

BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

PROCEEDINGS I OF THE 27TH STUDENT EEICT 2021

ELECTRICAL ENGINEERING, INFORMATION SCIENCE, AND COMMUNICATION TECHNOLOGIES **GENERAL PAPERS**

April 27, 2021 Brno, Czech Republic

> BRNO FACULTY OF ELECTRICAL UNIVERSITY ENGINEERING OF TECHNOLOGY AND COMMUNICATION



Title	Proceedings I of the 27 th Conference STUDENT EEICT 2021
Supervisor	Prof. Vladimír Aubrecht
Editor	Assoc. Prof. Vítězslav Novák
Publisher	Brno University of Technology, Faculty of Electrical Engineering and Communication
Year released	2021
	1 st Edition

Content and language issues are the responsibility of the authors. ISBN 978-80-214-5942-7







Honeywell

ON Semiconductor[®]



ThermoFisher SCIENTIFIC

BRNO FACULTY OF ELECTRICAL UNIVERSITY ENGINEERING OF TECHNOLOGY AND COMMUNICATION



SPONSORS

Silver tier



Honored tier











EXPERT GUIDANCE





MEDIA







Obsah

Předmluva

Středoškolské projekty

Jakub Bláha	
SMART HOME GATEWAY	2
Martin Štyndl, Patrik Kratochvíl, Ondřej Bárta	
REHABILITATION ASSISTANT	7
Jakub Jenáček	
ANALYSES OF STATIC PRESSURE IN OBLIQUE SHOCKWAVE ON THE PROBE TIP	11
Filip Maxa	
TUNING OF THE ANSYS FLUENT SYSTEM FOR THE ANALYSIS OF SUPERSONIC FLIGHT	
OF A 7.62X39 CALIBER PROJECTILE	15
Filip Szkandera	
PROCESSOR PINEAPPLE ONE	20
Filip Touš	
DEFEATING RANSOMWARE BY HOOKING SYSTEM CALLS ON WINDOWS OS	24
Mirek Trchalík, Michal Paluřík	
COOKER WITH THERMOELECTRIC CHARGING	28
Ondřej Kusník	
ADAPTIVE BREAKE LIGHT	33

Bakalářské projekty

B1 – Elektronika a komunikace, Mikroelektronika a technologie

Adam Budáč
COMPARISON OF MODERN VIDEO CODECS FOR 4K AND 8K VIDEO CONTENT
Kristýna Pijáčková
EVALUATION OF CNN AND CLDNN ARCHITECTURES ON RADIO MODULATION DATASETS .42
Matej Klement
RAY TRACING SOFTWARE FOR RADIATION ANALYSIS OF ELECTRONIC COMPONENTS46
Tomáš Stupka
AUTOMATED GARDEN SYSTEM FOR OUTDOORS
Adam Cagáň
A DESIGN AND DEVELOPMENT OF A UNIVERSAL POWER SUPPLY FOR LAPTOPS IN MOTOR
VEHICLES
Tomáš Kotian
DESIGN OF A SEMI-AUTOMATIC POSITIONING SYSTEM FOR MEASURING
OF SEMICONDUCTOR CHIPS
Hana Hálová
STUDY OF SURFACE CHANGES OF NEGATIVE ELECTRODE FOR LEAD-ACID BATTERIES
USING OPERANDO CONFOCAL LASER MICROSCOPY62
B2 – Biomedicínské inženýrství a bioinformatika, Zpracování signálů, obrazu a dat
T- 11 + 17 /

Eniko Vargova	
PPG SIGNAL QUALITY ASSESSMENT AND HEART RATE ESTIMATION	67
Kristýna Heřmánková	
THE EFFECT OF QUALITY TRIMMING ON JOINING PAIRED-END READS IN MICROBIOME	
DATA ANALYSIS	71
Ivan Medynskyi	
ADVANCED ANALYSIS OF MOVING OBJECTS IN THE IMAGE	75
Marek Sicha	
TRAFFIC SIGN CLASSIFICATION USING DEEP LEARNING	79
Táňa Kohúteková	
THE PROCESS OF CT IMAGE CLASSIFIER OPTIMIZATION FOR PEDATRIC SPINE	
SEGMENTATION	83
Libor Matějek	
HUMAN BODY SEGMENTATION USING R-CNN	87

B3 – Komunikační technologie a informační bezpečnost, Kybernetika a automatizace

Vladislav Válek	
IMPLEMENTATION OF AUTO-NEGOTIATION FOR ETHERNET INTERFACES RUNNING	
AT SPEEDS FROM 25 TO 100 GBPS	92
Ondřej Klíčník	
QUANTUM KEY DISTRIBUTION POLYGON	96
Jakub Jedlička	
CUSTOM INTELLECTUAL PROPERTY BLOCK FOR LBLOCK CIPHER	100
Dávid Kováč	
GRAPHIC USER INTERFACE FOR PASIVE OPTICAL NETWORK CONFIGURATION	104
Denis Jalovecky	
COMPUTER NETWORKS BASED ON VYOS AND CISCO IOS OPERATING SYSTEMS	107
Martin Dohnálek	
AUTOMATIC ACQUISITION OF VALUES FROM MEASUREMENT DEVICES WITHOUT	
COMMUNICATION INTERFACE	111

Josef Kolísek	
UAV SWARM COMMUNICATION AND MISSION PLANNING	115
Matúš Halgoš	
WIRELESS TELEMETRIC DATALOGGER	119
Otto Hrubý	
SOFTWARE FRAMEWORK FOR PLC PROGRAM STANDARDIZATION	123
František Rusnák	
MICRO HYDRO POWER PLANT	127

B4 – Silnoproudá elektrotechnika a elektroenergetika, Teoretická elektrotechnika, Fyzika a matematika

Jakub Súkeník	
THERMAL ANALYSIS OF SOLID ROTOR COATED WITH COPPER LAYER	132
David Hysek	
POTENTIAL DEVELOPMENT OF FAST CHARGING STATIONS IN SELECTED AREA	136
Enas Al Halabi	
A SELECTED WATER AREA FOR FLOATING PV POWER INSTALLATION-BASIC	
EVALUATION	140
Ivan Kirsanov	
DESIGN AND CONSTRUCTION OF MEASURING APPARATUS FOR LIGHTNING RESEARCH	144

Magisterské projekty

M1 - Biomedicínské inženýrství a bioinformatika

Radek Němec
A SYSTEM FOR CONTROLLING A COMPUTER PRESENTATION USING GESTURES15
Pavlina Sikorova
RESPIRATORY RATE ESTIMATION FROM ECG AND PPG SIGNALS VIA
CONTINUOUSWAVELET TRANSFORM154
Richard Ředina
ATRIAL FIBRILLATION MODEL
M2 – Elektronika a komunikace
Marek Šimka
LORA-BASED INDOOR LOCALIZATION
Alžbeta Kovaľová
QUANTIFICATION OF TURBULENCE BY THE EQUIVALENT TEMPERATURE GRADIENT16
Tomáš Sekanina
THE OCDM SYSTEM MODEL17
Dominik Krejčíř
2×2 PATCH ANTENNA ARRAY WITH SIW-BASED BEAM-SWITCHING NETWORK
Kristina Youssefová
RF TRANSMITTER CLASSIFICATION BASED ON FRONT-END IMPAIRMENTS
Kristián Levocký
OMNI-DIRECTIONAL ANTENNA FOR 60 GHZ BAND CHAN-NEL MEASURMENT

M3 – Komunikační technologie a informační bezpečnost

Antonín Bohačík	
POLYGON SCENARIOS OF THE CZECH DISTRIBUTION NETWORK	
David Hirš	
PROPOSAL OF CYBER THREAT DETECTOR USING RASPBERRY PI	192
Petr Mašek	
SUITABILITY OF THE LORAWAN TECHNOLOGY FOR MASSIVE MACHINE-TYPE	
COMMUNICATION	196
David Kohout	
IMPACT OF CYBER SECURITY ON DATA VOLUME IN SMART METERING	200
Štefan Krajanec	
PRIVACY-PRESERVING INSTANT MESSAGING APPLICATION	204
Jakub Širjov	
TESTING POLYGON FOR QUANTUM KEY DISTRIBUTION	208

M4 – Kybernetika a automatizace

Jan Novotný	
CONTROL SYSTEM FOR SMALL INJECTION MOLDING MACHINE	213
Tomáš Horeličan	
POSITION MEASUREMENT OF MOVING OBJECTS USING A ROBOTIC TOTAL STATION	218
Alexander Korotynskiy	
APPLIANCE FOR AUTOMATIC DISINFECTION OF INTERIORS	223
Petr Dvorský	
DATA CONCENTRATOR	227
Michal Ružička	
AUTOMATIC TABLE DISTILLATION COLUMN	231

Martin Šimek AUTOMATION OF THE HORTICULTURAL SYSTEM	235
M5 – Mikroelektronika a technologie	
Jan Hrabovský	
MATERIALS FOR BIODEGRADABLE BONES BASED ON Fe	240
Filip Smatlo	
CUPROUS OXIDE AS A SEMICONDUCTOR PHOTOCATALYST USED FOR ORGANIC	0.1.6
POLLUTANTS DEGRADATION	246
DEDECTMANCE OF DOLVAINVLIDENT FLUODIDE CARDON NANOTUDES COMPOSITE	251
PERFORMANCE OF FOLI VINILIDENE FLUORIDE-CARDON NANOI UDES COMPOSITE Detre Denečová	231
DETERMINATION OF THE MOST SUITABLE RATIO OF CATHODE MATERIALS FOR	
THE LITHIUM-SULFUR BATTERY SYSTEM	256
Tatiana Pisarenko	200
PVDF —AN IDEAL CANDIDATE FOR USE IN NANOGENERATORS	260
M6 – Silnoproudá elektrotechnika a elektroenergetika Vítězslav Halašta RADIAL BASIS FUNCTION ASSISTED INTERPOLATION FOR ELECTRICAL MACHINE ANALYSIS	266
DEVELOPMENT OF THE LOW POWER SYNCHRONOUS MACHINE	270
Martin Hemzal	270
SIMULATION OF ELECTROPORATION PROCESS IN STENT OCLUSSION THERAPY	274
Stanislav Navratil	
IMPLEMENTATION OF RES AND ELECTROMOBILITY ON ELECTRICAL PARAMETERS IN	MV
DISTRIBUTION NET	278
ENERGY FLOW ANALYSIS IN LV GRID WITH HIGH LEVEL PENETRATION OF PV INSTALATIONS	282
M7 – Zpracování signálů, obrazu a dat	202
Karolína Staňková	

COMPARISON AND EVALUATION SYSTEM FOR BEAT TRACKING ALGORITHMS	
Tomáš Suchánek	
BEAT TRACKING SYSTEM BASED ON A NEURAL NETWORK	291
Jan Malucha	
PROGRAM FOR ASSISTANCE IN LEARNING ENGLISH PRONUNCIATION	295

Doktorské projekty

D1 – Biomedicínské inženýrství a bioinformatika

Jana Schwarzerová	
OPERON-EXPRESSER: THE INNOVATED GENE EXPRESSION-BASED ALGORITHM FOR	
OPERON STRUCTURES IN-FERENCE	.301
Inna Zumberg	
MONITORING THE PARAMETERS OF CELLS MIGRATING IN PSEUDO-3D EXTRACELLULAR	
MATRIX	.306
Petra Novotna	
MULTIPLE INSTANCE LEARNING FRAMEWORK USED FOR ECG PREMATURE CONTRACTIO	ON
LOCALIZATION	.311

D2 – Komunikační technologie a informační bezpečnost I.

Daniel Paučo
COMPROMITATION OF NETWORK INFRASTRUCTURE THROUGH THE DEFAULT SETTINGS
OF WINDOWS OS
Marek Sikora
ANALYSIS AND DETECTION OF SLOWCOMM AND SLOWNEXT DOS ATTACKS
Michal Látal
VERIFICATION OF TRANSMISSION PARAMETERS OF MULTICORE OPTICAL FIBER
Adrián Tomašov
GPON ATTACKS AND ERRORS CLASSIFICATION
Jiri Jezek
SHARED DETECTION OF CYBERATTACKS ON VOIP EXCHANGES
Petr Dejdar
OPTICAL FIBER INFRASTRUCTURE PROTECTION DEMONSTRATION SYSTEM USING
DISTRIBUTED OPTICAL SENSING
Shujairi Murtadha
REVIEW COMPARISON AND ANALYSIS OF POX AND RYU CONTROLLER SDN
D3 – Komunikační technologie a informační bezpečnost II.
Lukáš Benešl

Lukas Denesi	
POSSIBLE REPLACEMENT OF BULK REMOTE SENSING AT THE POINT OF DELIVERY	353
Martin Rusz	
SMART ENERGY MONITORING SYSTEM BASED ON POWER LINE COMMUNICATION	
TECHNOLOGY	358
Karel Kuchař, Eva Holasová	
INDUSTRIAL NETWORK SECURITY MODULE	363
Ondřej Pospíšil	
IMPACT OF ACTIVE SCANNING ON THE INDUSTRIAL CONTROL NETWORKS	368
Tomas Slunsky	
SECURITY OF WEB APPLICATIONS IN PHP	373
Jan Pospisil	
IMPLEMENTATION OF FSK MODULATED SIGNAL RECEIVER USING SOFTWARE DEFINED	
RADIO	378
Tomas Stodulka	
START BUILDING YOUR OWN CYBER RANGE – CYBER ARENA	383
Shujairi Murtadha	
SOFTWARE DEFINED NETWORK FOR SMART HOME	388

D4 – Kybernetika a automatizace I.

Michal Kozubík
FINITE CONTROL SET BASED ON THE VOLTAGE VECTOR REPRESENTATION OF SWITCHING
STATE
David Michalík
A UNITY BASED INDUSTRIAL PROCESS SIMULATION CONTROLLED VIA VIRTUAL PLC 399
Dominik Friml
DERIVATION AND PRACTICAL COMPARISON OF RECURSIVE LS AND TLS SYSTEM
IDENTIFICATION METHODS
Michal Husák
IN-BED POSTURE CLASSIFICATION410
Simon Bilik
FEATURE SPACE REDUCTION AS DATA PREPROCESSING FOR THE ANOMALY
DETECTION
Tomas Sykora
SMART HOME WIRED ARCHITECTURE

D5 – Kybernetika a automatizace II.

Jakub Krejčí
INFLUENCE OF PIEZOELECTRIC MATERIAL PROPERTIES ON IMPEDANCE CHARACTERISTIC
AND IMPROVEMENTS OF CALIBRATION EQUIPMENT420
Vilem Karsky
COMPARISON OF METHODS FOR IMPULSE RESPONSE COMPUTATION
Ondrej Bostik
SEMI-SUPERVISED APPROACH TO TRAIN CAPTCHA LETTER POSITION DETETOR430
Vlastimil Mancl
MODELING OF MULTIPLE ELECTRODE SYSTEM FOR INHOMOGENEITY DETECTION
Michal Šindelář
CALIBRATION OF PRECISION ROTARY TABLE USING LASER INTERFEROMETER SYSTEM446
Lukas Kratochvila
TECHNIQUES FOR AVOIDING MODEL OVERFITTING ON SMALL DATASET

D6 – Mikroelektronika a technologie, Elektronika a komunikace I.

Iuliia Veselkova	
GEL POLYMER ELECTROLYTES BASED ON ETHYLENE GLYCOL DIMETHACRYLATE AND	
THEIR ELECTRO-CHEMICAL CHARACTERIZATION	458
Jan Smejkal	
XRD ANALYSIS OF CHEMICAL COMPOSITION ON LEAD SURFACE MICROELECTRODES	463
Martin Šedina	
HISTORY OF ELECTROMOBILITY, FUTURE BATTERY TRENDS AND HOWTO	
MEASURE THEM	468

D7 – Mikroelektronika a technologie, Elektronika a komunikace II.

Ondřej Klvač

IN-SITU SEM CHARACTERIZATION OF LITHIUM-ION BATTERY CREATED ON MEMS CHIP	474
Luděk Horák	
STUDY OF DIELECTRIC BEHAVIOR OF POLYMERS USING A COMPLEX ELECTRIC	
MODULUS	479
Radek Zavorka	
MEASUREMENT-BASED 60 GHZ TAPPED-DELAY-LINE CHANNEL MODEL	485

Stanislav Rozum
COMPARISON OF REMOTE AND SELF-POSITIONING APPROACHES FOR INDOOR BLE-BASED
LOCALIZATION

D8 – Silnoproudá elektrotechnika a elektroenergetika I.

Luděk Pelikán	
ISSUE OF MEASURING PARAMETERS OF DIELECTRIC MATERIALS	.496
Carina Beermann	
IMPLEMENTATION OF ENERGY MANAGEMENT IN THE INDUSTRIAL SECTORS	.502
Matěj Vrtal	
ASSESSMENT OF INVESMENTS IN LOCAL DISTRIBUTION SYSTEM USING COST-BENEFIT	
ANALYSIS	.508
Viktor Jurák	
HARDWARE-IN-THE-LOOP SIMULATION OF A DISTRIBUTION NETWORK WITH AN ON-LOA	4D
TAP CHANGER CONTROLLED BY AN INTELLIGENT ELECTRONIC DEVICE VIA DIGITAL	
COMMUNICATION	.513

D9 – Silnoproudá elektrotechnika a elektroenergetika II.

Jakub Piska	
BACK-COMMUTATION AND STICKING REDUCTION METHODS IN DC CONTACTOR	.519
Daniel Pribulla	
DESIGN OF A HIGH-SPEED GENERATOR FOR A HELIUM EXPANSION TURBINE	.524
Petr Klima	
COPPER LOSSES IN PARALLEL WIRES AND LITZ WIRE IN PERMANENT MAGNET	
SYNCHRONNOUS MACHINE	.529
Jakub Kaser	
TOPOLOGY AND BREAKING PROCESS OF HYBRID DC CIRCUIT BREAKER	.534
Tomáš Lažek	
ACCURACY ANALYSIS OF ROTOR FREQUENCY CALCULATION FOR INDUCTION MOTOR	
DRIVE	.539
Marek Toman	
ANALYSIS OF EQUIVALENT THERMAL CONDUCTIVITY OF WINDING USING FEM-BASED	
MODEL	.544
Martin Folprecht	
CONTROL UNIT FOR ELECTROPORATING GENERATOR	.549

D10 – Teoretická elektrotechnika, fyzika a matematika

Petr Raichl	
SIMULATING A UAV SWARM - OVERVIEW AND PROPOSAL	555
Fomáš Hejtmánek	
DESIGN OF THE PULSE GENERATOR FOR TESTING AND VALIDATING PARTIAL DISCHARGE	Ξ
MEASURING SYSTEMS	560

D11 – Zpracování signálů, obrazu a dat

Daniel Kováč	
MULTER DICULAT	

MULTILINGUAL ANALYSIS OF HYPOKINETIC DYSARTHRIA IN PATIENTS WITH	
PARKINSON'S DISEASE	566
Jiří Janoušek	
VISION-BASED AUTONOMOUS NAVIGATION OF AN UNMANNED AERIAL VEHICLE	571

Ondrej Mihálik
DEMYSTIFYING BAND-LIMITED EXTRAPOLATION576
Stepan Miklanek
NEURAL NETWORKS WITH DILATED CONVOLUTIONS FOR SOUND EVENT
RECOGNITION
Tomáš Bravenec
MULTI-CLASS WEATHER CLASSIFICATION FROM SINGLE IMAGES WITH CONVOLUTIONAL
NEURAL NETWORKS ON EMBEDDED HARDWARE586
Pavel Sikora
RIGHT CONVOLUTIONAL NEURAL NETWORK FOR CLASSIFICATION ILLUSTRATIONS
IN ARTWORKS
Ondřej Mokrý and Pavel Záviška
INCONSISTENT AUDIO DECLIPPING PERFORMANCE ENHANCEMENT BASED ON AUDIO
INPAINTING596
Anzhelika Mezina
IMPACT OF LOSS FUNCTION ON MULTI-FRAME SUPER-RESOLUTION
Ondrej Sladok, Alkanan Mohamad
PSEUDO-DIFFERENTIAL FRACTIONAL-ORDER FREQUENCY FILTER



OPENING WORD OF THE DEAN

These Proceedings contain papers presented during the **27**th **annual STUDENT EEICT conference**, held at the Faculty of Electrical Engineering and Communication, Brno University of Technology, on April 27, 2021. Despite the limitations due to COVID-19, the fruitful tradition of joining together creative students and seasoned science or research specialists and industry-based experts was not discontinued, providing again a valuable opportunity to exchange information and experience.

Supervised by the Electrical and Electronic Association of the Czech Republic, the EEICT involves multiple corporate partners, collaborators, and evaluators, whose intensive support is highly appreciated. Importantly, the competitive, motivating features of the conference are associated with a practical impact: In addition to encouraging students to further develop their knowledge, interests, and employability potential, the forum directly offers career opportunities through the affiliated PerFEKT JobFair, a yearly job-related workshop and exhibition complementing the actual EEICT sessions. In this context, the organizers acknowledge the long-term assistance from the Ministry of Education, Youth and Sports of the Czech Republic, which has proved essential for refining the scope and impact of the symposium.

In total, 188 peer-reviewed full papers distributed between 23 sessions were delivered interactively online, before examining boards with industry and academic specialists. The presenting authors exhibited a very high standard of knowledge and communication skills, and the best competitors received prize money and/or small gifts. Our sincere thanks go to the sponsors, experts, students, and collaborators who participated in, contributed to, and made the conference a continued success.

Considering all the efforts and work invested, I hope that the **27th STUDENT EEICT (2021)** has been beneficial for all the participants.

I believe that the inspiration gathered during the event will contribute towards a further rise of open science and research, giving all the attendees a chance to freely discuss their achievements and views.

Prof. Vladimír Aubrecht Dean of the Faculty of Electrical Engineering and Communication

Středoškolské projekty

SMART HOME GATEWAY

Jakub Bláha

Gymnázium Brno, Vídeňská, příspěvková organizace E-mail: jakub.blaha@zak.gvid.cz

Supervised by: Ondřej Krajsa

E-mail: krajsao@feec.vutbr.cz

Abstract: The Home Assistant software is a great open-source way of building an automatized smart home solution. It allows to connect smart devices of different manufacturers using the Zigbee protocol. Home Assistant can be installed on a Raspberry Pi which makes it a cheap solution. It provides a comfortable user interface and plenty of add-ons, which further extend its capabilities.

Keywords: EEICT, Home Assistant, Smart Home, Zigbee, Smart Devices, Automations

1 ÚVOD

Cílem této práce je popsat proces vytvoření chytré domácnosti. Centrální jednotkou chytré domácnosti bude Raspberry Pi s instalací otevřeného softwaru Home Assistant. Pro komunikaci s chytrými zařízeními využijeme protokolů Zigbee a MQTT. Ukážeme si, jak připojit chytrá zařízení a jak vytvořit automatizace.

2 REALIZACE CHYTRÉ DOMÁCNOSTI

Jako centrální jednotku chytré domácnosti využijeme Raspberry Pi se systémem Home Assistant. Home Assistant je také možné nainstalovat na Asus Tinkerboard nebo na počítač, jako je Intel NUC. Raspberry Pi je ale výrazně finančně dostupnější a má obrovskou komunitu. Spotřeba Raspberry Pi 4, do kterého je zapojen Zigbee koordinátor činí zhruba 2.9 W. To činí asi 25 kWh ročně.

Při výběru mezi Home Assistantem a alternativami jako jsou Domoticz a OpenHAB hraje velkou roli uživatelská přívětivost, jednoduchost používání a konfigurace a velikost komunity. Home Assistant je nejnovější z těchto tří řešení. Oproti Domoticz je instalace velmi jednoduchá a na rozdíl od OpenHAB lze téměř všechno nakonfigurovat přes UI. Home Assistant nám společně se Zigbee koordinátorem CC2531 umožní propojit chytrá zařízení od různých výrobců, umožní nám integraci s Google Assistentem, přizpůsobení dashboardu sloužícího k ovládání chytrých zařízení a mnoho dalších věcí. Chytrá domácnost postavená pomocí Home Assistanta může fungovat plně offline. Nikdo tedy nemůže sledovat uživatelská data, jak je zvykem u jiných řešeních (např. Google Home). Skrze komunitní rozšíření umožňuje pravidelné zálohování do služby Google Drive.

2.1 INSTALACE SYSTÉMU HOME ASSISTANT

Systém budeme instalovat na Raspberry Pi 4. Budeme proto potřebovat SD kartu. V dokumentaci Home Assistanta je doporučováno použití karty třídy *Application Performance Class 2*, která je optimalizována pro náhodné čtení a zápis. Konkrétně, její čtecí rychlost se pohybuje kolem 4000 IOPS a její zápisová rychlost kolem 2000 IOPS. Alternativně by se k Raspberry Pi 4 dalo připojit SSD, pokud bychom potřebovali vyšší spolehlivost. SD karta je ale opět mnohem dostupnější. Doporučovaná je SD karta alespoň o kapacitě 32 GB. SD kartu vložíme do počítače a následně použijeme nástroj, který umí na SD karty nahrávat .img soubory. Doporučuji nástroj balenaEtcher. Je jednoduchý a umí soubory nahrávat přímo z URL adres.

Soubory .img nalezneme v dokumentaci Home Assistanta v sekci *Installation*. Stáhneme soubor pro náš specifický model Raspberry Pi, otevřeme nástroj balenaEtcher, vybereme .img soubor a postupujeme podle pokynů. Po skončení procesu kartu z počítače vyjmeme a vložíme do Raspberry Pi.

2.2 PŘIPOJENÍ DO SÍTĚ A DOKONČENÍ INSTALACE

Pro dokončení instalace Home Asststanta je třeba připojit Raspberry Pi k internetu. Nejjednodušší způsob, jak toto udělat, je připojení ethernetovým kabelem k routeru, který je k internetu připojený. Po zapnutí Raspberry Pi se systém pokusí připojit k internetu a stáhnout nejnovější aktualizace. Může to nějakou dobu trvat. Pokud připojíme HDMI displej, systém pravděpodobně na obrazovku bude vypisovat stav aktualizace.

Po skončení aktualizace a dokončení instalace je možné se připojit k uživatelskému rozhraní Home Assistant pomocí jakéhokoliv moderního webového prohlížeče. Osobně rád používám nový Microsoft Edge, protože dokáže instalovat stránky jako aplikace a tedy umožnňuje snadnější a přehlednější přístup k rozhraní.

Adresa URL k webovému rozhraní je <u>http://homeassistant.local:8123</u>. Všimněme si, že rozhraní **neběží** na portu 80. Po připojení se zobrazí výzva k vytvoření nového účtu. Nový účet v systému vytvoříme podle instrukcí na obrazovce a přihlásíme se. Systém prohledá lokální síť, aby našel dostupné integrace. Pokud doma máme nějaká zařízení připojená do sítě, je dost možné, že k nim existují integrace.

2.3 KONFIGURACE ZIGBEE KOORDINÁTORU

Jako Zigbee koordinátor využijeme USB koordinátor CC2531. Jedná se o levné a lehce konfigurovatelné řešení s podporou nahrání vlastního firmwaru. Pro ovládání tohoto hardwaru využijeme rozšíření Zigbee2mqtt a Mosquitto broker. Nejprve ale musíme nahrát podporovaný firmware do Zigbee koordinátoru. Podrobný postup je popsán v dokumentaci Zigbee2mqtt. Po nahrání firmwaru můžeme koordinátor zapojit do jednoho z USB portů na Raspberry Pi.

Přidáme repozitář rozšíření Zigbee2mqtt a rozšíření nainstalujeme. Rozšíření Zigbee2mqtt nakonfigurujeme v několika jednoduchých krocích:

- 1. V nastavení našeho profilu zapneme tzv. *Advanced Mode*, abychom mohli přidat nového uživatele.
- 2. Přejdeme do *Configuration* \rightarrow *Users* a přidáme nového uživatele se jménem a heslem *mqtt*.
- 3. Přejdeme na stránku tohoto rozšíření v Add-on storu, přepneme na záložku *Configuration* a změníme následující položky.

server: 'mqtt://homeassistant', user: mqtt, password: mqtt

Ke správnému fungování Zigbee koordinátoru potřebujeme ještě jedno rozšíření. Má název *Mosquitto broker*. Rozšíření nainstalujeme a následně nakonfigurujeme:

- 1. Přejdeme do *Configuration* \rightarrow *Integrations* a spustíme konfiguraci integrace *MQTT*.
- 2. Zaškrtneme políčko Enable MQTT discovery a klikneme na tlačítko Submit.

Po konfiguraci obou rozšíření restartujeme celý systém.

2.4 PÁROVÁNÍ CHYTRÝCH ZAŘÍZENÍ

Před párováním zařízení je třeba změnit jednu položku v konfiguraci rozšíření *Zigbee2mqtt*. Tato položka otevře síť a povolí připojení nových zařízení.

permit_join: true

Seznam kompatibilních zařízení a podrobné instrukce k párování každého z nich jsou k nalezení v dokumentaci rozšíření *Zigbee2mqtt*. Po přidání zařízení do sítě by se měla položka nastavit zpět na *false*, aby se zamezilo nechtěnému připojování Zigbee zařízení.

Pro příklad si ukážeme, jak spárovat žárovku značky IKEA TRÅDFRI:

- 1. Zapneme žárovku.
- 2. Šestkrát žárovku vypneme a zapneme (celkem 12 stisknutí vypínače).
- 3. Žárovka se začne připojovat a blikat. Jakmile žárovka přestane blikat, je připojena.

Vypínač k žárovce stejné značky se páruje také velmi jednoduše.

- 1. Sundáme kryt vypínače.
- 2. Čtyřikrát rychle po sobě stiskneme párovací tlačítko.
- 3. Světelná dioda na vypínači začne blikat, svítit a jakmile zhasne, vypínač je připojen.

2.5 DASHBOARD

Tato sekce UI se nachází v záložce *Overview*. Skládá se z karet. Karty obsahují informace o stavu zařízení a integrací. Dashboard dovoluje ovládání chytré domácnosti např. z telefonu, i když se uživatel nachází mimo domov (pokud je zapnutá služba *Cloud Assistant*).

Ukážeme si, jak přidat kartu s prvkem pro ovládání jasu žárovky:

- 1. Klikneme na *Overview* \rightarrow : \rightarrow *Edit Dashboard* \rightarrow *Add Card*.
- 2. V záložce By Card vybereme Light.
- 3. Vybereme entitu, u které chceme jas ovládat, vyplníme nepovinné údaje a klikneme na *Save*.

Nyní můžeme ovládat jas žárovky pomocí tohoto UI prvku. Podobně můžeme např. přidat kartu pro zobrazení stavu baterie vypínače.



Obrázek 1: Dashboard v UI Home Assistanta

2.6 VYTVÁŘENÍ AUTOMATIZACÍ

Automatizace umožňují automatické reagování chytré domácnosti na podněty zaznamenané chytrou domácností. Umožňují uživateli komfortnější žití v domácnosti nebo třeba úsporu energie.

Využijeme entitu *Sun* (slunce), která nám umožní reagovat na události západu a východu slunce. Po západu automaticky rozsvítíme žárovku:

- 1. Přejdeme do *Configuration* \rightarrow *Automations* \rightarrow *Add Automation*.
- 2. Automatizaci můžeme pojmenovat a můžeme popsat, k čemu slouží.
- 3. V sekci Triggers, nastavíme Trigger type na Sun, Event nastavíme na Sunset.
- 4. V sekci *Actions* nastavíme *Action type* na *Device* a vybereme naši žárovku. *Action* nastavíme na *Turn on*.

Po západu slunce by se žárovka měla automaticky rozsvítit. Pokud chceme akce automatizace otestovat, můžeme automatizaci manuálně spustit.

Automatizace se dá kromě tlačítkového rozhraní upravovat i jako soubor YAML. U složitějších automatizací je tento způsob upravování přehlednější. Níže je vyobrazen soubor korespondující s výše popsanou automatizací.



Obrázek 2: YAML editor automatizací

3 ZÁVĚR

Po rozšíření chytré domácnosti o více zařízení, například o senzory pohybu, chytrá tlačítka, zásuvky nebo termostaty, lze vymyslet nespočet automatizací. Dále bychom mohli přidat například chytré multimediální stanice, propojit Home Assistant s Google Assistantem nebo třeba zahrnout nějaký bezpečnostní systém. Home Assistant je proměnlivé, lehce konfigurovatelné a celkem levné řešení chytré domácnosti. Má obrovskou komunitu a nabízí mnoho integrací a rozšíření. Určitě je složitější oproti řešením jako je Google Home nebo IKEA Smart home. Pro uživatele s menším množstvím zkušeností s moderními tehcnologiemi nebo pro uživatele, kteří nemají čas se o složitější systém starat, se určitě více vyplatí použít již zmíňená komerční řešení. Pro uživatele, kteří naopak chtějí mít nad svojí domácností plnou kontrolu a zahrnout senzory a složitější automatizace, je určitě Home Assistant řešení, které stojí za zvážení.

Existují dva hlavní konkurující softwary Domoticz a OpenHAB. Jsou ale starší a jejich komunita není tak aktivní. Instalace Domoticz je poněkud složitější. Uživatelské rozhraní Open HABu zase nezahrnuje včechny funkce, které systém nabízí a vyžaduje hodně konfigurace editováním souborů. Home Assistant je v tomto nejvíce uživatelský přívětivý. Snaží se vše udělat jednoduššeji, ale uživatelům, kteří potřebují více kontroly nad systémem, nechává možnost rozsáhlých konfigurací.

PODĚKOVÁNÍ

Tento příspěvek vznikl za finanční podpory JCMM.

Chtěl bych poděkovat panu Onřejovi Krajsovi za velkou ochotu a obětavou pomoc při řešení této práce.

REFERENCE

- [1] ASUS Tinker vs Raspberry Pi [online]. 2019 [cit. 2021-03-20]. Dostupné z: https://www.electromaker.io/blog/article/asus-tinker-board-vs-raspberry-pi
- [2] OpenHab vs Home Assistant vs Domoticz [online]. [cit. 2021-03-20]. Dostupné z: https://www.smarthomeblog.net/openhab-home-assistant-domoticz/
- [3] Home Assistant Documentation [online]. [cit. 2021-03-20]. Dostupné z: <u>https://www.home-assistant.io/docs/</u>
- [4] SD Card Application Performance Class [online]. [cit. 2021-03-20]. Dostupné z: https://www.sdcard.org/developers/sd-standard-overview/application-performance-class/
- [5] Zigbee2mqtt Documentation [online]. [cit. 2021-03-20]. Dostupné z: https://www.zigbee2mqtt.io/
- [6] Raspberry Pi Power Consumtion Benchmarks [online]. [cit. 2021-03-20]. Dostupné z: https://www.pidramble.com/wiki/benchmarks/power-consumption

REHABILITATION ASSISTANT

Martin Štyndl

SPŠ Lanškroun (4)

E-mail: martinstyndl@seznam.cz

Patrik Kratochvíl

SPŠE Pardubice (4) E-mail: Patrik.K.010@seznam.cz

Ondřej Bárta

SPŠ Lanškroun (3)

E-mail: Ondrasek150@gmail.com

Supervised by: Josef Němec

E-mail: Nemec@spslan.cz

Abstract: This work occupies exclusively the method of implementation of the rehabilition aid. The emphasis is placed on finding the optimal solution for automatic rehabilition asistence. Several possible solutions and their advantages or disadvantages are discribed here. The problem of detecting the patient's condition and calculating the expected development is also part of this work.

