalinami, ktoré mali počet výbojov 8-krát (3_SE) až 14-krát (1_SE) vyšší. So zvyšujúcou sa koncentráciou nanočastíc C₆₀, klesal počet strímrových výbojov, čo ale neplatí pre korónové výboje, ktoré boli najpočetnejšie pri najvyššej koncentrácii fullerénu. Nanočastice magnetitu výrazne navýšili počet výbojov v kladnej aj zápornej polperióde a v prípade vyššej koncentrácie sa zvýšil ich celkový počet skoro 13-krát.

Čistý PE v porovnaní s čistým SE má celkovo nižší počet výbojov, napriek tomu, že pri strímrových výbojoch zaostával za SE a počet mal 2-krát vyšší (rozdiel predstavoval 4 výboje). V prípade nanokvapalín na báze PE je zjavné, že fullerénové aj magnetitové nanočastice zlepšili vlastnosti základnej kvapaliny. Vzhľadom na nízku početnosť výbojov vzoriek PE a jeho nanokvapalín, boli vzorky merané aj pri hladine napätia 24 kV (Fig. 13).

Vyššia hladina napätia ukázala rozdiely medzi jednotlivými vzorkami. Čistý PE mal lepšie výsledky len v 2 prípadoch a to v porovnaní so vzorkou 1_PE v zápornej polperióde a so vzorkou 1_PE_FE v kladnej polperióde. Výsledky zvyšných vzoriek vykazujú zníženie celkového počtu čiastkových výbojov. Najlepšie výsledky zo všetkých meraných vzoriek mala vzorka s vyššou koncentráciou magnetických nanočastíc (2_PE_FE) a z pohľadu kombinácie s fullerénom, taktiež vzorka s najvyššou koncentráciou (3_SE) dosiahla najnižší počet čiastkových výbojov.

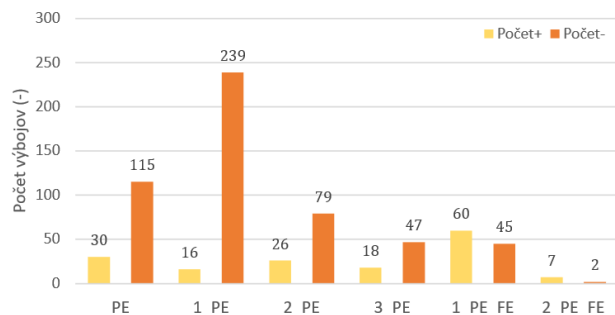


Fig. 13. Počet čiastkových výbojov pri hladine napätia 24 kV, kde Počet+ označuje počet čiastkových výbojov v kladnej polperióde a Počet- v zápornej

Z pohľadu priemerného zdanlivého náboja výsledky ukázali, že čistý PE a PE s fullerénovými nanočasticami, mali vo všeobecnosti vyššie hodnoty zdanlivého náboja ako čistý SE a jeho kombinácia s nanočasticami C₆₀ (Fig. 14 a Fig. 15). Tieto rozdiely môžu byť dané aj počtom výbojov, a pri vyššej početnosti výbojov v prírodných esteroch, by sa rozdiely medzi spomínanými vzorkami vyrovnali. Najvyšší rozdiel, okolo 65 %, zo spomínaných vzoriek bol medzi vzorkami 1_SE a 1_PE. V kladnej polperióde koncentrácia nanočastíc ovplyvňovala výšku priemerného náboja pre SE negatívne, čiže sa veľkosť zväčšovala s pribúdajúcou koncentráciou. Veľkosť náboja pri korónových výbojoch (záporná polperióda) sa výrazne nemenila.

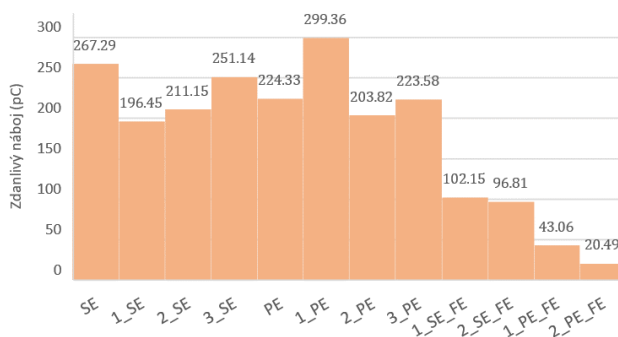


Fig. 14. Priemerné hodnoty zdanlivého náboja v kladnej polperióde priloženého napätia

Nanočastice magnetitu pri strímrových výbojoch (kladná polperióda) značne znížili hodnotu zdanlivého náboja pri oboch základných olejoch. Vzorky magnetických nanokvapalín na báze SE mali približne polovičnú hodnotu zdanlivého náboja v porovnaní s kombináciami esterov a fullerénu a kombinácia PE a magnetitových nanočastíc dosiahla 5-krát až 10-krát nižšie hodnoty.

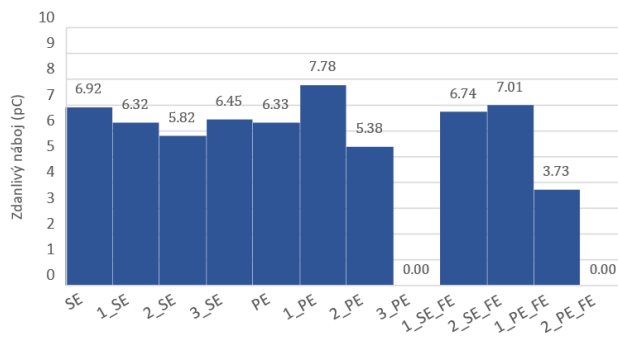


Fig. 15. Priemerné hodnoty zdanlivého náboja v zápornej polperióde priloženého napätia

Priemerný zdanlivý náboj v čistom PE a jeho nanokvapalinách pri napät'ovej úrovni 24 kV je na Fig. 16.

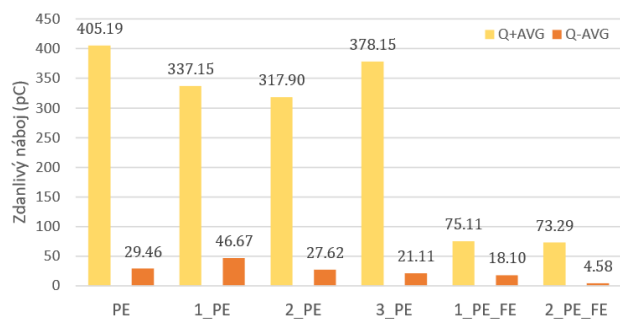


Fig. 16. Priemerný zdanlivý náboj pri hladine napätia 24 kV, kde Q+AVG označuje priemernú hodnotu v kladnej polperióde a Q-AVG v zápornej

V prípade väčšieho množstva výbojov pri vyššej napäťovej hladine sa priemerný zdanlivý náboj v kladnej polperióde znížil použitím nanočastíc. Fullerénové nanočastice znížili hodnotu náboja o 7,2 % až 27,5 % pri kladnej polarite napätia a o 6,7 % až 39,6 %. Vzorka 1_PE zvýšila hodnotu zdanlivého náboja v zápornej polperióde o 58,4 %. Potvrdilo sa, že magnetické nanočastice znižujú hodnotu zdanlivého náboja v oboch poleriódach. Koncentrácia magnetitových nanočastíc ovplyvnila veľkosť náboja výrazne len v zápornej polperióde, kde znížila hodnotu náboja viac ako 4-krát. Prehľad spracovaných výsledkov merania častkových výbojov je v TABLE IV.

TABLE IV. VÝSLEDKY MERANIA ČIASTKOVÝCH VÝBOJOV PRI RÔZNYCH HLADINÁCH NAPÄTIA

Vzorka	Zdanlivý náboj v kladnej polperióde (pC)	Zdanlivý náboj v zápornej polperióde (pC)	Počet výbojov v kladnej polperióde (-)	Počet výbojov v zápornej polperióde (-)
20 kV				
SE	267,29	6,92	4	44
1_SE	196,45	6,32	57	112
2_SE	211,15	5,82	38	55
3_SE	251,14	6,45	33	133
PE	224,33	6,33	8	6
1_PE	299,36	7,78	1	9
2_PE	203,82	5,38	5	6
3_PE	223,58	0,00	3	0
1_SE_FE	102,15	6,74	50	351
2_SE_FE	96,81	7,01	52	538
1_PE_FE	43,06	3,73	1	1
2_PE_FE	20,49	0,00	2	0
24 kV				
PE	405,19	29,46	30	115
1_PE	337,15	46,67	16	239
2_PE	317,90	27,62	26	79
3_PE	378,15	21,11	18	47
1_PE_FE	75,11	18,10	60	45
2_PE_FE	73,29	4,58	7	2

I. Vplyv externého magnetického poľa na parametre častkových výbojov

Zhodnotenie zmeny parametrov častkových výbojov po pridaní externého magnetického poľa pomocou statických magnetov s hodnotou 60 mT pri hrotovej elektróde je v TABLE V. Pre jednoduchšiu orientáciu vo výsledkoch je

v tejto tabuľke použité farebné označenie hodnôt zelenou a červenou farbou. Zelená farba značí zlepšenie vlastností po pridaní externého magnetického poľa a červená značí zhoršenie.

TABLE V. VÝSLEDKY MERANIA ČIASTKOVÝCH VÝBOJOV PRI RÔZNYCH HLADINÁCH NAPÄTIA A VPLYV EXTERNÉHO MAGNETICKÉHO POĽA

Vzorka	Zdanlivý náboj v kladnej polperióde (pC)	Zdanlivý náboj v zápornej polperióde (pC)	Počet výbojov v kladnej polperióde (-)	Počet výbojov v zápornej polperióde (-)
20 kV				
1_SE_FE	102,15	6,74	50	351
1_SE_FE_60mT	132,86	7,84	49	647
2_SE_FE	96,81	7,01	52	538
2_SE_FE_60mT	107,82	8,48	49	1196
1_PE_FE	43,06	3,73	1	1
1_PE_FE_60mT	13,24	3,91	10	3
2_PE_FE	20,49	0,00	2	0
2_PE_FE_60mT	22,28	16,77	2	2
24 kV				
1 PE FE	75,11	18,10	60	45
1_PE_FE_60mT	86,37	5,71	68	38
2 PE FE	73,29	4,58	7	2
2_PE_FE_60mT	86,47	6,37	4	8

Z výsledkov je možné vidieť, že externé magnetické pole zhoršuje parametre častkových výbojov vo väčšine prípadov. Zmeny v priemernom zdanlivom náboji nie sú výrazné s výnimkou vzorky 1_PE_FE, kde došlo k výraznému zníženiu hodnoty zdanlivého náboja, čo mohlo byť spôsobené nízkym počtom zaznamenaných výbojov tejto vzorky, pri meraní bez externého magnetického poľa. Výraznejšie zmeny nastali v zápornej polperióde, kde sa navýšil počet výbojov vzoriek 1_SE_FE a 2_SE_FE približne dvojnásobne. Zvýšenie počtu výbojov pri týchto vzorkách a pri vzorky 1_PE_FE v kladnej polperióde bolo zvýšenie počtu korónových výbojov, vzhľadom na zmenu priemernej hodnoty náboja, ktorá sa držala v nízkych číslach. Z týchto meraní je možné konštatovať, že externé magnetické pole vplyva na tvorbu korónových výbojov okolo hrotovej elektródy a mierne navyšuje hodnoty zdanlivého náboja.

Vplyv externého magnetického poľa na magnetitové nanočastice spôsobil zmenu v parametroch častkových výbojov. Konkrétne ide o jav, pri ktorom sa po aplikácii externého magnetického poľa v magnetickej kvapaline vytvárajú mikroreťazcové štruktúry, ktoré tvoria reťazovité klastre v smere magnetického poľa [32]. Podľa [33] klastre narastajú so zvyšujúcim sa magnetickým poľom do hodnoty 300 mT. Vzhľadom na to, že prítomnosť nanočastíc v kvapaline ovplyvňuje rozloženie intenzity elektrického poľa, vznik klastrov môže zmeniť rozloženie tohto poľa. V prípade individuálnych nanočastíc, s malým rozmerom je gradient elektrického poľa ostrejší a v oblastiach okolo nanočastíc je väčšia nehomogenita [26]. Existuje predpoklad, že pole v okolí klastrov je homogénnejšie a tým pádom sa mierne zníži počet výbojov (TABLE V). Z výsledkov možno konštatovať, že v okolí klastrov (vplyvom elektrického poľa) je potrebná vyššia energia na vytvorenie výbojov.

V. CELKOVÉ ZHODNOTENIE MERANÝCH VZORIEK A
ODPORÚČANIA PRE PRAX

Pre vyhodnotenie jednotlivých experimentov bola vytvorená tabuľka výkonnosti jednotlivých vzoriek, kde je zavedený pojem výkonnostná známka, ktorá predstavuje kvantitatívne hodnotenie meraných vzoriek pre jednotlivé experimentálne merania. Nanokvapaliny, alebo čisté kvapaliny s najvyššou kvalitou spomedzi všetkých vzoriek boli obodované nízkymi výkonnostnými známkami začínajúcimi číslom 1. Vzhľadom na to, že počet vzoriek je 12 udeľujú sa body podľa výsledkov meraní od 1 do 12, zoradené podľa výsledkov popísaných v predošlej kapitole. Pri meraní striedavého preskokového napätia a objemovej rezistivity boli kvalitatívne najlepšie vzorky s najvyššou hodnotou meranej veličiny, naopak pri zvyšných meraniach bola požadovaná čo najnižšia hodnota. Na určenie optimálnej koncentrácie a vyhodnotenie kvality meraných dielektrík, je na záver zostavená tabuľka so súčtom výkonnostných známok z jednotlivých meraní.

V TABLE VI sú výsledky meraní AC-PN hodnotené pomocou priemernej hodnoty a pre porovnanie nanokvapalín s ich základnou kvapalinou je v tabuľke znázornená *Zmena*. Striedavé preskokové napätie malo najvyššiu hodnotu pri vzorkách s nanočasticami magnetitu, okrem vzorky 2_PE_FE, ktorú prevýšili vzorky 3_PE a čistý SE. Najnižšie hodnoty zaznamenali vzorky s nízkou koncentráciou nanočastíc C₆₀.

TABLE VI. TABUĽKA VÝKONNOSTI VZORIEK PRE MERANIE AC-PN

Vzorka	Priemerná hodnota \bar{U} (kV)	Zmena (%)	Výkonnostná známka (-)
2_SE_FE	65,06	13,52	1
1_PE_FE	63,81	14,68	2
1_SE_FE	63,79	11,31	3
3_PE	58,57	5,27	4
SE	57,31	-	5
2_PE_FE	57,18	2,77	6
3_SE	56,71	-1,06	7
2_PE	56,36	1,29	8
PE	55,64	-	9
2_SE	55,06	-4,09	10
1_SE	53,22	-7,69	11
1_PE	34,61	-60,76	12

Pre celkové zhodnotenie teplotnej závislosti vzoriek budú merania dielektrickej výkonnosti (tgδ, relatívna permitivita a objemová rezistivita) započítavané ako priemerná hodnota výkonnostnej známky (TABLE VII). To zabezpečí, aby tieto 3 spolu súvisiace merania nemali neprimerane veľký vplyv na celkové hodnotenie.

TABLE VII. TABUĽKA VÝKONNOSTI VZORIEK PRE MERANIA DIELEKTRICKEJ VÝKONNOSTI (TGA, RELATÍVNA PERMITIVITA A OBJEMOVÁ REZISTIVITA)

Vzorka	Priemerná hodnota známky dielektrickej výkonnosti (-)	Výkonnostná známka (-)
3_PE	2	1
PE	2,33	2
2_PE	2,67	3
1_PE	3	4

2_SE	5,67	5
3_SE	7	6
SE	7,33	7
1_SE	7,33	8
1_SE_FE	8,67	9
2_SE_FE	9,67	10
1_PE_FE	10,33	11
2_PE_FE	11,67	12

Pri čiastkových výbojoch, bolo nutné brať do úvahy počet výbojov ale aj ich zdanlivý náboj. Preto pre určenie výkonnostnej známky, boli násobky hodnôt počtu a zdanlivého náboja kladnej polperiódy sčítané s násobkom v zápornej polperióde nasledovne:

$$PD = (Q_{AVG}^+ * n^+) + (Q_{AVG}^- * n^-) \quad (4)$$

kde Q_{AVG}^+ a Q_{AVG}^- značí priemerný zdanlivý náboj v kladnej a zápornej polperióde, n je počet výbojov a PD je vypočítaná hodnota. V prípade meraní na napäťovej hladine 24 kV čistého PE a nanokvapalín na jeho báze, došlo k zmene výkonnostnej známky iba pri vzorkách 3_PE a 1_PE, preto bude ich výkonnostná známka v TABLE VIII vymenená. Výrazne najlepšie hodnoty z meraných vzoriek dosiahli vzorky kombinácie PE a magnetických nanočastíc. Fulerénové nanočastice v PE zlepšili štatistiky čiastkových výbojov a to hlavne pri najvyššej a najnižšej meranej koncentrácii. Čistý SE v porovnaní s čistým PE dosiahol lepšiu výkonnostnú známku, kvôli nízkej početnosti výbojov s vysokou hodnotou zdanlivého náboja. Všetky nanokvapaliny na báze SE mali rádoovo vyššie hodnoty PD .

TABLE VIII. TABUĽKA VÝKONNOSTI VZORIEK PRE MERANIA ČIASTKOVÝCH VÝBOJOV

Vzorka	PD (pC)	Výkonnostná známka (-)
2_PE_FE	40,97	1
1_PE_FE	46,79	2
3_PE	670,75	3
1_PE	369,42	4
2_PE	1051,43	5
SE	1373,76	6
PE	1832,58	7
1_SE_FE	7474,56	8
2_SE	8343,84	9
2_SE_FE	8803,82	10
3_SE	9146,02	11
1_SE	11905,43	12

TABLE IX. TABUĽKA VÝKONNOSTI VZORIEK PRE MERANIA ČIASTKOVÝCH VÝBOJOV

Vzorka	Celková výkonnosť (-)	Poradie (-)
3_PE	8	1
1_PE_FE	15	2
2_PE	16	3
SE	18	4
PE	18	5
2_PE_FE	19	6
1_PE	20	7

1_SE_FE	20	8
2_SE_FE	21	9
3_SE	24	10
2_SE	26	11
1_SE	31	12

Celkové zhodnotenie jednotlivých meraní ukázalo, že vzorka 3_PE zlepšila dielektrické vlastnosti základnej kvapaliny. Táto vzorka vylepšila dielektrické vlastnosti v porovnaní s čistým PE pri všetkých meraniach okrem objemovej rezistivity, kde bol rozdiel len 1,8 %. Táto vzorka s koncentráciou 0,03 %w/V by bola vhodná pre experimentálne testovanie v reálnom transformátore. Vzhľadom na to, že z meraných vzoriek je to najvyššia koncentrácia, tieto experimenty by bolo vhodné zopakovať pre kombinácie PE MIDEL eN 1204 s fullerénovými nanočasticami pri vyšších koncentráciách, pokiaľ by nedošlo k poklesu celkovej kvality kvapalného dielektrika vplyvom vysokej koncentrácie.

Nanočastice fullerénu aj magnetitu v SE zhoršili dielektrické vlastnosti a vzorka s najnižšou koncentráciou fullerénu, podľa celkovej výkonnosti dosiahla najhoršie výsledky. So zvyšujúcou sa koncentráciou fullerénu v SE sa zlepšovali vlastnosti, takže pri vyšších koncentráciách, ako boli použité pre tieto experimenty, by nanokvapaliny mohli vykazovať zlepšenie celkových vlastností. Magnetitové nanočastice podľa týchto experimentov, nie sú vhodné na zlepšenie dielektrických vlastností SE MIDEL 7131.

Magnetické nanokvapaliny na báze PE, zlepšili hodnoty AC-PN výrazným spôsobom až do 14,68 %. Túto hodnotu dosiahla nižšia koncentrácia, ktorá z hľadiska všetkých experimentov dosahovala lepšie vlastnosti, ako vzorka s vyššou koncentráciou. Zlepšenie hodnôt nastalo aj pri meraní čiastkových výbojov, kde vzorky PE s nanočasticami Fe₃O₄ dosiahli najlepšie výsledky. Lepšie celkové výsledky pre túto kombináciu neboli možné, kvôli výsledkom merania dielektrickej výkonnosti (tgδ, relatívna permitivita a objemová rezistivita), kde pri všetkých teplotných závislostiach výrazne zaostávali za zvyšnými vzorkami.

VI. ZÁVER

Z pohľadu striedavého preskokového napätia nanokvapaliny na báze esterov s fullerénovými nanočasticami vyžadujú koncentráciu vyššiu ako 0,03 %w/V, aby sa stali prijateľnými pre praktické aplikácie z hľadiska priereznej pevnosti. Magnetitové nanočastice zlepšujú striedavé preskokové napätie syntetických aj prírodných esterov. Nanokvapalina na báze SE s koncentráciou nanočastíc 0,1 %w/V zvýšila hodnotu AC-PN o 13,52 % a nanokvapalina na báze PE s koncentráciou nanočastíc 0,05 %w/V zlepšila svoju hodnotu v porovnaní s čistým PE o 14,68 %.

Podľa meraní dielektrickej výkonnosti (stratový činiteľ, relatívna permitivita a objemová rezistivita) vzorka 3_PE vykazovala najlepšie a najkomplexnejšie výsledky v porovnaní so zvyškom skúmaných vzoriek. Vzorka 2_SE dosahovala lepšie výsledky v porovnaní s čistým SE a inými nanokvapalinami na báze SE z hľadiska dielektrickej výkonnosti. Všeobecne nanokvapaliny na báze PE a čistý PE

vykazovali lepšiu dielektrickú výkonnosť, ako nanokvapaliny na báze SE a čistý SE. Nanočastice magnetitu, výrazne zhoršujú dielektrickú výkonnosť v oboch skúmaných základných kvapalinách. Kombinácia magnetických nanočastíc a PE dosahovala najhoršie výsledky dielektrickej výkonnosti spomedzi všetkých vzoriek.

Z výsledkov merania čiastkových výbojov možno konštatovať, že čistý PE a nanokvapaliny na jeho báze dosahovali lepšie výsledky, ako vzorky obsahujúce SE. Najlepšie výsledky zo všetkých vzoriek mali vzorky 1_PE_FE a 2_PE_FE, ktoré výrazne znížili početnosť výbojov a zdanlivý náboj v porovnaní s čistým PE. Menej výrazné zlepšenie hodnôt čiastkových výbojov bolo zaznamenané pri fullerénových nanočasticách v PE. Nanokvapaliny na báze SE (nanočastice C₆₀ aj Fe₃O₄) v porovnaní s čistým SE vykazovali horšie výsledky.

Z celkového zhodnotenia vzoriek pomocou výkonnostných známkov, je možné odporučiť pre prax kombináciu PE a nanočastíc fullerénu s najvyššou koncentráciou 0,03 %w/V, ktorá dosiahla lepšie výkonnostné známky ako čisté estery. Ďalšie smerovanie výskumu by sa malo zamerať na túto kombináciu aj s vyššími koncentraciami, ktoré môžu mať výraznejší pozitívny vplyv na dielektrické vlastnosti. Je potrebné vykonať aj skúšky súvisiace so starnutím izolačných olejov a preveriť, či nanočastice fullerénu nebudú zhoršovať vlastnosti základnej kvapaliny po urýchlennom tepelnom starnutí.

Z ekonomického hľadiska je jasné, že vylepšenie dielektrických a tepelných vlastností transformátorových olejov pomocou nanočastíc zabezpečí vyššiu dielektrickú odolnosť a nižšiu pracovnú teplotu transformátora, čo môže viesť k nižším nákladom na výrobu transformátorov v dôsledku možnej redukcie materiálu (zmenšenie rozmerov). Zároveň prevádzka transformátora pri nižšej pracovnej teplote zaručí predĺženie jeho životnosti, čo predstavuje potenciálne úspory pre prevádzkovateľov distribučnej siete. Vylepšenie vlastností pomocou nanočastíc poskytuje príležitosť pre transformátory, ktoré neuspeli pri tepelných skúškach s bežným transformátorovým olejom. V takom prípade môže výrobca transformátorov použiť nanokvapaliny s efektívnejším odvodom tepla a lepšími dielektrickými vlastnosťami a tým predísť ekonomickým stratám spojeným s nevyužitými transformátormi, alebo nákladom na ich demontáž a úpravu.

ACKNOWLEDGMENT

Táto práca bola podporená vedeckou grantovou agentúrou Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky a Slovenskej akadémie v rámci projektu VEGA 2/0011/20 a 1/0154/21 a Agentúrou na podporu vedy a výskumu v rámci projektu APVV-15-0438, APVV-17-0372, a APVV-18-0160.