Keywords: rehabilitation, hand, health service, help, medical science

PODĚKOVÁNÍ

Tento projekt vznikl ve spolupráci se Sabinou Bartošovou ze společnosti IC Kliniky v Brně.

1 ÚVOD

Tento projekt se zaobírá návrhem a tvorbou rehabilitační pomůcky ruky. Pomůcka provede pacienta celou rehabilitací a bude snímat průběh léčby a správnost cviků. Na základě těchto dat pak upravuje obtížnost rehabilitace. Pomůcka detekuje pozici prstů pomocí speciální rukavice a zobrazuje průběh léčby a stav pacienta na displeji.

2 PRŮBĚH REHABILITACE

Rehabilitace probíhá pomocí mačkání míčků. Tímto způsobem je řešeno posilování ruky, procvičení motoriky a trénink různých úchopů s odlišnými velikostmi míčků (viz. Tabulka 1).

Při průběhu rehabilitace pomůcka nabádá pacienta k přemísťování míčků po podložce pro trénink úchopů. Po každém zvednutí míčku, bude pacient vyzván k provedení cviku.

Název pomůcky	Použití pomůcky
Gelový míček (d = 5cm)	3 různé druhy tvrdosti. Pro základní cvičení mačkání míčku.
Molitanový míček (d=5,7,9 cm)	Cvičení s nejmenší obtížností. A trénování různých velikostí úchopu.
Skleněná kulička	Pro trénink extrémně malých úchopů.

Tabulka 1:Popis použitých rehabilitačních pomůcek.

3 REHABILITAČNÍ CVIKY

Pro správný průběh rehabilitace je nutné dodržovat pravidla pro jednotlivé cviky (viz. Tabulka 2). Každý cvik má za úkol procvičit specifickou část ruky.

Název cvičení	Ideální pomůcka	Důvod cvičení	Popis cvičení	
Úchop míčků	Všechny míčky	Trénink úchopů	Pacient dle pokynů chytá a přemisťuje míčky na předem určená místa.	
Klasický stisk	Míček (d=5 cm)	Trénink síly stisku	Pacient standardně mačká míček všemi prsty stejně. Rychlost určuje pomůcka.	
Střední stisk	Molitanový míček (d=9 cm)	Trénink ohybu středního kloubu	Pacient chytne míček tak, aby se míček dotýkal dlaně. Poté mačká pouze středními klouby dle rytmu pomůcky.	
Bazální stisk	Molitanový míček (d=7 cm)	Trénink ohybu bazálního kloubu	Pacient chytne míček tak, aby měl zcela natažené prsty. Poté dle rytmu mačká míček pouze bazálními klouby.	
Jednotlivě	Molitanový míček (d=7 cm)	Trénink ohybu jednotlivých prstů	Pacient čeká na povel pomůcky, kterým prstem má zmáčknout míček. Poté tímto prstem míček zmáčkne.	

 Tabulka 2:
 Popis jednotlivých dostupných rehabilitačních cviků.

4 SNÍMÁNÍ RUKY

Snímání ruky v celém rozsahu pohybu je velmi složité. Pro tento účel stačí jednotlivé články snímat v pohybu flexe (ohyb prstu směrem do dlaně) a extenze (ohyb prstu směrem od dlaně).

Pro snímání ruky je potřeba snímat ohyb jednotlivých kloubů. Snímač ohybu je sestaven ze dvou částí. První část je připevněna ke hřbetu ruky blíže k metakarpální kosti (kost uvnitř dlaně). Druhá část je připevněna ke kosti za snímaný kloub. Tyto části jsou spojeny ramenem, které přenáší ohyb do trimru (proměnný rezistor).



Obrázek 1: Ukázka snímače ohybu kloubu



Obrázek 2: Rukavice s jednu řadou snímačů.

5 HARDWAROVÉ ŘEŠENÍ

Hardwarové řešení používá dvě výpočetní jednotky. HU (Hardwarová jednotka) má za úkol snímat data ze senzorů pomocí A/D převodníku. Tyto data poté odesílá do řídící jednotky CU (výpočetní jednotka). CU slouží pro analýzu dat a poskytování zpětné vazby pro pacienta. Pro snadnou interakci s pomůckou zařízení disponuje interakčním tlačítkem, který slouží pro jednoduché zobrazení nápovědy a přerušení cvičení.

Jako HU je použit mikroprocesor Atmega328p pro snadné připojení A/D převodníku a komunikaci. Pro dostatečný výpočetní výkon je pro jednotku CU použita základní deska z notebooku "Packard bell". Výhodou tohoto rozhodnutí je snadné připojení LCD o vysokém rozlišení a malé rozměry celkového zařízení.

Pro převod z analogového do digitálního signálu byl použit integrovaný obvod MAX11611EEE+. Jedná se o 10bitový, 12kanálový A/D převodník.

Ve snímači ohybu byl použit uhlíkový trimr s rozsahem 0 až 25 KΩ.

Podložka je snímána pomocí kamery Genius FaceCam 1000X v2 s rozlišením 1280*720 px a snímací frekvencí 30 fps.



Obrázek 3: Blokové schéma.

Název	Popis	komunikace
Senzory	10 senzorů pro snímaní polohy kloubů.	Spojitý signál
A/D převodník	Slouží k získaní digitálních hodnot pro následný výpočet.	I ₂ C
HU	Zpracovává hardware data.	I ₂ C + Sériová linka
CU	Výpočet průběhu rehabilitace.	Sériová linka + I ₂ C
Interakční tlačítko	Tlačítko pro ovládání.	Diskrétní signál
Displej	Zobrazovací LCD.	I ₂ C
Kamera	Kamera pro snímání situace na pomůcce.	USB

Tabulka 3:Popis bloků.

6 KONSTRUKCE

Celková konstrukce připomíná písmeno L. Na vodorovné desce jsou umístěny držáky na míčky. Na svislé desce je displej, nad kterým je umístěno interakční tlačítko s kamerou směřující na vodorovnou desku. Základní deska je skryta za displejem.

7 GRAFICKÉ UŽIVATELKÉ ROZHRANÍ

7.1 PŘI PRŮBĚHU CVIKU

Při začátku cviku je zobrazen jeho název, na levé straně pacient uvidí snímek ruky svírající míček v závislosti na pozici pacientovy ruky. Na pravé straně pacient uvidí model ruky, na které se mění barvy jednotlivých článků v závislosti na ohybu jednotlivých kloubů [viz. Obrázek 4]. Při potřebě pacient může zmáčknout interakční tlačítko pro zobrazení nápovědy [viz. Obrázek 5].



Obrázek 4: Nápověda pro cvik.



7.2 ΟVLÁDÁNÍ POMŮCKY

Ovládání pomůcky probíhá díky interakčnímu tlačítku a implementované kameře. Kamera snímá situaci na pomůcce a vyhodnocuje pozici jednotlivých míčku a ruky pacienta. Výběr činnosti v menu pacient uskutečňuje položením ruky na určitý míček. Dlouhým stiskem se lze vrátit do menu. Toto tlačítko je umístěno nahoře pro snadnou dostupnost.

8 ZÁVĚR

Díky spolupráci s IC klinikou v Brně bylo možné vytvořit tento projekt i pro praktické využití. Projekt je nyní ve fázi výroby. Věříme, že v konečné fázi tohoto projektu bude zařízení připraveno pomáhat lidem při rehabilitaci ruky a bude připraveno usnadňovat práci několika zdravotnickým asistentům.

REFERENCE

- [1] Fyzioterapie ruky. AEDITUS. Dostupné z: https://www.masaze-rehabilitace-praha.cz/fyzioterapie-ruky
- [2] Kostra ruky, Klouby ruky. Fakulta tělesné výchovy a sportu Univerzita Karlova. Dostupné z: https://ftvs.cuni.cz/FTVS-1451.html
- [3] ATMEGA328P Datasheet (PDF) ATMEL Corporation. Alldatasheet. Dostupné z: https://pdf1.alldatasheet.com/datasheet-pdf/view/241077/ATMEL/ATMEGA328P.html
- [4] MAX11611 Datasheet(PDF) Maxim integrated. Tme.eu. Dostupné z: https://www.tme.eu/Document/dbdc84a5a5f6554def59e1081f51d744/MAX11606-11611.pdf

ANALYSES OF STATIC PRESSURE IN OBLIQUE SHOCKWAVE ON THE PROBE TIP

Jakub JENÁČEK

Mathias Lerch Gymnasium Brno E-mail: JakubJenacek@seznam.cz

Supervised by: Pavla Šabacká

E-mail: hlavata.pavla@gmail.com

Abstract: This article deals with the debugging analyses of Ansys Fluent for supersonic flow analyzes. It is about analyzes on which the debugging of Ansys Fluent was made for analyzes of supersonic flow circumfluent the tip of the probe compared with the theory of isoentropic onedimensional flow as a basic step for further analyzes in low pressures. The results shown in this paper will further serve as a basis for experimental measuring of given problem.

Keywords: Ansys Fluent, Differentially pumped chamber, Pitot's tube, Shockwave

1 ÚVOD

V současné době je na Ústavu elektrotechnologie FEKT VUT Brno ve spolupráci s UPT AVČR vyrobena experimentální komora, na které budou mimo jiné měřeny rychlostní profily pomocí Pitotových sond, z nichž sonda pro statický tlak je opatřena hrotem, a to z důvodu zajištění šikmé rázové vlny u nadzvukového proudění, u které nevzniká tak velká tlaková ztráta jako u kolmé rázové vlny. Na tvar vlny má vliv závislost úhlu kužele sondy na Machove čísle viz. obrázek 1.

V tomto článku je popsán můj podíl na daném výzkumu při stanovení daného charakteru. Jedná se o analýzy, na kterých bylo provedeno odladění systému Ansys Fluent pro analýzy nadzvukového proudění obtékající hrot sondy srovnáním s teorií isoentropického jednorozměrového proudění, jako výchozí krok pro následné analýzy pro experimentální komoru uzpůsobenou pro proudění v nízkých tlacích.

2 TEORIE ISOENTROPICKÉHO JEDNOROZMĚROVÉHO PROUDĚNÍ

Závislost úhlu kužele sondy na úhlu rázové vlny vychází z teorie isoentropického jednorozměrového proudění. Jako vstupní parametry pro výpočty poměrů hodnot v experimentální komoře, podle vztahů 1-8 [1,2] bylo tlakovými poměry v komorách, průměru a tvaru dýzy dáno Machovo číslo M₁ na 2,58 a úhel kužele sondy na 18°, který zajišťuje tvar šikmé rázové vlny.

$$tg\alpha_{c} = \frac{2(M_{1}\sin\alpha_{s})^{2} - 2}{(\varkappa + 1)M_{1}^{2} - 2(M_{1}\sin\alpha_{s})^{2} + 2}ctg\,\alpha_{s} \tag{1}$$

Změna hodnot hustoty, tlaku, teploty a Machova čísla po průchodu rázovou vlnou je řešena teorií isoentropického jednorozměrného proudění pro kolmou (odtrženou) rázovou vlnu s daným úhlem.

$$M_{1n} = M_1 \sin \alpha_s \tag{2}$$

$$M_2^2 = \frac{2 + (\varkappa - 1)M_{1n}^2}{2\varkappa M_{1n}^2 - (\varkappa - 1)}$$
(3)

$$\frac{T_2}{T_1} = 1 + \frac{2(\varkappa - 1)}{(\varkappa + 1)^2} \cdot \frac{1 + \varkappa M_{1n}^2}{M_{1n}^2} \cdot (M_{1n}^2 - 1)$$
(4)

$$\frac{p_2}{p_1} = 1 + \frac{2\kappa}{\kappa+1} (M_{1n}^2 - 1)$$
(5)

$$\frac{\rho_2}{\rho_1} = \frac{V_1}{V_2} = \frac{(\varkappa + 1)M_{1n}^2}{2 + (\varkappa - 1)M_{1n}^2} \tag{6}$$

$$\frac{p_{02}}{p_{01}} = \left[1 + \frac{2\kappa}{\kappa+1} \left(M_{1n}^2 - 1\right)\right]^{-\frac{1}{\kappa-1}} \left[\frac{(\kappa+1)M_{1n}^2}{2+(\kappa-1)M_{1n}^2}\right]^{\frac{\kappa}{\kappa-1}}$$
(7)

$$\frac{p_{02}}{p_1} = \left[1 + \frac{2\varkappa}{\varkappa+1}(M_{1n}^2 - 1)\right]^{-\frac{1}{\varkappa-1}} \left[\frac{\varkappa+1}{2}M_{1n}^2\right]^{\frac{\varkappa}{\varkappa-1}}$$
(8)

; kde: $\varkappa-$, α_c- úhel kužele sondy, α_s- úhel rázové vlny, $M_{1n}-$ normálová složka Machova čísla, M_2- Machovo číslo za šikmou rázovou vlnou, T_2- Teplota za rázovou vlnou, T_1- teplota před rázovou vlnou, p_2- staticky tlak za rázovou vlnou, p_1- staticky tlak před rázovou vlnou, ρ_2- hustota za rázovou vlnou, ρ_1- hustota před rázovou vlnou, $p_{02}-$ celkový tlak za rázovou vlnou, $p_{01}-$ celkový tlak před rázovou vlnou, p_c- tlak na kuželu, ρ_c- hustota na kuželu, T_c- teplota na kuželu .



Obrázek 1: Závislost úhlu kužele sondy na Machově čísle [3]

Dle výše uvedených vztahů 1-8 pro vstupní parametry M_1 =2.58 a úhel kužele 18° vychází poměry tlaků, teploty a hustoty uvedené v tabulce č. 1.

p ₁ [Atm]	$\frac{p_2}{p_1}$	$\frac{p_c}{p_1}$	ρ ₁ [kgm ⁻³]	$\frac{\rho_2}{\rho_1}$	$\frac{\rho_c}{\rho_1}$	T ₁ [K]	$\frac{T_2}{T_1}$	$\frac{T_c}{T_1}$	p ₀₁ [MPa]	$\frac{p_{02}}{p_{01}}$
1	1,81	2,16	1,18	1,52	1,72	300	1,19	1,25	1,86	0,98

Tabulka 1:Poměry tlaků, teploty a hustoty.

Z daných poměrů byly ze vstupních údajů pro p_1 , ρ_1 , T_1 a p_{o1} vypočteny Teoretické hodnoty, které byly porovnány s hodnotami získanými v systému Ansys Fluent (tabulka 2). Bylo provedeno změření úhlu probíhajícího gradientu tlaku a odpovídá hodnotě úhlu rázové vlny 30.66 stupňů [4, 5].

3 ANALÝZA RÁZOVÉ VLNY V PODMÍNKÁCH ATMOSFERICKÉHO TLAKU

Na začátku byly provedeny matematicko – fyzikální analýzy v systému Ansys Fluent pro variantu s atmosférickým tlakem bez viskozity a s viskozitou plynu (obrázek 3). Pro tyto varianty jsme zvolili rychlost 2.58 Macha, úhel kužele sondy α_c 18° a úhel rázové vlny α_s 30.66° (obrázek 2). Variantu atmosférického tlaku bez viskozity plynu jsme srovnali s teorií fyziky isoentropického jednorozměrového proudění (rovnice 1-8). Výsledky tohoto srovnání lze najít v tabulce 2.

Dále jsme srovnali teoretické hodnoty fyziky isoentropického jednorozměrového proudění a hodnoty získané ze systému Ansys Fluent pro variant s atmosférickým tlakem bez viskozity s hodnotami pro atmosférický tlak s viskozitou, kde se již projevuje vliv tření na kuželu sondy (tabulka 2).

	p ₂ [Pa]	p _c [Pa]	ρ ₂	ρ _c	T ₂ [K]	$T_c[K]$	p ₀₂ [MPa]
			[kgm ⁻³]	[kgm ⁻³]			
Teoretická hodnota	183 264	218 737	1,79	2,03	357,5	376	1,81
Ansys hodnoty bez viskozity	183 555	221 633	1.77	2,03	358,3	376,4	1,82
Ansys hodnoty s viskozitou	191 261	221 782	1.83	1,18	363,3	649,3	1,855

Tabulka 2:Poměry tlaků, teploty a hustoty.

4 VÝSLEDKY A DISKUZE



Obrázek 2: Definované úhly sondy a gradientu



Obrázek 3: a) gradient pro atmosférický tlak bez viskozity plynu, b) gradient pro atmosférický tlak s viskozitou plynu

Veškeré výpočty, které byly v mé práci uvedeny, byly realizovány v systému Ansys Fluent. Dle srovnání gradientu pro atmosférický tlak (obrázek 3) je patrné, že varianta s viskozitou i bez viskozity je téměř stejná, avšak u varianty s viskozitou se již projevuje tření na kuželu sondy. Dané analýzy byly podkladem pro odladění systému Ansys Fluent pro analýzy nadzvukového proudění obtékající hrot sondy, jako výchozí krok pro následné analýzy pro experimentální komoru uzpůsobenou pro proudění v nízkých tlacích.

Jako vhodné nastavení systému Ansys Fluent, které prokázalo souhlasné výsledky s teorií isoentropického jednorozměrového proudění, byla nastaveno implicitní konfigurace solveru Density Based. Dále jsme použili výpočetní schéma AUSM (Advection Upstream Splitting Method), které se prokázalo jako vhodnější pro řešení nadzvukového proudění. Pro diskretizaci bylo použito schéma second order Upwind.

REFERENCE

- [1] Salga, J.; Hoření, B. Tabulky proudění plynu. UNOB, Brno, 1997.
- [2] ŠABACKÁ, Pavla. Analýza nadzvukového proudění v experimentální komoře při vložení tlakových a teplotních sond. Brno, 2020. Dostupné také z: https://www.vutbr.cz/studenti/zavprace/detail/126865. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav elektrotechnologie.
- [3] Dejč, M. J. Technická dynamika plynů. SNTL, 1967
- [4] Moran. M.; Shapiro, H. *Fundamentals of Engineering Thermodynamics*, 3rd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1996.
- [5] Škorpík, J. Proudění plynů a par tryskami, Transformační technologie. Brno, 2006.

TUNING OF THE ANSYS FLUENT SYSTEM FOR THE ANA-LYSIS OF SUPERSONIC FLIGHT OF A 7.62X39 CALIBER PROJECTILE

Filip Maxa

The St. Cyril and Methodius Comprehensive School and Pedagogical High School Brno (3) E-mail: filip1317maxa@gmail.com

Supervised by: Pavla Šabacká

E-mail: hlavata.pavla@gmail.com

Abstract: This article deals with the circumflow of the probe tips in external ballistic and with the comparison of results obtained using Ansys Fluent system with the results obtained from experiment of the given projectile according to the cited literature. Conclusions obtained will serve as a basis for further research in given research area.

Keywords: Shockwave, ANSYS Fluent, Ballistic, 7,62 x 39

1 ÚVOD

Vzhledem k tomu, že na Ústavu elektrotechnologie FEKT VUT Brno probíhá výzkum v oblasti proudění plynů při čerpání komor vakuového systému, jejíž součástí je obtékání hrotů statické sondy Pitotovy trubice nadzvukovou rychlostí, rozšířil se výzkum i na oblast vnější balistiky.

V článku je popsán můj podíl na daném výzkumu při stanovení odporové síly vzduchu působící na letící střelu pomocí systému ANSYS Fluent.

2 VNĚJŠÍ BALISTIKA

Výpočet balistické křivky vychází z celkové síly působící na střelu. Ta se skládá ze dvou složek:

$$F = F_g + F_{od} \tag{1}$$

; kde: Fg je vektor gravitační síly a Fod je vektor odporu vzduchu (Drag Force).

Pro velikost těchto sil platí:

$$F_g = m \cdot g \tag{2}$$

$$F_{od} = \frac{1}{2} \cdot C \cdot S \cdot \rho \cdot v^2 \tag{3}$$

; kde: m je hmotnost střely, C je součinitel odporu prostředí (Drag Coefficient), S je plocha průřezu střely, ρ je hustota vzduchu a v je rychlost střely.

Odporová síla vzduchu má na střelu mnohem větší vliv než gravitační síla. Vlivem přechodu do nadzvukového režimu není odporová síla přímo úměrná rychlosti, ale přechodem do nadzvukové rychlosti se tato úměra skokově zvyšuje.

V první kroku bylo provedeno srovnání odporové síly pro ráži 7,62 x 39 (Obrázek 1) publikovanou v [1] s výsledky matematicko-fyzikální analýzy v systému Ansys Fluent.



Obrázek 1: Odporová síla vzduchu F_{od} pro 7,62 x 39

	Odporová síla [N]			
Rychlost střely [m/s ⁻¹]	Teoretická hodnota	Ansys hodnota		
280	0,4	0,38		
300	0,5	0,53		
350	1,5	1,2		
400	2,3	2,1		
800	6,85	6,1		

Tabulka 1: Vyhodnocení výsledků závislosti rychlosti na odporové síle

Pro analýzy byl namodelován tvar střely z podkladu získaný v Practical Thoughts About Transonic Bullet Stability and Accuracy [2] a následné analýzy byly provedeny v systému Ansys Fluent. Pro řešení bylo zvoleno nastavení řešiče Density Based s implicitní formulací řešení rovnic, který je vhodný k řešení předpokládaných vysokých gradientů tlaků spojených s nadzvukovým prouděním.

Také bylo použito schéma výpočtu AUSM (Advection Upstream Splitting Method), který je výhodnější pro řešení nadzvukového toku. Pro diskretizaci jsme použili schéma s diskretizací druhého řádu.

Výpočty byly provedeny pro rychlost střely: 280, 300, 350, 400, 800 m.s⁻¹. Hodnoty byly úmyslně voleny převážně v okolí rozhraní nadzvukového proudění. Hodnoty vyšly velmi podobné, jen mírně nižší, což může znamenat idealizovaný stav (obrázek 1).





3 VÝSLEDKY A DISKUZE

Mým podílem na výzkumu bylo provedení srovnání rozložení rázových vln zobrazených metodou optického lomu (obrázek 2) publikovaných [2] a rozložení získané pomocí Ansys Fluent (obrázek 3). Čtyři základní tvary rázových vln jsou označeny na obou obrázcích 2 a 3 a jsou pro zvolenou rychlost 780 m.s⁻¹ totožné. Na obrázku 3 jsou ve spodní části obrázku zobrazeny pomocí gradientu tlaku, v horní části v zobrazení tlaku je patrný ostrý přechod hodnoty tlaku vznikající právě na hraně rázové vlny.



Obrázek 3: Ukázka rozložením rázových vln u letící střely pomocí Ansys Fluent.

Byl zkoumán průchod stavových veličin rázovou vlnou. Při srovnání s teorií jednorozměrového proudění došlo ke shodě [4,5,6]. Zde je v grafech uvedena změna veličiny tlaku. V prvním případě jde o oblast mírně nad osou střely, ve druhém jde o dráhu podél střely (obrázek 4).



Obrázek 4: Zkoumané dráhy.

Obrázek 5 popisuje stav proudění na čele střely, kde vzhledem k rychlosti střely 2,4 Mach dochází k silné kompresy vlivem vzniku rázové vlny. Dochází skokově k nárůstu tlaku až na 480 000 Pa. Obrázek 6 popisuje proudění těsně kolem střely na celém úseku rozrušeného proudu. Opět je patrný nárůst tlak na čele rázové vlny a další skokové změny způsobené rázovými vlnami 2 až 4.





Obrázek 5: Statický talk v dráze mírně and osou střely.

Obrázek 6: Statický tlak v dráze podél střely.

Výpočty sloužily jako odladění výpočtu střely zmapované pomocí metody optického lomu a experimentálně zjištěné odporové síly. Následně se budu podílet na dalších výpočtech optimalizující tvar střely pro úpravu čelní rázové vlny, která má vliv na velikost tlaku, která vznikán na čele a významně ovlivňuje odporovou sílu.

REFERENCES

- [1] Balistika. Vnější balistika [online]. Dostupné z: http://www.balistika.cz/vnejsi_teorie.html
- [2] Practical Thoughts About Transonic Bullet Stability and Accuracy. Accurate Shooter [online]. Dostupné z: http://bulletin.accurateshooter.com/tag/m118lr/?fbclid=IwAR1Bdiq6Uh7Vw760 dFtOAFknyZkwz1PU4QOhrc9olRU6JDlu9FxBnbHRZWo
- [3] Let z kulovnice [online]. ČVUT FJFI. Dostupné také z: http://fyzsem.fjfi.cvut.cz/2001-2002/Zima01/doprovod/lzk.pdf
- [4] Salga, J.; Hoření, B. Tabulky proudění plynu. UNOB, Brno, 1997.
- [5] Moran. M.; Shapiro, H. *Fundamentals of Engineering Thermodynamics*, 3rd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1996.
- [6] Dejč, M. J. Technická dynamika plynů. SNTL, 1967.

PROCESSOR PINEAPPLE ONE

Filip Szkandera

Dlouhodobá maturitní práce, Vyšší odborná škola a Střední průmyslová škola elektrotechnická E-mail: filip.szkandera@gmail.com

> Supervised by: Zuzana Veselá E-mail: vesela@spseol.cz

Abstract: This thesis deals with the design, simulation and making of a RISC-V based processor only out of descrete logic components. The final product is a macrocontroller that integrates a processor, program memory, data memory, graphics card and an input-output ports in a tower structure made of nine circuit boards. This thesis also describes a simple shell application programmed in a C language, that runs natively on this device.

Keywords: RISC, RISC-V, CPU, VGA, RV32I

1 ÚVOD

Procesory jsou nezbytnou součástí moderní elektroniky. Najdeme je nejenom v počítačích a mobilních telefonech, ale stále častěji také v malých elektronických zařízeních.

Cílem tohoto projektu je sestavit 32 bitový procesor s architekturou RISC-V [1]. Dal jsem si také za cíl, že v projektu nesmějí být žádné komerční procesory a hradlová pole, ale pouze základní číslicové obvody, jako jsou hradla a paměti. Protože cílem není emulace procesoru na jiné platformě, ale jeho zhotovení.

2 HARDWARE

Procesor je tvořen modulárním systémem skládajícím se z devíti desek plošných spojů o velikosti $(10 \cdot 10)$ cm². Jednotlivé moduly představují dílčí bloky procesoru [2].



Obrázek 1: Blokový diagram procesoru Pineapple ONE.

Tyto desky se poté navrší na sebe a vytvoří věžovitou strukturu. Desky jsou mezi sebou propojeny pouze konektory, díky kterým lze celé zařízení velmi jednoduše rozebrat, nebo vyměnit jakýkoliv z bloků. Všechny bloky jsou tvořeny z celkem 248 integrovaných obvodů. Nejčastěji použitým obvodem je sběrnicový buffer 74HCT245 s celkovým počtem 88 kusů v celém procesoru.

Obrázek 1 zobrazuje blokový diagram procesoru. Na obrázcích 4 a 5 lze vidět kompletní zařízení s víkem i bez.

3 PARAMETRY ZAŘÍZENÍ

Programová paměť procesoru je tvořena čtyřmi EEPROM organizovaných jako 128 K \cdot 8 b s označením 39SF010A. Celková velikost programové paměti je 512 KB. EEPROM čipy jsou umístěny v paticích, což umožňuje rozšíření programové paměti pouhou výměnou EEPROM například za kompatibilní čip 39SF040A organizovaný 512 K \cdot 8 b, datové paměti tedy lze rozšířit až na 2 MB.

Hodinový signál je generován 4 MHz krystalovým oscilátorem, jehož výstup je připojen na čítač, který slouží jako dělička frekvence, jenž je konfigurovatelná pomocí propojek. Výstupní frekvence může nabývat: 250 kHz, 500 kHz, 1 MHz a 2 MHz. Při frekvencích hodinového signálu, které jsou vyšší než 500 kHz již zařízení nepracuje spolehlivě.

Pro vygenerování validního monochromatického VGA výstupu s rozlišením $(200 \cdot 150)$ pixelů je nutný 10MHz hodinový signál [3]. Proto grafická karta obsahuje vlastní oscilátor pracující na kmitočtu 10 MHz.

Celé zařízení je napájeno napětím 5 V, které se připojuje prostřednictvím konektoru USB typu C. Při běžném provozu zařízení odebírá průměrně 450 mA.

4 PROGRAMOVÁNÍ ZAŘÍZENÍ

Pineapple ONE je RISC-V procesor plně kompatibilní s instrukční sadou RV32I, je proto možné využít jakýkoliv z volně dostupných RISC-V kompilátorů. Díky tomuto je možné psát programy v jazyce C, nebo jazyce symbolických adres.

Pro nahrání programu do programové paměti je vytvořen skript v jazyce Python, který pomocí převodníku USB/UART posílá data do programátoru, jenž programuje příslušné EEPROM. Programátor je vyobrazen na obrázku 3.

5 PERIFERIE

Procesor disponuje periferiemi, jako například VGA výstupem nebo dvěma vstupními a dvěma výstupními porty. Můžeme tak například k zařízení připojit externí monitor, PS/2 klávesnici a další moduly. Příslušné konektory pro tyto periferie jsou vyvedeny na přední panel, ten je možné vidět na obrázku 4 a 5. Detailní rozložení konektorů na předním panelu je vyobrazeno na obrázku 2.

6 PINESHELL

Pineshell je program, který byl napsán v programovacím jazyce C a po zkompilování běží na tomto zařízení. Jedná se o interaktivní shell poskytující textové rozhraní, které je zobrazováno na monitoru a uživatel s ním může komunikovat prostřednictvím klávesnice. Díky tomuto programu může uživatel měnit obsah datové paměti (viz. příkaz poke), nemůže však modifikovat program v programové paměti. Velikost tohoto programu je 3730 bytů. Obrázek 9 zachycuje textové rozhraní programu pineshell. Seznam všech příkazů v programu Pineshell:


Obrázek 2: Přední panel.

Obrázek 3: Programátor.

Příkaz HELLO a příkaz HI Slouží k demonstraci textového rozhraní, výstupem příkazu HELLO bude text "HI" a obdobně u příkazu HI bude výstupem text "HELLO".

Příkaz PEEK <ADDRESS> Vrátí hodnotu, která je uložena v datové paměti na adrese <ADD-RESS>. Všechny číselné hodnoty jsou v hexadecimální soustavě.

Příkaz POKE <ADDRESS> <DATA> Přepíše hodnotu v datové paměti na adrese <ADDRESS> hodnotou <DATA>. Všechny číselné hodnoty jsou v hexadecimální soustavě.

Příkaz SYSTEM INFORMATION Vypíše informace o konfiguraci procesoru. Tyto informace budou vypsány v novém okně, stisknutím klávesy "ESCAPE" se okno skryje.

Příkaz RUN SNAKE Spustí v novém okně hru s názvem "had". Cílem této hry je nasbírat co nejvíce otazníků po herním poli a tím vytvořit co nejdelšího hada. Had se po herním poli pohybuje automaticky a směr mu udává hráč stisknutím příslušné šipky na klávesnici. Po kontaktu hada s otazníkem se délka hada prodlouží o jedno políčko a vygeneruje se nová pozice otazníku. Pro ukončení hry je třeba stisknout klávesu "ESCAPE".

Příkaz CLEAR Vyčistí terminálové okno.



Obrázek 4: Zařízení s víkem.



Obrázek 5: Zařízení bez víka.



Obrázek 6: Modul s LED diodami.



Obrázek 8: Modul pro komunikaci s klávesnicí.



Obrázek 7: Modul s tlačítky.

PINESH > HI	IELL	V6	
HELLO			
HELLO)		
ні			
> PEEK	0000	00000	
88			
> POKE	0000)000A	AB
DONE			

Obrázek 9: Ukázka softwaru Pineshell.

7 ZÁVĚR

Výsledkem této práce je realizace procesoru RV32I pouze za použití základních logických hradel a pamětí. Výsledný procesor pracuje korektně až do frekvence 500 kHz. Dále jsem pro procesor vytvořil několik testovacích modulů, jenž se připojují ke GPIO portům, mezi tyto moduly patří zejména: LED modul, tlačítkový modul a PS/2 dekodér. Zařízení rovněž umožňuje výstup na monitor díky integrované grafické kartě pomocí VGA rozhraní. Pro demonstrační účely jsem napsal i jednoduchý shell, který umožňuje základní interakci s uživatelem pomocí klávesnice.

V budoucnosti plánuji rozšířit zařízení o další periferie, zejména zvukovou kartu. V dalších iteracích plánuji otestovat i zřetězené zpracování instrukcí.

PODĚKOVÁNÍ

Chtěl bych poděkovat Ing. Janu Vykydalovi za cenné rady, odborný dohled a konzultace, díky kterým jsem tuto práci dovedl do zdárného konce. Dále bych chtěl poděkovat Ing. Zuzaně Veselé, jako vedoucí práce.

REFERENCE

- Waterman, A.; Lee, Y.; Patterson, D. A.; aj.: The RISC-V Instruction Set Manual, Volume I: User-Level ISA, Version 2.0. Technická Zpráva UCB/EECS-2014-54, EECS Department, University of California, Berkeley, Květen 2014. URL http://www2.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-54.html
- [2] Patterson, D. A.; Hennessy, J. L.: Computer Organization and Design RISC-VEdition: The Hardware Software Interface. Oxford, England: MorganKaufmann, 2017.
- [3] Eater, B.: Let's build a video card! [online]. Červenec 2019, [Online; cit.2021-02-22]. URL https://eater.net/vga

DEFEATING RANSOMWARE BY HOOKING SYSTEM CALLS ON WINDOWS OS

Filip Touš

Bohumil Hrabal Grammar School Nymburk (4)

E-mail: smfitous@gmail.com

Abstract: This paper explains why ransomware needs to use the Windows API to encrypt files and how this can be utilized to protect sensitive data from ransomware. Critical API functions are examined on a low level and a generic method to monitor and possibly block their usage through system call hooks is presented. This approach is then demonstrated with a custom kernel mode driver which can keep protected files safe from any user mode malware. It is then compared to current ransomware protection in Windows 10.