REFERENCES

- [1] CHU, Steven a Arun MAJUMDAR. Opportunities and challenges for a sustainable energy future. Nature [online]. 2012, 488(7411), 294–303

- [cit. 2021-03-02]. ISSN 0028-0836, 1476-4687. Dostupné na internete: doi:10.1038/nature11475
- [2] RAFIQ, M., Y.Z. LV, Y. ZHOU, K.B. MA, W. WANG, C.R. LI a Q. WANG. Use of vegetable oils as transformer oils – a review. *Renewable and Sustainable Energy Reviews* [online]. 2015, 52, 308–324 [cit. 2020-10-29]. ISSN 13640321. Dostupné na internete: doi:10.1016/j.rser.2015.07.032
- [3] TOKUNAGA, Junko, Masanori NIKAIDO, Hidenobu KOIDE a Tomoyuki HIKOSAKA. Palm fatty acid ester as biodegradable dielectric fluid in transformers: a review. *IEEE Electrical Insulation Magazine* [online]. 2019, 35(2), 34–46 [cit. 2020-11-02]. ISSN 0883-7554, 1558-4402. Dostupné na internete: doi:10.1109/MEI.2019.8636104
- [4] FERNÁNDEZ, Inmaculada, Alfredo ORTIZ, Fernando DELGADO, Carlos RENEDO a Severiano PÉREZ. Comparative evaluation of alternative fluids for power transformers. *Electric Power Systems Research* [online]. 2013, 98, 58–69 [cit. 2021-03-08]. ISSN 03787796. Dostupné na internete: doi:10.1016/j.epr.2013.01.007
- [5] CIR. Transformer risk management. *M&I Materials* [online]. 2017, 38–43. Dostupné na internete: http://static.mimaterials.com/midel/documents/sales/MIDEL_CIR_roundtable_report.pdf
- [6] RAFIQ, Muhammad, Muhammad SHAFIQUE, Anam AZAM, Muhammad ATEEQ, Israr Ahmad KHAN a Abid HUSSAIN. Sustainable, Renewable and Environmental-Friendly Insulation Systems for High Voltages Applications. *Molecules* [online]. 2020, 25(17), 3901 [cit. 2020-11-04]. Dostupné na internete: doi:10.3390/molecules25173901M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
- [7] CHOI, S.U.S. a Jeffrey EASTMAN. Enhancing thermal conductivity of fluids with nanoparticles. In: *Proceedings of the ASME International Mechanical Engineering Congress and Exposition*. 1995.
- [8] FAL, Jacek, Omid MAHIAN a Gawel ŻYŁA. Nanofluids in the Service of High Voltage Transformers: Breakdown Properties of Transformer Oils with Nanoparticles, a Review. *Energies* [online]. 2018, 11(11), 2942 [cit. 2020-11-02]. Dostupné na internete: doi:10.3390/en11112942
- [9] KHALED, Usama a Abderrahmane BEROUAL. AC dielectric strength of synthetic ester-based Fe₃O₄, Al₂O₃ and SiO₂ nanofluids — conformity with normal and weibull distributions. *IEEE Transactions on Dielectrics and Electrical Insulation* [online]. 2019, 26(2), 625–633. ISSN 1558-4135. Dostupné na internete: doi:10.1109/TDEI.2018.007759
- [10] MÉNDEZ, C., I. FERNÁNDEZ, A. ORTIZ, F. ORTIZ, C. OLMO, A. SANTISTEBAN, F. DELGADO a C. J. RENEDO. Effect of Magnetic and Non-magnetic Nanoparticles on Insulation and Cooling Behaviour of a Natural Ester for Power Transformers. In: *2020 International Symposium on Electrical Insulating Materials (ISEIM): 2020 International Symposium on Electrical Insulating Materials (ISEIM)*. 2020, s. 111–114.
- [11] HUSSAIN, Md Rashid, Qasim KHAN, Asfar Ali KHAN, Shady S. REFAAT a Haitham ABU-RUB. Dielectric Performance of Magneto-Nanofluids for Advancing Oil-Immersed Power Transformer. *IEEE Access* [online]. 2020, 1–1. ISSN 2169-3536. Dostupné na internete: doi:10.1109/ACCESS.2020.3021003
- [12] OLMO, Cristian, Cristina MENDEZ, Felix ORTIZ, Fernando DELGADO, Rafael VALIENTE a Peter WERLE. Maghemite Nanofluid Based on Natural Ester: Cooling and Insulation Properties Assessment. *IEEE Access* [online]. 2019, 7, 145851–145860 [cit. 2021-04-07]. ISSN 2169-3536. Dostupné na internete: doi:10.1109/ACCESS.2019.2945547
- [13] BEROUAL, Abderrahmane a Usama KHALED. Statistical Investigation of Lightning Impulse Breakdown Voltage of Natural and Synthetic Ester Oils-Based Fe₃O₄, Al₂O₃ and SiO₂ Nanofluids. *IEEE Access* [online]. 2020, 8, 112615–112623. ISSN 2169-3536. Dostupné na internete: doi:10.1109/ACCESS.2020.3003246
- [14] KURIMSKÝ, Juraj, Michal RAJŇÁK, Pavol BARTKO, Katarína PAULOVICOVÁ, Roman CIMBALA, Dušan MEDVEĎ, Mária DŽAMOVÁ, Milan TIMKO a Peter KOPČANSKÝ. Experimental study of AC breakdown strength in ferrofluid during thermal aging. *Journal of Magnetism and Magnetic Materials* [online]. 2018, 465, 136–142 [cit. 2021-10-21]. ISSN 03048853. Dostupné na internete: doi:10.1016/j.jmmm.2018.05.083
- [15] RAJŇÁK, M., J. KURIMSKÝ, R. CIMBALA, Z. ČONKA, P. BARTKO, M. ŠUGA, K. PAULOVICOVÁ, J. TÓTHOVÁ, M. KARPETS, P. KOPČANSKÝ a M. TIMKO. Statistical analysis of AC dielectric breakdown in transformer oil-based magnetic nanofluids. *Journal of Molecular Liquids* [online]. 2020, 309. Dostupné na internete: doi:10.1016/j.molliq.2020.113243
- [16] KURIMSKÝ, Juraj, Michal RAJŇÁK, Miloš ŠÁRPATAKY, Zsolt ČONKA a Katarína PAULOVICOVÁ. Electrical and acoustic investigation of partial discharges in two types of nanofluids. *Journal of Molecular Liquids* [online]. 2021, 341, 117444 [cit. 2021-10-21]. ISSN 01677322. Dostupné na internete: doi:10.1016/j.molliq.2021.117444
- [17] PISLARU-DANESCU, Lucian, Alexandru M. MOREGA, Mihaela MOREGA, Victor STOICA, Oana Maria MARINICA, Florica NOURAS, Nicolae PADURARU, Istvan BORBATH a Tunde BORBATH. Prototyping a Ferrofluid-Cooled Transformer. *IEEE Transactions on Industry Applications* [online]. 2013, 49(3), 1289–1298 [cit. 2021-10-21]. ISSN 0093-9994, 1939-9367. Dostupné na internete: doi:10.1109/TIA.2013.2252872
- [18] RAJNAK, Michal, Milan TIMKO, Peter KOPCANSKY, Katarina PAULOVICOVA, Jozef KUCHTA, Marek FRANKO, Juraj KURIMSKY, Bystrik DOLNIK a Roman CIMBALA. Transformer oil-based magnetic nanofluid with high dielectric losses tested for cooling of a model transformer. *IEEE Transactions on Dielectrics and Electrical Insulation* [online]. 2019, 26(4), 1343–1349 [cit. 2021-10-21]. ISSN 1070-9878, 1558-4135. Dostupné na internete: doi:10.1109/TDEI.2019.008047
- [19] RAJNAK, Michal, Zan WU, Bystrik DOLNIK, Katarina PAULOVICOVA, Jana TOTHOVA, Roman CIMBALA, Juraj KURIMSKY, Peter KOPCANSKY, Bengt SUNDEN, Lars WADSÖ a Milan TIMKO. Magnetic Field Effect on Thermal, Dielectric, and Viscous Properties of a Transformer Oil-Based Magnetic Nanofluid. *Energies* [online]. 2019, 12(23), 4532 [cit. 2021-10-23]. Dostupné na internete: doi:10.3390/en12234532
- [20] SZCZEŚNIAK, Dominika a Piotr PRZYBYLEK. Oxidation Stability of Natural Ester Modified by Means of Fullerene Nanoparticles. *Energies* [online]. 2021, 14(2), 490 [cit. 2021-04-15]. Dostupné na internete: doi:10.3390/en14020490
- [21] HUANG, Z., F. WANG, Q. WANG, W. YAO, K. SUN, R. ZHANG, J. ZHAO, Z. LOU a J. LI. Significantly enhanced electrical performances of eco-friendly dielectric liquids for harsh conditions with fullerene. *Nanomaterials* [online]. 2019, 9(7). Dostupné na internete: doi:10.3390/nano9070989
- [22] BEROUAL, Abderrahmane a Hidir DUZKAYA. AC and Lightning Impulse Breakdown Voltages of Natural Ester Based Fullerene Nanofluids. *IEEE Transactions on Dielectrics and Electrical Insulation* [online]. 2021, 28, 1996–2003. Dostupné na internete: doi:10.1109/TDEI.2021.009772
- [23] KHELIFA, Hocine, Eric VAGNON a Abderrahmane BEROUAL. AC Breakdown Voltage and Partial Discharge Activity in Synthetic Ester-Based Fullerene and Graphene Nanofluids. *IEEE Access* [online]. 2022, 10, 5620–5634. ISSN 2169-3536. Dostupné na internete: doi:10.1109/ACCESS.2022.3140928
- [24] KHALED, Usama a Abderrahmane BEROUAL. AC dielectric strength of synthetic ester-based Fe₃O₄, Al₂O₃ and SiO₂ nanofluids — conformity with normal and weibull distributions. *IEEE Transactions on Dielectrics and Electrical Insulation* [online]. 2019, 26(2), 625–633. ISSN 1558-4135. Dostupné na internete: doi:10.1109/TDEI.2018.007759
- [25] PATTANADECH, Norasage a Michael MUHR. Comments on PDIV testing procedure according to IEC 61294. In: *2017 IEEE 19th International Conference on Dielectric Liquids (ICDL): 2017 IEEE 19th International Conference on Dielectric Liquids (ICDL)* [online]. 2017, s. 1–4. ISSN 2153-3733. Dostupné na internete: doi:10.1109/ICDL.2017.8124658
- [26] ŠÁRPATAKY, Miloš, Juraj KURIMSKÝ, Michal RAJŇÁK, Katarína PAULOVICOVÁ, Michal KRBAL a Luděk PELIKÁN. Synthetic and natural ester-based nanofluids with fullerene and magnetite nanoparticles – An experimental AC breakdown voltage study. *Journal of Molecular Liquids* [online]. 2022, 368, 120802 [cit. 2022-11-19].

- ISSN 0167-7322. Dostupné na internete: doi:[10.1016/j.molliq.2022.120802](https://doi.org/10.1016/j.molliq.2022.120802)
- [27] HWANG, J. George, Markus ZAHN, Francis M. O'SULLIVAN, Leif A. A. PETERSSON, Olof HJORTSTAM a Rongsheng LIU. Effects of nanoparticle charging on streamer development in transformer oil-based nanofluids. *Journal of Applied Physics* [online]. 2010, **107**(1), 014310 [cit. 2020-10-29]. ISSN 0021-8979, 1089-7550. Dostupné na internete: doi:[10.1063/1.3267474](https://doi.org/10.1063/1.3267474)
- [28] ŠÁRPATAKY, Miloš, Juraj KURIMSKÝ, Michal RAJŇÁK, Michal KRBAL a Marek ADAMČÁK. Dielectric Performance of Natural- and Synthetic-Ester-Based Nanofluids with Fullerene Nanoparticles. *Energies* [online]. 2023, **16**(1), 343 [vid. 2023-01-06]. ISSN 1996-1073. Dostupné z: doi:[10.3390/en16010343](https://doi.org/10.3390/en16010343)
- [29] HAVRAN, Peter, Roman CIMBALA, Juraj KURIMSKÝ, Bystrík DOLNÍK, Irida KOLCUNOVÁ, Dušan MEDVEĎ, Jozef KIRÁLY, Vladimír KOHAN a Ľuboš ŠÁRPATAKY. Dielectric Properties of Electrical Insulating Liquids for High Voltage Electric Devices in a Time-Varying Electric Field. *Energies* [online]. 2022, **15**(1), 391 [cit. 2022-05-27]. ISSN 1996-1073. Dostupné na internete: doi:[10.3390/en15010391](https://doi.org/10.3390/en15010391)
- [30] *Volume Resistivity: Unit, Formula & Test Methods of Plastics* [online]. [cit. 2023-03-06]. Dostupné na internete: <https://omnexus.specialchem.com/polymer-properties/properties/volume-resistivity>
- [31] KURIMSKÝ, Juraj, Michal RAJŇÁK, Roman CIMBALA, Jakub RAJNÍČ, Milan TIMKO a Peter KOPČANSKÝ. Effect of magnetic nanoparticles on partial discharges in transformer oil. *Journal of Magnetism and Magnetic Materials* [online]. 2020, **496**, 165923 [cit. 2023-03-13]. ISSN 0304-8853. Dostupné na internete: doi:[10.1016/j.jmmm.2019.165923](https://doi.org/10.1016/j.jmmm.2019.165923)
- [32] MOTOZAWA, Masaaki, Jia CHANG, Tatsuo SAWADA a Yasuo KAWAGUCHI. Effect of magnetic field on heat transfer in rectangular duct flow of a magnetic fluid. *Physics Procedia* [online]. 2010, **9**, 12th International Conference on Magnetic Fluids (ICMF12), 190–193 [cit. 2023-05-11]. ISSN 1875-3892. Dostupné na internete: doi:[10.1016/j.phpro.2010.11.043](https://doi.org/10.1016/j.phpro.2010.11.043)
- [33] AREKH, Kinnari a Hyo Sook LEE. Magnetic field induced enhancement in thermal conductivity of magnetite nanofluid. *Journal of Applied Physics* [online]. 2010, **107**(9), 09A310 [cit. 2023-05-11]. ISSN 0021-8979. Dostupné na internete: doi:[10.1063/1.3348387](https://doi.org/10.1063/1.3348387)