Keywords: ransomware, Windows API, system call, hooking

1 ÚVOD

Ransomware je typ malwaru, který na napadených zařízeních blokuje přístup k souborům jejich zašifrováním nebo celému počítači (například přepsáním master boot record sektoru na disku). Po infikování a šifrování je uživateli zobrazena zpráva o výkupném, které obvykle musí být zaplaceno ve virtuální měně jako Bitcoin nebo Ethereum. Klasický krypto ransomware (tedy ten, který šifruje jednotlivé soubory) sleduje typický vzorec chování:

- 1. infekce PC,
- 2. nalezení a šifrování souborů,
- 3. zobrazení zprávy požadující zaplacení pro dešifrování dat.

Tento druh malwaru se na OS Windows neobejde bez použití Windows API pro provedení zmíněných kroků [1]. To dokládá i analýza několika známých vzorků ransowmaru (WannaCry, CTB-Locker, Revenge a další) v online nástroji VirusTotal. Díky tomu byly v minulosti navrženy mechanismy pro detekování ransomwaru statickou nebo behaviorální analýzou volání specifických API funkcí [2] [3]. Cílem softwaru navrženého v rámci této práce je však kompletně ransomwaru zablokovat přístup k citlivým souborům a zastavit tak ransomware dříve, než je schopen napáchat škody.

2 WINDOWS API

Funkce pro I/O a manipulaci se soubory jsou součástí Win32 API. Téměř všechny vyžadují poskytnutí handlu na soubor jako parametr, a to s adekvátními právy (např. WriteFile vyžaduje vytvoření handlu s parametrem dwDesiredAccess "GENERIC_WRITE"). Výjimky jsou uvedeny dále v textu. Handle k souboru vrací funkce CreateFile, alternativně starší OpenFile. Přesněji se jedná o funkce s příponou -W. CreateFileA (druhá z používaných přípon) jen konvertuje cestu k souboru z ANSI do Unicode a volá CreateFileW, podobně funguje mnoho funkcí Win32 API.

Převzetím kontroly nad těmito funkcemi (CreateFileW a OpenFileW) by teoreticky šlo libovolně blokovat přístup k souborům a chránit je tak před ransomwarem. Tyto vysokoúrovňové funkce jsou ale ve skutečnosti jen wrappery pro funkce Native API s předponou Nt (NtCreateFile, NtOpenFile). K těm často neexistuje oficiální dokumentace. Ačkoliv jsou v systémové hierarchii o úroveň níže oproti Win32 API, ani hookování těchto funkcí není spolehlivé, neboť stále běží v uživatelském prostředí (ring 3) a tyto hooky lze obcházet. Z tohoto důvodu je nutné převzít kontrolu nad voláním API funkcí až na kernelové úrovni (ring 0), k čemuž je na Windows zapotřebí vytvořit ovladač.

Software navržený v rámci práce funkce hookuje až na úrovni systémových volání (system call) ve Windows. Jedná se o rozhraní umožňující aplikacím používat funkce operačního systému, klasicky skrze výše zmíněná API. V praxi funkce jako NtCreateFile vyvolají system call pomocí softwarového přerušení (v x86-64 instrukce syscall). Následně je dočasně zvýšena úroveň oprávnění CPU a kontrola je předána na předem definované místo v kernelu (konkrétně v ntoskrnl.exe, ve Správci úloh se jedná o proces "System"). Před vykonáním syscall instrukce je do EAX registru uložen index systémového volání / funkce OS, která má být volána a parametry jsou uloženy na zásobník. Po splnění požadavku a návratu z funkce OS je úroveň oprávnění CPU nastavena zpět na uživatelské prostředí. Celý proces ilustruje obrázek 1:



Obrázek 1: Cesta volání z CreateFileA do kernelu

Jak bylo zmíněno výše, některé API funkce na první pohled nevyžadují vytvoření handlu přes CreateFile nebo OpenFile, například CopyFile, DeleteFile, MoveFile, ReplaceFile nebo SetFileAttributes, protože jejich signatura neobsahuje handle parametr. Analýza takovýchto funkcí v disassembleru ale odhaluje, že odpovědnost za vytvoření handlu řeší sami a ve výsledku vždy dojde k systémovému volání NtCreateFile nebo NtOpenFile před provedením změn. Někdy skrze funkce vyšší úrovně (například CopyFileW volá CreateFileW skrze BasepCopyFileExW), v některých případech dochází rovnou k volání nativních funkcí (DeleteFileW volá NtOpenFile). To platí pro Win32 i Native API.

Jedinou skutečnou výjimkou je nativní funkce NtDeleteFile, ke které Microsoft ani neposkytuje kompletní informace v PDB. Nutno podotknout, že funkce vyšší úrovně DeleteFile nevolá NtDeleteFile, soubory jsou místo toho označeny k vymazání pomocí NtSetInformationFile. NtDeleteFile umožňuje soubor vymazat ihned (nečeká na uzavření všech handlů k souboru) a podle statické analýzy nevolá NtCreateFile nebo NtOpenFile i přesto, že jako parametr akceptuje i pouhou cestu k souboru. Některý ransomware manipuluje se soubory přímo, jindy je vytvořena kopie, která je poté zašifrována a původní soubor smazán. Navržený software se tedy zajímá celkem o tři systémová volání:

- NtCreateFile
- NtOpenFile
- NtDeleteFile

3 BLOKOVÁNÍ PŘÍSTUPU K SOUBORŮM

K hookování těchto systémových volání byla použita knihovna libfinityhook (součást InfinityHook). Software je sestaven jako primitivní ovladač (od Windows 10 verze 1903 takto musí být sestaveny všechny ovladače, které nejsou vázány na specifický hardware). Když dojde k systémovému volání, volá vlastní callback funkci, ve které je možné nahradit adresu cílené funkce v ntoskrnl.exe vlastní hook funkcí ještě před tím, než ji OS volá. Pokud adresa odpovídá nt!NtCreateFile, software ji nahradí adresou své vlastní funkce HookNtCreateFile (analogicky hookuje i další dvě funkce). Prvotní funkce ovladače (DriverEntry) získává adresy funkcí v ntoskrnl.exe pomocí MmGetSystemRoutine-Address a inicializuje callback funkci.

Uvnitř hook funkcí navržený software rozhoduje o zablokování nebo povolení požadavku. Pro testování jednoduše blokuje přístup ke všem souborům s příponou "protected". Alternativně je možné rozhodovat na základě umístění souboru pro ochranu celých složek. Tyto informace funkce obdrží v parametru ObjectAttributes. Software však není limitovaný jen funkcí samotnou a může volat jiné funkce, což umožňuje implementovat komplexnější nastavení ochrany. V případě, že je zachycena snaha o vytvoření handlu k chráněnému souboru (nebo jeho smazání), je kompletně popřena existence souboru (vrácená hodnota STATUS_NO_SUCH_FILE). Pokud nedojde k zablokování požadavku, software volá původní, pravou funkci v ntoskrnl.exe. Celý proces popisuje obrázek 2. Software pro ukázku blokuje i žádosti o read-only přístup, neboť i krádež dat představuje riziko pro firmy a možný peněžní zisk pro hackery [4]. Je ale možné jen manipulovat s parametrem DesiredAccess a zamezit pouze přepisování nebo mazání souboru a čtení vždy povolovat.



Obrázek 2: Vývojový diagram ovladače

Po načtení vytvořeného ovladače je jakákoliv manipulace z uživatelského prostředí s chráněnými soubory zablokována. Software byl otestován na Windows 10 s vypnutým Windows Defenderem proti ransomwaru WannaCry. I po zobrazení zprávy o výkupném zůstaly soubory obsahující v názvu řetězec "protected" nepoškozené (pro vynucení šifrování od WannaCry byla podmínka v ovladači upravena, aby skutečná přípona souborů mohla být ".docx").

Navržená metoda ochrany je na rozdíl od behaviorální detekce podezřelého používání API funkcí výpočetně i implementačně nenáročná a je 100% úspěšná při ochraně souborů před jakýmkoliv malwarem. Metoda v [3] průměrně ztratila 10 souborů před detekováním ransomwaru. Implementace v [2] hookovala API funkce přímo v paměti procesu v uživatelském prostředí. Tyto hooky lze odstranit a případně i kompletně obejít manuálním systémovým voláním pomocí assembly kódu. V obou případech analýza sekvencí API funkcí a chování programů spotřebovávala mnoho výpočetního času. V [3] došlo ke zvýšení latence operací se soubory v řádu milisekund. Metoda navržená v této práci nezpůsobuje žádné měřitelné zpomalení při tvorbě handlů a ostatní funkce nejsou ovlivněny (kromě NtDeleteFile, kterou by legitimní software používat neměl).

4 POROVNÁNÍ S CONTROLLED FOLDER ACCESS

Controlled folder access ("Řízený přístup ke složkám", dále jen CFA) je vestavěná součást protekce před viry ve Windows 10. Jejím účelem je chránit data před ransomwarem nebo jinými hrozbami. Uživatel může specifikovat složky, nebo celé disky, ke kterým CFA zablokuje přístup všem nepovoleným aplikacím.

CFA má však několik zásadních problémů, tím prvním je fakt, že uživatel nemůže vypnout ochranu výchozích složek. Konkrétně se jedná například o složky "Dokumenty" všech uživatelů. Jelikož spousta legitimních aplikací stále ukládá data do těchto míst, nemá uživatel jinou možnost než povolovat tyto aplikace a tím jim udělovat kompletní přístup ke všem chráněným složkám a diskům.

Některé systémové aplikace / procesy jsou povoleny vždy, Windows je označuje jako "přátelské aplikace". Jednou z nich je explorer.exe (Průzkumník Windows), aby uživatel nezablokoval přístup ke složkám "sám sobě". Mezi přátelské aplikace ale patří například i notepad.exe (Poznámkový blok). V grafickém rozhraní pro povolení aplikace je také napsáno, že "Řízený přístup ke složkám bude povolovat většinu aplikací, aniž byste je sem museli přidat". Uživatel tak ani nad tímto nemá absolutní kontrolu.

Všechny povolené aplikace jsou zneužitelné k získání přístupu ke všem chráněným složkám. Útočník má několik možností:

- Provádět změny přes již spuštěný (a povolený) proces, například vytvořením nového vlákna nebo pomocí thread hijackingu.
- Skrytě spustit povolenou aplikaci a ihned nad procesem převzít kontrolu.
- Nahradit .exe soubor svým vlastním (pokud není také chráněn) a soubor spustit.

Všemi třemi metodami se podařilo ochranu obejít. CFA nekontroluje, zda bylo se souborem nebo procesem manipulováno. Stačí, aby kompletní cesta (včetně názvu) ke spouštěnému souboru byla v seznamu povolených aplikací. Ten je navíc volně přístupný v registru jako názvy hodnot specifického klíče, k jehož přečtení není zapotřebí administrátorských privilegií. Malware tak má k dispozici mnoho jednoduchých a 100% funkčních možností k obcházení CFA. Pokud se malwaru podaří získat administrátorská práva, může dokonce manipulovat přímo s jeho nastavením pomocí Powershell příkazů Add-MpPreference a Set-MpPreference. Tyto změny vejdou v platnost až po restartu systému, ale v obou případech není uživatel o změnách informován.

Navržený software těmito problémy netrpí a bez přístupu ke kernelu ho nelze nijak obejít nebo s ním manipulovat. Jeho aktuální implementace však nedovoluje povolovat specifickým aplikacím přístup ke chráněným souborů. Tato varianta po vzoru CFA (a komerčních řešení jako Avast Ransomware Shield) by pro spolehlivé zabezpečení vyžadovala nejen ověřování pravosti .exe souborů, ale i integrity spuštěného procesu, všech jeho vláken a funkcí (a ukazatelů na ně) v paměti. Přísná ochrana použitá v této implementaci by mohla být vhodná pro protekci záloh nebo velmi citlivých souborů. S chráněnými soubory je možné manipulovat skrze ovladač, ten může případně poskytovat zabezpečené rozhraní pro komunikaci z uživatelského prostředí.

5 ZÁVĚR

V práci byly představeny funkce Windows API, které ransomware musí použít k získání přístupu k jednotlivým souborům před jejich šifrováním. Bylo vysvětleno, jak lze jejich používání monitorovat nebo blokovat z kernelu pomocí hookování systémových volání. Na základě toho byl vytvořen vlastní ovladač, který touto metodou blokuje přístup ke všem souborům s koncovkou "protected" pro demonstraci navrženého ochranného systému. Ten úspěšně zabránil poškození těchto souborů ransomwarem WannaCry. Použitá metoda byla porovnána s aktuální ochranou proti ransomwaru ve Windows 10 (Controlled folder access), kterou lze na rozdíl od navrženého systému obcházet i z uživatelského prostředí. Ochrana souborů hookováním systémových volání je také oproti dříve navrženým metodám detekce ransomwaru na základě používání Windows API výpočetně i implementačně nenáročná a 100% úspěšná při ochraně souborů.

REFERENCE

- [1] Hampton, N., Baig, Z., Zeadally, S.: Ransomware behavioural analysis on windows platforms. In: Journal of Information Security and Applications. 2018, s. 44-51. ISSN 22142126
- [2] Ki, Y., Kim, E., Kim, H.: A Novel Approach to Detect Malware Based on API Call Sequence Analysis. In: International Journal of Distributed Sensor Networks. 2015, ISSN 1550-1477
- [3] Scaife, N., Carter, H., Traynor, P., Butler, K.: CryptoLock (and Drop It): Stopping Ransomware Attacks on User Data. In: 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS). 2016, s. 303-312. ISBN 978-1-5090-1483-5. ISSN 1063-6927
- [4] Porter, J.: Cyberpunk 2077 studio falls victim to ransomware attack, data leak threatened: Hackers claim to have accessed source code for Cyberpunk 2077 and Witcher 3. In: The Verge [online]. Vox Media, LLC [cit. 2021-02-22]. Dostupné z: https://www.theverge.com/2021/2/9/22274035/cd-projekt-hack-source-code-cyberpunk-2077-witcher-3-encrypt-data-ransom

COOKER WITH THERMOELECTRIC CHARGING

Mirek Trchalík

Michal Paluřík

Electrical engineering 4th year, SŠPHZ Uherské Hradiště

E-mail: <u>17Etrchalikmi@student-uh.cz</u> E-mail: <u>17Epalurikmi@student-uh.cz</u>

Supervised by: Petr Hanáček

E-mail: Hanacek@ssphzuh.cz

Abstract: Our project deals with the development and production of ecological cooker, that, in addition to heating food or water, will also be used to produce elektricity for recharging batteries. The product is intended for those who often camp, fish or stay in nature outside civilization for a long time. The cooker can be used as a heat source, also in the preparation of meals, in addition to the production of electricity after recharging the batteries and powering various electrical devices via USB ports.

Keywords: cooker, heat, thermoelectric generator, eletricity generator

1 ÚVOD

Termoelektrický vařič je určen pro výrobu tepla a zároveň i elektrické energie. Tento vařič lze využít při kempování při přípravě jídla a může sloužit jako zdroj elektrické energie na místech, kde jiný zdroj elektrické energie není. Jako palivo se dá použít dřevěné třísky, papír, piliny, dřevěné uhlí, pelety, nebo jiný hořlavý materiál, který je ekologický a na daných místech dostupný. Tyto vařiče se už v dnešní době vyrábějí a dají se zakoupit. Podle konstrukčního provedení a parametrů vařiče se jejich cena pohybuje v rozmezí od 3000 do 10000 Kč. Ve svém projektu jsme se snažili vyrobit vařič podle vlastních požadavků a cenově výhodněji.

2 KONSTRUKCE VAŘIČE

Při návrhu jsme preferovali požadavek, aby vařič byl rozměrově malý, přenosný, ale prakticky využitelný pro účely ohřevu jídla a výroby elektrické energie při pobytu v přírodě. Dalším požadavkem, který jsme se snažili dodržet byla bezpečnost výrobku, aby při svém používání nezpůsobil požár a nebyla ohrožena ani obsluha vařiče. Tyto požadavky jsme brali v úvahu při samotném návrhu, při volbě materiálů a konstrukci výrobku.

Při realizaci elektrické části pro dobíjení akumulátorů jsme vycházeli ze základního principu vzniku termoelektrického napětí u termočlánků [1]. V našem případě jsme použili termočlánky SP1848-27145 [3]. Tyto termočlánky jsou vyráběny speciálně pro výrobu energie, ale lze je také použít pro chlazení nebo topení. Svými parametry jsou tyto termočlánky mnohem lepší než peltierovy články, které se používají výhradně pro výrobu elektrické energie [2]. U použitého článku jsme potřebovali na styčných plochách teplotní rozdíl minimálně 60°C.

Ohniště vařiče má válcový tvar s průměrem 140 mm a hloubkou 250 mm. Z důvodu bezpečnosti jsme se rozhodli kolem ohniště vytvořit celkem tři stěny (obrázek č. 1). Dvě stěny jsou vyrobeny plechu tloušťky 2 mm a poslední ochrannou stěnu tvoří plechová mřížka tloušťky 2 mm. V první stěně je vyvrtáno několik otvorů pro zajištění proudění vzduchu. Celý výrobek je posazen na tři kovové nožičky, které mají délku asi 140 mm. Nožičky zajišťují stabilitu vařiče a pro převoz jsou sklápěcí. Celková výška vařiče i s nožkami je 300 mm.



Obrázek 1: Konstrukční části vařiče

- 1. Hliníková tyč zakončená hliníkovou deskou slouží k přenosu tepla z ohně k termočlánkům
- 2. Elektronický obvod pro kontrolu nabíjení se sedmisegmentovým displejem a USB porty
- 3. Termočlánky umístěny mezi chladič a hliníkovou desku
- 4. Chladič slouží k ochlazování termočlánků
- 5. Elektronické obvody s akumulátory daná část je tvořena stabilizačními obvody na plošných spojích, akumulátorem, dvěma tlačítky na řízení nabíjení, zapínání a regulaci ventilátoru
- 6. Ventilátor slouží k chlazení a zajištění proudění vzduchu do ohniště
- 7. Horní kryt ohniště
- 8. Ochranná kovová mřížka
- 9. Druhá vnější stěna ohniště uzavírá vzduchový prostor mezi ohništěm a zároveň zajištuje správné proudění vzduchu z ventilátoru a zvyšuje bezpečnost a stabilitu výrobku.
- 10. První vnitřní stěna ohniště jsou v ní vyvrtány otvory pro správné proudění vzduchu z ventilátoru a je připevněna na spodní nosnou část
- 11. Otvor pro hliníkovou tyč
- 12. Otvor pro přívod vzduchu z ventilátoru do ohniště
- 13. Tři nosné nožky vařiče

Na obrázku č.2 je ukázka vyrobeného vařiče při pohledu se shora a), z boku b) a při činnosti.



a)

Obrázek 2: Ukázka vyrobeného vařiče

c)

ELEKTRONICKÁ ČÁST VAŘIČE 3

Elektronická část vařiče zajišťuje výrobu elektrické energie pro dobíjení akumulátorů. Základem jsou dva termočlánky, které jsou vloženy mezi nahřívanou hliníkovou desku a chladič. Nahřívaná hliníková deska je spojena s hliníkovou tyčí průměru 20 mm, která je vložena do ohniště. Průměr tyče byl zvolen tak, aby přenášené teplo nemělo velké ztráty a tyčka nahřívala rychle. Termočlánky pro zajištění většího napětí jsou spojeny do série. Termočlánky při teplotním rozdílu 100 °C produkují společně elektrické napětí 9,6 V a dobíjí dva akumulátory s celkovou kapacitou 6700 mAh, následně taky napájí ventilátor, který je potřebný k činnosti zařízení.

Ventilátor vtahuje chladný vzduch skrze přední otvor v krabičce a chladí chladič. Zahřátý vzduch vyfukuje přes speciální otvor do ohniště, čímž podporuje hoření a také výrazně pomáhá hlavně při roztápění. Ventilátor je stavěný pro napětí 5 V a odebírá proud 300 mA. Elektronická část obsahuje stabilizační a transformační obvody, protože odběr napětí a proudu z termočlánků se mění podle teploty, která se neustále mění. Konstantní napětí a proud jsou zejména důležité pro ventilátor, aby udržel své otáčky.

Elektronický obvod pro kontrolu nabíjení na displeji ukazuje procento nabití akumulátoru. Tento obvod obsahuje 2x port USB-A, 1x micro USB, 1x USB-C a připojují se k němu zařízení, které se z něho nabíjejí.

Celá elektronická část vařiče je vložena do plastové krabičky vytisknuté na 3D tiskárně. Protože materiál krabičky má svou určitou tepelnou odolnost (200 °C), jsou pro jistotu části, které jsou nejvíce vystavovány teplu chráněny izolační nehořlavou deskou CDZ (betonově-třísková nehořlavá deska), která odolá působení teploty až 1000°C.



+

Obrázek 3: Schéma zapojení elektronické části vařiče

4 CENOVÁ KALKULACE VÝROBKU

Na výrobu vařiče bylo nutno objednat a nakoupit potřebný materiál a komponenty.

Položka	Cena
Kovový materiál	382 Kč
3D tisk	1 078 Kč
Barva/lak	367 Kč
Elektronické součástky	1 618 Kč
Akumulátory	300 Kč
Hliníková tyč a deska	637 Kč
Izolační deska	100 Kč
Spojovací materiál	100 Kč
Celkem	4 582 Kč

Tabulka 1: Vynaložené náklady na výrobu vařiče

5 ZÁVĚR

Podařilo se nám navrhnout a vyrobit plně funkční vařič, který slouží k ohřevu a k výrobě elektrické energie pro dobíjení akumulátorů. Při své činnosti dokáže účinně nabíjet mobilní telefon, powerbanku nebo jiné zařízení.

Výrobek je navržen jako přenosný, dá se rozložit a zmenšit rozměry pro převoz. Výhodou je taky, že se dá oddělit elektronická část od ohniště. Rozložení na části je výhodné nejenom při přepravě, ale taky při čištění a opravách.

Vyrobený vařič je zcela bezpečný. Díky kovovým stěnám okolo ohniště je téměř nemožné se popálit, avšak celý výrobek má díky tomu větší hmotnost.

Výrobek se dá upravovat a modifikovat podle potřeb. Elektronická část by mohla být doplněna modulem pro dobíjení pomocí fotovoltaických panelů.

Náklady na materiál a komponenty vařiče vyšly 4582 Kč. V této ceně není zahrnuta naše práce a čas. Díky výrobě tohoto zařízení jsme získaly cenné zkušenosti. Věříme, že dané zařízení plně využijeme a bude nám sloužit při cestování a rybaření.

REFERENCE

- [1] Termočlánek [online]. [cit. 2021-03-27]. Dostupné z: https://cs.wikipedia.org/wiki/Termo%C4%8Dl%C3%A1nek
- [2] Termoelektrické jevy [online]. 2021 [cit. 2021-03-26]. Dostupné z: https://www.fzu.cz/vyzkum/vyzkumna-temata/termoelektricky-jev
- [3] Termočlánek SP1848-27145 [online]. [cit. 2021-03-27]. Dostupné z: https://www.researchgate.net/publication/329222572_Experimental_and_Analytical_ Simulati-

on_Analyses_on_the_Electrical_Performance_of_Thermoelectric_Generator_Modul es_for_Direct_and_Concentrated_Quartz-Halogen_Heat_Harvesting

ADAPTIVE BREAKE LIGHT

Ondřej Kusník

SŠPHZ Uherské Hradiště, obor Elektrotechnika, 4. ročník

E-mail: O.Kusnik@seznam.cz

Petr Hanáček

E-mail: Hanacek@ssphzuh.cz

Abstract: The adaptive brake light system represents an evolutionary improvement of the brake light function. In principle, the more the driver brakes the vehicle, the larger the brake lights are on. As a result, other road users receive more information about your current journey, thus preventing further accidents.

Keywords: brake, light systém, Arduino

1 ÚVOD

Systém adaptivního brzdového světla představuje evoluční vylepšení funkce brzdových světel. V základu jde o to, že čím více řidič daného vozidla brzdí, tím větší plocha brzdových světel je rozsvícena. Díky tomu se dalším účastníkům silničního provozu dostává informace o intenzitě brždění vozidla a tím se zvyšuje možnost adekvátní reakce na danou situaci.

1.1 Příklad

Jako příklad je možno uvést jízdu automobilu z kopce, při které má řidič neustále nohu na brzdě a jemně přibržďuje, nebo nákladní vozidlo s aktivovaným retardérem, který udržuje rychlost. V obou případech nám u automobilu prakticky neustále svítí brzdová světla. Pokud na silnici nastane situace, kvůli které je třeba začít intenzivně brzdit, po zahájení tohoto úkonu se stav světel nezmění. Jediná změna pro druhé vozidlo, které jede za ním bude ta, že se začne k prvnímu vozidlu velmi rychle a nebezpečné přibližovat, aniž by bylo varováno brzdovými světly a tím se značně prodlužuje reakční doba řidiče druhého vozidla.

2 KONSTRUKCE SYSTÉMU ADAPTIVNÍHO BRZDOVÉHO SVĚTLA

Pro předváděcí model systému adaptivního brzdového světla bylo potřeba vyrobit mechanismus brzdového pedálu, který v našem případě tvoří základ celého systému (Obrázek 4). Převod úhlu sešlápnutí pedálu na elektrický signál zajišťuje potenciometr 1 k Ω . Elektrický signál je zpracováván a vyhodnocován pomocí desky Arduino Uno R3. Výstupní signály z Arduina ovládají spínací relé a přes ně se spínají dané obvodové větve s diodovými svítidly.

2.1 ZÁKLADNÍ KOMPONENTY

Systém adaptivní brzdy tvoří:

- Programovací deska Arduino Uno R3 (Obrázek 1)
- Spínacích relé 5 V (Obrázek 2)
- USB adaptér do auta (Obrázek 3)
- Potenciometr 1 k Ω
- Diodové svítidla

Všechny komponenty kromě soustavy diodových svítidel jsou vloženy do těla modelu s pohyblivým pedálem.



Obrázek 1: Deska Arduino Uno R3



Obrázek 3: [3] USB adaptér



Obrázek 2: [2] Spínací relé





2.2 Zjednodušené zapojení systému pro dva světelné obvody

Zde je možno vidět zjednodušený nákres vnitřního zapojení, navrhnutý pro dva světelné obvody.



Obrázek 5: Zapojení

2.3 PROGRAMOVÁNÍ

Pro řízení spínání jednotlivých obvodů s diodovými svítidly byl vytvořen jednoduchý program. Mikroprocesor vyhodnocuje velikost vstupního elektrického signálu z potenciometru od brzdového pedálu. Na základě určitých zvolených hodnot posílá mikroprocesor výstupní signály pro spínací relé, které připojují obvody s diodovými svítidly. Zde je možno zvolit libovolný počet světelných obvodů a jednoduše přenastavit hodnoty pro spínání.

```
void setup() {
  Serial.begin(9600);
  pinMode(2, OUTPUT);
  pinMode(2, OUTPUT);
  pinMode(4, OUTPUT);
  pinMode(5, OUTPUT);
  pinMode(6, OUTPUT);
  pinMode(7, OUTPUT);
  pinMode(8, OUTPUT);
  pinMode(9, OUTPUT);
3
void loop() {
  int sensorValue = analogRead(A0);
  Serial.println(sensorValue);
  delay(10);
  if(sensorValue < 530){digitalWrite(10, HIGH);}
  else{digitalWrite(10, LOW);};
  if(sensorValue < 510){digitalWrite(9, HIGH);}
  else{digitalWrite(9, LOW);};
  if(sensorValue < 490){digitalWrite(8, HIGH);}
  else{digitalWrite(8, LOW);};
  if(sensorValue < 470){digitalWrite(7, HIGH);}
  else{digitalWrite(7, LOW);};
  if (sensorValue < 450) {digitalWrite(6, HIGH);}
  else{digitalWrite(6, LOW);};
  if(sensorValue < 430) {digitalWrite(5, HIGH);}
  else{digitalWrite(5, LOW);};
  if(sensorValue < 415){digitalWrite(4, HIGH);}
  else{digitalWrite(4, LOW);};
  if(sensorValue < 415){digitalWrite(3, HIGH);}</pre>
  else{digitalWrite(3, LOW);};
  if(sensorValue < 405){digitalWrite(2, HIGH);}
  else{digitalWrite(2, LOW);};
```

}

Obrázek 6: Ukázka programu

2.4 ZKOUŠKY VÝROBKU

Na následujících obrázcích (Obrázek 7, 8) jsou ukázky modelu systému adaptivního brzdového světla, kde se spíná 8 světelných obvodů.



Obrázek 7: Hotový výrobek

Obrázek 8: Zkoušení

3 ZÁVĚR

Podařilo se navrhnout, sestavit a naprogramovat systém adaptivního brzdového světla. Daný model pracuje podle původních představ. Výhodou je, že model se dá jednoduše modifikovat a vylepšovat. Také je vhodný k naučení programování s Arduinem. Pokud by systémy adaptivního brzdového světla mohly být využívány v praxi, určitě by to pomohlo pro zvýšení bezpečnosti silničního provozu.

REFERENCE

- Voda Z., tým HW Kitchen: Průvodce světem Arduina, Bučovice, Nakladatelství Martin Stříž 2017, ISBN 978-80-87106-93-8
- [2] Elektro-hofman.cz [online]. [cit. 28. 03. 2021] Dostupné z: https://www.elektro-hofman.cz/
- [3] Czc.cz [online]. [cit. 28. 03. 2021] Dostupné z: https://www.czc.cz/

Bakalářské projekty

Elektronika a komunikace, Mikroelektronika a technologie

COMPARISON OF MODERN VIDEO CODECS FOR 4K AND 8K VIDEO CONTENT

Adam Budáč

Bachelor Degree Programme (3), FEEC BUT E-mail: xbudac02@stud.feec.vutbr.cz

Supervised by: Jan Kufa

E-mail: kufa@feec.vutbr.cz

Abstract: The purpose of this paper is to compare modern video codecs, that ensure delivery of 4K and 8K video content in high quality at reasonable bitrates. There are compared three video codecs: HEVC, VVC and the AV1. Video quality assessment is ensured by using four objective video quality metrics, which are: PSNR, SSIM, VIF and VMAF. For tests, there were chosen two from total of seven videosequencies with different parameters: very high resolutions, high framerates, wide range of spatial and temporal indexes and 3D and 360° projection.

Keywords: 4K, 8K, modern video codecs, video quality assessment, objective quality metrics

1 ÚVOD

Najnovšie zariadenia určené na sledovanie multimediálneho obsahu sú vyrobené s vysokým rozlíšením (HD) zobrazovacích panelov, ktorými sú televízie (TV) s 8K rozlíšením a smartfóny a virtuálna realita (VR) s rozlíšením 4K. Zvlášť pre VR je potrebný obsah vo vysokom rozlíšením, pretože pri sledovaní 360° foto a video obsahu dochádza k výrezu v obraze. Tento obsah je potrebné zakódovať a dopraviť ho konečnému užívateľovi v dostatočnej kvalite pri čo najmenších nárokoch na distribúciu dát. Obsah môže mať pridanú hodnotu vo forme vysokého dynamického rozsahu (HDR), stereoskopickej projekcie (3D), alebo 360° uhlu záberu. Preto je nutné použiť video kodeky s vysokou efektivitou kódovania a podporou všetkých týchto rozšírení. Použité videosekvencie majú dĺžku 10 sekúnd a sú zakódované širokým rozsahom dátových tokov v rozmedzí 5 – 50 Mbit/s, preto je možné použiť výsledky vo viacerých oblastiach, akými sú TV vysielanie DVB-T2, Blu-Ray a streaming a to aj v 3D a VR.

2 ZDROJOVÉ KÓDOVANIE

Obraz je zložený z pixelov, kde každý je zložený z 3 subpixelov – červeného (R), zeleného (G) a modrého (B) – RGB sústava. RGB hodnoty sú prepočítané na luminančný (Y) a 2 rozdielové chrominančné signály (UV) – YUV sústava. YUV signál je chromaticky podvzorkovaný, typicky na hodnotu 4:2:2 alebo 4:2:0. Snímky sú medzi sebou premiestnené a je vytvorená skupina snímok Group of Pictures (GOP). V nej sa nachádzajú 3 typy snímok: I - referenčný, P – dopredne predikované a B – obojsmerne predikované. Snímok je rozdelený na kódové stromové jednoty Coding Tree Unit (CTU), ktoré sú rozdelené na menšie predikčné a transformačné jednotky a bloky. Intrapredikciou je znížená priestorová redundancia. Na bloky je aplikovaná diskrétna kosínusová transformácia (DCT) a diskrétna sínusová transformácia (DST), čím sú hodnoty pixelov v bloku prevedené na maticu frekvenčných koeficientov. Matica je vydelená kvantizačnou tabuľkou, koeficienty sú vyčítané diagonálne (cik-cak). Ďalším krokom je run-length kódovanie a entropické kódovanie [1], [2], [3].

2.1 KODEK HIGH EFFICIENCY VIDEO CODING (HEVC)

Kodek HEVC [1] (H.265) bol vyvinutý v roku 2013. Je použitý v TV vysielaní DVB-T2. Kodek bol po svojom vzniku rozšírený funkciami HDR, 3D alebo 360° video.

2.2 KODEK VERSATILE VIDEO CODING (VVC)

Kodek VVC [1], [2] (H.266), bol vyvinutý v roku 2020. Pri zachovaní kvality je dosiahnuté zníženie dátového toku o 30-50% oproti HEVC. Obsahuje podporu pre HDR, 3D a 360° video. Nové algoritmy sú spojenie hrán 360° videa, predikcia farby z jasu, škálovanie a rotácia obrazu a rozpoznávanie geometrických tvarov a hrán.

2.3 KODEK AOMEDIA VIDEO 1 (AV1)

Kodek AV1 [1], [3] bol vyvinutý v roku 2018. Je použitý vo webových aplikáciách, napríklad streaming videa cez platformu YouTube. Je konkurenciou pre VVC. Obsahuje podporu pre HDR, 3D aj 360° video. Nové algoritmy sú identifikácia transformačných blokov, použitie viacero kvantizačných tabuliek v rámci jedného snímku, kruhové vyčítavanie frekvenčných koeficientov.

3 OBJEKTÍVNE METRIKY KVALITY

Objektívne metriky sú použité na vyjadrenie kvality videosekvencií pomocou matematických modelov a algoritmov. Všetky použité metriky sú plne referenčné, na ich výpočet je použitý rozdiel medzi referenčnou a skreslenou videosekvenciou. V presnejších metrikách je brané do úvahy aj vnímanie ľudským zrakom (HVS model). Vyššie číslo u výsledkov použitých metrík znamená vyššiu kvalitu.

3.1 PEAK-SIGNAL-TO-NOISE-RATIO (PSNR)

Najjednoduchšia a najpoužívanejšia objektívna metrika. Vyjadruje špičkový pomer medzi signálom a šumom. Nepočíta s HVS, preto je to najmenej presná metrika. Z dôvodu vysokého dynamického rozsahu vyhodnocovaných signálov je vyjadrená v logaritmickej miere [4].

3.2 STRUCTURAL SIMILARITY INDEX (SSIM)

Metrikou je vyjadrená podobnosť 2 obrazov. Sú zohľadnené princípy HVS, ako štrukturálne zmeny, štrukturálne skreslenia, posun jasov a zmeny kontrastu v scéne. Existujú rozšírenia zamerané na 3D a 360° video, 3D-SSIM a SSIM360. Je dosiahnutá vysoká korelácia s ľudským vnímaním [4].

3.3 VISUAL INFORMATION FIDELITY (VIF)

Metrika je založená na štatistike prírodnej scény v spojení s modelom skreslenia na výpočet informácie zdieľanej medzi referenčným a testovacím obrazom. Sú v nej zohľadnené princípy HVS a je dosiahnutá vysoká korelácia s ľudským vnímaním [5].

3.4 VIDEO MULTIMETHOD ASSESSMENT FUSION (VMAF)

V metrike je odhadovaná subjektívna kvalita videa na základe strojového učenia. Na odhad sú použité trénované modely, ktorými je simulované hodnotenie ľudským respondentom. Sú tu zlúčené aj ďalšie objektívne metriky, ako VIF, Detail Loss Metric, Mean Co-Located Pixel Difference a Anti-Noise Signal To Noise Ratio [6]. Výsledky VMAF najviac odpovedajú subjektívnemu vnímaniu.

4 DOSIAHNUTÉ VÝSLEDKY

Pre porovnanie kvality kodekov boli použité videosekvencie v rozlíšení 4K a 8K so snímkovou frekvenciou 60 snímok za sekundu (FPS). Na spracovanie, kódovanie a dekódovanie videosekvencií bol použitý software (SW) FFmpeg¹, VVenC a VVdeC² [1]. Pre výpočet PSNR, SSIM a VIF bol použitý SW VQMT³ a pre výpočet VMAF SW FFmpeg s použitým modelom 4K vo verzii 0.6.1.json. Ďalej sú uvedené príklady príkazov na zakódovanie videosekvencií v poradí: HEVC, AV1, VVC.

¹ FFmpeg v4.3.1-2021-01-01 - https://www.gyan.dev/ffmpeg/builds/

² VVenC a VVdeC v0.2.1.0 - https://github.com/fraunhoferhhi/vvenc, https://github.com/fraunhoferhhi/vvdec

³ VQMT - https://www.epfl.ch/labs/mmspg/downloads/vqmt/

ffmpeg -i original.mp4 -c:v libx265 -r 60 -preset medium -b:v 10M -pix_fmt yuv420p hevc.mp4

ffmpeg -i original.mp4 -c:v libaom-av1 -r 60 -row-mt 1 -tiles 3x4 -threads
12 -cpu-used 8 -b:v 7M -minrate 1M -maxrate 10M -pix_fmt yuv420p av1.mkv
vvencapp -i original.yuv -s 3840x2160 -c yuv420 -r 60 --preset faster -b
10000000 --internal-bitdepth 8 --qpa 0 -ip 128 -t 6 -o vvc.266

Na obrázku 1 je uvedené porovnanie výsledkov metrík pre 2 videosekvencie: Balkón Canon 4K 2D 60 a Balkón Samsung 8K 2D 60. Obrazový obsah a FPS oboch týchto videosekvencií je zhodný, rozdielne je len rozlíšenie. V budúcej práci bude celkový počet videosekvencií 7, ako je možné vidieť aj na obrázku 2, kde sú uvedené základné informácie o videosekvenciách vo forme časovo-priestorového indexu, ktorým sú vyjadrené časové zmeny a priestorové informácie jednotlivých videosekvencií. Na obrázku 3 je screenshot z videosekvencie Balkón Canon 4K 2D 60. Pre HEVC bol použitý profil rýchlosti medium, čo je rozumný kompromis medzi kvalitou a rýchlosťou kódovania. Z dôvodu dlhého kódovania AV1 a VVC boli použité najvyššie profily rýchlostí, cpu-used 8 (AV1) a faster (VVC). Čas kódovania 4K videosekvencie kodekom HEVC bol priemerne 1 minútu, 10 minút kodekom AV1 a 20 minút kodekom VVC. Pre 8K bol čas kódovania 4-násobne dlhší. Pri nízkych dátových tokoch dôležitých pre TV vysielanie je najefektívnejší kodek VVC, kde je dosiahnutá úspora toku 20 - 30% oproti AV1 a 50% oproti HEVC. Pri vysokých dátových tokoch vhodných pre streaming alebo Blu-Ray sú výsledky AV1 a VVC pri 4K podobné a pri zvyšovaní dátového toku je AV1 efektívnejší ako VVC, ale pri 8K je znova VVC efektívnejší o 20 – 30% oproti AV1. Pri VMAF v rozmedzí 85 – 95 bodov je možné kvalitu považovať za dostatočnú až vysokú. 8K VMAF model ešte nie je dostupný, preto bol pre vyhodnotenie kvality 8K videosekvencie použitý 4K model. Preto môžu byť výsledky trochu odlišné od budúceho 8K modelu.



Obrázok 1: Výsledky objektívnych metrík pre rozlíšenie 4K60 (vľavo) a 8K60 (vpravo).





Obrázok 2: Časovo-priestorový index videosekvencií.

Obrázok 3: Screenshot z videosekvencie.

5 ZÁVER

Kódovanie kodekmi AV1 a VVC je efektívnejšie ako kodekom HEVC. Dátový tok 5 Mbit/s pri 4K je pre kodek HEVC minimum pre zachovanie dostatočnej kvality, zatiaľ čo AV1 a VVC majú značnú rezervu. Pre 8K je potrebný tok omnoho vyšší, 30 Mbit/s pre HEVC (6-krát viac), 18 Mbit/s pre AV1 (3,5-krát viac) a 13 Mbit/s pre VVC (2,5-krát viac). Rovnaké testy boli vykonané aj na videosekvenciách s polovičnou snímkovou frekvenciou 30 FPS, kde výsledky boli o 0,5 - 1 dB PSNR, respektíve 1,5 - 5 VMAF bodov lepšie. Avšak z dôvodu stúpajúcich nárokov dnešných používateľov na kvalitu a vernosť obrazu a videa som sa zameriaval hlavne na vyššiu snímkovú frekvenciu, v ktorej nový obsah stále pribúda a je dôležitý hlavne v oblastiach ako streaming a VR. Vzhľadom na malé zmeny rozdielov medzi tokmi 25 a 50 Mbit/s nebola použitá hodnota 35 Mbit/s, pri nižších tokoch boli rozdiely značne väčšie, preto sú hodnoty dátových tokov hustejšie. Cieľom do budúcna je tvorba laboratórnej úlohy zameranej na kódovanie a analýzu kvality videa vhodného pre VR.

POĎAKOVANIE

Tento príspevok vznikol s podporou projektu interného grantu VUT v Brne FEKT-S-20-6325.

REFERENCIE

- [1] Mansri I., Doghmane N., Kouadria N., Harize S. and Bekhouch A.: "Comparative Evaluation of VVC, HEVC, H.264, AV1, and VP9 Encoders for Low-Delay Video Applications," 2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA), Valencia, Spain, 2020, pp. 38-43, doi: 10.1109/MCNA50957.2020.9264275
- [2] H.266. BITMOVIN [online]. San Francisco (CA), 2013, 14.2.2020 [cit. 2021-02-25]. Dostupné z: https://bitmovin.com/compression-standards-vvc-2020/
- [3] AV1. Alliance for Open Media: Next Generation, Open-Source Digital Media Technology for Everyone [online]. Wakefield (MA), 2015, 18.1.2019 [cit. 2021-03-01]. Dostupné z: http://aomedia.org/av1/specification/
- [4] Nový spôsob hodnotenia kvality videa. In: Technická univerzita Ostrava [online]. Ostrava (ČR), 2017, 6.12.2017 [cit. 2020-12-08]. Dostupné z: https://homel.vsb.cz/~voz29/files/Frnda_Dizertacna_Praca.pdf
- [5] VIF. In: LIVE: Laboratory for Image & Video Engineering [online]. Austin (TX), 2006 [cit. 2020-12-08]. Dostupné z: https://live.ece.utexas.edu/research/Quality/VIF.htm
- [6] VMAF. THE NETFLIX TECH BLOG [online]. San Francisco (CA): Medium, https://netflixtechblog.com/, 6.6.2016 [cit. 2021-03-01]. Dostupné z: https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652

EVALUATION OF CNN AND CLDNN ARCHITECTURES ON RADIO MODULATION DATASETS

Kristýna Pijáčková

Bachelor Degree Programme (3), FEEC BUT E-mail: xpijac02@stud.feec.vutbr.cz

Supervised by: Tomáš Götthans E-mail: gotthans@feec.vutbr.cz

Abstract: This paper presents an evaluation of deep learning architectures designed for modulation recognition. The evaluation inspects, whether the architectures behave in the same way as they did on the dataset they were designed on. The architectures are trained and tested on two different radio modulation datasets. This results in proposing additional binary classification as a method to reduce misclassification of QAM modulation types in one of the datasets.

Keywords: Radio modulation, classification, neural network, deep learning, CNN, CLDNN

1 INTRODUCTION

Deep learning (DL) has been on the rise over the last ten years. Even though many may have heard about DL in combination with computer vision or natural language processing, it has been used in other fields as well, including wireless communications [1].

One of the applications of DL in wireless communications is automatic modulation recognition (AMR), which is also the topic of this work. A decision-theoretic approach and a statistical pattern recognition approach are commonly used for the AMR. Both of them rely on extracting features from received data before classification. Meanwhile, with the deep learning approach, the received data can be pasted directly into the DL model. This is due to the ability of a neural network to find patterns (features) in data during training.

A design of a convolutional neural network (CNN) and a convolutional long-short-term memory deep neural network (CLDNN) architectures for AMR were described in [2]. This paper, therefore, includes only a brief overview of these architectures and focuses on an evaluation of achieved accuracies on two radio modulation datasets.

2 ARCHITECTURES

Architectures designed for the AMR introduced in papers such as [3,4] very often use either convolutional neural network, recurrent neural network (RNN), or their combination. This led me to design CNN and CLDNN architectures. Even though the architectures often have over a million trainable parameters, the parameters of the proposed CNN and CLDNN were kept low. While a high number of parameters means, that the model can learn more features, it can cause overfitting on smaller datasets, requires more computational power and results in a larger model size. A more detailed description of the proposed CNN and CLDNN architectures can be found in [2]. An overview of the used architectures is shown in Figures 1 and 2.



Figure 1: CNN architecture

Figure 2: CLDNN architecture

3 EXPERIMENTAL RESULTS

In [2], the CLDNN achieved better results in accuracy than the CNN, and both networks were able to achieve high accuracy even on a smaller dataset. This section presents results achieved by the CNN and CLDNN architectures and shows, that these conclusions apply to other datasets as well.

3.1 MIGOU-MOD DATASET

The MIGOU-MOD dataset [5] includes 8.8 million samples of over-the-air measurements with 11 different modulations - QPSK, QAM16, QAM64, CPFSK, BFSK, GFSK, 8PSK, PAM4, AM-SSB, AM-DSB, and WBFM. The measurements were taken indoors at two distances from a transmitter at 1 and 6 meters. This corresponds to the average signal-to-noise ratio (SNR) of 37dB and 22dB respectively. More details on this dataset can be found in [6].

Figure 3 shows confusion matrices of the dataset. Voice records used for the signal generation contain pauses, during which the only present signal is a carrier frequency. The authors in [6] mention, that this causes confusion between the AM-SSB and AM-DSB. Why these pauses do not affect the third analog modulation WBFM the same way as they did in [2] on a RadioML dataset, is, however, unclear to me.

96



Figure 3: MIGOU-MOD dataset - Confusion Matrices

3.2 VUT DATASET

This dataset was provided to me by my supervisor and it was generated using MATLAB. This dataset is rather small and contains 120,000 samples of 6 modulations - QAM4, QAM16, QAM64, OFDM,

GFDM, and FBMC. The data are again stored as 2x128 vectors of I/Q signals and the SNRs are in a range from -20dB to 18dB.



Figure 4: VUT dataset - Confusion Matrices

Figure 4 shows confusion matrices of CNN and CLDNN architectures both at low and high SNRs. It is easy to see, that both of the networks have troubles telling the QAM modulation types apart at low SNRs. The FBMC modulation, which uses overlapping QAM, gets confused at low SNRs for QAM as well. Since the QAM64 may contain the same constellation points as QAM16, it can be challenging to tell them apart at times, especially with 128 sample points only.

The accuracy improved by more than 30% at low SNRs when the CNN and CLDNN architectures were used as binary classifiers. The significant improvement can be seen in Figure 5. Adding binary classification after the current categorical classification has, however, its downsides. Binary classifiers can only tell if the input sample is a certain modulation type or not. This means, that three additional models would be needed for the QAM modulation types in this dataset, and would increase the overall model size. Two possible options to solve this problem are either designing another architecture for the binary classification, that is small in size, or using quantization on the binary classifiers to reduce their size.



Figure 5: Accuracy improvement of QAM with binary classification

4 CONCLUSION

This paper presented an evaluation of the CNN and CLDNN on two additional datasets. The results confirmed expectation from [2], that the CLDNN architecture achieves higher accuracy and both models can work well with smaller datasets. The CLDNN outperformed the best result in [6] by 2.35% on the MIGOU-MOD dataset, even though it was trained on a 10 times smaller training set. The accuracy of the CLDNN was better at SNRs above 0dB for the VUT dataset generated by MATLAB as well. The quadrature amplitude modulations at SNR below 0dB were challenging for both architectures. Proposed binary classifiers at the end of the current architectures might increase the classification accuracy by more than 30%. This option should be further examined, as the extra classifiers would increase the overall size of the model.

REFERENCES

- [1] DeepSig's publications on ML applications in wireless communications: https://www.deepsig.ai/publications
- [2] K. Pijackova, T. Gotthans, "Radio Modulation Classification Using Deep Learning Architectures," 2021 31st International Conference Radioelektronika (RADIOELEKTRONIKA), 2021, pp. 1-5, doi: 10.1109/RADIOELEKTRONIKA52220.2021.9420195.
- [3] O'Shea T.J., Corgan J., Clancy T.C. (2016) Convolutional Radio Modulation Recognition Networks. In: Jayne C., Iliadis L. (eds) Engineering Applications of Neural Networks. EANN 2016. Communications in Computer and Information Science, vol 629. Springer, Cham. Available at: https://doi.org/10.1007/978-3-319-44188-7_16
- [4] A. Emam, M. Shalaby, M. A. Aboelazm, H. E. A. Bakr and H. A. A. Mansour, "A Comparative Study between CNN, LSTM, and CLDNN Models in The Context of Radio Modulation Classification," 2020 12th International Conference on Electrical Engineering (ICEENG), Cairo, Egypt, 2020, pp. 190-195, doi: 10.1109/ICEENG45378.2020.9171706.
- [5] R. Utrilla, "MIGOU-MOD: A dataset of modulated radio signals acquired with MIGOU, a low-power IoT experimental platform", Mendeley Data, V1, doi: 10.17632/fkwr8mzndr.1
- [6] R. Utrilla, E. Fonseca, A. Araujo, and L. A. Dasilva, "Gated Recurrent Unit Neural Networks for Automatic Modulation Classification With Resource-Constrained End-Devices" in IEEE Access, vol. 8, pp. 112783-112794, 2020, doi: 10.1109/ACCESS.2020.3002770.

RAY TRACING SOFTWARE FOR RADIATION ANALYSIS OF ELECTRONIC COMPONENTS

Matej Klement

Bachelor Degree Programme (3), FEEC BUT E-mail: xkleme14@vutbr.cz

Supervised by: Marek Bohrn

E-mail: bohrn@feec.vutbr.cz

Abstract: This article presents software for calculation of ionising and non-ionising radiation received by a component. The software evaluates surrounding shielding provided by user-defined model of a shielding box. It utilises ray tracing method to simulate and measure the trajectories of particles, which are assumed to be traveling in a straight line.

Keywords: Ray tracing, Radiation, Total Ionising Dose, TID

1 ÚVOD

Radiačná analýza je veľmi dôležitý aspekt návrhu zariadení pre špecializované účely. Vyhodnocuje sa najmä v aplikáciách, ktoré sú vystavené veľkému množstvu radiácie a obsahujú citlivé zariadenia ako sú polovodičové tranzistory a integrované obvody. Prípadne sa môže jednať o rozsiahle systémy obsahujúce veľké množstvo integrovaných obvodov aj pri pôsobení malého množstva radiácie. Typicky sa jedná o vesmírne misie, medicínske prístroje, armádne prístroje a iné.

V oblasti radiačnej analýzy neexistuje veľké množstvo nástrojov pre návrhára. Cieľom vytvoreného softwaru je poskytnúť užívateľsky prívetivý nástroj pre zjednodušenie práce s radiačnou analýzou.

2 RADIAČNÁ ANALÝZA

Existuje niekoľko prístupov k radiačnej analýze, ktoré sa menia v závislosti od špecifickosti projektu. Medzi dôležité efekty radiácie patrí TID (*Total Ionising Dose*) a TNID (*Total Non-Ionising Dose*). Sú to kumulatívne efekty degradácie parametrov súčiastky vplyvom dopadajúcej radiácie. Vytvorený software slúži na zistenie úrovne týchto dvoch efektov.

Množstvo dopadajúcej radiácie sa vyhodnocuje pomocou miery tienenia, ktoré je v okolí analyzovaného bodu. Dávka radiácie sa obecne zmenšuje s pridaným tienením. V určitých prípadoch, konkrétne pri použití tenkých vrstiev materiálov s malou hustotou, môže dôjsť z dôvodu sekundárneho žiarenia k nárastu radiácie.

Na vyhodnotenie množstva dopadajúcej radiácie na súčiastku potrebujeme niekoľko dôležitých informácií. Medzi najdôležitejšie z nich patria parametre prostredia, v ktorom sa daný systém bude nachádzať. Informácie o prostredí sú typicky zadané v požiadavkách projektu, pre ktorý sa radiačná analýza vyhodnocuje. Pre vesmírne misie je možné použiť modely ako AE-8 a AP-8 (založené na meraných dátach zo satelitov NASA [1]) ako externý zdroj pre zistenie tokov radiačných častíc v prostredí. Na tieto účely existujú aplikácie ako SPENVIS alebo OMERE, ktoré dokážu spočítať presné toky v závislosti na trajektórii misie. Zo získaných tokov a dĺžky trvania misie je možné následne pomocou aplikácie SHIELDOSE vypočítať takzvanú Dose-Depth-Curve (ďalej len DDC). DDC predstavuje krivku závislosti dopadajúcej radiácie od hrúbky materiálu krytia, ktorú musia častice preletieť. Táto krivka je definovaná pre umiestnenie detektoru radiácie v 4 rôznych elementárnych typoch geometrie. Medzi tieto geometrie patrí konečný plát, semi-nekonečný plát, plné sférické tienenie a prázdne sférické tienenie. [2]

Ak je možné nahradiť analyzovanú geometriu jednou z týchto 4 geometrií, stačí pre danú hrúbku z DDC odčítať množstvo radiácie.

V opačnom prípade je potrebné prepočítať množstvo dopadajúceho žiarenia niektorou z metód simulácie uvedených v nasledujúcej kapitole.

3 METÓDY SIMULÁCIE

Existuje viac spôsobov ako spočítať množstvo radiácie v špecifickej geometrii. Doporučené postupy sú popísané v normách ECSS a riadia sa určitými pravidlami. [3]

Sektorová analýza spočíva v manuálnom rozdelení analyzovanej geometrie na niekoľko sektorov, kde sa každý sektor aproximuje do jednej z uvedených 4 elementárnych geometrií. Dávka žiarenia z každého sektoru sa prepočíta cez podiel celkového priestorového uhlu a priestorového uhlu, ktorý daný sektor predstavuje. Presnosť tohto prístupu závisí na presnosti aproximácie analyzovanej geometrie.

Monte Carlo analýza predstavuje úplne odlišný prístup k tejto problematike. Jedná sa o najpresnejší druh analýzy, v ktorej sa simulujú dráhy konkrétnych častíc a ich interakcie s atómovou mriežkou objektov. Táto analýza je presná ale výpočtovo veľmi náročná a zdĺhavá.

Ray trace analýza je princípom podobná sektorovej analýze. Pri ray trace analýze sa celkový priestorový uhol rozdelí na rovnomerne rozložené časti bez ohľadu na umiestnenie detektoru. Do každej časti je vyslaný lúč, ktorý predstavuje pomyselnú trajektóriu častice. Dĺžka dráhy, ktorú lúč prešiel vnútrom materiálu zadanej geometrie, je cez DDC prepočítaná na dávku radiácie a vydelená celkovým počtom vyslaných lúčov. Výsledné množstvo radiácie je dané súčtom jednotlivých dielčích príspevkov. Výhodou tohto prístupu je výrazne väčšia výpočetná rýchlosť oproti Monte Carlo analýze a vyššia presnosť oproti sektorovej analýze.

4 IMPLEMENTÁCIA METÓDY RAY TRACE

Pre softwarovú implementáciu bola zvolená metóda ray trace. Táto metóda predstavuje kompromis medzi rýchlosťou, jednoduchosťou a presnosťou.

Podľa normy ECSS je potrebné vyslať do analyzovaného bodu (detektoru) minimálne 1800 v priestore rovnomerne rozložených lúčov. [2] Príručka k tejto norme popisuje možnosť použiť rozloženie pomocou krokovania uhlových súradníc alebo sieťové rozloženie po ploche obklopujúcich obdĺžnikov. Oba prístupy vedú k väčšej hustote lúčov v okolí pólov (osí). Aby sa tomuto nežiadúcemu efektu predišlo, bol zvolený iný prístup.

Rovnomerné rozloženie N-bodov na guli je doposial' nevyriešený matematický problém. Pre dosiahnutie čo najlepšieho rozloženia lúčov bolo preto použité rozloženie pomocou takzvanej Zlatej špirály. [4] Uhlové súradnice lúčov sú dané rovnicami (4.1)

$$lat_i = \arcsin\left(\frac{2i}{2N+1}\right) \qquad \qquad lon_i = 2\pi i \phi^{-1} \qquad (4.1)$$

Detektory radiácie, do ktorých sú vysielané lúče, sú zadané ako jednotlivé body v trojrozmernom karteziánskom priestore.

Na výpočet dráhy lúčov je potrebné, aby bola geometria zadaná pomocou povrchovej trojuholníkovej siete (mesh). Každý trojuholník siete je zapísaný pomocou troch bodov v priestore. Tento zápis zodpovedá aj popisu geometrie vo formáte STL, ktorý je využívaný ako formát dát pre export geometrie z bežných CAD systémov.

Pre každý lúč z každého detektoru sú analyticky vypočítané priesečníky s trojuholníkmi popisujúcimi geometriu. Rozdiel vzdialeností zistených priesečníkov určuje dĺžku dráhy lúča vo vnútri materiálu tienenia. Vypočítaná dĺžka je prepočítaná na úroveň radiácie cez DDC zadanú

používateľom. Aby výpočet prebehol správne, musí sa na prepočet použiť DDC popisujúca tienenie modelu elementárneho plného sférického tienenia.

Takto získaná radiácia daného lúča zodpovedá celkovému priestorovému uhlu a preto je vydelená počtom vyslaných lúčov. Výsledná úroveň radiácie je získaná súčtom dielčích príspevkov radiácií jednotlivých lúčov.

Popísaný algoritmus bol implementovaný do programu MATLAB[®], ktorý umožňuje jednoduchú vizualizáciu výsledkov. Vďaka podpore výpočtov s využitím grafickej karty pomocou integrovaného Parallel Computing Toolbox[™] môže byť tento časovo náročný výpočet výrazne zrýchlený.

4.1 POUŽITIE VYTVORENÉHO SOFTWARU

Základom TID a TNID analýzy je výpočet tienenia poskytnutého objektmi obklopujúcimi detektor, typicky skrinka s citlivým zariadením vo vnútri. Používateľ si v preferovanom CAD programe vytvorí 3D model geometrie, ktorý následne rozdelí na objekty podľa typu materiálu. Tieto objekty z návrhového programu exportuje do formátu STL a následne importuje do popisovaného simulačného programu.

Po importe všetkých STL súborov musí používateľ pre každý objekt určiť materiál, ktorého vlastnosti pre účel výpočtu popisuje hustota $[g/cm^3]$. (**Obrázok 1**)



Obrázok 1: Užívateľské rozhranie softwaru.

V poslednom kroku pred spustením simulácie používateľ importuje špecifickú DDC a určí umiestnenie súčiastky (detektoru). Ak používateľ nie je rozhodnutý o umiestnení citlivej súčiastky, môže použiť plošné zobrazenie radiácie v rovine, na základe ktorého vyberie najvhodnejšie miesto. (**Obrázok 2**)



Obrázok 2: Plošné zobrazenie radiácie v modeli hliníkovej skrinky.

Po zvolení vhodného umiestnenia súčiastky môže používateľ spustiť simuláciu znova s väčším počtom lúčov a upresniť radiáciu v danom bode.

Pre odhalenie najslabších miest v tienení analyzovanej geometrie je možné zobraziť lúče, ktorých príspevok radiácie bol najvyšší. (**Obrázok 3**)



Obrázok 3: Hliníková skrinka s analyzovaným integrovaným obvodom v plastovom puzdre BGA. Červeno-žlté spektrum farieb znázorňuje 100 najhorších z celkových 7200 lúčov.

5 ZÁVER

Popisovaný software pre ray tracing slúži na simuláciu a výpočet úrovne radiácie typicky vnútri prístrojovej skrinky. Výstup simulácie je možné použiť ako podklad pre radiačnú analýzu pri návrhu elektronických prístrojov a modulov. Software využíva postup výpočtu popísaný v norme ECSS [2]. Naviac používa rovnomernejšie rozloženie lúčov, čo môže viesť k presnejším výsledkom.

Používateľské rozhranie vytvoreného programu je intuitívne a každý kľúčový krok od importu dát až po vizualizáciu výsledkov má vyhradenú samostatnú záložku. Oproti iným nástrojom používateľovi umožňuje analyzovať aj rozloženie radiácie na ploche, vrátane vizualizácie výsledkov. Príklady výstupov z analýzy hliníkovej skrinky sú zobrazené na obrázku 2 a 3.

REFERENCIE

- [1] NASA. *Space Physics Data Facility* [online]. 11 Mar 2021 [cit. 2021-03-11]. Dostupné z: https://spdf.gsfc.nasa.gov/index.html
- [2] ECSS SECRETARIAT. ECSS-E-ST-10-12C: Methods for the calculation of radiation received and its effects, and a policy for design margins. First issue. Noordwijk, The Netherlands: ESA, 2008. Dostupné také z: https://ecss.nl/standards/active-standards/
- [3] ECSS SECRETARIAT. ECSS-E-HB-10-12A: Calculation of radiation and its effects and margin policy handbook. First issue. Noordwijk, The Netherlands: ESA, 2010. Dostupné také z: https://ecss.nl/hbs/active-handbooks/
- [4] GONZÁLEZ, Álvaro. Measurement of Areas on a Sphere Using Fibonacci and Latitude– Longitude Lattices. *Mathematical Geosciences* [online]. 2010, (1), 49-64 [cit. 2020-12-07]. ISSN 1874-8961. Dostupné z: doi:10.1007/s11004-009-9257-x

AUTOMATED GARDEN SYSTEM FOR OUTDOORS

Tomáš Stupka

Bachelor Degree Programme (3), FEKT VUT E-mail: xstupk04@stud.feec.vutbr.cz

Supervised by: Tomáš Frýza

E-mail: fryza@vut.cz

Abstract: The aim of this article work is to design and implement an automated irrigation system for use in a family house. The proposed system is based on the Raspberry Pi platform and AVR microcontroller with the possibility of automatic adaptation to outdoor temperature and humidity conditions. The design also includes a web interface using the NodeRED for manual control and monitoring of collected data from individual wireless sensor modules

Keywords: Garden automation, Wireless communication, Raspberry Pi, Atmega328P, C, Python, IoT, nRF24L01

1 ÚVOD

Práce se zabývá návrhem a realizací automatizovaného zavlažovacího systému pro venkovní použití. Zmíněný systém má konkrétní uplatnění v rodinném domě, kde se využívá na zavlažování travních ploch i truhlíkových a záhonových rostlin. Systém je navržený pro 9 zavlažovacích okruhů, ale je myšleno i na případné rozšíření, kdy do řídící místnosti jsou přivedeny i záložní hadice na rozvod vody.



Obrázek 1: Blokové schéma systému

Systém je postavený na řídící jednotce Raspberry Pi[1], která svým programem ovládá devět zavlažovacích okruhů zahrady. Každý okruh je spínaný prostřednictvím relé a elektromagnetického ventilu. Systém by měl mít i schopnost přizpůsobit se aktuálním podmínkám, díky informacím ze dvou modulů s řídícími mikrokontroléry Atmega328P. Svá data budou sbírat pomocí snímačů teploty/vlhkosti vzduchu a vlhkosti půdy. Tato data následně budou odeslána řídící jednotce bezdrátovou komunikací za pomocí modulu nRF24L01, kde budou uloženy do SQL databáze. Vlhkost půdy a teplota se zobrazí ve webovém rozhraní, naprogramovaném v programovacím nástroji NodeRED, kde i technicky nezkušený uživatel dokáže ručně nastavit spínání zavlažovacích okruhů a dobu zavlažování.

1.1 BEZDRÁTOVÁ KOMUNIKACE

Pomocí bezdrátové komunikace můžeme změřená data z jednotlivých snímačů poslat do řídící jednotky, která je následně vyhodnotí. V této problematice se často setkáváme s pojmem "Internet věcí". Je definován Mezinárodní telekomunikační unií jako globální infrastruktura pro IT společnosti. Definují také rozdíl mezi zařízením a věcí. Zařízení musí být schopno komunikovat s ostatními zařízeními a mělo by poskytovat minimálně snímání, sběr dat, ovládání nebo zpracování dat. Zástupci těchto technologií jsou například Lora nebo Sigfox.[2][3]

Můžeme také využít levné a lehce implementované WiFi moduly. Jedním takový zástupcem je NRF24L01. Modul komunikuje na frekvenci 2,4 GHz s přenosovou rychlostí od 250 kb/s do 2 Mb/s. Dosah se liší v závislosti na modelu a na použité přenosové rychlosti a také na umístění modulu. Základní model disponuje integrovanou anténou na desce, která má udávaný dosah výrobcem 100 metrů ve volném prostoru (v reálných podmínkách se počítá s nižším dosahem). Komunikaci mezi řídícím procesorem obstarává sběrnice SPI. Napájecí napětí modulů je 3,6 V. A to konkrétně za pomoci tužkových lithiových baterií od firmy SAFT. Při plné kapacitě baterie, by měl modul vydržet dostatečnou dobu, i s ohledem na to že modul bude pracovat pouze v letním ročním období. Teoretické hodnoty výdrže jsou uvedeny níže. V práci byl zvolen modul NRF24L01 s integrovanou anténou z důvodů možnosti volby tří variant a zároveň nízké spotřeby, rozměrům a nízké pořizovací ceně.

1.2 Řídící jednotky

Celé zařízení je řízeno jednodeskovým počítačem Raspberry Pi, který byl původně vytvořen pro výukové účely. Poslední verze 4 obsahuje 4-jádrovým procesorem ARM Cortex A7 s frekvencí 900MHz, 1 GB RAM. Pro vytvářenou aplikaci byl vybrán model 2, který dosahuje dostačujícího výkonu.

Jelikož byl pro komunikaci vybrán modul nRF24L01, bude využita HW periférie pro sběrnici SPI, spínací relé pro elektromagnetické ventily se jednoduše připojí na GPIO piny a sepnou pomocí napětí 3,3 V nebo 5 V. Program pro jednotlivé spínací relé bude programován v jazyce Python, který je již součástí operačního systému Rasbian. Je založen na tzv. flow based přístupu a pracuje s uzly (nodes), z nichž každému lze zadat určitou funkci a postupně tak vytvořit posloupnost, která zpracovává příchozí data a posílá je dále do dalšího uzlu. NodeRed také nabízí webové prostředí, ve kterém můžeme zobrazovat data ze snímačů, spínání tlačítek, ventilů a mnoho více.

Pro sběr dat z jednotlivých senzorů (vlhkost půdy, teplota a vlhkost vzduchu) byly navrženy desky plošných spojů s mikrokontroléry ATmega328P[4] a komunikačními moduly NRF24L01. Napájení zajišťuje baterie na napětí 3,6V a mikrokontrolér je možné programovat v systému prostřednictvím AVR_ISP konektoru.



Obrázek 2: Schéma zapojení pro bezdrátové moduly

1.3 SPÍNACÍ ČÁST

Spínání jednotlivých okruhů je naplánováno pomocí dvou spínacích součástek, a to pomocí relé a elektromagnetických ventilů. Ventily jsou spínány střídavým napětím 19-24 V. Pro buzení relé je využit integrovaný obvod ULN2801a, který obsahuje tranzistory v Darligtonově zapojení na jejichž vstupy je přivedeno napětí 3,3V z řídící jednotky Raspberry Pi.



Obrázek 1: 3D model desky pro spínací část

Pro spínání potrubí byly zvoleny elektromagnetické ventily EZP-22-54[5] s regulací průtoku. Jedná se o ventil spínaný střídavým napětím 24 V. Přívod vody je možno ručně vypnout/zapnout centrálním ventilem. A zároveň lze každý okruh uzavřít samostatně kulovým ventilem. Celý ventilový systém je umístěn do vnitřní místnosti, tak aby se k ventilům nedostala vlhkost. Při venkovním uložení často dochází k zamrzání, korozi a celkovému poškození systému. Proto je výhodnější systém umístit do vnitřního prostoru.

2 ZÁVĚR

V článku byl popsán koncept realizace systému pro automatizovaný zahradní systém s využitím v rodinném domě. V současné době jsou vyrobeny desky plošných spojů pro spínací část a bezdrátové moduly. V realizované zahradě jsou již kompletně rozvedeny podzemní hadice pro zavlažování travních ploch a záhonů a hadice pro zavlažování okenních truhlíků jsou zazděné do omítky domu. A každý okruh je ukončen příslušnou technologií zavlažování (postřikovače, hadicí s integrovanými kapkovači, nebo kapkovači). Všechny okruhy byly přivedeny do řídící místnosti, kde jsou připojeny přes spínací ventily k přívodu vody z vodovodního řádu. V práci nyní zbývá dokončit programovou část a její následné ověření v reálných podmínkách.