Metódy a prostriedky pre návrh, modelovanie a realizáciu distribuovaných systémov riadenia veľkých fyzikálnych experimentov

Milan Tkáčik
Katedra kybernetiky
a umelej inteligencie
Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach
milan.tkacik@tuke.sk

Ján Jadlovský
Katedra kybernetiky
a umelej inteligencie
Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach
jan.jadlovsky@tuke.sk

Slávka Jadlovská
Katedra priemyselného
inžinierstva a informatiky
Fakulta výrobných technológií
Technická univerzita v Košiciach
slavka.jadlovska@tuke.sk

Abstrakt—Tento príspevok predstavuje sumarizáciu dosiahnutých výsledkov počas doktorandského štúdia, ktoré sú detailne spracované v dizertačnej práci *Metódy a prostriedky pre návrh, modelovanie a realizáciu distribuovaných systémov riadenia veľkých fyzikálnych experimentov*. Príspevok prezentuje navrhnutú metodiku pre modelovanie a analýzu distribuovaných systémov riadenia. Cieľom navrhutej metodiky je vytvorenie zložených modelov distribuovaných systémov riadenia, na základe ktorých je možné určiť parametre modelovaných systémov z hľadiska priepustnosti a odozvy, ktoré je možné použiť pre určenie limitov navrhnutého distribuovaného systému riadenia alebo pri optimalizácii štruktúry samotného systému. Navrhnutá metodika využíva koncept hybridných systémov za účelom vytvorenia čiastkových modelov systému s využitím konečno-stavových automatov a farbených časovaných Petriho sietí. Metodika je prezentovaná na dvoch scenároch, zaoberajúcich sa modelovaním a analýzou riadiaceho systému detektorov experimentu ALICE v CERNe a aplikácie mobilnej robotiky v rámci distribuovaného systému riadenia v CMMRaPI na školiacom pracovisku KKUI.

Kľúčové slová—kyber-fyzikálny systém, hybridný systém, konečno-stavový automat, Petriho sieť, distribuovaný systém riadenia, riadiaci systém detektora

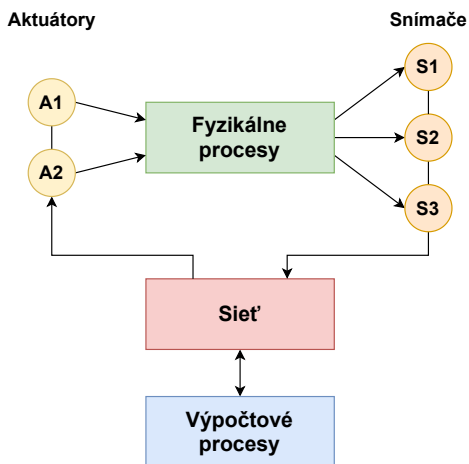
Abstract—This paper presents an overview of the results achieved during the doctoral studies, which are elaborated in detail in the dissertation *Methods and Tools for Design, Modeling and Realisation of Distributed Control Systems of Large Physical Experiments*. The paper describes the developed methodology for modeling and analysis of distributed control systems. The goal of the proposed methodology is to create complex models of distributed control systems, based on which it is possible to determine the parameters of the modeled systems in terms of throughput and response, which can be used to determine the limits of the designed distributed control system or to optimize the structure of the system itself. The proposed methodology uses the concept of hybrid systems in order to create partial models of the system using finite-state automata and colored timed Petri nets. The methodology is presented on two scenarios dealing with the modeling and analysis of the detector control system of the ALICE experiment at CERN and the application of mobile robotics within the distributed control system at CMMT&II at the DCAI department.

Index Terms—Cyber-Physical System, Hybrid system, Finite State Machine, Petri net, Distributed Control System, Detector Control System

I. ÚVOD

Distribuovaný systém riadenia je v priemysle často používaný koncept sieťových riadiacich systémov vyznačujúci sa viacúrovňovou architektúrou, kde jednotlivé úrovne riadenia sú prepojené rôznymi typmi komunikačných sietí [1], [2]. V prípade potreby riadenia a zberu dát z riadeného procesu v reálnom čase je potrebné brať do úvahy obmedzenia vyplývajúce z použitých komunikačných a výpočtových procesov z hľadiska priepustnosti a odozvy, keďže tieto parametre systému môžu mať výrazný vplyv na kvalitu a stabilitu implementovaného riadenia. Odhad týchto parametrov nemusí byť triviálnou úlohou v prípade komplexnejších distribuovaných systémov riadenia s premenlivou veľkosťou prenášaných dát. Za účelom odhadu parametrov priepustnosti a odozvy distribuovaného systému riadenia je tento systém možné považovať za kyber-fyzikálny systém, ktorý je možné modelovať a analyzovať pomocou konceptu hybridných systémov. Výpočtové procesy a komunikačné siete v rámci distribuovaných systémov riadenia je možné považovať za systémy s diskretnými udalosťami, ktoré je vhodné modelovať s využitím konečno-stavových automatov a Petriho sietí [3].

Architektúru distribuovaných systémov riadenia implementuje aj riadiaci systém detektorov DCS (*Detector Control System*) experimentu ALICE (*A Large Ion Collider Experiment*) v CERNe. Riadiaci systém detektorov zaisťuje stabilnú a bezpečnú prevádzku detektorov, pričom zabezpečuje úlohy riadenia, monitorovania a zberu dát z detektorovej elektroniky [4]. Experiment ALICE a jeho detektory prebiehali modernizáciou od roku 2018 do roku 2022, ktorej následkom je aj zvýšenie frekvencie monitorovania detektorovej elektroniky a rádové zväčšenie množstva dát, ktoré je potrebné spracovať a distribuovať v reálnom čase [5]. Takéto zvýšenie požiadaviek na spracovávanie dát v reálnom čase by malo za následok neúnosné zaťaženie SCADA systémov v rámci riadiaceho systému detektorov. Z toho dôvodu bola do riadiaceho systému detektorov experimentu ALICE zaradená nová softvérová vrstva ALFRED (*ALICE Low Level Front End Device*), ktorá



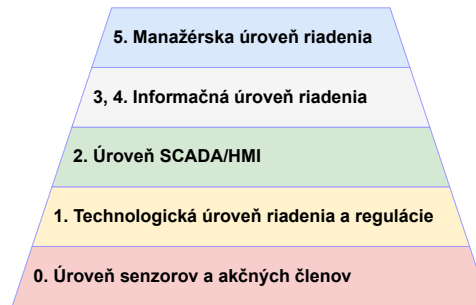
Obr. 1. Schematické znázornenie konceptu kyber-fyzikálneho systému

zabezpečuje spracovanie dát pred ich postúpením SCADA systému, a zároveň vytvára vrstvu abstrakcie pre detektorovú elektroniku z pohľadu SCADA systému [6]. Distribúcia softvérových a hardvérových prostriedkov v rámci riadiaceho systému detektorov ALICE je väčšinou určovaná experimentálne, čo je zdĺhavý a iteratívny proces, z čoho vyplýva potreba modelovania riadiaceho systému detektorov za účelom určenia vhodnej distribúcie použitých prostriedkov.

V tomto príspevku sú distribuované systémy riadenia v kontexte kyber-fyzikálnych systémov vrátane možností ich modelovania predstavené v prvej časti tohto príspevku. Ako nástroje pre modelovanie diskretných udalostí v rámci kyber-fyzikálnych systémov je použitý formalizmus konečno-stavových automatov a Petriho sietí. Ďalšia časť príspevku sa venuje predstaveniu infraštruktúr distribuovaných systémov riadenia uvažovaných v rámci scenárov pre prezentáciu navrhutej metodiky. Popisu navrhutej metodiky pre modelovanie a analýzu distribuovaných systémov sa venuje posledná časť príspevku, v rámci ktorej je predstavená metodika aplikovaná na riadiaci systém detektorov experimentu ALICE a aplikáciu mobilnej robotiky v kontexte distribuovaného systému riadenia v CMMRaPI (*Centrum Moderných Metód Riadenia a Priemyselnej Informatiky*) na KKUI FEI TU v Košiciach.

II. KYBER-FYZIKÁLNE SYSTÉMY

Kyber-fyzikálne systémy (*Cyber-Physical Systems*) sú počítačové systémy, ktorých fyzikálne procesy sú monitorované, riadené a koordinované výpočtovými a komunikačnými prostriedkami. Z hľadiska konceptu je možné kyber-fyzikálne systémy definovať ako celok skladajúci sa zo vzájomne prepojených výpočtových a fyzikálnych procesov [7]. Nevyhnutným prvkom kyber-fyzikálnych systémov sú siete, ktoré zabezpečujú komunikáciu medzi jednotlivými výpočtovými procesmi, akčnými členmi a snímačmi. Konceptnú schému štruktúry kyber-fyzikálneho systému je možné vidieť na Obr. 1. Kyber-fyzikálne systémy sú zároveň podstatnou technológiou používanou v rámci konceptu štvrtej priemyselnej revolúcie *Industry 4.0*.

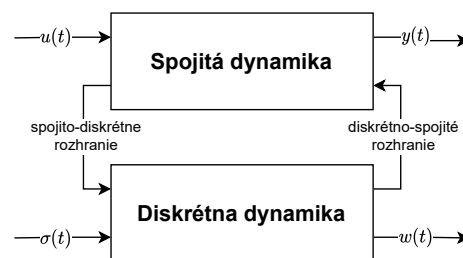


Obr. 2. Viacúrovňová architektúra distribuovaných systémov riadenia

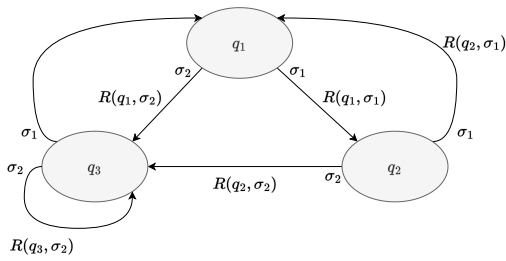
Distribuovaný systém riadenia (DSR) je možné považovať za kyber-fyzikálny systém. Jedná sa o počítačový riadiaci systém pre zložité procesy, v ktorom sú jednotlivé regulačné a koordinačné procesy distribuované v rámci systému bez centrálného uzla riadenia. Na rozdiel od centralizovaného systému riadenia sú jednotlivé riadiace procesy umiestnené bližšie k riadeným procesom, vďaka čomu je riadiaci systém spoľahlivejší a vyznačuje sa nižšími počiatočnými nákladmi pre implementáciu. Nadradené systémy zároveň monitorujú a dohliadajú na jednotlivé podsystemy, čím získavajú komplexný prehľad o stave riadených procesov. Architektúru distribuovaného systému riadenia je možné zaradiť do niekoľkých úrovní riadenia [8], ktoré sú ilustrované na Obr. 2.

Z hľadiska modelovania je možné kyber-fyzikálny systém modelovať s využitím konceptu hybridných systémov [9]. Hybridné systémy sa vyznačujú tým, že ich správanie vykazuje ako spojitú tak aj diskretnú dynamiku. Hybridný systém môže mať jednu alebo viacero spojitých dynamík, ktoré je možné popísať diferenciálnymi rovnicami s uvažovaním spojitého vstupu $u(t)$ a spojitého výstupu $y(t)$, čím je možné vyjadriť správanie sa fyzikálnych procesov v rámci kyber-fyzikálneho systému. Z hľadiska diskretnéj dynamiky hybridných systémov je možné uvažovať diskretný vstup $\sigma(t)$ a diskretný výstup $w(t)$ systému, čím je možné reprezentovať správanie sa výpočtových procesov a komunikačných sietí kyber-fyzikálneho systému [10].

Spojité a diskretné dynamiky hybridného systému navzájom interagujú, ako je možné vidieť na Obr. 3, kde spojitodiskretné rozhranie (*event generator*) transformuje spojitý signál na diskretnú udalosť, ktorá môže mať za následok prepnutie stavu podsystemu diskretnéj dynamiky. Zároveň



Obr. 3. Prepojenie spojitej a diskretnéj dynamiky hybridných systémov



Obr. 4. Príklad grafického zobrazenia konečno-stavového automatu

diskrétno-spojité rozhranie (*injector*) prirad'uje diskrétnemu signálu spojitú hodnotu, ktorá pôsobí ako vstup subsystému spojitej dynamiky [11]. Za účelom modelovania diskkrétnej dynamiky je vhodné použiť konečno-stavové automaty alebo Petriho siete.