REFERENCE

- [1] *T*each, Learn, and Make with Raspberry Pi. Teach, Learn, and Make with Raspberry Pi [online]. Dostupné z: https://www.raspberrypi.org
- [2] sigfox.cz Connecting Things. [online]. Copyright © 2016 [cit. 13.03.2021]. Dostupné z: <u>https://sigfox.cz/cs</u>
- [3] Lora The Lora aliance [online]. Dostupné z: <u>https://lora-alliance.org</u>
- [4] Smart | Connected | Secure | Microchip Technology. Smart | Connected | Secure | Microchip Technology [online]. Copyright © Copyright 1998 [cit. 13.03.2021]. Dostupné z: <u>https://www.microchip.com</u>
- [5] Lawn Mowers, Golf Equipment, Landscape Equipment, Irrigation | Toro. [online]. Copyright
 © 2021 The Toro Company. All Rights Reserved. [cit. 13.03.2021]. Dostupné
 z: <u>https://www.toro.com/en</u>
- [6] Node-RED. Node-RED [online]. Dostupné z: <u>https://nodered.org</u>

A DESIGN AND DEVELOPMENT OF A UNIVERSAL POWER SUPPLY FOR LAPTOPS IN MOTOR VEHICLES

Adam Cagáň

Bachelor Degree Programme (3), FEEC BUT E-mail: xcagan00@vutbr.cz

Supervised by: Martin Šťáva

E-mail: stavam1@feec.vutbr.cz

Abstract: This thesis deals with design of DC power supply with adjustable output voltage, working in buck-boost topology. Thesis explains selection of buck-boost controller integradet circuit and calculation of values of external components. Designed converter is simulated in LTspice software, where parasitic properties of used components are included. There are summarized parameters of converter in different working states. Further there are mentioned planned expansions of converter, including microcontroler integration with user interface.

Keywords: DC-DC converter, buck-boost topology, high efficiency, synchronous rectifier, Arduino

1 ÚVOD

V dnešnej dobe sa čoraz častejšie využívajú prenosné počítače aj vďaka možnosti ich použitiu na miestach, kde nie je prístup k napájacej sieti. Z toho dôvodu boli vyvinuté napájacie zdroje pracujúce s palubnou sieťou automobilu vhodné nielen pre profesionálnych vodičov, ale aj širokú verejnosť. V tejto práci sa zameriam na návrh zariadenia, ktoré bude kompatibilné z väčšinou typov noteboo-kov. Napájacie napätie si zvolí užívateľ na základe svojho modelu. Navyše bude toto zariadenie schopné pracovať nielen s 12 V palubnou sieťou, ale s rôznymi veľkosťami napätí v rozsahu 6 až 48 V. Zároveň sa zameriam na funkčnosť zariadenia z pohľadu užívateľa, teda z hľadiska kompaktnosti a jednoduchosti používania.

2 NÁVRH DC-DC MENIČA

Spočiatku bolo potrebné zvoliť vhodný riadiaci obvod. Požadované parametre na riadiaci obvod tvorili z veľkej časti parametre samotného zariadenia. Predovšetkým pracovné napätie v rozsahu 6 až 48 V, výstupné napätie v rozsahu 15 až 30 V a minimálny výkon 60 W. Z dôvodu väčšej kompaktnosti zariadenia, je výhodné použiť obvod pracujúci v topológii buck-boost.

Zvolil som riadiaci obvod LT8390 od firmy Analog Devices. Výrobca v katalógovom liste uvádza vhodné zapojenie externých súčiastok vrátane príkladu výpočtu ich hodnôt. Medzi súčiastky, ktorých hodnoty a vlastnosti sú špecifické pre konkrétnu aplikáciu patrí cievka, spínacie tranzistory a vstupné a výstupné kondenzátory. Dôležitým parametrom meniča pri výpočte je spínacia frekvencia. Voľba spínacej frekvencie je kompromis medzi účinnosťou obvodu a veľkosťou akumulačných prvkov v obvode, ako sú kondenzátory a cievky [1]. Obvod LT8390 je schopný spínať frekvenciou v rozsahu 150 až 650 kHz. Keďže navrhujem prenosný napájací zdroj, zvolil som frekvenciu 600 kHz z dôvodu menšej veľkosti potrebných súčiastok, napriek menšej účinnosti.

2.1 VÝBER SÚČIASTOK

Synchrónny usmerňovač využíva 4 spínacie tranzistory typu MOSFET. Maximálny spínací prúd, na ktorý je potrebné dimenzovať cievku a spínacie tranzistory vypočítam pre buck a boost režim zvlášť. Z predpokladu účinnosti meniča minimálne 80 % môžem vypočítať maximálny spínací prúd, pre

boost režim so vstupným napätím 6 V a výstupným napätím 15 V. Pre zachovanie výstupného výkonu 60 W počítam s výstupným prúdom 4 A. Najprv vypočítam striedu spínania obvodu v tomto stave

$$D = 1 - \frac{U_{IN} \cdot \eta}{U_{OUT}} = 1 - \frac{6 \cdot 0.8}{15} = 0,68,$$
(1)

ďalej vypočítam zvlnenie prúdu cievkou, kde použijem indukčnosť s hodnotou 47 μH, ktorú som zvolil ako väčšiu hodnotu, než je minimálna požadovaná hodnota indukčnosti, vypočítaná v rovnici (7).

$$\Delta I_{\rm L} = \frac{U_{\rm IN} \cdot D}{f \cdot L} = \frac{6 \cdot 0.68}{6 \cdot 10^5 \cdot 47 \cdot 10^{-6}} = 0,144 \, \text{A}.$$
(2)

Vypočítam maximálnu veľkosť spínacieho prúdu, ktorá je rovná

$$I_{SW} = \frac{\Delta I_L}{2} + \frac{I_{OUT}}{1 - D} = \frac{0.144}{2} + \frac{4}{1 - 0.68} = 12,57 \text{ A.}$$
(3)

Rovnakým spôsobom vypočítam hodnotu spínacieho prúdu pre boost režim s výstupným napätím 30 V a prúdom 2 A. V tomto prípade je $I_{SW} = 12,58$ A. Ďalej vypočítam podobným postupom hodnotu spínacieho prúdu pre buck režim so vstupným napätím 48 V, výstupným napätím 15 V a prúdom 4 A.

$$D = \frac{U_{OUT}}{U_{IN} \cdot \eta} = \frac{15}{48 \cdot 0.8} = 0.39,$$
(4)

$$\Delta I_{L} = \frac{(U_{IN} - U_{OUT}) \cdot D}{f \cdot L} = 0,94 \text{ A},$$
(5)

$$I_{SW} = \frac{\Delta I_L}{2} + I_{OUT} = \frac{0.94}{2} + 4 = 4,47 \text{ A}.$$
 (6)

Rovnako vypočítam veľkosť spínacieho prúdu pre buck režim s výstupným napätím 30 V a prúdom 2 A. v tomto prípade je $I_{SW} = 2,25$ A. Z predošlých výpočtov vyplýva, že tranzistory treba dimenzovať na spínací prúd $I_{SW} = 12,58$ A, ktorý nimi bude tiecť v boost režime s výstupným napätím 30 V a prúdom 2 A.

Zvyšné požadované parametre na tranzistory sú zhrnuté v Tabuľka 1.

UGS(min RDSon)	$U_{BR(DSS)}$	R_{DSon}	I_D
< 5 V	>48 V	Najmenší možný	> 12,58 A

Tabul'ka 1: Požadované parametre spínacích tranzistorov

Zvolil som tranzistory BSC070N10LS5 s parametrami $U_{BR(DSS)} = 100$ V, $U_{GS(min RDSon)} = 4,5$ V, $R_{DSon} = 7$ m Ω , $Q_g = 16$ nC a $I_D = 79$ A.

Požadované hodnoty indukčnosti cievky a kapacity výstupných kondenzátorov sa líšia podľa toho, či menič práve pracuje v buck režime, alebo boost režime. Požadovanú hodnotu týchto prvkov určíme spôsobom, že vypočítame požadované hodnoty pre jednotlivé režimy a minimálna požadovaná hodnota bude väčšia z nich.

$$L_{BUCK} > \frac{U_{OUT} \cdot (U_{IN(MAX)} - U_{OUT})}{f \cdot I_{OUT\max} \cdot \Delta I_L \% \cdot U_{IN\max}} = \frac{30 \cdot (48 - 30)}{6 \cdot 10^5 \cdot 2 \cdot 0.3 \cdot 48} = 31,25 \,\mu H \tag{7}$$

$$L_{BOOST} > \frac{U_{IN\max}^2 \cdot (U_{OUT} - U_{IN\max})}{f \cdot I_{OUT\max} \cdot \Delta I_L \% \cdot U_{OUT}^2} = \frac{6^2 \cdot (30 - 6)}{6 \cdot 10^5 \cdot 2 \cdot 0.3 \cdot 30^2} = 2,67 \,\mu H$$
(8)

Potrebujem indukčnosť väčšiu ako 31,25 µH, zvolím bežne dostupnú hodnotu 47 µH. Cievka musí byť dimenzovaná na prúd väčší ako I_{sw}, teda minimálne 12,58 A.

Kondenzátory znižujú výstupné zvlnenie napätia. Toto zvlnenie je spôsobené dvoma javmi, a to nabíjaním a vybíjaním kapacity a úbytkom napätia na vnútornom sériovom odpore kondenzátoru vplyvom nabíjacieho prúdu. Zvolím maximálne prípustné zvlnenie spôsobené jednotlivými javmi ΔU_{CAP} = ΔU_{ESR} = 50mV. Maximálne výsledné zvlnenie je rovné súčtu zvlnení spôsobených jednotlivými javmi, teda ΔU_{OUTmax} = 100 mV. Následne vypočítam potrebné hodnoty kapacity a maximálny vnútorný sériový odpor výstupných kondenzátorov pre jednotlivé režimy tak, aby boli splnené mnou zadané hodnoty maximálneho zvlnenia napätia.

$$C_{OUT(BOOST)} = \frac{I_{OUTmax} \cdot (U_{OUT} - U_{INmax})}{\Delta U_{CAP} \cdot U_{OUT} \cdot f} = \frac{4 \cdot (15 - 6)}{0.05 \cdot 15 \cdot 6 \cdot 10^5} = 80 \,\mu F \tag{9}$$

$$C_{OUT(BUCK)} = \frac{U_{OUT} \cdot (1 - \frac{U_{OUT}}{U_{INmax}})}{8 \cdot L \cdot f^2 \cdot \Delta U_{CAP}} = \frac{30 \cdot (1 - \frac{30}{48})}{8 \cdot 68 \cdot 10^{-6} \cdot (6 \cdot 10^5)^2 \cdot 0.05} = 1.15 \,\mu F \tag{10}$$

$$ESR_{BOOST} = \frac{\Delta U_{ESR} \cdot U_{IN(MIN)}}{U_{OUT} \cdot I_{OUT(MAX)}} = \frac{0.05 \cdot 6}{30 \cdot 2} = 5 \ m\Omega \tag{11}$$

$$ESR_{BUCK} = \frac{\Delta U_{ESR} \cdot L \cdot f}{V_{OUT} \cdot (1 - \frac{U_{OUT}}{U_{IN(MAX)}})} = \frac{0.05 \cdot 68 \cdot 10^{-6} \cdot 6 \cdot 10^{5}}{30 \cdot (1 - \frac{30}{48})} = 181 \, m\Omega \tag{12}$$

Pre neprekročenie vopred určeného zvlnenia napätia $U_{OUTmax} = 100 \text{ mV}$, je potrebná kapacita väčšia ako 80 µF a sériový odpor menší ako 181 m Ω . Vhodné je použiť viacero kondenzátorov zapojených paralelne, čím dosiahneme menší ESR ako pri použití jedného kondenzátora.

2.2 SIMULÁCIA OBVODU

Funkciu navrhnutého obvodu meniča som overil v simulačnom programe LTspice. Využil som možnosť zadať parazitné vlastnosti súčiastok ako ESR výstupných kondenzátorov a odpor vinutia cievky. Nepodarilo sa mi nájsť model tranzistoru, ktorý som zvolil v kapitole 2.1, preto som v simulácii použil dostupný model BSC100N06LS3 s podobnými parametrami.



Obrázok 1: Schéma navrhnutého meniča

Sledoval som celkovo štyri stavy z hľadiska zmeny vstupného napätia na výstupné. Využil som časovú analýzu a sledoval som časové priebehy vstupných a výstupných napätí a prúdu cievkou od spustenia obvodu až po ustálený stav. Dôležitým parametrom meniča je jeho účinnosť. Účinnosť som vypočítal ako pomer výstupného výkonu ku vstupnému v ustálenom stave.

	ts [ms]	D [%]	η [%]	ΔU_{OUT} [mV]
$6 \Rightarrow 15$	11,4	64	87,88	59
$6 \Rightarrow 30$	22	81,8	83,65	64
$48 \Rightarrow 15$	9,1	31,9	90,94	1,05
$48 \Rightarrow 30$	10.2	62.61	94.79	1.14

Zistené parametre sú zhrnuté v nasledujúcej tabuľke.

Tabul'ka 2: Parametre obvodu v jednotlivých režimoch zistené pri simulácii

2.3 ROZŠÍRENIE OBVODU

Súčasťou napájacieho zdroja bude mikrokontrolér na platforme Arduino doplnený o displej a rotačný enkodér. V jednoduchom menu bude možnosť zvoliť značku notebooku s prednastavenou hodnotou výstupného napätia, prípadne si nastaviť ľubovoľné napätie z rozsahu 15 až 30 V. Po zvolení požadovaného napätia mikrokontrolér odošle informáciu cez I2C zbernicu do DAC prevodníku. DAC prevodník zmenou pomerov v obvode spätnej väzby zmení výstupné napätie meniča.



Obrázok 2: Vľavo - zapojenie DAC prevodníku, vpravo – ukážka menu

3 ZÁVER

Cieľom práce bolo navrhnúť napájací zdroj schopný pracovať v rozsahu vstupných napätí 6 až 48 V s možnosťou nastavenia výstupného napätia v rozsahu 15 až 30 V. Práca sa venuje návrhu DC-DC meniča ako základného bloku tohto zdroja. V prvej časti som zvolil vhodný riadiaci obvod a následne som vypočítal potrebné parametre externých súčiastok. Funkčnosť navrhnutého zapojenia som overil pomocou simulácie v programe LTspice. Simuloval som štyri pracovné stavy, a to zvyšovanie napätia z 6 na 15 V a na 30 V a znižovanie napätia z 48 na 15 V a na 30 V. Účinnosť tohto meniča sa pohybuje v rozmedzí 83 až 94 % v závislosti na pracovnom režime a hodnote vstupného napätia. Maximálne zvlnenie výstupného napätia dosahovalo 64 mV, čo je menej ako mnou stanovená maximálna prípustná hodnota 100 mV. V štádiu vývoja je rozšírenie meniča o mikrokontrolér a užívateľské rozhranie pre jednoduché nastavenie požadovaného výstupného napätia. Výstupne napätie meniča bude riadené pomocou mikrokontroléru a DAC prevodníku. V poslednej fáze vývoja bude zdroj doplnený o vstupné a výstupné ochrany, ako je ochrana proti prepólovaniu alebo skratu.

REFERENCIE

 COUFAL, M. DC/DC měnič pro světelný zdroj s LED. Brno : Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2019. s. 48. Vedoucí bakalářské práce prof. Ing. Jaroslav Boušek, CSc.
DESIGN OF A SEMI-AUTOMATIC POSITIONING SYSTEM FOR MEASURING OF SEMICONDUCTOR CHIPS

Tomáš Kotian

Bachelor Degree Programme (3), FEEC BUT E-mail: xkotia00@vutbr.cz.

Supervised by: Alexandr Otáhal

E-mail: Alexandr.Otahal@vutbr.cz

Abstract: The aim of this work is to design a semi-automatic positioning system for the test of semiconductor chips. The work solves the problems with the implementation of the construction to achieve the highest possible accuracy for contacting with probe card. Next, the work deals with the selection of a suitable control unit for stepper motors and finally to create a program to control the system and display image from the camera to the computer. The last requirement for the program is detection of chips with the camera image and automatic centering with a probe card.

Keywords: Semi-automatic positioning system for chip test, TMCM-6110, LabWIEW, probe cards

1 ÚVOD

Neoddeliteľnou súčasťou pri výrobe integrovaných obvodov je ich verifikácia a testovanie. Toto testovanie sa realizuje pomocou špecializovaných prístrojov ku kontaktovaniu čipov na kremíkovej doske (wafer). Využívajú sa k tomu karty so sondami (probe cards), alebo mikro posuvy so sondami. Aktuálnym trendom integrácie sú požiadavky na presnosť týchto prístrojov, ktoré sa stále zväčšujú a tým aj ich obstarávacia cena. Väčšina z týchto prístrojov ma obmedzené využitie pre pár druhov aplikácii a tým sa stávajú nedostupné pre menšie laboratória, alebo školy k testovaniu malého počtu kusov čipov.

Cieľom mojej práce je realizovať cenovo dostupnejšiu alternatívu prístroja k testovaniu a verifikácii integrovaných obvodov, ako aj vizuálnej inšpekcii, testu menších plošných spojov a manipuláciu s vysokou presnosťou. Zabezpečenie pohybu v troch osiach, zobrazenie obrazu z kamery, ovládanie v manuálnom a poloautomatickom režime vo vytvorenom užívateľskom prostredí za pomoci počítača.

2 NÁVRH

Konštrukcia je prispôsobená viacerým aplikáciám podľa komerčných prístrojov. Medzi ktoré patria napríklad test čipov na kremíkovej doske, test zapuzdrených súčiastok, vizuálna inšpekcia, meranie homogenity nanesenej vrstvy a test vodivých ciest pre plošne spoje. Pracovná plocha je 10 000 mm² s možnosťou upevnenia mikrometrického rotačného stolčeku k ručným korekciám. Pohyb kamery v osi Z je 140 mm a maximálne zaťaženie je 9 kg. Kontaktovacia hlava sa pohybuje v rozmedzí 100 mm a jej maximálne možné zaťaženie je 4 kg, čo zabezpečuje možnosť využitia nie len pre kartu so sondami, ale aj mikro posuvy so sondami alebo iné zariadenia.

2.1 POŽIADAVKY NA PRÍSTROJ

Využitie súčiastok z nefunkčného prístroja vizuálnej inšpekcie pre zníženie nákladov. Pristroj musí zabezpečovať pohyb v osiach X, Y a Z plus rotáciu okolo osi Z. Os Z by mala zabezpečiť dostatočnú pevnosť pri uchytení mikroskopu s kamerou. Ďalším dôležitým faktorom je tuhosť celej kon-

štrukcie od ktorej záleží presnosť pohybu. Rozlíšenie kroku 1 μ m, presnosť väčšia ako ±5 μ m. Riadiaca doska schopná ovládať minimálne 4 krokové motory s funkciou mikrokrokovania a snímania elektromagnetického poľa motorov. Pracovná plocha 10 000 mm² s možnosťou upevnenia mikrometrického rotačného stolčeku. Ako poslednou požiadavkou je univerzálna kontaktovacia hlava s uchytením pre karty so sondami a mikro posuvy so sondami.

2.2 POŽIADAVKY NA PROGRAM

Program zabezpečujúci príjemné a jednoduché užívateľské prostredie v ktorom bude zobrazený obraz z kamery a možnosť ovládania prístroja. Ovládanie prístroja v manuálnom a automatickom móde. Automaticky mód identifikuje čip podľa obrazu kamery a zarovná ho s kartou so sondami. Ďalšie funkcie nájdenie referenčného bodu, posuv do stredu a zastavenie pohybu. Ochrana pred stratou kroku s upozornením a automatickým zastavením pri prekážke v pohybe. Komunikácia cez USB.

2.3 MODEL

Na obrázku 2 je návrh modelu pristroja a na obrázku 1 je detail na kontaktovaciu hlavu. Pre kontaktovanie je využitý systém kariet so sondami so šírkou 80 mm, dĺžkou 100 mm a s hrúbkou do 2 mm. Plocha na okolí hlavy je dostatočne veľká pre minimálne 4 mikro posuvy so sondami. V aktuálnej zostave je využitá kamera s objektívom pre väčšie čipy, ale je možne ju vymeniť za mikroskop pri ktorom je vyžadované väčšie zväčšenie. Ako riadiaca jednotka bola zvolená TMCM-6110, ktorá je schopná ovládať 6 motorov, z ktorých budú využite len 4 [2]. To prináša možnosť ďalšej automatizácie pohybu do budúcnosti, napríklad pre rotáciu čipu.



Obrázok 2: Model prístroja

Obrázok 1: Detail kontaktovacej hlavy

3 POROVNANIE CENY A PARAMETROV

Na obrázku 3 môžeme vidieť poskladanú základnú konštrukciu realizovaného prístroju a v tabuľke 1 rozpísane náklady na jednotlivé časti. Na obrázku 4 je jeden z najlacnejších poloautomatických prístrojov k testu čipov na kremíkovej doske s názvom Signatone CM460-22 [1]. Ich základné parametre sú porovnané v tabuľke 2 . Realizovaný pristroj má oproti komerčnému prístroju motorizovaný pohyb v osi Z pre mikroskop. Ďalšou výhodou je aj motorizovaný pohyb pre kontaktovanie za pomoci kariet so sondami. Parameter v ktorom sa realizovaný pristroj odlišuje od komerčného je presnosť. Hodnota uvedená v tabuľke 2 predstavuje najväčšiu možnú vôľu medzi guľôčkovou skrutkou a maticou. Presné hodnoty budú známe až po teste presnosti za pomoci odchylkometra.



Obrázok 4: Aktuálny stav pristroja



Obrázok 3: Signatone CM460-22 [1]

Názov	Cena [Kč]
Hliník	7000
Spojovací materiál	500
Lineárna technika	16000
Krokové motory	2100
Riadiaca doska	5300
Zdroj	500
Elektronika	500
Kamera	1800
Mikrometrický stolček	15250
Výsledná cena	48950

Tabuľka	1:	Roz	písane	náklad	lv na	rea	lizova	ıný	prísti	·oi
I uoui nu	••	TOL	pibulle	mannav	ay ma	I Cu.	112010	ury	priou	· vj

Tabul'ka 2: Porovnanie parametrov [1] [3]

	Realizovaný prístroj	Signatone CM460-22	
Pohyb [mm]	100	150	
Rozlíšenie [µm]	1	1	
Opakovateľnosť [µm]	±1	±1	
Presnosť [µm]	±25	±3	

4 UŽIVATELSKE PROSTREDIE

Na obrázku 5 je zobrazené aktuálne užívateľské prostredie vytvorené v program LabVIEWTM, ktoré zabezpečuje obraz z kamery, údaje o polohe, nastavenie veľkosti kroku, ovládanie smeru pohybu a tlačidlo pre zastavenie. Veľkosť kroku (SET STEP) je možné nastavovať za pomoci vpísania do okna Set value, alebo posunutím jazdca v rozsahu 1 až 100 mm, alebo μm. Údaje o polohe (PO-SITON) sú zobrazované v μm od referenčného bodu. Program bude doplnený o funkcie nájdenia referenčného bodu, automatickú detekciu a zarovnanie karty so sondami s testovaným čipom a ovládanie pohybu kontaktovacej osi.



Obrázok 5: Užívateľské prostredie

5 ZÁVER

V práci bol navrhnutý model na obrázku 1, vybrané vhodne krokové motory a riadiaca doska. Pri skladaní bol braní veľký dôraz na presné vycentrovanie, minimalizáciu odporu pri pohyb a kolmosť plôch voči sebe. Týmito opatreniami bol zabezpečený spoľahlivý pohyb v troch osiach bez zvýšeného zaťaženia motorov a predišlo sa možnému opotrebeniu vodiacich tyči.

Aktuálny stav pristroja je zobrazený na obrázku 3. Zostáva pripevniť os pre kontaktovanie, upevniť mikrometrický rotačný stolček a koncové spínače. Potrebné časti sú objednané a čaká sa na ich doručenie.

Užívateľské prostredie má dokončený manuálny mód so zobrazením obrazu z kamery, ktorý bude doplnení o funkcie nájdenia referenčného bodu, posuv do stredovej pozície, detekciu čipu a zarovnanie čipu s kartou so sondami.

POĎAKOVANIE

Tento príspevok vznikol na Ústavu mikroelektroniky VUT v Brne za podpory projektu FEKT-S-20-6215.

REFERENCE

- [1] Signatone. CM460 Semi-Automatic Probe Station [online]. 12 Mar 2021 [cit. 2021-03-12]. Dostupné z: https://www.signatone.com/products/probestations/cm460.asp
- [2] Trinamic. TMCM-6110 [online]. 12 Mar 2021 [cit. 2021-03-12]. Dostupné z: https://www.trinamic.com/products/modules/details/tmcm-6110
- [3] Cncshop. SFI matice jednoduchá [online]. 12 Mar 2021 [cit. 2021-03-12]. Dostupné z: http://www.cncshop.cz/sfi-matice-jednoducha

STUDY OF SURFACE CHANGES OF NEGATIVE ELECTRO-DE FOR LEAD-ACID BATTERIES USING OPERANDO CON-FOCAL LASER MICROSCOPY

Hana Hálová

Bachelor(3), FEEC BUT E-mail: xhalov01@vutbr.cz

Supervised by: Ladislav Chladil

E-mail: chladil@feec.vutbr.cz

Abstract: Our contribution deals with the laser microscopy observation of the negative electrode for lead-acid batteries during their life-span. We evaluate the surface changes during both deep cycling and PSoC (Partial State of Charge) cycling. The mode leads to a gradual degradation of the negative electrode – mostly because of irreversible sulfation. The speed of the sulfation can be influenced by additives added already during the production process. The active mass was firstly examined by electron microscope and X-ray diffractometer. After that, the crystal growth during cycling was observed using confocal microscope Olympus Lext OLS4100.

Keywords: Negative electrode, Premature Capacity Loss, cycle life, PSoC, lead sulfate, confocal laser microscopy.

1 ÚVOD

Přestože jsou olověné akumulátory známy již více než 150 let, dochází k jejich neustálému vývoji z důvodu výskytu nových problémů způsobených dalšími oblastmi využití a prodlužováním celkové životnosti akumulátorů. Aktuální vývoj olověných akumulátorů se týká především optimalizace pro hybridní vozidla, kde vlivem dlouhodobého setrvávání akumulátorů pouze v částečně nabitém stavu (mezi 40 - 60% celkové kapacity) dochází k nevratné sulfataci [4].

Za normálních podmínek dochází v olověném akumulátoru ke vratné sulfataci – vzniklé krystalky síranu olovnatého se při nabíjení snadno rozkládají na původní aktivní materiály elektrod a ionty kyseliny sírové. Vlivem setrvávání článku pouze v částečně nabitém stavu ovšem nedochází k dostatečnému rozpadu krystalů při nabíjení, a jelikož dochází k přednostnímu ukládání síranu na již dříve vzniklé krystalky, dochází časem k jejich růstu. Vzniklá vrstva velkých krystalů ucpává póry v aktivních hmotách a snižuje účinnou plochu elektrod – postupně dochází jak ke snížení kapacity akumulátoru, tak ke zvýšení vnitřního odporu a namáhání článků, což má za následek postupné roztahování mřížek elektrod [1] [3].

Rychlost sulfatace může být do jisté míry ovlivněna aditivy přidávaných do záporné hmoty – využití např. BaSO₄, lignosulfátů, či uhlíku [4]. Tyto aditiva mají určitý pozitivní vliv na velikost a následný růst krystalů, ale zároveň mohou vést k intenzivnějšímu vývoji vodíku na elektrodě během nabíjení a tím ovlivňovat plynování článku a ztrátu vody [3][4].

Aby mohlo dojít k určitému omezení rychlosti sulfatace, je důležité se zaměřit a porozumět samotnému vývoji krystalů síranu, jak při hlubokém cyklování, tak při PSoC režimu. Z hlediska podmínek experimentu se v tomto směru nabízí využití konfokálního mikroskopu, pomocí kterého může být povrch záporné aktivní hmoty nepřetržitě sledován s rozlišovací schopností až 10 nm [2]. Pomocí postupného laserového snímání dochází při každém snímku k postupnému skenování reliéfu a vytvoření realistických snímků záporné elektrody uložené v In-situ.

2 PŘÍPRAVA EXPERIMENTU

Cílem první přípravné etapy bylo připravit elektrody z olověného plechu o tloušťce 0,5 mm a pastovanou oblast optimalizovat tak, aby bylo možné elektrodu cyklovat a současně pozorovat povrch pomocí konfokálního laserového mikroskopu ve speciální elektrochemické cele od firmy EL-CELL. Jako záporná elektroda byl využit kruhový disk o průměru 1,4 cm potažený tenkou vrstvou epoxidového lepidla určeného pro fixaci součástek na DPS. Tím bylo zabráněno tomu, aby docházelo k reakci na povrchu olova mimo oblast pastované hmoty. Po vytvrdnutí lepidla byla elektroda ve středu provrtána – vznik průchodu o průměru 1,5 mm. Na tento otvor byla nanesena aktivní hmota, která vznikla smícháním práškové hmoty (olověný prach, síran barnatý a borosilikát) a kyseliny sírové v poměru 10 g hmoty na 0,5 g H_2SO_4 (hmotnost aktivní hmoty vybraného vzorku m = 0,0104 g). Proti-elektroda byla vytvořena stočením olověného plechu o délce 11 cm a výšce pásku 2 mm, což zajistilo dostatečně velkou aktivní plochu proti-elektrody.



Obrázek 1: Příklad proti-elektrody (vlevo) a kruhového disku (vpravo)



Obrázek 2: Postup přípravy záporné elektrody – olověný plech s vytvrzeným lepidlem a otvorem (vlevo), s nanesenou aktivní hmotou (uprostřed), uložené v In-situ (vpravo)

Vytvořené elektrody oddělené separátorem byly uloženy do hermeticky uzavřené cely určené pro optická pozorování, zaplaveny 27% kyselinou sírovou ($\rho = 1,19$ g/cm³), upevněny do svěrky a umístěny do konfokálního mikroskopu. Cela byla nakontaktována a ovládána potenciostatem Bilogic VSP-VMP3.



Obrázek 3: Připravené měřicí pracoviště, detail upevněné elektrochemické cely v konfokálním mikroskopu (vpravo)

3 EXPERIMENT

Prvotní formátování olověného článku probíhalo 15 hodin rychlostí 0,1 C – tedy velikostí proudu I = 0,104 mA (uvažujeme-li 100 mA/g aktivní hmoty), přičemž samotné nabíjení trvalo 10 hodin a vybíjení 5 hodin. Z důvodu neustálého plynování článku při nabíjení během formování, nebylo možné z této části pořizovat konfokálním mikroskopem validní série 3D snímků.



Obrázek 4: Snímek původní aktivní hmoty (vlevo) a po naformování (vpravo)

Poté již byla započata série prvních hlubokých cyklů, tentokrát již s dvojnásobnou velikostí proudu, tedy I = 0,208 A (0,2 C), při které bylo zahájeno také sledování konfokálním mikroskopem. Nabíjení článku bylo omezeno napětím záporné elektrody vůči referenční kadmiové elektrodě na 0,05 V (přičemž po dosáhnutí potřebného maxima bylo napětí 0,05 V 30 minut udržováno) a při vybíjení na 0,3 V.

Z laserových snímků pořízených konfokálním mikroskopem během prvních cyklů lze pozorovat změny v aktivní hmotě při nabitém a vybitém stavu. V nabitém stavu byly pozorovány oblasti zvýšeného výskytu šedého olova – lesklejších míst a zároveň postupná redukce míst s krystalky síranu olovnatého, přičemž při vybíjení byl proces opačný – docházelo k postupnému poklesu množství šedého olova oproti tvorbě síranů.



Obrázek 5: Snímek článku ve vybitém stavu (vlevo) a nabitém stavu (vpravo)

4 ZÁVĚR

Využití konfokálního mikroskopu při pozorování růstu krystalů přináší nové dosud nerealizované možnosti sledování procesu sulfatace a díky tomu také rozsáhlejší porozumnění dané problematice.

Aby však mohl být samotný experiment realizován, bylo zapotřebí přizpůsobení cely koroznímu prostředí kyseliny sírové a hermetické utěsnění systému i pro korozní vodné prostředí. Jelikož pozorování probíhalo pouze na svrchní části elektrody bylo rovněž nutné navrhnout vhodnou geometrii elektrod a přizpůsobit režim cyklování vyhýbající se přílišnému přepolarizování elektrod a následnému nadměrnému plynování článku. Správnou cestou se ukázala výrazná redukce aktivní hmoty záporné elektordy, kdy oproti předchozím návrhům došlo k odstranění až 90 % hmotnosti.

Již z první série hlubokých cyklů byl patrný vliv nabíjení a vybíjení na povrch elektrody. Na snímcích byla patrná místa s výrazější redukcí vznikajících krystalů oproti jiným částem vzorku, kde docházelo k obtížnější redukci velkých krystalů. Na těchto málo aktivních místech lze předpokládat, že bude docházet během dalších experimentů k rychlejší sulfataci.

Pro lepší představu o vzniku krystalů síranu olovnatého je zapotřebí hluboké cykly opakovat a vyhodnotit delší časový úsek během cyklování. Současným předpokladem pro další fázi experimentu je po dostatečném prozkoumání hlubokého cyklování a ustálení kapacity článku zapojení režimu částečného nabití, kde budou intervaly mezi cykly mnohonásobně zkráceny a bude tedy docházet k postupnému rozšiřování neaktivních míst. V tomto režimu je také důležité zapojení vybraných aditiv.

REFERENCE

- [1] CENEK, M. *Akumulátory od principu k praxi*. Praha: FCC Public, 2003. ISBN 80-86534-03-0.
- [2] LEXT OLS4100 *Industrial Laser Confocal Microscopes*. OLYMPUS [online]. [cit. 2021-03-16]. Dostupné z: https://www.olympus-ims.com/cs/metrology/ols4100/
- [3] MOSELEY, P. T. Journal of Energy Storage 19: Understanding the functions of carbon in the negative active-mass of the lead-acid battery: A review of progress. 2018, 272 290. Amsteram: Elsevier Ltd.. ISSN 2352-152X.
- [4] PAVLOV, Detchko. *Lead-acid batteries: science and technology : a handbook of lead-acid battery technology and its influence on the product.* Amsterdam: Elsevier Science, 2011, 656 s. ISBN 978-0-444-52882-7.

Bakalářské projekty

Biomedicínské inženýrství a bioinformatika, Zpracování signálů, obrazu a dat

PPG SIGNAL QUALITY ASSESSMENT AND HEART RATE ESTIMATION

Enikö Vargová

Bachelor Degree Programme (3), FEEC BUT E-mail: xvargo02@vutbr.cz

Supervised by: Andrea Němcová

E-mail: nemcovaa@feec.vutbr.cz

Abstract: The presented paper describes an algorithm for signal quality assessment based on cluster analysis and also an algorithm for heart rate estimation from PPG signals. This work includes a database which comprises 48 PPG signals collected by a smartphone and 48 ECG signals recorded by an ECG recorder. The accuracy of the quality assessment is 97,5 % on the training set and 87,5 % on the test set. The average deviation of the estimated heart rate is 1,39889 bpm.

Keywords: mobile devices, PPG, photoplethysmography, ECG, heart rate, signal quality

1 ÚVOD

Tep je jedním ze základních projevů srdeční činnosti. Tepová frekvence (TF) poskytuje informace o zdraví pacienta a výkonnosti jeho srdce. Kromě toho je jedním z ukazatelů kardiovaskulárních onemocnění, která patří k nejčastějším příčinám úmrtí. [1][2]

TF je možné stanovit z chytrých telefonů, které jsou každodenní součástí našich životů. Chytrý telefon má k dispozici několik senzorů, které je možné využít pro stanovení TF (akcelerometr, mikrofon, fotoaparát). Výhoda použití mobilního telefonu spočívá v jeho přenosnosti a miniaturizaci. Práce se zabývá stanovením TF pomocí metody zvané fotopletysmografie, což je neinvazivní metoda detekující změny objemu krve, přičemž je využito fotoaparátu a LED diody na zadní straně mobilního telefonu.

Cílem práce bylo vytvořit databázi fotopletysmografických (PPG) signálů, navrhnout algoritmus pro určení jejich kvality a následně z kvalitních signálů stanovit TF.

2 SNÍMÁNÍ DAT

Pro účely této práce byla vytvořena databáze obsahující 48 10s PPG signálů a 48 10s referenčních EKG signálů. Bylo měřeno 12 subjektů (6 mužů a 6 žen) ve věku 21-61 let. Všechny měřené osoby podepsaly informovaný souhlas. Zisk biosignálů je součástí výzkumného projektu (Snímání a zpracování signálů za účelem monitorování zdravotního stavu a aktivity člověka), který byl schválen Etickou komisí UBMI FEKT VUT pod č. j. EK 05/2018. Součástí práce jsou anotace PPG signálů získané od pěti anotátorů, které slouží pro vyhodnocení úspěšnosti obou navržených algoritmů. Kompletní oanotovaná databáze byla nahrána na Physionet [4], kde je podrobněji popsána.

PPG signál byl nasnímán pomocí chytrého telefonu Xiaomi MI9. Měřený subjekt přiložil svůj prst na fotoaparát nacházející se na zadní straně telefonu tak, aby zároveň překrýval objektiv fotoaparátu i svítící LED diodu. Snímkovací frekvence byla nastavena na 30 Hz a rozlišení videozáznamu na 720 x 1280 px. Z videozáznamu načteného do prostředí Matlab byl extrahován průměr červené složky odpovídající surovému PPG signálu. Následně byl surový PPG signál invertován, jelikož fotoaparát pracuje se světlem odraženým. Současně s PPG signály byly snímány referenční EKG signály pomocí EKG záznamníku – Bittium Faros 360. Vzorkovací frekvence byla nastavena na 1000 Hz. Referenční signály byly též načteny do prostředí Matlab.

3 STANOVENÍ KVALITY PPG SIGNÁLU A URČENÍ TEPOVÉ FREKVENCE

Blokové schéma zahrnující zisk PPG signálu, stanovení jeho kvality i určení tepové frekvence je znázorněno na Obrázek 1. Jednotlivé bloky jsou barevně odlišeny. Oranžový blok odpovídá načtení videozáznamu a zisku invertovaného PPG signálu. Zelená část schématu popisuje algoritmus pro stanovení kvality signálu (Obrázek 3). Předzpracování a filtraci popisuje modrozelená část schématu. Samotná modrá část se věnuje stanovení TF pouze z kvalitních signálů (Obrázek 4).



Obrázek 1: Blokové schéma zahrnující zisk PPG signálu, stanovení jeho kvality i určení TF

3.1 ALGORITMUS STANOVENÍ KVALITY PPG SIGNÁLU

Databáze obsahující 48 signálů byla náhodně rozdělena na trénovací sadu o 40 signálech a testovací sadu obsahující 8 signálů. Obě sady obsahují kvalitní i nekvalitní signály. Pro odlišení obou zmíněných skupin signálů byly vypočítány tři příznaky: perfuze, poměr signálu k šumu a Shannonova entropie. Jelikož pro výpočet některých příznaků je třeba znát i filtrovaný signál, byla ještě před samotným stanovením kvality provedena filtrace.

Perfuze: poměr pulzující a nepulzující krve v periferní tkáni. Je dána vzorcem:

$$P_{SQI} = [(y_{max} - y_{min})/|\bar{x}|] \times 100, \tag{1}$$

kde y je filtrovaný PPG signál a $|\bar{x}|$ je statistický průměr surového PPG signálu. [3]

<u>Shannonova entropie</u>: udává míru s jakou se rozdělení pravděpodobnosti odchyluje od rovnoměrného rozdělení. Vyjadřuje ji vztah:

$$E_{SQI} = -\sum_{n=1}^{N} x[n]^2 \log_e(x[n]^2),$$
(2)

kde N je délka signálu a x je surový PPG signál. [3]

Poměr signálu k šumu: je vyjádřen jako poměr rozptylu signálu k rozptylu šumu. Je dán vztahem:

$$N_{SQI} = \sigma_{signal}^2 / \sigma_{noise}^2, \tag{3}$$

kde σ_{signal} je směrodatná odchylka absolutní hodnoty filtrovaného PPG a σ_{noise} je směrodatná odchylka filtrovaného signálu. [3]

Pro každý příznak byly vykreslené krabicové grafy sloužící k vizualizaci rozložení hodnot daného příznaku (Obrázek 2). V případě, že by se krabicové grafy pro kvalitní a nekvalitní signály v rámci jednoho příznaku příliš překrývaly, nemělo by smysl je používat pro klasifikaci. Krabicové grafy byly nejdříve vytvořeny pro trénovací sadu. Perfuze i Shannonova entropie se ukázaly jako vhodné příznaky. Následně byla tato volba ověřena na testovací sadě dat.



Obrázek 2: Krabicové grafy pro perfuzi a SNR (trénovací sada)

Rozřazení do skupiny kvalitní/nekvalitní bylo provedeno pomocí nehierarchické metody shlukování zvané k-means, která shlukuje data do předem definovaného počtu shluků (v tomto případě 2) dle podobnosti, případně rozdílnosti (Obrázek 3). Rozdílnost (podobnost) je brána jako vzdálenost mezi každým objektem a centroidem, což je střed reprezentující každý shluk. Objekt je přiřazen do toho shluku, jehož vzdálenost od centroidu je nejmenší.



Obrázek 3: Detailněji popsaný zelený blok z Obrázku 1 – stanovení kvality PPG signálu

3.2 ALGORITMUS STANOVENÍ TEPOVÉ FREKVENCE Z PPG SIGNÁLU

Algoritmus zahrnuje předzpracování signálů (normalizaci a filtraci). Nejdříve byl zredukován šum použitím Buttherworthovy pásmové propusti a následně byl použit Savitzky-Golay FIR filtr pro vyhlazení signálu [5]. Ve filtrovaných signálech byly detekovány peaky. Z pozic peaků byla vypočítána diference (Obrázek 4), která slouží pro určení okamžité TF. Celková stanovená TF byla určena jako medián z vektoru okamžité TF.



Obrázek 4: Detailněji popsaný modrý blok z Obrázku 1 – stanovení TF z PPG signálu

4 VÝSLEDKY A DISKUSE

Shluková analýza pro stanovení kvality PPG je znázorněna na Obrázek 5. Výsledky určení kvality signálu pomocí shlukové analýzy byly porovnány s anotacemi. Dále byla pro obě sady vypočítána senzitivita, specificita a přesnost. U trénovací sady přesnost odpovídá 97,5 %, senzitivita 100 % a specificita 90 %, protože byl jeden signál mylně označen jako kvalitní. U testovací sady dat je přesnost 87,5 %, senzitivita 83,33 % a specificita 100 %, jelikož byl jeden signál chybně klasifikován jako nekvalitní.



Obrázek 5: Výsledek shlukovací analýzy

Stanovené hodnoty TF byly porovnány s hodnotami TF určenými z referenčního EKG signálu. Průměrná odchylka vypočítaná pouze z kvalitních signálů je 1,3889 tepů/minutu. Nejvyšší odchylka u kvalitních signálů je 4 tepy/minutu a nejmenší 0 tepů/minutu. Dále byla pouze pro porovnání spočítaná průměrná odchylka pro všechny signály (resp. i nekvalitní), která odpovídá 6,1458 tepům/minutu. Záměrem je poukázat na fakt, že nekvalitní signály ovlivní výslednou TF, proto je nutné od sebe v první části práce kvalitní a nekvalitní signály odlišit.

5 ZÁVĚR

Algoritmus pro stanovení kvality PPG signálů dosahuje dobrých výsledků na trénovací i validační sadě, kde byl v obou případech špatně zařazen jeden signál. Součástí práce je také algoritmus pro stanovení TF z kvalitních PPG signálů. Algoritmy byly testovány na námi vytvořené databázi, kterou jsme oanotovali a publikovali na PhysioNetu [4]. Průměrná odchylka stanovené TF je 1,3889 tepů/minutu, což je hodnota nižší než 1,98 tepů/minutu, kterou uvádějí autoři článku [6]. Naopak průměrná odchylka z článku [7] je 0,4093 tepům/minutu, nízká odchylka je nejspíše dána jiným přístupem (použití stacionární vlnkové transformace).

REFERENCES

- [1] MOUREK, Jindřich. *Fyziologie: učebnice pro studenty zdravotnických oborů*. 2., dopl. vyd. Praha: Grada, 2012. Sestra (Grada). ISBN 978-80-247-3918
- [2] ROSOLOVÁ, Hana. *Preventivní kardiologie: v kostce*. Praha: Axonite CZ, 2013. Asclepius. ISBN 9788090489950.
- [3] ELGENDI, Mohamed. Optimal Signal Quality Index for Photoplethysmogram Signals. *Bioengineering* [online]. 2016, 3(4) [cit. 2021-01-26]. ISSN 2306-5354. Dostupné z: doi:10.3390/bioengineering3040021
- [4] Nemcova, A., Smisek, R., Vargova, E., Maršánová, L., Vítek, M. & Smital, L. (2021). Brno University of Technology Samrthphone PPG Database (BUT PPG) (version 1.0.0). *Physio-Net*. https://doi.org/10.13026/7vy8-av04
- [5] CHATTERJEE, Ayan a Uttam Kumar ROY. PPG Based Heart Rate Algorithm Improvement with Butterworth IIR Filter and Savitzky-Golay FIR Filter. In: 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) [online]. IEEE, 2018, 2018, s. 1-6 [cit. 2021-01-26]. ISBN 978-1-5386-5550-4. Dostupné z: doi:10.1109/IEMENTECH.2018.8465225
- [6] SIDDIQUI, Sarah Ali, Yuan ZHANG, Zhiquan FENG a Anton KOS. A Pulse Rate Estimation Algorithm Using PPG and Smartphone Camera. *Journal of Medical Systems* [online]. 2016, 40(5) [cit. 2021-01-26]. ISSN 0148-5598. Dostupné z: doi:10.1007/s10916-016-0485-6
- [7] NĚMCOVÁ Andrea, Martin VÍTEK. *Application of SWT for heart rate monitoring using smartphone camera*. [online]. Mikulov: IEEE Student Branch Conference, 2017.

THE EFFECT OF QUALITY TRIMMING ON JOINING PAIRED-END READS IN MICROBIOME DATA ANALYSIS

Kristýna Heřmánková

Bachelor Degree Programme (3), FEEC BUT E-mail: xherma30@stud.feec.vutbr.cz

Supervised by: Karel Sedlář

E-mail: sedlar@feec.vutbr.cz

Abstract: The main goal of microbiome research is to determine the microbial composition of a target sample. In successfully performed research, a cause of a disease can be found or pathogens in an environmental sample can be revealed. The correct evaluation of microorganisms present in the analysed sample is therefore required, but unfortunately not that often reached. The accuracy of the resulting composition can be affected already in pre-processing steps of the analysis. The frequent issue that can negatively affect the reliability of research is data loss. This loss is most common in the case of the paired-end reading method. The occurrence of data loss can be caused by the need of joining this pair of sequences into one continuous sequence. The need of joining reads with low-quality ends together can lead to counts reduction of sequences in an available dataset, and therefore unrealistic results can be obtained. This paper shows, how quality trimming can affect the number of lost sequences during the reads joining step.

Keywords: NGS, quality trimming, 16S rRNA, amplicon

1 INTRODUCTION

For purpose of the microbial study, which aims to determine the taxonomic diversity of a sample, the amplicon sequencing approach is used. The principle of this approach lies in sequencing only a target gene or its regions instead of the whole genome. This gene, called the marker gene, should be specific enough, to distinguish ideally each species. For the domain of *Bacteria*, the 16S rRNA gene is used. The suitability of this gene for taxonomy classification lies in its presence in every organism and its hypervariable, and also conserved regions. The most significant information on taxonomic diversity would be reached by sequencing the whole 16S rRNA gene. Unfortunately, current sequencing technologies, which are capable sequence its length of approximately 1600 bp, are still in the development process because of their high sequencing error rate. In this case, the next-generation sequencing technologies (NGS), which generate reads of significantly shorter length, are used. This means, that only a few regions of the 16S rRNA gene can be sequenced.

2 METHODS

The widely used sequencing technology for amplicon sequencing purposes is Illumina. This technology offers a few platforms for various applications. For reaching the longest possible reads, Illumina also offers paired-end sequencing. This method generates two reads from both ends of one sequence. Each read can be of length up to 300 bp, depends on the selected platform, so the target sequence can be almost twice longer than sequencing with the single-end method.

For a longer resulting sequence obtained from the paired-end method, overlapping of generated reads is required. The length of an overlapping region can vary. If the whole length of reads overlaps, the resulting sequence is not longer, but can only reaches higher quality. If reads overlap only partially, the long continuous sequence is reached. This requires the process of joining these two reads together.

Because of the different orientation of the sequencing process, reads are joined with their ending tails. The longer the sequenced region is, the higher probability of error occurrence is. With the higher error rate, the number of low-quality bases increases. This can lead to the failure of the joining step. The following text describes, how reads can be treated to prevent unsuccessful reads joining.

2.1 QUALITY TRIMMING

Before the quality trimming was done, specific primers were removed. This step was performed in QIIME 2 (Version 2020.8) [1] (Quantitative Insight Into Microbiome Environment), which is the most used open-source platform for microbiome analysis.

A Phred quality score defines the accuracy of base calling. This accuracy is measured for each base in the read and the corresponding score is recorded. FASTQ format then stores both, the sequence of nucleotides and the related score. Generally, bases with the quality under the value of 20 are considered low-quality, but the value can vary depending on usage. Table 1 shows the selected Phred scores, and corresponding accuracies calculated using formula:

$$A = 1 - 10^{\frac{-Q}{10}} \tag{1}$$

where A is the base call accuracy of the Phred score Q.

Phred quality score	Base call accuracy
10	90 %
20	99 %
30	99,9 %
40	99,99 %

Table 1:	The base call accuracy and the corresponding score.
----------	---

For trimming low-quality ends of paired-end reads, Trimmomatic (Galaxy Version 0.38.0) [2] [3] was used. This tool is designed for trimming sequences from Illumina platforms. Trimmomatic offers trimming of single-end or paired-end reads in FASTQ format and allows a few options, how to trim these reads. In this paper, the TRAILING method was used. The method only needs target files with paired-end reads and a quality threshold. The TRAILING method begins at the end of reads and tracks each base quality. If the quality is under the selected threshold, the base is removed. Trimming continues towards the beginning of the read and stops when the base quality is at least equal to the threshold. This is done for both paired reads.

2.2 JOINING READS

Data from Trimmomatic were imported back to the QIIME 2 and a joining step was performed. A few parameters can be set for joining reads. The –p-maxdiffs parameter can be important in the case of successful reads joining. The default value of this parameter is 10, which means that only 10 mismatches in the overlapping region are allowed. If a match is recorded, the resulting sequence has at this position higher quality. If there is a mismatch between reads, the base with higher quality is chosen and the resulting quality of the base decreases [4]. Because of the subsequent analysis, sequences still need to be of high quality, and thus the default value of this parameter was used.

2.3 MATERIALS

Sequencing data used in this paper were sequenced with the Illumina MiSeq platform for generating paired-end reads of length 2x300 bp. These data consist of three bacterial communities artificially prepared in Veterinary Research Institute with the purpose to provide testing datasets for a chimera detection tool. Sequences are amplicon sequence variants (ASVs) of 16S rRNA of bacteria present in the sample. The overlapping region is approximately 140 bp long, so the resulting continuous sequence is of length about 460 bp.

3 RESULTS

Table 2 contains numbers of paired-end reads in available datasets before the mentioned exact process was performed.

Dataset	Number of reads
P2	103 606
P3	71 610
P4	64 029

Table 2:Datasets and their counts of reads.

In Figure 1A, the impact of allowed mismatches in overlapping regions is shown. The default value of 10 was chosen, because of a relatively high number of joined sequences, and also high-quality sequences for further analysis.

As Figure 1B shows, sequences without the quality trimming step have a bigger issue with joining pairs together. Unlike the untrimmed reads where only 68 % of all reads in available datasets passed the joining step, 90 % of quality trimmed reads were successfully joined. The reason why this happens can be, as mentioned previously, that low-quality ends in overlap region can mismatch. If lots of mismatches are recorded, sequences are not allowed for passing the joining step. On the opposite, if the match between bases is recorded, the resulting sequence is of high quality.





Figure 2 shows a few quality thresholds used in the Trimmomatic tool for quality trimming and their impact on counts of successfully joined reads. Value of threshold ensures which bases at the

end of reads should be retained and which are supposed to be trimmed. The orange highlighted threshold was selected as it achieved the highest number of joined reads.



Figure 2: The difference between counts of successfully joined reads with the chosen threshold.

4 CONCLUSIONS

The further analysis of microbiome studies lies in subsequent processing steps as dereplication, where every ASV is identified, chimera detection, which should reveal spurious organisms, and operational taxonomic unit (OTU) picking, where ideally every OTU represents one species. In all these steps, the removal of insufficient sequences is performed, and thus the unnecessary data loss needs to be inhibited.

Also, the result of microbial research, which aims to diversity analysis of a sample, depends on the number of sequences that went through the whole analysis. The higher the number of lost sequences is, the probability of incorrectly revealed microbial community increases. On the opposite, if the number of appropriately treated sequences is the highest possible, results will be probably more realistic.

As results showed, quality trimming has a positive impact on obtained reads for further analysis, and therefore should be included as the pre-processing step on Illumina paired-end reads.

REFERENCES

- [1] BOLYEN, Evan, Jai Ram RIDEOUT, Matthew R. DILLON, et al., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* [online]. **37**(8), 852-857. ISSN 1087-0156. DOI: 10.1038/s41587-019-0209-9
- BOLGER, Anthony M., Marc LOHSE a Bjoern USADEL, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15), 2114-2120. ISSN 1460-2059. DOI: 10.1093/bioinformatics/btu170
- [3] AFGAN, Enis, Dannon BAKER, Bérénice BATUT, Marius VAN DEN BEEK, Dave BOU-VIER, Martin ECH, John CHILTON, Dave CLEMENTS, Nate CORAOR, Björn A. GRÜNING, Aysam GUERLER, Jennifer HILLMAN-JACKSON, Saskia HILTEMANN, Vahid JALILI, Helena RASCHE, Nicola SORANZO, Jeremy GOECKS, James TAYLOR, Anton NEKRUTENKO a Daniel BLANKENBERG, 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research.* 46(W1), W537–W544. ISSN 13624962. DOI:10.1093/nar/gky379
- [4] EDGAR, Robert C. a Henrik FLYVBJERG, 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*. **31**(21), 3476-3482. ISSN 1367-4803. DOI:10.1093/bioinformatics/btv401

ADVANCED ANALYSIS OF MOVING OBJECTS IN THE IMAGE

Ivan Medynskyi

Bachelor Degree Programme (3), FEEC BUT E-mail: xmedyn00@stud.feec.vutbr.cz

Supervised by: Martin Kiac

E-mail: xkiacm00@stud.feec.vutbr.cz

Abstract: This work is focused on the image processing using the OpenCV library and detecting moving objects in video using convolutional neural networks. The created application can detect moving objects in the video and contains additional functionality. The application includes the YOLO convolutional model, which helps to detect moving objects.

Keywords: OpenCV, CNN, neuronové sítě, počítačové vidění, detekce pohybujících se objektů.

1 ÚVOD

Člověk dostává očima hodně informací o vnějším světě a poté je umí velmi efektivně zpracovávat. V současnosti se naučily také počítače analyzovat obraz přijatý kamerou, což získalo název počítačové vidění. Postupem času si počítačové vidění začalo získávat velkou popularitu a dnes se používá v různých oblastech jako automatizace, medicina, ochrana, ve výrobě pro zlepšení a kontrolu kvality výrobku a ve spoustě dalších oblastí.

Tato práce je věnována analýze pohybujících se objektů v obraze. Cílem práce je implementace GUI aplikace, která bude schopná detekovat objekty s využitím konvolučních neuronových sítí. Navíc bude aplikace podporovat instalace videa a pomocí dalších nastavení bude možné rozpoznat objekty: auta, lidi, motorky, autobusy apod.

2 ZPRACOVÁNÍ OBRAZU

Zpracování obrazu je část počítačového vidění. Celý proces je možné následovně rozdělit:

- snímání a digitalizace obrazu,
- předzpracování obrazu,
- segmentace obrazu,
- popis objektů,
- klasifikace objektů [1].

2.1 SNÍMÁNÍ A DIGITALIZACE OBRAZU

Dnes má každý člověk ve svém mobilu kameru a může v jakýkoliv moment něco vyfotit nebo natočit video. Všechny fotografie nebo videa se ukládají do paměti telefonu a lze je chránit dlouhou dobu. V minulosti, když se využívaly kazety, se video zapisovalo pomocí analogového signálu. Ale kvůli demagnetizaci kazety mohou být některé soubory ztraceny bez možnosti obnovení. Pro uložení a obnovu souboru po dlouhé době je vhodným řešením využití digitálního signálu. Proto byl vymyšlen proces digitalizace, tzn. proces převodu spojitého analogového signálu na signál digitální [2].

2.2 PŘEDZPRACOVÁNÍ OBRAZU

Cílem předzpracování obrazu je jeho zlepšení pro další zpracování. Existuje několik metod, které můžeme potřebovat pro opravu nebo zlepšení původního obrázku:

- filtrace šumu,
- jasová transformace,
- geometrická transformace.

2.3 SEGMENTACE OBRAZU

Segmentace slouží k analýze obrazu a popisu jeho obsahu pro další zpracování. Segmentace může být kompletní, nebo částečná. Rozdílem je jen to, že při částečné segmentaci nemusí části obrazu (segmenty) přímo korespondovat s objekty. Existuje několik metod pro provedení segmentace:

- prahování obrazu,
- detekce hran.

2.4 POPIS A KLASIFIKACE OBJEKTU

Po segmentaci je potřeba objekt co nejlépe popsat. Pro jeho popis může posloužit charakteristika objektu. Může to být například jas, kontrast, rotace objektu apod. Je velmi důležité najít co nejvíce příznaků, které budou sloužit pro klasifikaci objektu.

Klasifikace objektu probíhá podle nalezených příznaků. Jejím principem je rozdělení příznaků do soukromých tříd, s jejichž pomocí je neuronová síť schopná se učit. Úkolem klasifikace obrázků je pořídit původní obrázek a přiřadit mu jeho třídu (kočka, pes apod.).

3 NEURONOVÁ SÍŤ

Teoreticky můžeme říct, že neuronová síť (NS) je jako mozek člověka, kde jsou miliony neuronů, jež si sdělují informace pomocí elektronických impulzů. Neurony se spojují mezi sebou pomocí synapsí. O neuronech a synapsích bude pojednáno níže v práci. Struktura neuronových sítí je téměř stejná jako v biologii. Pomocí této struktury mohou stroje analyzovat, pamatovat si a reprodukovat informaci z paměti. Na obrázku 1 je vidět struktura neuronové sítě [3].



Obrázek 1: Struktura neuronové sítě.

3.1 NEURON A SYNAPSE

Neuron je buňka, která přijímá informaci a provádí nad ní základní výpočty a rozpracovanou informaci posílá dál. Když je využito velké množství neuronů, používají se vrstvy. První – vstupní, vrstva přijímá informaci. Druhá – skrytá, zpracovává předanou informaci a posílá ji dál. Třetí - výstupní, zobrazuje výsledek. V případě zpracování velkého množství informací může být využito několik skrytých vrstev.

Jak bylo uvedeno výše, synapse je spoj mezi dvěma neurony. Každá synapse má různou váhu, a ta je také jediný a nejdůležitější parametr. Čím přesněji je vybraná váha, tím lepší je výsledek. První váhy se vybírají náhodně a pomocí výpočtu chyby NS se můžeme dozvědět funkčnost NS a zvolit vhodnější váhy. Vhodným algoritmem pro výpočet chyby je tzv. algoritmus backpropagation.

3.2 UČENÍ NEURONOVÝCH SÍTÍ

Učení sítě je proces výběru správných parametrů – váhy synapse, který se obyčejně provádí pomocí adaptačních algoritmů, jež můžeme rozdělit na dvě skupiny: učení s učitelem a učení bez učitele [3].

Pro učení s učitelem potřebujeme dataset, ve kterém budou rozpracovány příklady, tzn. že budou připravené úkoly a správné odpovědi. NS probírá všechny správné odpovědi a nachází mezi nimi závislosti. Dataset slouží jako učitel a pomáhá trénovat NS. Po době, kdy se data z datasetu dostala na vstup a byla zpracována neuronovou sítí, se vypočítá chyba NS. Při velké hodnotě chyby se mění parametr váhy synapse a stejný proces probíhá znovu, dokud hodnota chyby nebude v námi požadovaném rozsahu.

V případě učení bez učitele, probíhá trénování pomocí shlukování objektů do různých tříd. Například na vstup pošleme 10 000 různých objektů s charakteristikou, budou to například studenti. Program je rozdělí na různé třídy podle známých parametrů získaných při popisu a klasifikaci objektu. Tento typ se nevyužívá k učení neuronové sítě často.

3.3 KONVOLUČNÍ NEURONOVÁ SÍŤ

Konvoluční neuronová síť (CNN) je část hloubkového učení (anglicky deep learning) neuronové sítě. Díky své metodě konvolucí a struktuře je vhodná pro rozpoznání obrazu, jeho detekci a klasifikaci. Struktura CNN se skládá z mnoha vrstev: konvoluční vrstva, pooling vrstva a plně propojená vrstva.

Hlavní funkcí CNN je proces konvoluce. Tento proces je doprovázen pomocí tzv. filtru. Tímto filtrem je matice s váhami, která hraje nejdůležitější roli v konvoluci a učení CNN. Metoda konvoluce pracuje pomocí násobení každého elementu matice obrazu na matici filtru [4]. Jsou několik modelů CNN: R-CNN, SSD, YOLO.

4 ZÁVĚR

Pro zpracování obrazu jsem si vybral knihovnu OpenCV. Jsou v ní hotové nástroje a metody určené pro zpracování obrazu a jeho analýzy.

Pro řešení problémů detekcí pohybujících se objektů v obraze jsem vybral CNN model *yolov4-tiny* a připravil jsem vlastní datovou sadu. Vlastní datová sada byla vytvořena kvůli tomu, že po trénovaní známe datove sady COCO byla metrika modelu mnohem horší. Trénování probíhalo na grafické kartě NVIDIA GTX 1050, která nepodporuje trénování většího modelu CNN jako je *yolov4*, na kterém bych dosáhl lepších metrických hodnot. Snímky pro datovou sadu byly pořízeny z dopravních kamer a anotovány programem *Yolo_mark*. Celkem bylo použito kolem 1000 snímků pro trénování a 100 pro testování NS. Pro trénování modelu YOLO jsem využil framework *Darknet*. Výhodou tohoto frameworku je možnost provádět výpočty na CPU a GPU, což je velmi důležité a zrychlí to práci. Ve výsledku byl vytvořen program, který dokáže rozpoznat: lidi, auta, autobusy, jízdní kola, motocykly, dodávky a kamiony. V programu můžete otevřít jakékoli video ve formátu mp4, vybrat typ objektu, který má být detekován, a zobrazit hodnotu pravděpodobnosti, že se jedná o stejný objekt, který jste vybrali. Nakonec můžete vybrat ROI (Region Of Interest) a program bude detekovat objekty pouze v této oblasti.

Následující tabulka znázorňuje výsledné metriky natrénovaného modelu YOLO. Nejvhodnější natrénovaný model dosahuje průměrnou přesnost mAP 61.24%, recall 0.75, f-score 0.73 a průměrnou hodnotu IoU 55.12%. Z tabulky je možné vidět ze nejlepší hodnoty přesnosti dosahuje třída "Auta". Třída dosáhla hodnotu TP 411 a zároveň hodnota FP dosáhla 128. Zároveň nejmenší hodnoty přesnosti dosáhla třída "Jízdní kola". Je to pravděpodobně z důvodu, že vytvořený dataset obsahuje nízký počet snímků, kde se zmíněná třída "Jízdní kola" vyskytuje. Natrénovaný model dosahuje průměrnou hodnotu 63.7 FPS na grafické kartě NVIDIA GTX 1050.

Název třídy	AP [%]	TP	FP
Auta	77,55	411	128
Lidi	58,76	22	28
Kamiony	64,31	22	14
Dodávky	46,14	19	22
Jízdní kola	30,90	2	4
Motocykly	60,00	3	3
Autobusy	91,02	16	3

Tabulka 1:Výsledky trénování YOLO modelu.

REFERENCE

- [1] KIAC, M. Využití moderních metod zpracování obrazu při kontrole laboratorních procesů. Brno, 2019. Diplomová práce. FEKT VUT Brno.
- [2] FIŘT, J., HOLOTA R. Digitalizace a zpracovaní obrazu. Plzeň. Nové technologie výzkumné centrum. Dostupné z: http://home.zcu.cz/~holota5/publ/DigZprO.pdf
- [3] HAYKIN, Simon. Neural networks: a comprehensive foundation. Second Edition. Ontario: Prentice Hall PTR, 1994. ISBN 978-0132733502
- [4] NIELSEN, Michael. Neural Networks and Deep Learning. Online kniha. Dostupné z: http://neuralnetworksanddeeplearning.com/index.html

TRAFFIC SIGN CLASSIFICATION USING DEEP LEARNING

Marek Sicha

Bachelor Degree Programme (3), FEEC BUT E-mail: xsicha02@stud.feec.vutbr.cz

Supervised by: Tomáš Bravenec E-mail: xbrave01@vutbr.cz

Abstract: The thesis focuses on the classification of traffic signs in images and video sequences. The goal is real-time processing and usage of software in the vehicle. Neural networks and the Python programming language were chosen to solve the problem. To solve the problem a machine learning method was chosen, more precisely a convolutional neural network. A neural network in the Python programming language was created for the classification of traffic signs, using the Keras and Tensorflow libraries. The neural network architecture is chosen for optimization for use on a single-board computer with limited performance.

Keywords: classification, neural networks, traffic signs

1 ÚVOD

Tato práce se zabývá rozpoznáváním dopravních značek v obraze. Detekce a rozpoznávání dopravních značek je jeden z hlavních předpokladů pro vytvoření autonomního řízení automobilů. Samozřejmě detekce a rozpoznávání dopravních značek není jenom čistě pro autonomní řízení aut. Může být použito i jako jeden za základních asistenčních prvků pro dopravní prostředky, ve kterých je stále ještě řidič člověk. Tento asistenční prvek by mohl upozornit řidiče např. na překročení maximální rychlosti nebo přehlednutí zákazové značky a včas zabránit dopravní nehodě.

Cílem této práce je navrhnout systém pro klasifikaci dopravních značek využívající konvolučních neuronových sítí se zaměřením na efektivitu výpočtu, pro možnost využití systému v automobilu.

Pro vytvoření softwaru byl zvolen programovací jazyk Python, který obsahuje nespočet knihoven pro úpravu vstupních datových sad. Datové sady jsou nedílnou součástí pro vytvoření neuronové sítě klasifikátoru. Vytvoření klasifikátoru bylo za pomocí neuronových sítí, které jsou v dnešní době stále populárnější a to díky stále rostoucímu výpočetnímu výkonu.

2 KLASIFIKACE DOPRAVNÍCH ZNAČEK

Klasifikátory slouží k rychlé identifikaci v předem vyznačené části obrázku. Pokud tedy chceme klasifikovat dopravní značky v reálném čase například z dopravní kamery, je nutné nejprve daný obrazový signál zpracovat vhodným detektorem, který nám určí oblast, ve které se dopravní značka nachází. Klasifikátory jsou tedy takový podpůrný prostředek pro detektory a z pravidla mají jednoduchou architekturu, ale vysokou pravděpodobnost klasifikace daného objektu, tedy dopravní značky.

2.1 METODY KLASIFIKACE DOPRAVNÍCH ZNAČEK

Způsoby jak klasifikovat dopravní značky můžeme rozdělit do dvou kategorií.

První kategorie jsou metody, které jsou založeny na klasifikaci podle tvaru nebo barev výhodou této metody je, že dopravní značky mají své specifické barvy a tvary, které se běžně v přírodě nevysky-

tují. Dalším způsobem je porovnání dopravní značky s již existující databází značek, nevýhoda této metody je její časová náročnost a problém s rozpoznáváním deformovaných dopravních značek. [6]

Jako další metody jsou metody strojového učení, neboli vytvoření neuronové sítě. Při vytváření neuronové sítě je potřeba dataset, podle kterého se síť naučí specifické prvky pro dané třídy. Neuronová síť dostane vstupní data, podle výstupních dat sítě a dat, které jsme očekávali, určí chybu a provede korekci. Nově nastaví hodnoty prahů a vah. Celý tento cyklus opakuje do chvíle, dokud nenajde takovou konfiguraci prahových hodnot a vah, která odpovídá naší stanovené chybě. [1, 5] Nevýhoda této metody je pokud chceme síť rozšířit na více tříd, je nutné provést celý proces učení znovu, zatímco při porovnávacích metodách se připíše jen pár podmínek.