A. Modelovanie systémov s diskrétnymi udalosťami

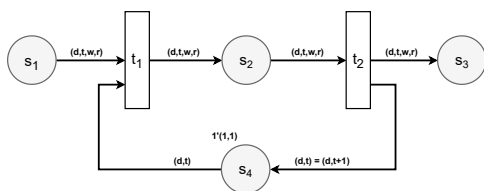
Konečno-stavové automaty (*Finite-State Machines*) je možné považovať za podmnožinu hybridných automatov, keď neberieme do úvahy spojitú dynamiku popisovaného systému a automatom vyjadrujeme len diskrétnu dynamiku modelovaného systému. Konečno-stavový automat je abstraktný matematický model popisujúci systém, ktorý sa v konkrétnom čase nachádza v práve jednom diskrétnom stave z konečnej množiny stavov. Privádzanie vstupov do automatu má za následok prechod medzi jednotlivými stavmi na základe zadefinovanej prechodovej funkcie [12]. Príklad grafického zobrazenia konečno-stavového automatu je možné vidieť na Obr. 4.

Konečno-stavový automat je možné vyjadriť ako kolekciu (1), kde jednotlivé prvky majú nasledujúci význam [12]:

$$M = (Q, \Sigma, Init, R, F) \tag{1}$$

- Q je konečná množina diskrétnych stavov systému nadobúdajúca hodnoty $\{q_1, q_2, \dots, q_m\}$
- Σ je konečná množina diskrétnych vstupov systému nadobúdajúca hodnoty $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$
- $Init \in Q$ je počiatočný stav konečno-stavového automatu
- $R : Q \times \Sigma \rightarrow Q$ určuje funkciu prechodov, teda určuje nový stav q_s na základe predchádzajúceho stavu q_i a vstupu σ_j
- $F \subseteq Q$ je množina koncových stavov automatu (môže byť prázdna)

Petriho sieť je matematický model pre popis systémov s diskrétnymi udalosťami, vďaka čomu sú Petriho siete vhodným kandidátom pre modelovanie distribuovaných systémov



Obr. 5. Príklad grafického zobrazenia farebnej Petriho siete

riadenia z hľadiska toku dát [3], [13]. Petriho sieť je možné opísať ako bipartitný orientovaný graf obsahujúci dva typy uzlov - miesta a prechody. Jednotlivé miesta a prechody sú spojené orientovanými hranami, kde platí, že dve miesta alebo dva prechody nemôžu byť priamo spojené. Teda medzi dvoma rovnakými typmi uzlov sa musí nachádzať druhý typ uzla. Na každé miesto je možné umiestniť nezáporné množstvo *tokenov* (značka, reprezentujúca napríklad prenášané dáta), pričom *tokeny* sa môžu medzi miestami presúvať na základe zadefinovaných pravidiel cez aktivované, resp. odpálené prechody [14].

Formálnu definíciu Petriho sietí je možné vyjadriť ako kolekciu (2), kde jednotlivé prvky majú nasledujúci význam [14]:

$$PN = (S, T, E, V, C, M_0) \tag{2}$$

- S je konečná množina miest nadobúdajúca hodnoty s_1, s_2, \dots, s_m
- T je konečná množina prechodov nadobúdajúca hodnoty t_1, t_2, \dots, t_n , kde platí, že $S \cap T = \emptyset$
- $E \subseteq (S \times T) \cup (T \times S)$ je množina hrán, zjednotenie množín hrán orientovaných od miest k prechodom a od prechodov k miestam
- $V : E \rightarrow \mathbb{N}^*$ je funkciou pre ohodnotenie hrán siete kladnými váhami
- $C : S \rightarrow \mathbb{N}^* \cup \infty$ je funkciou určujúcou maximálnu kapacitu *tokenov* jednotlivých miest
- $M_0 : S \rightarrow \mathbb{N}^0 \cup \infty$ je počiatočné rozmiestnenie *tokenov* v sieti rešpektujúce obmedzenie $M_0(p) \leq C(p)$ pre $\forall p \in S$

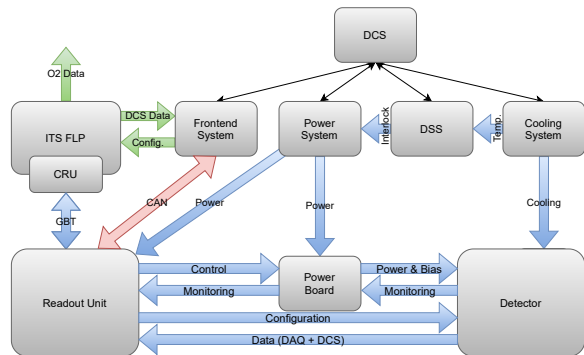
Za účelom rozšírenia možností aplikácie Petriho sietí boli vytvorené viaceré rozšírenia, ako napríklad farbené Petriho siete, kde je možné jednotlivým *tokenom* priradiť prenášané dáta. Ukážku grafického zobrazenia farebnej Petriho siete je možné vidieť na Obr. 5. Ďalším rozšírením Petriho sietí sú časované Petriho siete, kde je možné zadefinovať dobu trvania jednotlivých prechodov, čím je možné modelovať aj dobu trvania vykonávaných procesov. Pomocou týchto rozšírení je možné hodnoverne modelovať správanie sa distribuovaných systémov riadenia z hľadiska toku dát cez komunikačné rozhrania a siete.

III. DISTRIBUOVANÉ SYSTÉMY RIADENIA

Hlavným cieľom tohto príspevku je prezentovať aplikáciu navrhutej metodiky na dvoch infraštruktúrach distribuovaných systémov riadenia. Jedná sa o riadiaci systém detektorov experimentu ALICE v rámci organizácie CERN a o aplikáciu mobilnej robotiky v kontexte distribuovaného systému riadenia v CMMRaPI. Aplikácia navrhutej metodiky na uvažované systémy bude prezentovaná v dvoch scenároch v kapitolách IV-A a IV-B.

A. Infraštruktúra riadiaceho systému detektorov experimentu ALICE

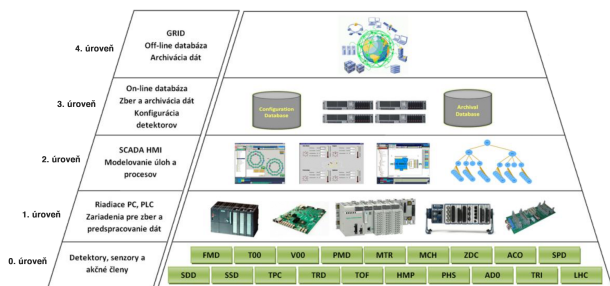
Experiment ALICE je komplex 18-tich detektorov na veľkom hadrónovom urýchl'ovači LHC v CERNe, ktorý je



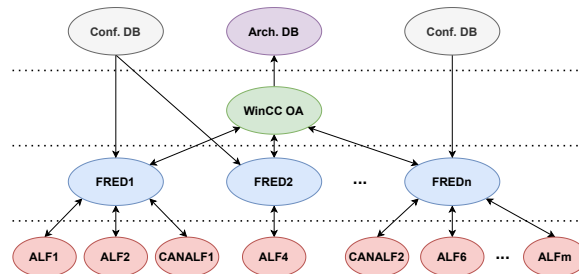
Obr. 6. Architektúra riadiaceho systému detektora ITS

zameraný na štúdium ultrarelativistických zrážok ťažkých iónov [15]. Miesto zrážok častíc v experimente ALICE je obklopené viacerými vrstvami časticových detektorov, ktoré sa nachádzajú vo vnútri L3 magnetu schopného vyvinúť magnetické pole s indukciou 0,5 T. Vytvorené magnetické pole zabezpečuje zakrivenie trajektórií nových častíc vzniknutých pri zrážkach urýchľovaných zväzkov. Experiment ALICE je zameraný na štúdium kvark-gluónovej plazmy vznikajúcej pri zrážkach ťažkých iónov, čo je piaty stav hmoty, ktorý existoval krátko po vzniku vesmíru, kedy sa kvarky a gluóny môžu voľne pohybovať [16]. Výskum priebehu vzniku vesmíru je jedným z cieľov experimentu ALICE.

Riadenie a monitorovanie detektorov experimentu ALICE zabezpečujú riadiace systémy detektorov. Aj keď architektúra riadiacich systémov jednotlivých detektorov využíva rovnaký koncept, ich štruktúry sú mierne odlišné z dôvodu rôznej detektorovej elektroniky jednotlivých detektorov [17]. Ako príklad architektúry riadiaceho systému detektorov je možné uviesť architektúru riadiaceho systému detektora ITS (*Inner Tracking System*), ktorú je možné vidieť na Obr. 6. Štruktúru systému je možné rozdeliť na štyri základné podsystémy: distribuovaný systém ALFRED (*Frontend System*), napájací systém (*Power System*), bezpečnostný systém detektora (*Detector Safety System*) a chladiaci systém (*Cooling System*) [6]. Architektúra riadiaceho systému detektora ITS bola navrhnutá pracovníkmi z CERNU, zatiaľ čo niektoré softvérové moduly boli implementované členmi riešiteľského tímu v rámci projektu *Experiment ALICE na LHC v CERNe: Štúdium silno*



Obr. 7. Zaradenie komponentov DCS systému do konceptu DSR



Obr. 8. Distribuovaná architektúra ALFRED systému

interagujúcej hmoty v extrémnych podmienkach.

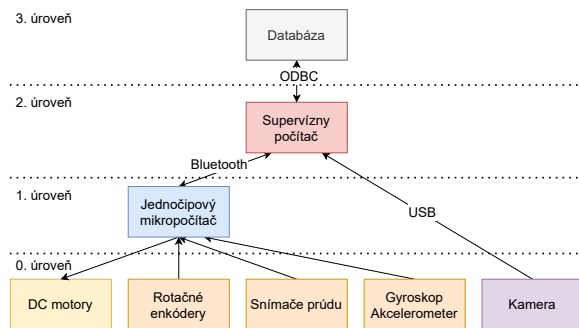
Riadiaci systém detektorov experimentu ALICE je možné zovšeobecniť na viacúrovňovú architektúru distribuovaného systému riadenia predstavenú v kapitole II, pričom jednotlivé komponenty riadiaceho systému detektorov je možné zaradiť do príslušných úrovní riadenia podľa predstavenej architektúry, ako je možné vidieť na Obr. 7.

Zavedený ALFRED systém sa tiež vyznačuje distribuovanou architektúrou, pričom jednotlivé komponenty je možné zaradiť do jednotlivých úrovní architektúry distribuovaného systému riadenia, ako je možné vidieť na Obr. 8 [6]. Najnižšia úroveň ALFRED systému je tvorená detektorovou elektronikou, v prípade detektora ITS konkrétne jednotkami *Readout Unit* a *Power Board*. Jednotky *Readout Unit* zabezpečujú zber dát a ovládanie samotného detektora, pričom s nadradenou úrovňou serverových aplikácií ALF a CANALF komunikujú prostredníctvom gigabitovej optickej linky GBT (*GigaBit Transceiver*), resp. pomocou zbernice CAN bus [18]. Serverové aplikácie ALF a CANALF zabezpečujú preklad správ do formátu protokolu DIM (*Distributed Information Management System*), prostredníctvom ktorého je realizovaná komunikácia so softvérovou vrstvou serverových aplikácií FRED [19].

Serverová aplikácia FRED vytvára abstraktný pohľad na detektorovú elektroniku pre SCADA/HMI systém z hľadiska zjednotenia komunikačných protokolov a formátov správ, zároveň odľahčuje SCADA/HMI systém od výpočtovo náročného prvotného spracovania dát a zabezpečuje riadenie a monitorovanie detektorov na najnižšej úrovni [20]. SCADA/HMI systém WinCC OA zabezpečuje supervízne riadenie detektorov, ktoré je implementované pomocou konečno-stavových automatov, a zároveň poskytuje možnosť ovládania detektorov prostredníctvom operátorských panelov [21]. Najvyššiu úroveň ALFRED systému uzatvára vrstva konfiguračných a archivačných databáz, ktoré poskytujú parametre pre konfiguráciu detektorovej elektroniky a archivujú fyzikálne a technické dáta získané počas behu experimentu [17].

B. Aplikácie mobilnej robotiky v rámci infraštruktúry distribuovaného systému riadenia v CMMRaPI

Množstvo aplikácií mobilnej robotiky sa vyznačuje distribuovanou architektúrou, kde komponenty aplikácie mobilnej robotiky je možné zaradiť do príslušných úrovní distribu-



Obr. 9. Zaradenie aplikácie robotického futbalu do konceptu DSR v CMMRaPI na KKUI

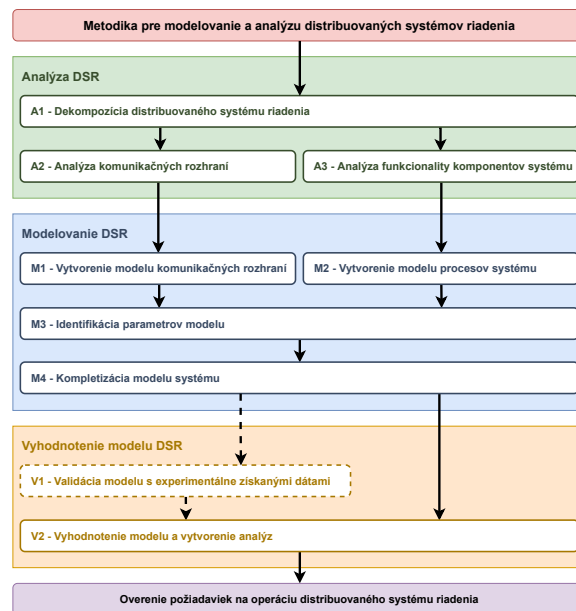
ovaného systému riadenia podľa konceptu predstaveného v kapitole II. Často sa jedná o aplikácie na báze multiagentových systémov, kde je v jednej aplikácii použitých viacero mobilných robotov [22]. V tomto prípade je každý mobilný robot nezávislý funkčný celok zahŕňajúci nižšie úrovne distribuovaného systému riadenia vrátane monitorovania svojho okolia a riadenia vlastného pohybu. Vyššie úrovne riadenia sú spravidla implementované mimo samotných mobilných robotov svyžitím externých počítačov, ktoré môžu zabezpečovať supervízne riadenie, taktické plánovanie alebo zber a archiváciu dát.