3 KLASIFIKÁTOR DOPRAVNÍCH ZNAČEK

Pro vytvoření klasifikátoru byla zvolena metoda strojového učení, přesněji konvoluční neuronová síť. Model neuronové sítě klasifikátoru byl vytvořen pomocí knihovny Tensorflow, jejíž novější verze již obsahují knihovnu Keras v základním balíčku. Výhoda knihovny Tensorflow je její podpora pro jednodeskové počítače.

3.1 ARCHITEKTURA KLASIFIKÁTORU DOPRAVNÍCH ZNAČEK

Zvolen byl sekvenční model, který umožnuje skládat vrstvy lineárně za sebe, každá vrstva je propojena jen s předchozí a následující vrstvou. Mimo to také obsahuje jen jeden vstup a výstup, což pro klasifikaci dopravních značek není žádné omezení. [2]



Obrázek 1: Architektura klasifikátoru.

Vstupní obrázek klasifikátoru je o rozměrech 32x32x1, obrázek je ve stupních šedi. Následuje série konvolučních a sdružovacích vrstev viz obrázek 1. V architektuře jsou použity dvě dropout vrstvy, které se uplatňují jenom při trénování sítě. Zabraňují přetrénování sítě, tím že náhodně nastaví vstup pro další vrstvu na 0. Poslední část architektury jsou plně propojené vrstvy, poslední vrstva obsahuje 43 uzlů, což odpovídá počtu tříd. Aktivační funkce poslední vrstvy je softmax, které se používá pro jedno třídovou identifikaci, určuje pravděpodobnost, že objekt spadá do určité třídy, součet všech pravděpodobností tříd je 1. [2] Až na poslední vrstvu, která má funkci softmax, jsou použity aktivační funkce ReLU, jejíž výhodou je její jednoduchost, oproti jiným funkcím neuvažuje záporné hodnoty.

Optimalizátor je algoritmus, který mění hodnoty neuronové sítě, jako jsou váhy a míra učení, aby ztráty byly co nejmenší. V této architektuře byl použit optimalizátor Adam, který používá metodu gradientního sestupu. Metoda spočívá ve výpočtu první derivace ze ztrátové funkce a podle toho předpovídá, jak by se měli váhy měnit, aby funkce mohla dosáhnout minima. [4] Metoda je výpočetně efektivní, má malý požadavek na paměť.

Ztrátová funkce vyjadřuje aktuální ztrátu sítě během trénování. Výstupem funkce jsou ztráty, které určují rozdíl mezí předpokládaným výstupem a daným výstupem neuronové sítě. Pro klasifikaci do více než dvou tříd se používá funkce Categorical Cross-entropy, kdy každé třídě odpovídá jeden výstupní neuron.

3.2 TRÉNOVÁNÍ KLASIFIKÁTORU

Trénování klasifikátoru bylo provedeno na Německém datasetu pro klasifikaci značek, který obsahuje kolem 30 tisíc obrázků značek ve 43 různých třídách. Dataset byl rozdělen na tři množiny, pro trénování bylo použito 20 tisíc obrázků, testovací množina obsahuje 3 tisíce snímků a ověřovací množina 5 tisíc obrázků. Trénování bylo provedeno pro 30 cyklů, kdy do každého cyklu vstupovaly všechny obrázky z množiny pro trénování. Trénování bylo provedeno, jak pro model ve stupních šedi, tak i pro model v RGB. Dále je použit model pracujících ve stupních šedi, a to z důvodu menšího počtu parametrů, a tedy menší náročnosti na výpočetní výkon. Snížením výpočetního výkonu jsme ale nesnížili přesnost klasifikace, protože oba modely dosahovali stejných výsledků. Pro model byly v průběhu trénování tvořeny grafy pro zobrazení jeho přesnosti a ztrát pro každý cyklus viz obrázek 2.



Obrázek 2: Přesnost a ztráty klasifikátoru v průběhu trénování.

Přesnost a ztráty byly zaznamenávány pro každý cyklus trénování, podívejme se nejdříve na přesnost, přesnost klasifikátoru trénovací sady je již po třech cyklech přes 90 %, pro ověřovací sadu je to již po prvním cyklu. Průběh odpovídá rostoucí logaritmické funkci, která dosahuje po 30 cyklech přesnosti až 99 %, jak pro testovací, tak i pro ověřovací sadu. Průběh ztrát má opačný charakter, průběh pro trénovací sadu je klesající logaritmická funkce, pro kterou jsou ztráty na konci trénování menší než 0,3 pro trénovací sadu, pro ověřovací sadu jsou ztráty pod 0,2.

3.3 TESTOVÁNÍ KLASIFIKÁTORU

Vstupní obrázek nejprve prochází úpravami, tak aby se shodoval s obrázky, které byly použity při testování. Nejprve je změněn jeho rozměr na 32x32, poté je převeden do stupňů šedi a je změněn jeho kontrast.



Class: Hlavni cesta (v obci) Probability : 99.9%

Obrázek 3: Příklad správné klasifikace.

Na obrázku 3 je uveden příklad správné klasifikace. Dopravní značka pro klasifikaci je značně ovlivněna světelným vlivem v tomhle případě se jedná o sluneční svit a v jejím středu se nachází objekt, který není součástí tohohle typu dopravní značky, i tak ale klasifikátor určil správně typ značky s pravděpodobností blížící se ke 100 %. Ve srovnání s přesnostmi ostatních systémů testovaných nad

	Metoda	Úspěšnost klasifikace
Navržené řešení	Konvoluční Neuronová síť	99,54 %
IDSIA	Konvoluční Neuronová síť	99,46 %
sermanet	Konvoluční Neuronová síť	98,31 %
CAOR	Náhodný les	96,14 %

Tabulka 1: Srovnání přesností navrženého systému klasifikace s ostatními řešeními

stejným datasetem vychází navržená neuronová síť nejlépe. Srovnání úspěšnosti klasifikace ostatních systémů je v Tabulce 1.

Nesprávná klasifikace se v testovacím setu vyskytovala u míň než 2 % případů. A ve většině případů se správný typ značky umisť oval s druhou nejvyšší pravděpodobností, která se od nesprávného typu lišila jen o pár desítek procent.

Testování bylo prováděno na grafické kartě Nvidia GeForce GTX 1050 Ti, kde čas pro klasifikaci jednoho obrázku se pohyboval kolem 0,1 sekundy.

4 ZÁVĚR

Cílem práce bylo prostudovat možné způsoby ke zpracování obrazových signálů, se zaměřením na klasifikaci dopravních značek v obrazcích a videosekvencích. Pro dosažení tohohle cíle byla použita neuronová síť a programovací jazyk Python. Pomocí konvoluční neuronové sítě byl vytvořen klasifikátor dopravních značek, pro model ve stupních šedi. Klasifikátor byl trénován i testován na německém datasetu, určený pro klasifikaci dopravních značek. Úspěšnost klasifikace modelu testovacího datasetu se pohybuje kolem 99,5% a čas pro klasifikaci jednoho snímku je kolem 0,1 sekundy.

Datová sada použita pro trénování klasifikátoru obsahovala ve 43 třídách přes 30 tisíc obrázků, za různých světelných podmínek a pro různé deformace značek. Množina datové sady byla rozdělena do tří sad: testovací, trénovací a validační.

REFERENCE

- [1] O'Shea, Keiron & Nash, Ryan. (2015). An Introduction to Convolutional Neural Networks. Ar-Xiv e-prints.
- [2] Keras API reference. Keras [online]. [cit. 2020-12-05]. Dostupné z: https://keras.io/api/
- [3] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In Proceedings of the IEEE International Joint Conference on Neural Networks, pages 1453–1460. 2011.
- [4] Overview of various Optimizers in Neural Networks. Towardsdatascience [online]. June 9 [cit. 2020-12-05]. Dostupné z: https://towardsdatascience.com/overview-of-various-optimizersin-neural-networks-17c1be2df6d5
- [5] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9
- [6] Saadna, Yassmina & Behloul, Ali. (2017). An overview of traffic sign detection and classification methods. International Journal of Multimedia Information Retrieval. 6. 1-18. 10.1007/s13735-017-0129-8.

THE PROCESS OF CT IMAGE CLASSIFIER OPTIMIZATION FOR PEDATRIC SPINE SEGMENTATION

Táňa Kohúteková

Bachelor Degree Programme (1.), FEEC BUT E-mail: xkohut10@stud.feec.vutbr.cz

Supervised by: Jan Mikulka

E-mail: mikulka@vut.cz

Abstract: Computed tomography (CT) is a method commonly used when generating images of a spine diagnosed with scoliosis. The main reason for this common employment is found in its practicality: the spine in the image is clearly visible. However, while portraying, a child is exposed to a dose of radiation. To eliminate a risk of children being negatively affected, this article is trying to find the best technique of segmentation, that could be applied also while using magnetic resonance imaging (MRI). It implements a 3D Slicer equipped with *Support Vector Machines (SVM) Classifier* plug-in to search for the best feature combination. Afterwards, segmentation results are compared to ground truth image, where the area of hard tissue is determined. The aim is to get *Dice* coefficients as close as possible to 1.

Keywords: segmentation, spine, scoliosis, image processing

1 ÚVOD

Skolióza je patologické zakrivenie chrbtice. Zobrazovacie metódy v diagnostike, predoperačných a pooperačných úkonoch pri skolióze majú kľúčovú úlohu.[1] Snímky, zhotovené počítačovou tomografiou (CT) poskytujú lepšie zobrazenie tvrdých tkanív, preto sú výhodnejšie a segmentácia chrbtice je presnejšia. Nevýhodou počítačovej tomografie však je rádioaktívne žiarene, ktorému sú detskí pacienti počas snímania vystavení. Použitie magnetickej rezonancie (MRI) úplne eliminuje rádioaktívne žiarenie, no neposkytuje dobrý obraz na jednoduchú a presnú segmentáciu. Cieľom článku je popis optimalizácie vstupného vektoru pre *Support Vector Machines (SVM) Classifier* zloženého z obrazových charakteristík pre získanie čo najkvalitnejšieho výsledku segmentácie tvrdých tkanív v programe *3D Slicer*.

2 POSTUP SPRACOVANIA OBRAZU

Pri vypracovávaní tejto práce boli k dispozícii CT a MRI snímky detského pacienta z Fakultnej nemocnice Brno. Jednotlivé záznamy z oboch typov vyšetrení sú zložené z 231 rezov. Pri spracovaní bol využitý *SVM Classifier*, ktorý bol vytvorený v rámci bakalárskej práce vo forme plug-inu.[2] Na samotnú segmentáciu využíva metódu podporných vektorov, ktorá dokáže rozdeliť dáta do 2 tried, respektíve nájsť priaznivú deliacu hranicu medzi dvoma rôznymi hodnotami. V tomto prípade ide o rozdelenie pixelov na pixely, ktoré obsahujú kosť a tie, ktoré obsahujú okolie.**Chyba! Nenalezen zdroj odkazů.**

Spracovanie obrazu prebieha v niekoľkých krokoch (Obrázok 1). Po zhotovení CT alebo MRI záznamu lekár vytvorí na základe svojej expertnej znalosti binárnu masku – predlohu (ground truth image) a určí, kde sa na snímke nachádza kosť. Následne v programe *3D Slicer* prebieha spracovanie vstupného obrazu na základe vybraných segmentačných metód do binárnych masiek, v ktorých každý pixel nadobúda hodnoty 1 (pre kosť) alebo 0 (pre okolie). Z týchto hodnôt sa potom vytvára vstupný vektor, ktorý obsahuje len binárnu informáciu, s ktorou dokáže modul *SVM*

Classifier pracovať. Výsledný segmentovaný obraz bol následne porovnaný s predlohou a zhoda bola určovaná pomocou *Dice* koeficientu.



Obrázok 1: Schéma spracovania obrazu

2.1 BINÁRNA MASKA – PREDLOHA

Do doby zaslania tohto článku ešte nebol žiaden lekár účastný na projekte a binárna predloha (**Obrázok 2**) bola vytvorená laikom v programe *Adobe Illustrator*. Aby mohla optimalizácia byť čo najpresnejšia, je úlohou lekára v budúcnosti jednoznačne určiť z určitého rezu oblasti, kde sa nachádza kosť, a kde okolie (Obrázok 2). Predloha je využitá pri porovnávaní výslednej segmentácie pri stanovení *Dice* koeficientu.





Obrázok 2: Zobrazenie rezu z CT vyšetrenia a) originálny obraz; b) binárna predloha

2.2 SPRACOVANIE CT OBRAZOV DO BINÁRNEJ MASKY

Po nahratí súboru z vyšetrenia do modulu *SVM Classifier*, je možné nastaviť charakteristické parametre, ktoré budú spracovávať obraz do binárnej podoby. V tejto práci boli postupne vystriedané všetky dostupné charakteristiky z modulu (Tabuľka 1) na vytvorenie vstupného vektora pre klasifikátor jednotlivo (Obrázok 3), aby bolo možné do ďalších segmentácií vybrať len tie, ktoré majú dobré výsledky následnej segmentácie. Následne boli zvolené kombinácie metód, u ktorých bol predpokladaný priaznivý výsledok (Obrázok 4). Príliš veľa zvolených metód pre spracovanie predlžuje čas segmentácie.

Názov charakteristických parametrov	Význam segmentačnej metódy
Mean	Priemeruje intenzitu pixelov v okolí pixelov
Median	Určuje medián hodnôt intenzity pixelov
Variance	Počíta rozptyl hodnôt v okolí pixelu
Gaussian	Rozmazáva obraz okolo pixelu
Sobel	Hranový detektor
Gradient matrix	Hranový detektor
Laplacian	Hranový detektor

Tabuľka 1: Prehľad charakteristických obrazov použitých pre segmentáciu

2.3 STANOVENIE *DICE* KOEFICIENTU

Dice koeficient je celosvetovo uznávaný štatistický údaj, ktorý opisuje podobnosť dvoch porovnávaných výsledkov. [4] V tejto práci pôjde o porovnanie takzvaného *ground truth* obrazu a výslednej segmentácie. *Ground truth* obraz je binárna predloha (Obrázok 2), na ktorej je manuálne a presne vymedzená časť kosti a okolia. Výpočet *Dice* koeficientu bol realizovaný v programe Matlab porovnaním oboch obrazov.

3 VÝSLEDKY



Obrázok 3: Výsledné segmentácie 8 rôznymi charakteristikami obrazu a) Voxel; b) Mean; c) Variance; d) Gaussian; e) Median; f) Sobel; g) Gradient Matrix; h) Laplacian



Obrázok 4: Výsledné segmentácie kombináciami charakteristík obrazu a) Voxel, Median; b) Voxel, Median, Mean; c) Voxel, Madian, Maan, Sabal; d) Voxel, Median, Mean, Laplacian; c) Věst

c) Voxel, Median, Mean, Sobel; d) Voxel, Median, Mean, Laplacian; e) Všetky

Použité charakteristiky	Hodnota Dice koeficientu
Voxel, Median	0,825
Voxel, Median, Mean	0,818
Voxel, Median, Mean, Sobel	0,821
Voxel, Median, Mean, Laplacian	0,821
Všetky	0,756

Tabul'ka 2: Hodnoty Dice koeficientov pre kombinácie charakteristík

4 ZÁVER

Tento článok sa zaoberá spresnením segmentácie snímok zhotovených počítačovou tomografiou. Neskôr by mal byť *SVM Classifier* schopný presne spracovať aj snímky zhotovené magnetickou rezonanciou (MRI) u detských pacientov.

V tejto práci boli vyhodnotené viaceré charakteristiky obrazov jednotlivo aj ich kombinácie s cieľom vylúčiť tie neefektívne, aby nespomaľovali proces segmentácie. Medzi neefektívne patrili metódy Gradient matrix, ktorá nesprávne segmentovala vnútorne plochy kosti a Variance, ktorá nesprávne rozlišovala prechod medzi okolím a kosťou a spájala ich. Z analyzovania výsledku segmentácie možno konštatovať, že metóda Sobel je zaťažená šumom a je nevyhnutná jej úprava.

Z kombinácii segmentačných metód podľa výsledkov Dice koeficientov (**Chyba! Nenalezen zdroj odkazů**.) vyplýva, že použitie všetkých metód naraz nevedie k uspokojivému výsledku. Ako najpresnejšia vyšla kombinácia Voxel, Median, ktoré tvorili základ aj ďalších kombinácií. Po pridaní hranových detektorov ako napríklad Sobel alebo Laplacian bola výsledkom menej presná segmentáciu obrazu.

Pre zvýšenie presnosti plug-inu *SVM Calssifier* bude v budúcnosti možné prestaviť vstupné parametre jednotlivých metód pre spracovanie pôvodného obrazu pred samotnou segmentáciou.

REFERENCIE

- [1] Repko, M. Skolióza komplexní diagnostické a terapeutické postupy. Solen, 2010, vol. 11, iss. 4, p. 218-222
- [2] Svoboda, J. Přesná segmentace obrazových dat [online]. Brno, 2020 [cit. 2021-03-12]. Dostupné z: https://dspace.vutbr.cz/xmlui/handle/11012/190449. Bakalářská práce. Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací. Vedoucí práce Jan Mikulka
- [3] Tináková, B. Základy Support Vector Machines. *Umelá inteligencia* [online]. 2019 [cit. 2021-03-23]. Dostupné z: https://umelainteligencia.sk/zaklady-support-vector-machines/
- [4] Milletari, F. et al. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation [online] Fourth International Conference on 3D Vision (3DV), 2016. IEEE, 2016, 2016, s. 565-571 [cit. 2021-03-12]. ISBN 978-1-5090-5407-7. Dostupné z: doi:10.1109/3DV.2016.79

HUMAN BODY SEGMENTATION USING R-CNN

Libor Matějek

Bachelor Degree Programme (3), FEEC BUT E-mail: xmatej53@stud.feec.vutbr.cz

Supervised by: Tomáš Bravenec E-mail: xbrave01@vutbr.cz

Abstract: The article deals with basic concepts in the sector of image processing using neural networks. It describes the basic principles and methods of object recognition and image segmentation using neural network architectures based on R-CNN architecture. Specifically the work focuses on human body segmentation from static images, resulting in individual segments in the output mask corresponding to each limb of a human body.

Keywords: Image segmentation, Body part recognition, CNN, R-CNN, deep learning, object recognition

1 ÚVOD

V práci jsou přiblíženy základní informace o aktivních prvcích celého konceptu zpracování obrazu jakožto konvoluční neuronové sítě a dále navazující technologie pro zefektivnění rozpoznávání objektů pomocí architektury R-CNN. Konkrétní náplní práce jsou architektury neuronových sítí Mask R-CNN a Keypoint R-CNN postavené na základu neuronové sítě Resnet-50. Práce také obsahuje detailní popis úpravy keypoint R-CNN, která je schopna rozpoznávat jednotlivé končetiny lidského těla. Z mého průzkumu současného stavu algoritmů pro rozpoznávání lidských těl se všechna řešení, výcházející z datasetu COCO, zabývají segmantací celého těla a nezaměřují se na rozsegmentování končetin.

2 KONVOLUČNÍ NEURONOVÉ SÍTĚ

Konvoluční neuronová síť, je typ Deep learning algoritmu, který je schopen přijmout vstupní obrázek, zpracovat jeho znaky, rozpoznat objekty na obrázku a dokonce jim přiřadit význam. CNN oproti jiným druhům neuronových sítí nevyžaduje téměř žádný pre-processing. [1]

3 ARCHITEKTURA POUŽITÉ CNN

3.1 RESNET

Od architektury CNN AlexNet, která zvítězila v soutěži ImageNet 2012, využívá každá další vítězná architektura více vrstev v hluboké neuronové síti ke snížení chybovosti. Po překonání určité hranice vrstev se však můžeme setkat s problémem zvaným Vanishing gradient. Tento problém má za následek zvýšení míry chyb. Vanishing gradient problém je založen na ztrátě učící informace při derivaci hodnoty aktivační funkce. Pro eliminaci Vanishing gradient problému byla v roce 2015 představena Kaimingem He architektura ResNet (Residual Network). V síti využíváme techniku zvanou přeskakování spojů. Tato technika umožňuje při tréninku přeskočit některé vrstvy a připojit se rovnou na výstup. [2]

4 POČÍTAČOVÉ VIDĚNÍ

Rozdílem mezi detekčními algoritmy a jednoduššími klasifikačními algoritmy je konkretizace a ohraničení objektů ve vstupním obraze. Důvodem, proč nelze použít standardní konvoluční sítě s plně propojenou vrstvou (fully connected layer) je to, že délka výstupní vrstvy je proměnlivá, ne konstantní. Tato proměnlivost je způsobena nekonstantním počtem objektů. Nabízí se rozdělení obrazů na různé oblasti a použít klasickou konvoluční neuronovou síť (CNN) postupně na každou oblast. Objekty v obraze mohou mít však různá prostorová umístění a různé poměry stran, tzn. bylo by nezbytné vybrat obrovské množství oblastí (regionů), což je výpočetně náročné. Pro snazší nalezení výskytů objektů byly vyvinuty další algoritmy. [3]

4.1 R-CNN (REGION BASED CONVOLUTIONAL NEURAL NETWORKS)

Abychom se vyhnuli zpracovávání velkého množství oblastí (regionů), můžeme využít metodu Rosse Girsicketa, která využívá selektivního vyhledání a extrahování pouze 2000 regionů z obrázku. Tyto oblastí se nazývají region proposals (návrhy regionů). [4]

4.2 FAST R-CNN

Jedná se o optimalizaci a zrychlení R-CNN. Rozdílem je prohození kroků zpracování, tedy nejprve se vstupní obraz zpracuje v konvoluční neuronové síti CNN, čímž dostaneme mapu vlastností (feature map). Z této mapy dále algoritmus identifikuje návrhy oblastí a znormuje je na čtverec. Aplikací RoI (Region of interest) vrstvy, která svojí funkcí sdružuje hodnoty z CNN do pevně dané velikosti, dostaneme validní vstup pro plně propojenou neuronovou vrstvu (fully connected layer). Z upraveného vektoru z RoI vrstvy se nakonec pomocí softmax vrstvy předpoví třídy a hodnoty posunutí navrhovaných oblastí. [3, 4, 5]

4.3 FASTER R-CNN

Oba předchozí algoritmy využívají pomalé vyhledávací algoritmy pro určení navržených oblastí. Nová metoda je inovativní ve využívání rozpoznávacích neuronových sítí, které se jsou schopny učit. Tyto sítě se nazývají Region Proposal Network (Síť pro návrh oblastí). Faster R-CNN je prvním algoritmem, který je schopen pracovat v reálném čase. [3, 4, 6]

4.4 MASK R-CNN

Nadstavbou Faster R-CNN jménem Mask R-CNN získáváme jednotlivé masky kopírující přesně tělo objektu, tedy nejen pouhé hranaté oblasti. Mask R-CNN lze vidět na obrázku 1. [4, 7, 9]

4.5 **KEYPOINT R-CNN**

Detekce klíčových bodů zahrnuje oproti předchozím metodám detekci bodů, jež definují lidské tělo, popřípadě I konstrukci jiných objektů. Body jsou neměnné vůči rotaci, změně velikosti nebo zkreslení obrazu. Metoda čerpá z definovaných klíčových bodů ze speciálních anotací, kterými je učena neuronová síť. Jedná se o další nástavbu konvoluční a klasifikační vrstvy Faster-RCNN (Mask-RCNN). Ve své práci vycházím z datasetu COCO, jež disponuje potřebnými anotacemi. Keypointy R-CNN vidíme na obrázku 2. [4, 8]

5 BODYPARTS KEYPOINT R-CNN

Modifikací výstupu neuronové sítě v podobě keypointů bylo dosaženo schopnosti rozpoznání jednotlivých částí končetin lidského těla. Tato modifikace je založena na výpočtu konstanty za pomoci vzdálenosti určitých keypointů a rozšíření spojnice mezi nimi, popřípadě vytvoření určitého geometrického tvaru na správné pozice. Moji modifikaci rozpoznávání lidských končetin lze vidět na obrázku 3.

Předtrénovaná neuronová síť ResNet-50 na základě vstupního obrázku vrací 3 sady souřadnic keypointů, které by měly být umístěny v klíčových bodech definujících lidské tělo. Každá sada je zároveň ohodnocena tzv. keypoint score, což nám říká, jak přesně jsou keypointy umístěny vzhledem k "vědomostem", které neuronový model při tréninku získal. Algoritmus dále využívá tu nejlepší sadu, která musí zároveň splňovat keypoint score větší, než 0.9.

Máme-li tedy vhodnou sadu souřadnic 17 keypointů, můžeme pokročit k metodám, které vhodně rozšíří spojnice těchto bodů na vhodném místě a označí tak jednotlivé lidské končetiny.

Šířka spojnice keypointů rukou je definována vektorovou délkou začátek - konec předloktí dále vynásobenou konstantou 0.4, která je vypozorována testováním na více obrázcích

Šířka spojnice nohou je řešena stejným algoritmem, pouze měříme vektorovou vzdálenost bodů začátek - konec stehna a násobíme ho konstantou 0.3.

Označení oblasti hlavy vychází ze spojnice uší. Nalezneme střed spojnice uší, ve středu vztyčíme kolmici a na kolmici vyznačíme dva body symetricky kolem středu, jejichž vektorová vzdálenost je 0.8 vzdálenosti uší. Mezi těmito body vykreslíme spojnici o tloušť ce vektorové vzdálenosti uší.

Trup je rozeznáván pomocí 4 keypointů, 2 v ramenou a 2 na začátku nohou. Naleznou se středy těchto dvojic a vektorová vzdálenost se symetricky zmenší na 0.6 původní vzdálenosti. Poté se vykreslí spojnice mezi těmito středy s tloušť kou, která se rovná vektorové vzdálenosti bodů v ramenou.

6 VYHODNOCENÍ MODIFIKACE BODYPARTS METODOU IOU

K vyhodnocení úspěšnosti detekce navrženého algoritmu bylo využito metody IoU (Intersection over Union), což je poměr mezi opravdovou mapou pixelů, na kterých je rozpoznávaný objekt a mapou pixelů, kterou rozpoznal testovaný algoritmus. Vyhodnocení bylo realizováno na testovacím subsetu COCO datasetu s upravenými maskami skutečné pozice končetin lidského těla. Ne této testovací sadě bylo dosaženo průměrné úspěšnosti IoU 0,88. V celé testovací sadě tedy algoritmus označil jednotlivé končetiny správně.

7 ZÁVĚR

Tento článek přibližuje oblast rozpoznávání objektů pomocí natrénovaného modelu neuronové sítě ResNet. Současně popisuje základní metody R-CNN. V článku je dále popsána modifikace Keypoint R-CNN, která má za úkol rozpoznávat jednotlivé končetiny lidského těla. Každá z metod je demonstrována obrázkem, ne kterém je aplikována. Celý proces rozpoznávání končetin, vycházející z keypointů, je detailně popsán. Modifikace keypointů pro rozeznávání končetin funguje velmi dobře a na vzorku testovacích snímků bylo dosaženo výsledků bez výrazné chybovosti - kolem 0,88 IoU na testovací sadě. Jediným nedostatkem je označování hlavy na lidském těle z profilu což je možné řešit získáním na úhlu nezávislé konstanty pro velikost označení hlavy. Dalším potenciálním krokem vpřed by mohla být úprava programu do podoby pro embedded systémy a dosáhnout tak vyšší rychlosti zpracování obrazu vhodné k real-time využití. Vzhledem k absenci podobných algoritmů na rozpoznávání končetin, která by vycházela z datasetu COCO, není možné srovnání úspěšnosti rozpoznání končetin. Ostatní algoritmy rozpoznávají například celé masky lidského těla. Tato absence byla motivací pro právě moji úpravu rozpoznávají samotných končetin.



Obrázek 1: Mask R-CNN Obrázel

Obrázek 2: Keypoint R-CNN

Obrázek 3: Bodyparts

REFERENCE

- S. Saha, "A Comprehensive Guide to Convolutional Neural Networks the ELI5 way," [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutionalneural-networks-the-eli5-way-3bd2b1164a53. [Přístup získán 18 11 2020].
- [2] "Residual Networks (ResNet) Deep Learning," [Online]. Available: https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/. [Přístup získán 24 11 2020].
- [3] R. Gandhi, "R-CNN, Fast R-CNN, Faster R-CNN, YOLO Object Detection Algorithms," [Online]. Available: https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yoloobject-detection-algorithms-36d53571365e. [Přístup získán 25 11 2020].
- [4] R. Girshick, J. Donahue, T. Darrell a J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," [Online]. Available: https://arxiv.org/pdf/1311.2524.pdf. [Přístup získán 26 11 2020].
- [5] R. Girshick, "Fast R-CNN," [Online]. Available: https://arxiv.org/pdf/1504.08083.pdf. [Přístup získán 27 11 2020].
- [6] R. Shaoqing, H. Kaiming, Sun a J. Sun, "Faster R-CNN: Towards Real-Time Object," [Online]. Available: https://arxiv.org/pdf/1506.01497.pdf. [Přístup získán 20 11 2020].
- [7] K. He, G. Gkioxari, P. Dollár a R. Girshick, "Mask R-CNN," [Online]. Available: https://arxiv.org/pdf/1703.06870.pdf. [Přístup získán 25 11 2020].
- [8] S. R. Rath, "Human Pose Detection using PyTorch Keypoint RCNN," [Online]. Available: https://debuggercafe.com/human-pose-detection-using-pytorch-keypoint-rcnn/. [Přístup získán 1 12 2020].
- X. Zhang, "Simple Understanding of Mask RCNN," [Online]. Available: https://alittlepain833.medium.com/simple-understanding-of-mask-rcnn-134b5b330e95. [Přístup získán 25 11 2020].

Bakalářské projekty

Komunikační technologie a informační bezpečnost, Kybernetika a automatizace

IMPLEMENTATION OF AUTO-NEGOTIATION FOR ETHERNET INTERFACES RUNNING AT SPEEDS FROM 25 TO 100 GBPS

Vladislav Válek

Bachelor Degree Programme (3), FEEC BUT E-mail: xvalek14@vutbr.cz

Supervised by: Adrián Tomašov E-mail: xtomas32@vutbr.cz

Abstract: Auto-negotiation is the capability of the interface to automatically determine the link abilities, like link speed and the FEC ability, according to information sent by the link partner. This article addresses the design of auto-negotiation component working on Xilinx Virtex UltraScale+FPGAs using GTY transceivers. Main guidelines were taken from the Clause 73 IEEE 802.3 standard for Ethernet interfaces. This design focuses on the interface with single line running at 25 Gbps using optical cable as medium. The special description is provided considering design changes on FPGA against the IEEE standard.

Keywords: auto-negotiation, FPGA, Ethernet, 802.3, GTY

1 ÚVOD

Současná infrastruktura páteřních sítí vyžaduje především vysokou propustnost pro zpracování kumulovaných datových toků a speciálně upravené obvody pro provoz na těchto rychlostech. Tyto obvody nenacházejí uplatnění v běžných LAN sítích provozovaných v domácím prostředí, kde je vyžadována velká univerzálnost, jako paralelní provoz LAN a WLAN, firewallu, přepínače nebo směrovače. Pro použití ve velkých sítích si však naleznou uplatnění i některé funkce používané v sítích malého rozsahu. Jednou z takových běžně využívaných je auto-negociace. Ta slouží ke vzájemné dohodě parametrů, jako komunikační rychlosti, schopnosti provozovat *Forward Error Correction* (FEC) nebo dočasné pozastavení provozu; mezi dvěma stranami komunikačního kanálu. Tento článek pojednává o způsobu návrhu komponenty pro auto-negociaci a její implementaci pomocí jazyka VHDL na síťových kartách řízených hradlovými poli (*Field Programmable Gate Array*, FPGA).

2 STANDARD PRO ETHERNET

Při návrhu jsou uváženy zásady uvedené ve standardu IEEE 802.3-2018, klauzule 73. Ta zavádí způsob chování auto-negociace na Ethernetových rozhraních pro rychlosti 10–100 Gb/s. Návrh je zaměřen zejména na technologie 40GBASE-CR4, 100GBASE-CR4, 25GBASE-CR-S a 25GBASE-CR. V případě, že komunikace probíhá po více linkách, je funkce vzájemné dohody parametrů provozována po lince 0. Tímto se redukuje pouze na rychlosti 10 a 25 Gb/s. Standard zařazuje vrstvu pro provoz auto-negociace na nejnižší místo v hierarchii fyzické vrstvy, tj. pod *Physical Medium Dependent* (PMD).

Negociace zahajuje činnost ještě před započetím užitečného provozu po komunikačním médiu a blokuje tím jakýkoliv vycházející provoz. Pro komunikaci během auto-negociace jsou ve standardu stanoveny komunikační jednotky zvané stránky, které lze nalézt ve dvou variantách. První je tzv. *Base Page*, která slouží pro primární činnost, při které dochází k dohodě parametrů mezi stranami přenosového kanálu. Obsahuje pole pro informaci o provozovaných typech rozhraní, o schopnosti provádět FEC nebo schopnost pozastavit provoz. Stránka obsahuje rovněž kontrolní sekvence bitů pro potvrzení úspěšného přijetí na protistraně. Druhým typem jsou stránky typu *Next Page*, jež slouží pro přenos informací nad rámec Base Page. Tímto způsobem je možné informovat protistranu o *Organizationally Unique Identifier* (OUI) nebo specifikovat chybu, ke které došlo během procesu vzájemné dohody parametrů. Zpracování stránek Next Page, jež může existovat ještě jako formátovaná (s identifikační hlavičkou) a neformátovaná (bez hlavičky), není zajištěno v logice auto-negociační komponenty a musí být vyřešeno ve vyšších vrstvách. Oba typy stránek mají délku 48 bitů. Každá strana zašle svou kombinaci nastavenou ve stránce Base Page a v případě, že alespoň jedna strana potřebuje zaslat dodatečné informace, je umožněna výměna stránek Next Page.

Stránky jsou pro přenos komunikačním kanálem zakódované pomocí pravidel *Diferenciálního kódování Manchester*, které disponuje výhodami toho, že přenášený signál neobsahuje stejnosměrnou složku, při výpadku synchronizace umožňuje sám ji obnovit a data jsou přenášena v přechodech, nikoliv hladinách. Sériová datová rychlost je, dle standardu, stanovena na 156,25 Mb/s, uvážíme-li ještě přenos hodinových přechodů, je výsledná rychlost dvojnásobná, tj. 312,5 Mb/s. Zakódovaná stránka je uvozena osmicí bitů tvořících oddělovač pro identifikaci začátku stránky, a datové přechody se v ní poté pravidelně střídají s hodinovými[1].