V rámci distribuovaného systému riadenia v CMMRaPI je uvažovaná aplikácia robotického futbalu kategórie MiroSot v *Laboratóriu výrobných liniek a rozpoznávania obrazov* [23]. Aplikácia využíva diferenciálne riadené dvojkoľosové mobilné roboty MiroSot a nadradený počítač zabezpečujúci lokalizáciu robotov a taktické plánovanie ich pohybu za účelom realizácie samotnej hry robotického futbalu [24]. Samotnú aplikáciu robotického futbalu je možné považovať za distribuovaný systém riadenia, kde je možné jednotlivé komponenty aplikácie zaradiť do architektúry distribuovaného systému riadenia predstavenej v kapitole II, ako je možné vidieť na Obr. 9.

V kontexte distribuovaného systému riadenia v CMMRaPI je možné uvažovať aj modulárnu robotickú platformu ModBot, ktorá sa vyznačuje vysokou modularitou z hľadiska konfigurácie robotického platformy pre potreby konkrétnych aplikácií. Na rozdiel od mobilných robotov MiroSot je možné platformu ModBot jednoducho rozširovať o ďalšie senzory a akčné členy pomocou viacerých slotov pre prídavné moduly, vďaka čomu je možné platformu ModBot využiť pre širokú škálu distribuovaných aplikácií [25].

IV. NÁVRH METODIKY PRE MODELOVANIE A ANALÝZU DISTRIBUOVANÝCH SYSTÉMOV RIADENIA

Navrhnutá metodika pre modelovanie a analýzu distribuovaných systémov riadenia prezentuje postup pre vytvorenie komplexného modelu distribuovaného systému riadenia a následnú analýzu vlastností modelovaného distribuovaného systému riadenia. Modely vytvorené pomocou tejto metodiky je následne možné implementovať v prostredí *MATLAB Simulink* pomocou nástroja *MATLAB Stateflow*, čím je možné vykonať simulácie pre zistenie vlastností distribuovaného



Obr. 10. Navrhnutá metodika pre modelovanie a analýzu distribuovaných systémov riadenia

systému riadenia na základe privedených vstupov. Metodika pozostáva z troch modulov, z ktorých každý je členený na niekoľko podmodulov, ako je zobrazené na Obr. 10.

Vytvoreniu modelu distribuovaného systému predchádza analýza systému z hľadiska štruktúry, komunikačných rozhraní a funkcionality výpočtových a technických procesov. Tento proces popisuje modul **Analýza DSR** navrhovanej metodiky, ktorý je delený na tri podmoduly.

A1 - Dekompozícia distribuovaného systému riadenia

- zahŕňa dekompozíciu distribuovaného systému riadenia na komunikačné rozhrania a výpočtové procesy, ktoré je možné nezávisle modelovať.

A2 - Analýza komunikačných rozhraní

- zahŕňa analýzu funkcionality komunikačných rozhraní z hľadiska toku dát a princípu ich fungovania.

A3 - Analýza funkcionality komponentov systému

- zahŕňa analýzu funkcionality výpočtových a technických procesov z hľadiska zložitosti spracovávania dát.

Druhý modul navrhovanej metodiky **Modelovanie DSR** je zložený zo štyroch podmodulov a venuje sa vytvoreniu modelov komponentov systému analyzovaných v prvom module. Modely komunikačných sietí sú vytvárané pomocou Petriho sietí a modely výpočtových a technických procesov sú vytvárané vo forme konečno-stavových automatov.

M1 - Vytvorenie modelu komunikačných rozhraní

- zahŕňa vytvorenie modelov komunikačných sietí v tvare farbených časovaných Petriho sietí na základe analýzy funkcionality vykonanej v podmodule A2.

M2 - Vytvorenie modelu procesov systému

- zahŕňa vytvorenie modelov výpočtových a technických procesov v tvare konečno-stavových automatov na základe analýzy funkcionality vykonanej v podmodule A3.

M3 - Identifikácia parametrov modelu - zahŕňa určenie parametrov vytvorených modelov z hľadiska doby trvania prenosu dát, resp. vykonávania podprocesov na základe analýzy vykonanej v podmoduloch **A2**, **A3** a na základe experimentálne získaných dát.

M4 - Kompletizácia modelu systému - zahŕňa vytvorenie komplexného modelu distribuovaného systému zloženého z modelov komunikačných rozhraní a výpočtových procesov a zadefinovanie prepojení jednotlivých modelov na základe analýzy vykonanej v podmodule **A1**.

Posledný tretí modul navrhutej metodiky **Vyhodnotenie modelu DSR** sa venuje validácii vytvoreného modelu a vykonaniu analýz vlastností distribuovaného systému riadenia. Modul je zložený z dvoch podmodulov.

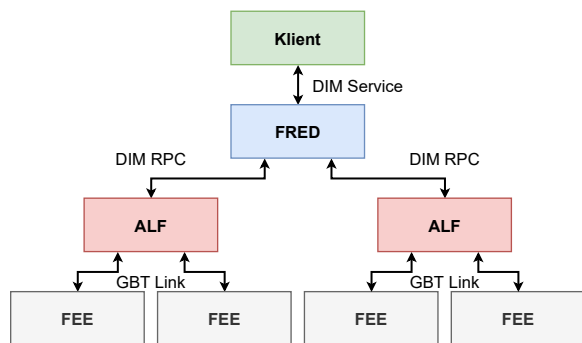
V1 - Validácia modelu s experimentálne získanými dátami - zahŕňa validáciu vytvoreného modelu distribuovaného systému riadenia na základe experimentálne získaných dát, pričom tento podmodul je možné aplikovať v prípade, ak je už distribuovaný systém riadenia aspoň z časti implementovaný.

V2 - Vyhodnotenie modelu a vytvorenie analýz - zahŕňa vytvorenie analýz distribuovaného systému riadenia na základe vytvoreného modelu, napríklad určenie priepustnosti a odozvy systému vzhľadom na rôzne privádzané vstupy.

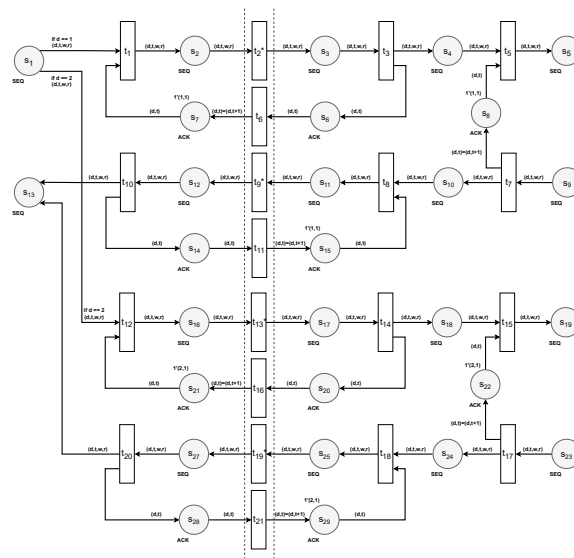
Výstupom metodiky sú analýzy správania sa systému pri rôznych vstupoch, čo je možné použiť pre optimalizáciu štruktúry distribuovaného systému riadenia, určenie jeho limitov alebo modifikáciu navrhnutých algoritmov pre dosiahnutie lepších výsledkov. Navrhnutá metodika je v rámci tohto príspevku prezentovaná v dvoch scenároch.

A. Modelovanie a analýza priepustnosti ALFRED systému

V rámci prvého scenára bola navrhnutá metodika aplikovaná na distribuovaný systém ALFRED. Prezentovaný scenár vychádza z *Prípadovej štúdie 1: Modelovanie a analýza priepustnosti ALFRED linky s využitím detektorovej elektroniky* a *Prípadovej štúdie 2: Modelovanie a analýza priepustnosti paralelného ALFRED systému*, ktoré sú detailne popísané v dizertačnej práci.



Obr. 11. Blokový diagram modelovaného ALFRED systému

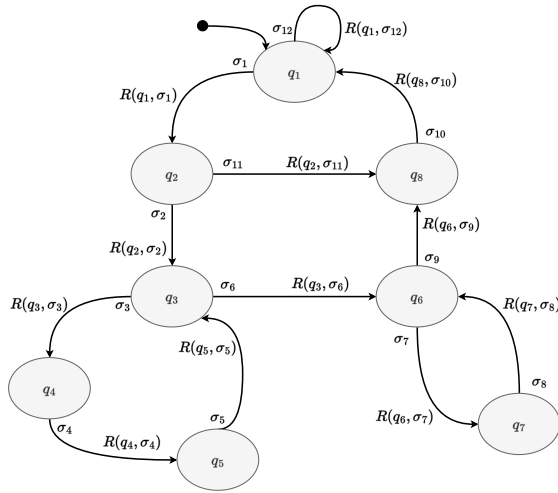


Obr. 12. Model prenosu dát cez DIM RPC rozhranie medzi klientom a dvoma servermi v tvare Petriho siete

V rámci tohto scenára boli uvažované štyri jednotky detektorovej elektroniky (FEE), dve inštancie serverovej aplikácie ALF, jedna inštancia serverovej aplikácie FRED a testovací klient zabezpečujúci generovanie sekvencií príkazov pre detektorovú elektroniku. Pri vytváraní modelu je uvažované aj paralelné spracovávanie dát v aplikácii FRED, ako aj sekvenčné posielanie dát prostredníctvom DIM rozhrania [20]. Blokový diagram modelovaného systému je možné vidieť na Obr. 11, ktorý definuje dekompozíciu modelovaného systému na základe podmodulu **A1** navrhutej metodiky. Model systému je zároveň navrhnutý tak, že je možné ho ľubovoľne horizontálne rozširovať na základe množstva paralelných vetiev modelovaného systému.

Obr. 12 zobrazuje ukázkový model komunikačného rozhrania vytvorený v rámci tohto scenára, ktorý bol vytvorený na základe podmodulov **A2** a **M1** navrhutej metodiky, pomocou formalizmu predstaveného v (2) v kapitole II-A. Model v tvare farebnej časovanej Petriho siete reprezentuje funkcionality komunikačného rozhrania DIM RPC prepájajúceho klientskú aplikáciu s dvoma serverovými aplikáciami. Smerovanie *tokenov* príslušnej serverovej aplikácii je realizované na základe identifikačného čísla cieľového servera uloženého v rámci *tokenu* a obmedzenia priechodnosti hrán na základe farebných vlastností siete. Sekvenčné posielanie dát po sieti medzi klientom a serverom je zabezpečené pomocou spätnej slučky v jednotlivých vetvách modelu. Prechody reprezentujúce prenos dát po sieti majú zároveň priradenú dobu trvania závislú na veľkosti prenášaných paketov, určenú podľa podmodulu **M3** predstavenej metodiky, pričom veľkosť prenášaného paketu je uložená v rámci dátovej štruktúry *tokenu*.

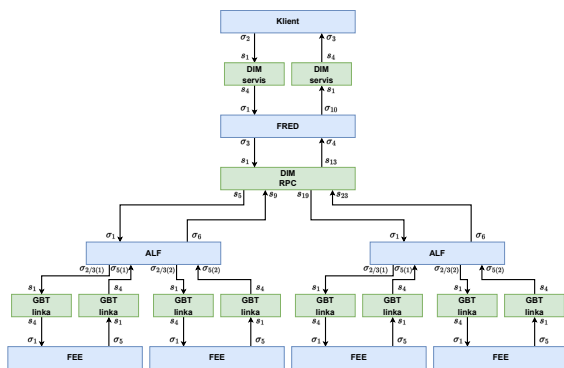
Výpočtové a technické procesy ALFRED systému sú v rámci tohto scenára modelované v tvare konečno-stavových automatov na základe podmodulov **A3** a **M2** navrhutej metodiky. Prvky vytvorených modelov boli určené na základe



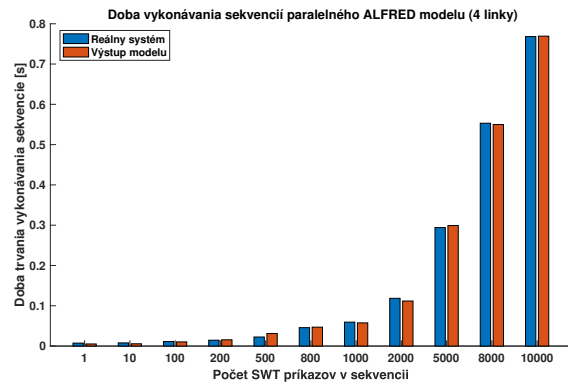
Obr. 13. Model funkcionality komunikačnej fronty aplikácie FRED v tvare konečno-stavového automatu

formulácie predstavenej v (1) v kapitole II-A. Ako príklad je možné uviesť vytvorený model funkcionality komunikačnej fronty serverovej aplikácie FRED, ktorého grafická reprezentácia je zobrazená na Obr. 13. Komunikačná fronta aplikácie FRED zabezpečuje spracovanie dát a komunikáciu s jednou jednotkou detektorovej elektroniky, pričom v rámci aplikácie FRED beží viacero komunikačných frontov paralelne vzhľadom na počet obsluhovaných jednotiek. Konečno-stavový automat komunikačnej fronty aplikácie FRED obsahuje viacero stavov reprezentujúcich rôzne štádiá spracovávania a preposielania sekvencií príkazov, od prijatia žiadosti od nadradeného systému, cez generovanie sekvencií, posielanie žiadostí aplikácii ALF, až po spracovanie odpovedí na sekvencie a vygenerovanie odpovede pre nadradený systém. Príslušné prechodové funkcie medzi stavmi automatu majú zároveň priradenú dobu trvania prechodu podľa podmodulu M3 na základe množstva príkazov v sekvencii, čím je emulovaná doba potrebná na spracovanie dát aplikáciou FRED.

Obr. 14 zobrazuje vytvorené modely komunikačných rozhraní a výpočtových procesov, a ich prepojenie v rámci výsledného modelu ALFRED systému, vytvoreného v rámci



Obr. 14. Prepojenie modelov podsystémov v rámci modelu ALFRED systému

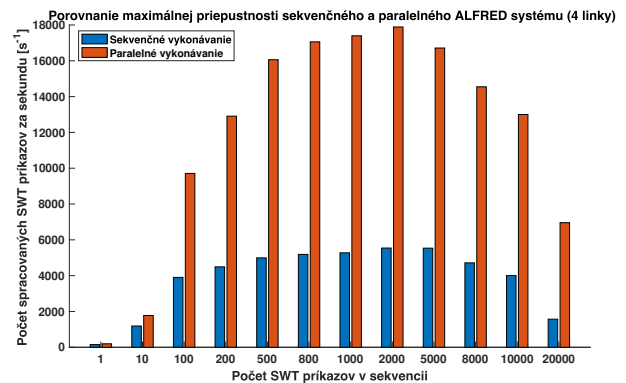


Obr. 15. Porovnanie odozvy modelu paralelného ALFRED systému s experimentálne získanými dátami

podmodulu M4. Prepojenie modelov v tvare Petriho sietí a konečno-stavových automatov je realizované mechanizmom, kde *token* vo výstupnom mieste Petriho siete má za následok aktiváciu vstupu konečno-stavového automatu, a naopak, aktivácia vstupu konečno-stavového automatu má za následok pridanie *tokenu* do vstupného miesta príslušného modelu Petriho siete. Výsledný model bol zároveň implementovaný v prostredí *MATLAB Simulink* pomocou *MATLAB Stateflow* nástroja za účelom vykonania simulácií s rôznymi vstupnými dátami.

Na Obr. 15 je zobrazené porovnanie doby trvania vykonávania sekvencií príkazov pre detektorovú elektroniku na štyroch paralelných linkách na základe simulačného modelu s experimentálne získanými dátami, ktoré bolo vykonané na základe podmodulu V1 predstavenej metodiky. Vytvorený model vykazuje strednú absolútnu percentuálnu odchýlku $MAPE_{ALFRED} = 11,36\%$ a koeficient determinácie $R^2_{ALFRED} = 0,9997$, teda je možné usúdiť, že model správne kopíruje chovanie sa reálneho systému.

Na základe vytvoreného modelu je zároveň možné získať predikciu maximálnej priepustnosti paralelného ALFRED systému podľa podmodulu V2 navrhutej metodiky. Maximálna



Obr. 16. Porovnanie priepustností sekvencného a paralelného ALFRED systému vychádzajúcich zo simulácií nad vytvorenými modelmi

priepustnosť bola počítaná v závislosti od počtu príkazov pre detektorovú elektroniku v rámci jednej sekvencie v rozsahu od 1 do 20 000. Maximálne priepustnosti vychádzajúce z vytvoreného simulačného modelu je možné vidieť na Obr. 16, pričom je zobrazené porovnanie doby trvania paralelného a sekvenčného vykonávania sekvencií na štyroch linkách. Najvyššia priepustnosť systému je dosahovaná pri sekvenciách o veľkosti 2 000 príkazov v hodnote približne 17 900 príkazov za sekundu pre každú linku. Pre porovnanie, pri sekvenčnom vykonávaní príkazov je na štyroch linkách možné dosiahnuť maximálnu priepustnosť približne 5 500 príkazov za sekundu.

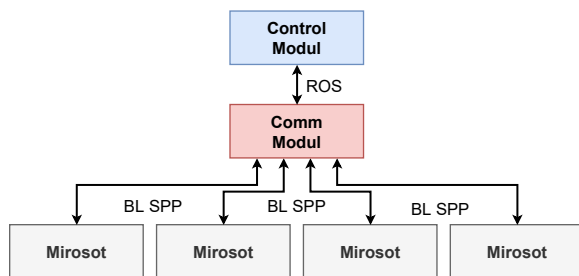
B. Modelovanie a analýza odozvy aplikácie mobilnej robotiky

Druhý scenár je venovaný aplikácii navrhutej metodiky na aplikáciu mobilnej robotiky s využitím mobilných robotov MiroSot v kontexte distribuovaného systému riadenia v CMM-RaPI na KKUI FEI TU v Košiciach. Tento scenár vychádza z *Prípadovej štúdie 3: Modelovanie a analýza odozvy aplikácie mobilnej robotiky v rámci distribuovaného systému riadenia v CMMRaPI na KKUI*, ktorá je detailne popísaná v dizertačnej práci.

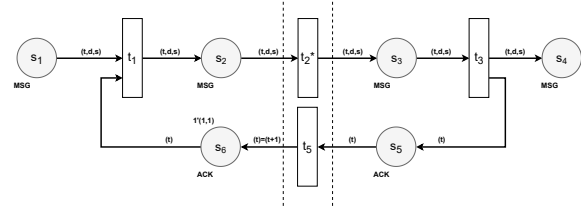
Na základe podmodulu **A1** bola vykonaná dekompozícia modelovaného procesu, kde modelovaná aplikácia je zložená zo štyroch mobilných robotov MiroSot, ktoré komunikujú s komunikačným modulom (Comm Modul) bežiacom na nadradenom počítači prostredníctvom rozhrania *Bluetooth SPP* [24], ako je možné vidieť na Obr. 17. Komunikačný modul zabezpečuje preklad správ medzi rozhraniami ROS a *Bluetooth SPP*, pričom každému mobilnému robotu MiroSot je v rámci komunikačného modulu priradená jedna z paralelne bežiacich komunikačných frônt [26]. Riadiaci modul (Control Modul) zabezpečuje generovanie trajektórií pre mobilných robotov, avšak v zmysle tohto scenára je riadiaci modul považovaný za testovacieho klienta, ktorý generuje správy pre mobilných robotov.

V rámci tohto scenára bola použitá aplikácia robotického futbalu využívajúca mobilné roboty MiroSot v rámci robotického pracoviska v *Laboratóriu výrobných liniek a rozpoznávania obrazov* (<http://kyb.fei.tuke.sk/laboratoria/miest/V147.php>) v CMMRaPI [24], [23].

Na Obr. 18 je zobrazený model prenosu dát cez rozhranie ROS v tvare farebnej časovanej Petriho siete vytvorený na základe podmodulov **A2** a **M1** navrhutej metodiky, ktorý



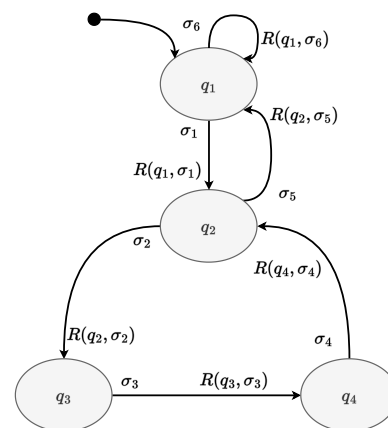
Obr. 17. Blokový diagram modelovanej aplikácie mobilnej robotiky



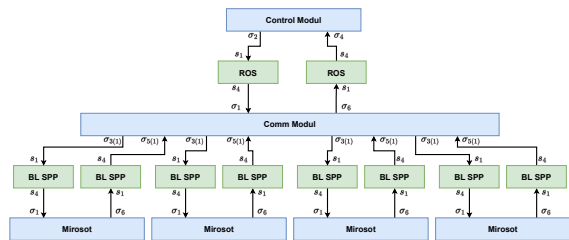
Obr. 18. Model prenosu dát cez rozhranie ROS v tvare Petriho siete

je možné uviesť ako príklad modelu komunikačného rozhrania vytvoreného v rámci tohto scenára pomocou formalizmu uvedeného v (2) v kapitole II-A. Prenášané *tokens* obsahujú informáciu o veľkosti prenášaných dát cez rozhranie, ako aj identifikačné číslo cieľového robota, na základe ktorého sú správy smerované v rámci komunikačného modulu. Spätná slučka v modeli reprezentuje potvrdzovanie prijatých správ, keďže rozhranie ROS využíva protokol TCP na prenos dát po sieti. Prechody reprezentujúce prenos dát po sieti majú zároveň priradenú dobu trvania na základe veľkosti prenášaných dát a experimentálne získaných informácií o závislosti trvania prenosu od veľkosti prenášanej správy na základe podmodulu **M3** predstavenej metodiky.

Ako príklad modelu funkcionality výpočtových procesov v rámci tohto scenára je možné uviesť model funkcionality fronty komunikačného modulu v tvare konečno-stavového automatu zobrazeného na Obr. 19, ktorý je vytvorený na základe podmodulov **A3** a **M2** navrhutej metodiky. Modely výpočtových procesov boli vytvorené v tvare konečno-stavových automatov podľa (1) v kapitole II-A. Stavov konečno-stavového automatu reprezentujú etapy preposielania dát medzi riadiacim modulom a mobilným robotom MiroSot, ako napríklad prijatie žiadosti od riadiaceho modulu, preklad žiadosti, odoslanie žiadosti mobilnému robotovi alebo vygenerovanie odpovede pre nadradený systém. Prechodové funkcie medzi jednotlivými stavmi majú priradenú dobu trvania na základe veľkosti prenášaných správ, pričom závislosť medzi dobou potrebnou na spracovanie správy a veľkosťou prenášaných správ bola



Obr. 19. Model funkcionality fronty komunikačného modulu v tvare konečno-stavového automatu



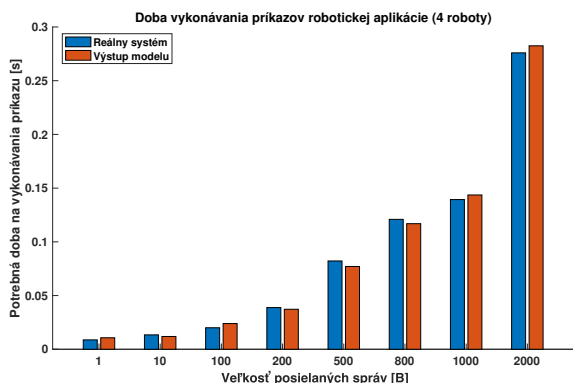
Obr. 20. Prepojenie modelov podsystémov v rámci modelu aplikácie mobilnej robotiky

určená na základe experimentálne získaných dát v rámci podmodulu **M3** navrhutej metodiky.

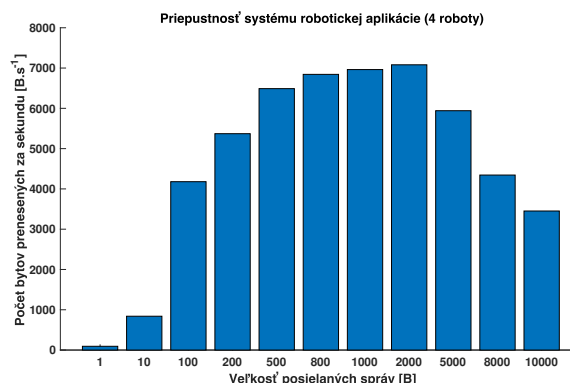
Obr. 20 zobrazuje prepojenie modelov komunikačných rozhraní v tvare Petriho sietí a výpočtových procesov v tvare konečno-stavových automatov, na základe podmodulu **M4**, ktoré je realizované rovnakým mechanizmom ako v predchádzajúcom scenári. Výsledný model bol implementovaný v prostredí *MATLAB Simulink* pomocou *MATLAB Stateflow* nástroja za účelom zistenia vlastností modelovaného distribuovaného systému vzhľadom na rôzne vstupné dáta.

Na Obr. 21 je možné vidieť porovnanie odozvy reálneho systému a výsledkov simulácií nad vytvoreným modelom pri rôzne veľkých správach preposielaných medzi riadiacim modulom a mobilnými robotmi MiroSot podľa podmodulu **V1** predstavenej metodiky. Validácia modelu zahŕňa paralelnú komunikáciu so štyrmi mobilnými robotmi, pričom vo výstupe modelu sú uvažované aj procesy, ktoré sú vykonávané sekvenčne. V rámci experimentu boli použité rôzne veľkosti správ od 1 po 2 000 bytov. Zo štatistického hľadiska vykazuje vytvorený model strednú absolútnu percentuálnu odchýlku $MAPE_{RAPP} = 7,45\%$ a koeficient determinácie $R^2_{RAPP} = 0,9994$, z čoho vyplýva, že model vhodne kopíruje správanie sa reálneho systému robotickej aplikácie.

Vytvorený a implementovaný model je možné na základe podmodulu **V2** použiť aj na určenie maximálnej priepustnosti modelovaného systému za účelom zistenia optimálnej veľkosti prenášaných dát pre dosiahnutie čo najvyššej priepustnosti.



Obr. 21. Porovnanie odozvy modelu robotickej aplikácie s experimentálnymi získanými dátami



Obr. 22. Priepustnosť systému robotickej aplikácie vychádzajúca zo simulácií nad vytvoreným modelom

Ako vstupy modelu boli použité veľkosti posielaných správ od 1 po 10 000 bytov, kde Obr. 22 zobrazuje závislosť maximálnej priepustnosti systému od veľkosti prenášaných správ. Ako je možné vidieť, systém robotickej aplikácie s využitím štyroch jednotiek mobilných robotov MiroSot dosahuje najvyššiu priepustnosť pri posielaní správ o veľkosti 2 000 bytov, kedy je dosahovaná priepustnosť približne 7 000 bytov za sekundu.

Výsledky týchto simulácií je možné použiť pri návrhu zloženia posielaných správ a algoritmov riadenia na supervíznej úrovni pre dosiahnutie čo najvyššej výkonnosti a spoľahlivosti systému. Zároveň je možné posúdiť vhodnosť implementovaného algoritmu riadenia vzhľadom na možnú minimálnu periódu vzorkovania, ktorá vychádza z predikovanej odozvy a priepustnosti modelovaného systému.

V. ZÁVER

Tento príspevok sumarizuje výsledky dosiahnuté v rámci riešenia dizertačnej práce, ktorá bola zameraná na návrh metodiky pre modelovanie a analýzu distribuovaných systémov riadenia. Navrhnutá metodika je zložená z troch modulov, z ktorých každý z nich je členený do niekoľkých podmodulov. Metodika zahŕňa kroky pre analýzu komunikačných rozhraní a výpočtových procesov a prezentuje spôsob vytvárania modelov komponentov distribuovaného systému riadenia v tvare farbených časovaných Petriho sietí a konečno-stavových automatov. Výsledkom aplikovania metodiky sú analýzy budovaného distribuovaného systému, ako napríklad predikcie priepustnosti a odozvy systému vzhľadom na rôzne vstupné faktory. Modely vytvorené aplikáciou tejto metodiky je taktiež možné použiť na optimalizáciu štruktúry alebo funkcionality skúmaného distribuovaného systému riadenia. Pre optimalizáciu štruktúry skúmaného systému je možné použiť rozličné optimalizačné metódy, kde ako príklad je možné uviesť evolučné algoritmy, čo môže byť predmetom ďalšieho výskumu.

Navrhnutá metodika bola demonštrovaná a overená na scenároch venujúcich sa aplikácii metodiky na distribuovaný systém ALFRED, čo je podsystém riadiaceho systému detektorov experimentu ALICE v CERNe a na aplikáciu mobilnej robotiky v kontexte distribuovaného systému riade-

nia v CMMRaPI na školiacom pracovisku KKUI. Vďaka univerzálnosti navrhutej metodiky je možné metodiku aplikovať na distribuované systémy riadenia s rôznou štruktúrou. Vytvorené modely je možné implementovať v prostredí *MATLAB Simulink* pomocou nástroja *MATLAB Stateflow* a vykonaním simulácií nad vytvorenými modelmi je možné predikovať správanie sa systému pri rôznych vstupoch.

V rámci dizertačnej práce boli navrhnuté a implementované serverové aplikácie FRED a CANALF, ako aj viaceré podporné knižnice pre komunikáciu v rámci riadiaceho systému detektorov, ktoré sú už v čase písania tohto príspevku používané viacerými detektormi experimentu ALICE a zabezpečujú riadenie a monitorovanie príslušnej detektorovej elektroniky.

POĎAKOVANIE

Dizertačná práca bola riešená v rámci projektov:

- **Experiment ALICE na LHC v CERNe: Štúdium silno interagujúcej hmoty v extrémnych podmienkach - ALICE KE TUKE 0222/2016 (2016 - 2020)**
- **Experiment ALICE na LHC v CERNe: Štúdium silno interagujúcej hmoty v extrémnych podmienkach - ALICE KE TUKE 0195/2021 (2021)**
- **Experiment ALICE na LHC v CERNe: Štúdium silno interagujúcej hmoty v extrémnych podmienkach - ALICE KE TUKE 0410/2022 (2022 - 2026)**
- **Implementácia výsledkov vedeckého výskumu v oblasti modelovania a simulácie kyber-fyzikálnych systémov do výučby – tvorba moderných vysokoškolských učebníc - KEGA 072TUKE-4/2018 (2018 - 2020)**
- **Modulárne multifunkčné kontrolné pracovisko s využitím techník výpočtovej inteligencie - projekt podporovaný Agentúrou na podporu výskumu a vývoja APVV-19-059 (2019 - 2024)**

ZOZNAM POUŽITEJ LITERATÚRY

- [1] A. Jadlovská, S. Jadlovská, and D. Vošček, "Cyber-physical system implementation into the distributed control system," *IFAC-PapersOnLine*, vol. 49, no. 25, pp. 31–36, 2016.
- [2] J. JADLOVSKÝ, M. ČOPÍK, and P. PAPCUN, "Distribuované systémy riadenia, elfa, košice 2013."
- [3] O. Baldellon and J.-C. F. M. ROY, "Modeling distributed real-time systems using adaptive petri nets," *Saint-Malo, France*, pp. 7–8, 2011.
- [4] A. Augustinus, P. Chochula, L. Jirdén, M. Lechman, P. Rosinsky, O. Pinazza, G. De Cataldo, A. Kurepin, and A. Moreno, "Computing architecture of the ALICE detector control system," in *13. International Conference on Accelerator and Large Experimental Physics Control Systems - ICALEPCS 2011*, 2011, pp. 156–158.
- [5] M. Bernardini and K. Foraz, "Long shutdown 2@ lhc," *CERN Yellow Reports*, vol. 2, no. 00, p. 290, 2016.
- [6] J. Jadlovský, A. Jadlovská, S. Jadlovská, M. Oravec, D. Vošček, M. Kopčík, J. Čabala, M. Tkáčik, P. Chochula, and O. Pinazza, "Communication architecture of the Detector Control System for the Inner Tracking System," *Proceedings, 16th International Conference on Accelerator and Large Experimental Physics Control Systems (ICALEPCS 2017): Barcelona, Spain, October 8-13, 2017*, p. THPHA208, 2018.
- [7] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Design automation conference*. IEEE, 2010, pp. 731–736.
- [8] J. Jadlovský, A. Jadlovská, S. Jadlovská, J. Čerkala, M. Kopčík, J. Čabala, M. Oravec, M. Varga, and D. Vošček, "Research activities of the center of modern control techniques and industrial informatics," in *2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMII)*. IEEE, 2016, pp. 279–285.
- [9] Y. Yalci and Z. Xingshe, "Cyber-physical systems modeling based on extended hybrid automata," in *2013 International Conference on Computational and Information Sciences*. IEEE, 2013, pp. 1871–1874.
- [10] A. J. Van Der Schaft and J. M. Schumacher, *An introduction to hybrid dynamical systems*. Springer London, 2000, vol. 251.
- [11] J. Lunze and F. Lamnabhi-Lagarigue, *Handbook of hybrid systems control: theory, tools, applications*. Cambridge University Press, 2009.
- [12] J. E. Hopcroft, R. Motwani, and J. D. Ullman, "Introduction to automata theory, languages, and computation," *Acm Sigact News*, vol. 32, no. 1, pp. 60–65, 2001.
- [13] M. Jamro, D. Rzonca, and W. Rzaša, "Testing communication tasks in distributed control systems with SysML and Timed Colored Petri Nets model," *Computers in Industry*, vol. 71, pp. 77–87, 2015.
- [14] W. Reisig, *Petri nets: an introduction*. Springer Science & Business Media, 2012, vol. 4.
- [15] K. Aamodt, A. A. Quintana, R. Achenbach, S. Acounis, D. Adamová, C. Adler, M. Aggarwal, F. Agnese, G. A. Rinella, Z. Ahammed *et al.*, "The ALICE experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 002–S08 002, aug 2008. [Online]. Available: <https://doi.org/10.1088/2F1748-0221/2F3%2F08%2Fs08002>
- [16] A. Collaboration, "Performance of the ALICE experiment at the CERN LHC," *International Journal of Modern Physics A*, vol. 29, no. 24, p. 1430044, 2014.
- [17] P. Chochula, L. Jirden, A. Augustinus, G. De Cataldo, C. Torcato, P. Rosinsky, L. Wallet, M. Boccioni, and L. Cardoso, "The ALICE detector control system," *IEEE Transactions on Nuclear Science*, vol. 57, no. 2, pp. 472–478, 2010.
- [18] P. Moreira, R. Ballabriga, S. Baron, S. Bonacini, O. Cobanoglu, F. Faccio, T. Fedorov, R. Francisco, P. Gui, P. Hartin, K. Kloukinas, X. Llopert, A. Marchioro, C. Paillard, N. Pinilla, K. Wyllie, and B. Yu, "The GBT Project," *Topical Workshop on Electronics for Particle Physics*, pp. 342–346, 2009. [Online]. Available: <https://cds.cern.ch/record/1235836>
- [19] C. Gaspar, M. Dönszelmann, and P. Charpentier, "Dim, a portable, light weight package for information publishing, data transfer and inter-process communication," *Computer Physics Communications*, vol. 140, no. 1, pp. 102 – 109, 2001, cHEP2000. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010465501002600>
- [20] M. Tkáčik, J. Jadlovský, S. Jadlovská, L. Koska, A. Jadlovská, and M. Donadoni, "FRED — Flexible Framework for Frontend Electronics Control in ALICE Experiment at CERN," *Processes*, vol. 8, no. 5, p. 565, 2020.
- [21] P. Chochula, A. Augustinus, P. Bond, A. Kurepin, M. Lechman, J. LĀ, O. Pinazza *et al.*, "Challenges of the alice detector control system for the lhc run3," *ICALEPCS Barcelona, Spain*, 2017.
- [22] J. M. Santos, D. Portugal, and R. P. Rocha, "An evaluation of 2D SLAM techniques available in robot operating system," in *2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 2013, pp. 1–6.
- [23] J. Jadlovský and M. Kopčík, "Distributed control system for mobile robots with differential drive," in *2016 Cybernetics & Informatics (K&I)*. IEEE, 2016, pp. 1–5.
- [24] A. Jadlovská, J. Jadlovský, S. Jadlovská, J. Čerkala, M. Kopčík, J. Čabala, M. Oravec, M. Varga, D. Vošček, M. Tkáčik, and A. Březina, "Návrh metodiky pre modelovanie, riadenie, simuláciu a nedeštruktívnu diagnostiku mobilných robotov," *Strojárstvo / Strojrenstvĭ: Engineering Magazine*, vol. XXI, no. 11, pp. 1–9, 2017.
- [25] M. Tkáčik, A. Březina, and S. Jadlovská, "Design of a Prototype for a Modular Mobile Robotic Platform," *IFAC-PapersOnLine*, vol. 52, no. 27, pp. 192–197, 2019.
- [26] M. Tkáčik, "Embedded systémy a ich implementácia v rámci distribuovaných riadiacich a monitorovacích systémov," 2019, diplomová práca (školicel' doc. Ing. Ján Jadlovský, CSc.), *KKUI FEI, Technická univerzita v Košiciach*.

QuoVadis Research @ FEI

Vydavateľ: Technická univerzita v Košiciach Fakulta elektrotechniky a informatiky

Sídlo: Letná 9, 042 00 Košice

Elektronická adresa: <http://quovadis.fei.tuke.sk>

Periodicita: 2-krát ročne

Založenie časopisu: September 2018.

Jazyk: V časopise sú články v anglickom a slovenskom jazyku.

ISSN: 2585-9587