3 PLATFORMA PRO VÝVOJ

Pro návrh bylo zvoleno hradlové pole Xilinx Virtex UltraScale+. Jedná se o jeden z nejvýkonnějších čipů vyráběných touto společností. Pro využití ve vysokorychlostní komunikaci jsou vybavené specializovanými bloky nazvanými GTY transceivery[2].

GTY v základu pracuje jako serializér/deserializér s diferenciálními vývody na sériové straně a paralelní sběrnicí na druhé s konfigurovatelnou šířkou. Zpracování paralelních dat nevyžaduje tak vysokou frekvenci, sběrnice pracuje na kmitočtu 390,6 MHz, což umožňuje standardní postup pro návrh synchronních logických systémů. Transceivery dosahují vysokých pracovních frekvencí díky zabudovaným *smyčkám fázového závěsu*, vzhledem k vylepšení těchto obvodů u řady UltraScale+ je možné dosáhnout rychlosti až 32,75 Gb/s. GTY lze spojovat do skupin až dvaceti transceiverů, přičemž nad 4 je maximální dosažitelná frekvence postupně omezována.

Transceivery jsou dále schopny vykonávat funkci ekvalizace, aby vyrovnaly ztrátovost přenosového kanálu. Používaným typem je *Decision Feedback Equalization* (DFE). Během provozu dochází k automatické kalkulaci koeficientů ekvalizace, což usnadňuje konfiguraci vzhledem k prudkému navyšování počtu koeficientů s rostoucí datovou rychlostí. Pakliže je užito komunikační médium s krátkým vedením nebo zanedbatelnou ztrátovostí, je možné ekvalizaci výrazně omezit režimem *Low-power mode* (LPM). Komunikačním médiem je buď optický nebo DAC (měděný) kabel[3].

4 NÁVRH KOMPONENTY

Vzhledem k odlišnému uspořádání struktury hradlového pole je třeba změnit umístění komponenty auto-negociace oproti tomu specifikovaném ve standardu 802.3. GTY transceivery reprezentují vrstvu *Phasical Medium Attachement* (PMA) a na FPGA se jedná o nejnižší vrstvu, kterou lze instanciovat (tudíž konfigurovat). Vrstva zajišť ující auto-negociaci je proto přesunuta mezi PMA a *Physical Co-ding Sublayer* (PCS), čímž zaujímá, v logice hradlového pole, místo za GTY transceivery. Zde je napojena na paralelní sběrnici, jejíž přenos je během procesu vzájemné dohody parametrů blokován.

4.1 ÚPRAVA RYCHLOSTI

První aspekt návrhu se týká úpravy odesílané bitové posloupnosti. Vzhledem k pevně nastavené rychlosti transceiverů dochází k rozdílu s komunikační datovou rychlostí stanovenou pro auto-negociaci.
Každý odeslaný bit je tudíž naklonován tolikrát, aby délka skupiny klonů odpovídala periodě, se kterou je třeba auto-negociační data odesílat, tj. $83 \times$ pro 25 Gb/s a $33 \times$ pro 10 Gb/s. Standard definuje jednotnou periodu (konkrétně 3,2 ns) pro všechny linkové rychlosti. Řada naklonovaných bitů je následně dělena podle délky paralelní sběrnice, která je nastavena na 66 bitů, kde 64 bitů je vyhrazeno pro data a 2 bity pro jednoduchou hlavičku. Počet klonů lze vypočítat podle vzorců $L = L' \cdot \frac{D+H}{D}$ a $N = T_{AN} \cdot L$, kde D je bitová délka dat, H délka hlavičky, L' linková rychlost v Gb/s udávaná (25 nebo 10 Gb/s), L linková rychlost v Gb/s skutečná, T_{AN} perioda AN přenosu v ns a N počet klonů.

4.2 USPOŘÁDÁNÍ KOMPONENT

Obsluha funkce auto-negociace je zřízena spoluprací komponent přijímače, vysílače a komponenty arbitráže. Vysílač slouží k zakódování odesílané stránky pomocí pravidel Diferenciálního kódování Manchester, následnému klonování každého bitu a nakonec dělení řad klonů podle šířky sběrnice pro odesílání dat (zde 66). Navíc je přítomen jednoduchý čítač odeslaných stránek, jenž nalézá využití v jednom ze stavů arbitráže (viz dále). Komponenta arbitráže zajišť uje logiku potřebnou pro průběh auto-negociačního procesu. Implementuje stavový automat popsaný ve standardu IEEE 802.3. Kromě přímého napojení na vysílač i přijímač zajišť uje rovněž přístup ke konfiguračním registrům, jimiž lze jednak celý proces řídit z uživatelské perspektivy, a zároveň informovat vyšší vrstvy o průběhu auto-negociace.

Třetí komponentou je přijímač, který je z uvedené trojice nejsložitější. Jelikož přechody v řadách bitů, přicházející v přijímací sběrnici, neustále mění svou polohu, je třeba zajistit adaptabilní funkci při zachytávání vstupních slov. V každém hodinovém taktu je tudíž neustále zkoumán počet přechodů ve vstupním slově (zde vzhledem k počtu klonů 83 a šířce sběrnice 66 je maximální počet přechodů 1) a, v případě jejich zvýšeného počtu, je postup přijmu stránek resetován. Za normálních okolností je tedy nalezen přechod a spočítána jedna vzdálenost k začátku slova a druhá k jeho konci. Tyto hodnoty jsou potom ve stejném pořadí, společně s příslušnými hodnotami bitů v daných částech, zapsány do registru. Více zapsaných hodnot potom tvoří tzv. *redukovanou posloupnost bitů*. Následně je z těchto řad spočtena celková délka jednotlivých hodnot bitů. Na základě celkových délek je potom zapsán adekvátní počet bitů do výstupního registru, jehož hodnota již odpovídá zakódované auto-negociační stránce před klonováním jejích bitů.

V přijímači je následně přijatá stránka validována, kdy je kontrolována přítomnost hodinových přechodů na pozici mezi každým lichým bitem a bitem o pozici výše, mimo oddělovač na začátku stránky. Při splnění kontroly dochází k dekódování přijatého slova. V přijímači dochází k rozdělení hodinových signálů na přijímací a odesílací straně (RX a TX kanál). V RX směru dochází k tzv. *rekonstrukci hodinového signálu*, což spočívá v odvození frekvence hodinového signálu z rychlosti, s jakou data přicházejí na rozhraní GTY transceiverů[3]. Vznikají tudíž dvě hodinové domény a přechod mezi nimi je nutné řešit pomocí speciálních komponent pro asynchronní přechody, které navíc zavádějí nutná pravidla pro syntézu v nástrojích pro návrh hardware.

5 PRŮBĚH AUTO-NEGOCIACE

Proces vzájemné dohody parametrů je spuštěn nastavením bitu AN_ENABLE v hlavním řídícím registru. Stavový automat je v tuto chvíli přepnut do stavu TRANSMIT_DISABLE, který slouží pro zastavení obousměrného provozu v přenosovém kanálu. Standard definuje časovou periodu 65–75 ms, po kterou je zastaven provoz vysílače, na což protistrana reaguje rovněž zastavením stejné komponenty. Po uplynutí časového okamžiku dochází k přepnutí do stavu ABILITY_DETECT. Zde je spuštěn vysílač a zahájen příjem, kdy je očekáván příjem třech stejných stránek za sebou (oznámeno z přijímače nastavením signálu ABILITY_MATCH). Pakliže je tato podmínka splněna, je ještě nutné zkontrolovat, zda přijatá stránka neobsahuje totožnou kombinací bitů *Transmitted Nonce* jako stránka odesílaná. Tato kombinace je generována z komponenty *Linear Feedback Shift Register* (LFSR)

a určuje jedinečnost stránek Base Page během jednoho přenosu. V případě rozdílných hodnot je přepnuto do stavu ACKNOWLEDGE_DETECT, kde je nastaven bit ACK odeslaných stránek. Tímto je komponentou signalizováno úspěšně přijetí validní auto-negociační stránky. Z přijaté stránky je rovněž zkopírována hodnota Transmitted Nonce do pole Echoed Nonce stránky odesílané. Tímto si strany navzájem zajišť ují své "podání ruky". Pro vyhodnocení na konci stavu je zde rovněž nutná podmínka přijmu alespoň tří validních stránek, navíc s nastaveným ACK bitem (oznámeno z přijímače nastavením signálu ACKNOWLEDGE_MATCH). Následně je zkontrolována rovnost pole Echoed Nonce v přijaté stránce a Transmitted nonce ve stránce odesílané a při splnění je přepnuto do stavu COMPLETE_ACKNOWLEDGE. Zde dochází k odeslání posledních ACK stránek, jejichž maximum určuje přetečení čítače ve vysílači, a následně je přepnuto do stavu AN_GOOD_CHECK. Zde je již provoz auto-negociační komponenty zastaven a automat vyčkává na zahájení užitečného provozu po přenosovém kanálu, což signalizuje PCS vrstva. Vyčkávání trvá, na základě udávaného ve standardu IEEE 802.3, po dobu 500–510 ms. Pokud do konce časového okamžiku není zahájen úspěšný provoz na lince, potom je přepnuto do stavu TRANSMIT_DISABLE. Při pozitivní signalizaci z vrstvy PCS o přijmu validních bloků je přepnuto do závěrečného stavu AN_GOOD. Zde je provoz komponenty kompletně zastaven. K přerušení může dojít jen v případě signalizování z vrstvy PCS o zastavení provozu, kdy je automat přepnut opět do stavu TRANSMIT_DISABLE[1].

6 ZÁVĚR

Vysokorychlostní komunikační systémy přinášejí výzvu pro adaptaci starších technologií a jejich implementace, jakožto plně rekonfigurovatelných, je obrovskou výhodou v dnešní dynamicky se měnící infrastruktuře. Tento článek posoudil možnost řešení auto-negociační komponenty na programovatelných logických obvodech firmy Xilinx. Komponenta po syntéze využívá 1103 LUT, 1312 registrů a 47 multiplexerů, což je pro použité hradlové pole minimální obsazenost. V současném stavu vývoje je funkce komponenty již ověřena pomocí simulací a probíhá její implementace na fyzickém zařízení. První řešení problémů při implementaci vedlo k odstranění kontroly hlaviček přicházejících stránek, která způsobovala zahazování neformátovaných Next Page. V aktuálním stavu nedochází, při vyčkávání ve stavu AN_GOOD_CHECK, ke spolehlivému zahájení užitečného provozu po přenosovém kanálu. Tento problém souvisí se spoluprací GTY transceiverů a PCS komponenty.

PODĚKOVÁNÍ

Chci poděkovat svému vedoucímu ve sdružení CESNET, panu inženýru Štěpánu Friedlovi, za poskytnutí potřebné pomoci při návrhu hardware a dále svému vedoucímu bakalářské práce, panu inženýru Adriánu Tomašovi. Oběma chci rovněž poděkovat za cenné připomínky a četné vylepšující poznámky při zpracování mé bakalářské práce.

- [1] IEEE Computer Society. *IEEE Standard for Ethernet*. IEEE Std 802.3-2018 (Revision of IEEE Std 802.3-2015) [online]. 2018. USA: IEEE, 31.8.2018 [cit. 10.2.2021]. ISBN 978-1-5044-5090-4. Dostupné z: https://doi.org/10.1109/IEEESTD.2018.8457469
- [2] Xilinx, DS890: UltraScale Architecture and Product Data Sheet: Overview. *Xi-linx* [online]. USA: Xilinx, ©2013–2020, 14.9.2020 [cit. 10. 2. 2021]. Dostupné z: https://www.xilinx.com/support/documentation/data_sheets/ds890-ultrascale-overview.pdf
- [3] Xilinx. UG578: UltraScale Architecture GTY Transceivers: User Guide. Xi-©2013–2020, 20.9.2017 [cit. linx [online]. USA: Xilinx, 10.2.2021]. Dostupné <https://www.xilinx.com/support/documentation/user_guides/ug578-ultrascale-gtyz: transceivers.pdf>

QUANTUM KEY DISTRIBUTION POLYGON

Ondřej Klíčník

Bachelor Degree Programme (3), FEEC BUT E-mail: xklicn02@vutbr.cz

Supervised by: Petr Münster

E-mail: munster@feec.vutbr.cz

Abstract: This article presents basics of quantum key distribution BB84-like protocols. Briefly explains principles of their operations and impact of QBER on intrusion detection. Afterwards results of QKD simulations are presented as well as own QKD polygon draft (on multiplexed optical fibre).

Keywords: BB84, Quantum key distribution (QKD), QKD networks (QKDN), QKD polygons

1 ÚVOD

Pro přenos šifrovaných dat je nejdříve nutné ustanovit mezi dvěma uzly symetrický šifrovací klíč. V současných počítačových sítích slouží k tomuto účelu tzv. asymetrická kryptografie, založená na matematických problémech. Její algoritmy jsou však náchylné na útoky tzv. kvantových počítačů. Tedy technologii, která by dle odhadů mohla být v blízké době dostupná. V roce 1984 přišla dvojice fyziků Gilles Brassard a Charles Bennett s novým způsobem ustanovení klíče, postaveném na jevech kvantové mechaniky. Tato metoda přenosu klíče se tak podle nich nazývá protokolem BB84. Ačkoliv dnes již existuje velké množství nejrůznějších protokolů založených na rozličných kvantových principech, BB84 zůstává stále jedním z nejpoužívanějších a existuje v bezpočtu modifikací. [1]

2 RODINA PROTOKOLŮ BB84

Původní verze protokolu je postavená na tzv. polarizačním kódování a počítá s ideálním zdrojem jednotlivých fotonů (SPS). Stručně lze protokol shrnout takto. BB84 využívá čtyř stavů polarizace, tedy dvou bází $\{|\uparrow\rangle, |\rightarrow\rangle\}$ a $\{|\rangle, |\rangle\}$. Jak je možné vidět níže na obrázku 1, jsou tyto báze na polarizátoru vzájemně pootočeny o 45°. Protokol začíná tím, že si Alice vygeneruje náhodný návrh klíče. Následně náhodně přepíná mezi bázemi a bitům tak přiděluje vybranou polarizaci fotonu. [1]



Obrázek 1: Schéma protokolu BB84 s odposlechem. [2]

Pro QKD protokoly je nutné využívat kvantového generování čísel (QRNG), to zaručuje distribuci 50:50 a pravou náhodnost. Fotony jsou optickým vláknem zaslány k Bobovi. Ten také náhodně volí báze. Pokud zvolí stejnou bázi jako Alice, je foton vždy správně rozklíčován. Pokud zvolí bázi opačnou, je foton rozklíčován v 50 % případů. Bob však v tuto chvíli netuší, které bity v jeho řetězci

neodpovídají původním Aliciným hodnotám. Bob proto odešle Alici po libovolném kanálu pořadí použitých bází. Alice je porovná s vlastními bity. Ty, u kterých se báze shodují, prohlásí za surový klíč (raw key), ostatní bity zahodí. Bobovi pak pošle pořadí vhodných bitů. Tak získá klíč i Bob. [1]

Surový klíč však ještě nemusí být kvůli chybám zcela shodný. To může znamenat odposlech. K jeho detekci je tak potřeba klíč "prosít". Z tohoto důvodu se část klíče zveřejní. Tím lze určit chybovost přenosu neboli QBER (Quantum-Bit Error Rate). Vysoké QBER je příznakem odposlechu. Pokud by se zde totiž Eva vyskytovala, byla by nucena stavy fotonů měřit a volbou bází by sama dělala chyby. Ty by se následně projevily na QBER. V ideálním případě by jediným zdrojem chyb byla Eva. Kvůli nedokonalostem přístrojů ale vznikají chyby i jiným způsobem (zdroje, přenosová soustava, detektory). Z těchto důvodů je nízká chybovost tolerována (u BB84 do cca 11 %). [1]

V praxi se ovšem SPS nepoužívá a místo něj jsou nasazovány lasery se slabými koherentními pulzy (WCP). Pravděpodobnost odeslání jednoho fotonu je dána Poissonovým rozložením. Teoreticky tak může být současně odesláno fotonů i více. To je příležitostí pro Evu, která si může nadbytečný foton ponechat a zbytek poslat dál Bobovi. Jednofotonové pulzy zahodí. Jedná se o tzv. PNS útok a je rovněž zobrazen na obrázku 1. K obraně proti němu slouží tzv. návnadové stavy (decoy states). [1]

Pro zvýšení dosahu lze však BB84 najít i s fázovým kódováním. Výraznou modifikací je pak protokol T12 (Toshiba 2012). Jedná se o efektivní variantu BB84, jenž většinou implementuje fázové kódování. Hlavním rozdílem je asymetrické používání bází. Tedy minoritní báze je vybrána např. ve 20 % případů a majoritní báze v 80 % případů. Tento poměr byl využit i v simulaci. [1]

3 KVANTOVÁ BITOVÁ CHYBOVOST (QBER)

Polygony všech tří výše zmíněných variant protokolu byly vymodelovány pomocí programu VPItransmissionMaker[™] Optical Systems a knihovny VPItoolkit QKD. Simulace slouží k měření QBER, nejdůležitějšího parametru v QKD. QBER je vypočítávána v průběhu prosévání klíče. [3]

$$QBER = \frac{\nu_{\rm chyby}}{\nu_{\rm cil}} \tag{1}$$

QBER-Kvantová bitová chybovost [-]

 v_{chyby} – Počet chyb za sekundu [bit/s]

 v_{cil} – Bitová (přenosová) rychlost u cíle [bit/s]

Hodnota QBER bývá u většiny QKD protokolů zásadní pro detekci odposlechu. To znamená, že Eva bude v ideálním případě (SPS) vytvářet chybovost okolo 25 %. Za bezpečnou chybovost je u BB84 protokolů považována hranice 11 %. QBER rovna 50 % znamená nulovou korelaci mezi Alicí a Bobem. Bobovy výsledky jsou tedy zcela náhodné a neodpovídají tomu, co Alice poslala. [3]

4 SIMULACE QKD A MĚŘENÍ QBER

4.1 ZÁVISLOST QBER NA DÉLCE KVANTOVÉHO KANÁLU

V reálném světě je QBER generována nejen Evou, ale i prodlužující se délkou kvantového kanálu (útlum). Obecné zapojení polygonů lze najít níže (pro daný protokol je vždy použit specifický vysílač a přijímač). Jedná se o WCP protokoly využívající návnadových stavů. Měrný útlum kvantového kanálu je nastaven na $\alpha = 0,24 \ dB/km$. S rostoucí délkou $l \ [km]$ tak dochází k postupnému zvýšení útlumu kanálu podle vzorce $a = \alpha l \ [dB]$. Čím je útlum kanálu vyšší, tím vyšší je i pravděpodobnost



Obrázek 2: Obecné zapojení testovaných protokolů.

falešné detekce fotonu (dark count). Délka kvantového kanálu s každou iterací simulace roste o 5 km. Celkem je tak měření provedeno 23krát pro vzdálenosti od 0 do 110 km. Chybovost přístrojů byla taktéž změřena a osciluje kolem 1 %.

4.2 VÝSLEDKY

Výsledné grafy popisují závislost QBER na délce kvantového kanálu, a tedy i na útlumu. Hodnoty QBER pro vybrané délky jsou zapsány v tabulce. Výsledky měření odpovídají tvrzení, že fázové kódování a asymetrické využívání bází má pozitivní vliv na dosah.



Obrázek 3: Výsledky měření QBER.

5 ODHALENÍ ODPOSLECHU POMOCÍ QBER

V následující simulaci už bude vystupovat také odposlouchávající Eva. Eva změří od Alice pocházející fotony v náhodné bázi. Jejich hodnoty následně namoduluje na nové fotony pomocí báze, kterou měřila. Dojde-li při detekci k chybě, je hodnota nového bitu automaticky určena jako 0.



Obrázek 4: Zapojení s odposlouchávající Evou pro protokol BB84-P.

Oproti předchozím měření lze vyčíst prudký nárůst QBER. V ideálním případě s SPS by měla tvořit cca 25 %. Využití WCP stejně jako nedokonalosti zařízení zvyšují QBER až na cca 48 %. To značí velmi nízkou míru korelace mezi Alicí a Bobem (nulová korelace odpovídá 50 % QBER). Drtivá většina kvantových signálů tak k Bobovi dorazí poškozená.



Obrázek 5: Odhalení odposlechu pomocí QBER.

6 NÁVRH TESTOVACÍHO QKD POLYGONU

Pro stavbu vlastního polygonu (QKDN) se počítá s využitím zařízení Clavis3 firmy IDQ. Toto zařízení využívá tzv. COW protokol (Coherent One-way protocol). Obecně se ale předpokládá, že návrh není na QKD protokolu závislý. Pro kompletní šifrovanou komunikaci jsou potřeba tři logické linky. Kvantový kanál pro přenos kvantových signálů, klasický kanál (2 fyzické linky pro duplex), pro výměnu servisních a signálů mezi QKD servery a libovolný kanál pro šifrovanou komunikaci.



Obrázek 6: Návrh testovacího QKD polygonu s multiplexory.

Návrh počítá s testováním vlivu multiplexorů na kvantové signály. Z tohoto důvodu se plánuje využít OADM s nižším útlumem než u běžných multiplexorů. Nekvantové signály lze cestou obnovit, a tak mohou být agregovány pomocí běžných DWDM. Navržené vlnové délky QKD serveru jsou nastaveny na $\lambda_k = 1551,72 nm$ pro kvantový kanál a $\lambda_{s1} = 1563,047 nm$ a $\lambda_{s2} = 1563,863 nm$ pro servisní data (C-pásmo). Šifrovaný kanál může mít přidělenu libovolnou vlnovou délku dle potřeby. Rychlost generování klíče je v = 1,4 kbit/s s maximální dosahem cca 58 km. To odpovídá útlumu a = 14 dB. Pro správu QKD serveru mohou být dočasně připojeny počítače. [4]

Jakožto testovací šifrátor bylo navrženo open-source řešení QKD in OpenSSL, poskytující L4 datovou ochranu. Tento typ QKDN bývá kvůli připojení bod-bod taktéž nazýván QVPN. Pro komunikaci mezi QKD serverem a šifrátorem je nutné, aby měla obě zařízení jednotné standardizované rozhraní – ETSI Key Delivery API. [5]

7 ZÁVĚR

QKD sítě začínají být v současnosti testovány čím dál častěji a současně začínají být nasazovány v ostrém provozu. Příkladem mohou být testovací sítě DARPA (USA), SwissQuantum (Švýcarsko), mezinárodní evropský projekt SECOQC nebo japonská městská QKDN v Tokiu. V provozu jsou rovněž rozsáhlé integrované čínské páteřní (např. Peking-Šanghaj) a městské (např. Peking, Wuchan, Šanghaj) sítě. V lednu roku 2021 tvořila celková délka vláken čínských sítí cca 4600 km. Kromě nich jsou rovněž testovány free-space (bezdrátové) verze QKD protokolů. Nejznámějším zástupcem je patrně síť QUESS, vyžívající satelit Mozi/Micius. [6]

- PIRANDOLA, Stefano, Ulrik ANDERSEN, Leonardo BANCHI, et al. Advances in Quantum Cryptography. Advances in Optics and Photonics [online]. [cit. 2021-03-10]. ISSN 1943-8206. Dostupné z: doi:10.1364/AOP.361502
- [2] PROTOCOLEBB84 [online]. Université de Nice, 2015 [cit. 2021-03-10]. Dostupné z: http://physique.unice.fr/sem6/2014-2015/PagesWeb/PT/Tomographie/?page=bb84
- [3] MUGA, Nelson J., Mário F. S. FERREIRA a Armando N. PINTO. QBER Estimation in QKD Systems With Polarization Encoding. Journal of Lightwave Technology [online]. 2011, 29(3), 355-361 [cit. 2021-03-10]. ISSN 0733-8724. Dostupné z: doi:10.1109/JLT.2010.2099643
- [4] IDQ [online]. Genève: IDQ, 2020 [cit. 2021-03-10]. Dostupné z: https://www.idquantique.com
- [5] RIJSMAN, Bruno. Quantum Key Distribution (QKD) in OpenSSL [online]. 2019 [cit. 2021-03-10]. Dostupné z: https://brunorijsman.github.io/openssl-qkd/doc/the-etsi-qkd-api.html

CUSTOM INTELLECTUAL PROPERTY BLOCK FOR LBLOCK CIPHER

Jakub Jedlička

Bachelor Programme (3), FEEC BUT E-mail: xjedli24@stud.feec.vutbr.cz

Supervised by: David Smékal E-mail: smekald@feec.vutbr.cz

Abstract: This paper presents the implementation of a lightweight cryptographic cipher for a hardware-constrained device. Describes the basic problems of lightweight cryptography on the field programmable gate array (FPGA) and its one representative cipher LBlock. Furthermore deals with the implementation of the LBlock cipher in a very high speed integrated circuit hardware description language (VHDL) on the FPGA. The LBlock cipher is used with a custom advanced extensible interface (AXI) wrapper for the creation of a custom intellectual property (IP) block. This IP block will be used to cipher files on the development board ZYBO Z7-20 powered by Zynq-7000. The final part of the paper describes testing of the IP block with a defined set of the inputs and the outputs are validated with the correct outputs.

Keywords: FPGA, LBlock, lightweight cipher, ZYBO Z7-20, Zynq-7000, IP block

1 ÚVOD

Kryptografie se využívá v různých formách již několik let, ale její expanze začala až s příchodem výpočetních zařízení. Kvůli potřebám šifrování na výkonově omezených zařízeních jako jsou mikrokontroléry, čipy využívající identifikace na rádiové frekvenci (RFID) a jim podobná zařízení, se vytvořila nová kategorie šifer nazývaná lehká kryptografie [1]. Standardní šifry by nemusely být pro tyto zařízení vhodné a to z důvodu jejich vyšších nároků na výkon. Proto pro kategorii lehkých šifer byly vytvořeny nové šifry jako je šifra PRESENT, vycházející ze šifry AES, dále čínská šifra LBlock, která vychází z šifry DES a mnoho dalších [2].

V posledních letech se zvýšila implementace kryptografie na programovatelných hradlových polích (FPGA) kvůli jejich čím dál většímu komerčnímu rozšíření. Hlavním důvodem, který k tomu vedl, je jejich výkon oproti procesorům a možnost přeprogramovaní na rozdíl od integrovaných obvodů určených pro aplikaci (ASIC) čipů. Lehká kryptografie se implementovala převážně na ASIC čipy, které byly výhodné, pokud se vytvářely ve velkém množství. Hlavní nevýhodou ASIC čipů je nemožnost přeprogramování po jejich výrobě. Z tohoto důvodu se čím dál více snaží prosadit implementace šifer lehké kryptografie na FPGA čipy, které jsou nízkoenergetické a v dnešní době cenově dostupné. Přeprogramování FPGA umožňuje opravit případné nově nalezené chyby v šifrách po implementaci. Poslední výhodou je možnost přidat kryptografický IP blok do návrhu nebo upravit aktuální návrh ke zmenšení celkové spotřeby plochy a tím i nákladů s tím spojených [3].

2 ŠIFRA LBLOCK

Šifra LBlock spadá do blokových šifer lehké kryptografie. Je založena na Feistelově síti, kde vstupem i výstupem je blok bitů o velikosti 64. Využívá klíč o délce 80 bitů, který je stanoveným minimem pro šifry lehké kryptografie. LBlock se skládá z 32 kol, kdy se pracuje s dvěma bloky od délce 32 bitů. Levá strana je vstupem do funkce kola a novou pravou stranou pro následující kolo. Na pravou stranu je využit bitový posun o 8 bitů a exkluzivní disjunkce s výstupem funkce kola. Funkce kola se

skládá ze tří částí. První částí je exkluzivní disjunkce vstupu a klíče, druhou funkce konfúze starající se o záměnu 4bitových bloků a třetí funkce difúze využívající 4bitové S-boxy. Dále je každé kolo upravován klíč pomocí bitových posunů, dvou S-boxů a operace exkluzivní disjunkce [4].

Šifra LBlock byla vybrána na základě srovnání šifer lehké kryptografie. Výběr probíhal z porovnávaní výkonu a hardwarových požadavků a byl proveden na šifrách LBlock, LED, mCRYPTON, PRESENT a SIMON. Jako nejoptimálnější se jeví šifry LBlock a PRESENT, které mají velmi podobné nároky. Z uvedených šifer byla vybrána šifra LBlock, jejíž popis implementace pomocí VHDL byl nalezen pouze v odborných článcích ale ne v podobě opensource.

3 IMPLEMENTACE ŠIFRY LBLOCK

Pro implementaci šifry byl vytvořen vývojový diagram, který je na obr. 1. Tento diagram popisuje průběh při procesu šifrovaní.



Obrázek 1: Vývojový diagram šifry LBlock

Důležité části z vývojového diagramu byly rozčleněny na moduly. Těchto modulů je celkově 17, z toho 10 modulů je pouze pomocných k funkcionalitě S-boxů a jsou rozděleny kvůli přehlednosti kódu. Zbylých 7 modulů tvoří hlavní logiku, kde nejvyšší modul nese název LBlockTOP. Celý návrh využívá 142 vyhledávacích tabulek (LUT) a 216 vstupních a výstupních (I/O) portů.

Modul LBlockTOP má v sobě jako jediný sekvenční logiku, která využívá hodin (clk). Řídí levou a pravou stranu Feistelovy sítě a také hodnoty klíče kola a počet kol opakovaní. Modul v každém kole volá dva submoduly, z nichž jeden má na starosti výpočet klíče a druhý průběh celého kola.

4 IMPLEMENTACE VLASTNÍHO IP BLOKU

Při vytváření IP bloku se vzala v potaz architektura cílové vývojové desky ZYBO Z7-20, která obsahuje 40 I/O portů dostupných pro uživatele. Jak vstupní i výstupní blok k šifrovaní, tak klíč mají velikost větší než je možný počet I/O portů. Z tohoto důvodu musel být použít tzv. "AXI stream wrapper", využívající sekvenční logiku. AXI stream wrapper má za úkol načtení dat k šifrovaní ve dvou cyklech, po kterých následují tři cykly pro čtení klíče. Každý z těchto cyklů načítá blok 32 bitů dat. U načítání klíče se ve třetím cyklu vybírá pouze 16 bitů z 32 bitů, jelikož by to způsobilo přetečení signálu. Hodnota 32 bitů byla vybrána z důvodu potřeby využití řídících signálů pro fáze načítaní a výstupu dat, které jsou nad rámec 32 bitů. Druhým důvodem je, že hodnota 32 je mocnina čísla dva, stejně jako hodnoty 64 a 16. V důsledku této skutečnosti lze efektivně rozdělit vstupní a výstupní bloky, aniž by byla potřebná nadbytečná režie. Po načtení obou hodnot modul předává hodnoty modulu LBlockTOP, který provádí šifrovaní a vrací šifrovaná data. Tyto data poté načítá a posílá na výstup ve dvou cyklech. V případě potřeby úpravy AXI stream wrapper, například z důvodu nedostatečného počtu I/O portů, který má druhá verze vývojové desky ZYBO Z7 stanoven na 32, by musela být snížena velikost vstupu a výstupu na 16 bitů a navýšen počet cyklů pro vstup a výstup dat. Velikost 16 bitů je zvolena z důvodu zmíněných v předchozím odstavci.

5 VALIDACE FUNKČNOSTI MODULŮ

Pro každý modul byly napsány testy, aby byla ověřena validita výstupů dle vstupních dat. Nejdůležitější částí při validaci bylo ověřit správnost modulu LBlockTOP a jeho AXI stream wrapper. Tato validace byla provedena pomocí simulace ve vývojovém prostředí Vivado Design Suite. Pro uvedené dva moduly byly použity vstupní data z návrhové publikace Wenling a Zhang (2011)[4], které jsou uvedeny v tabulce 1.

	Tubulka 1. Vysteane vektory pro Ebioek v nexadeennamini tvara [4].							
	Zpráva	Klíč	Šifrovaný text					
1	00 00 00 00 00 00 00 00 00	00 00 00 00 00 00 00 00 00 00 00	c2 18 18 53 08 e7 5b cd					
2	01 23 45 67 89 ab cd ef	01 23 45 67 89 ab cd ef fe dc	4b 71 79 d8 eb ee 0c 26					

Tabulka 1	: V	ýsledné	vektorv	pro	LBlock	v	hexadecimalním tvaru [41
I abama I	L• • .	ysicane	ventory	pro	LDIOCK	v	nexaucennamin tvara	-

Simulace probíhaly pomocí tzv. "testbench", přes které jsou do modulů postupně načítány hodnoty z tabulky 1. Výsledky zprávy 1 a zprávy 2 pro modul LBlockTOP jsou zobrazeny na obrázku 2, kde vstupní hodnota zprávy je označena *data_in*, hodnota klíče *key_in* a výstupní hodnota *data_out*.

😼 clk	1							
> 😻 data_in[63:0]	000000000000000	00000000000000						
> 😻 key_in[79:0]	000000000000000000000000000000000000000	000000000000000000000000000000000000000						
> 😼 data_out[63:0]	c218185308e75bcd	00000000000000000000000000000000000000						
18 clk	0							
> 😻 data_in[63:0]	0123456789abcdef	0123456789abcdef						
> 😻 key_in[79:0]	0123456789abcdeffedc	0123456789abcdeffedc						
> 😽 data out[63:0]	4b7179d8ebee0c26	00000000000000	4b7179d8ebee0c26					

Obrázek 2: Testovaní vstupů pro LBlockTOP

Výsledky zprávy 1 a zprávy 2 pro modul axi_stream_wrapper jsou zobrazeny na obrázku 3. Je zde zobrazen výstup obou dvou šifrovaných zpráv a je také vidět přechodná fáze načítaní z *M_AXIS_DATA* do vytvořeného pomocného logického vektoru *ciphertext*. V pomocném vektoru se nachází jako první hodnota šifrovaného textu pro zprávu 1 a po dvou cyklech hodin hodnota pro zprávu 2.

ACLK	0							
M_AXIS_DATA[31:0]	0000000	00000000	4b71	79 d8	ebeel	0c26	00000000	
> 🕏 ciphertext[63:0]	4b7179d8ebee0c26	c218185308	Be75bcd	4b7179d80	8e75bcd	4b7179d	8ebee0c26	

Obrázek 3: Testovaní výstupu pro axi_stream_wrapper

6 POROVNÁNÍ IMPLEMENTACE ŠIFRY LBLOCK S OSTATNÍMI ZÁSTUPCI

Tabulka 2 porovnává požadovaný počet LUT tabulek vybraných šifer, kde hodnoty LUT jsou převzaty z článku DIEHL (2017) [5]. Výstupem této implementace je v tabulce uvedená hodnota LUT pro šifru LBlock.

Název šifry	LBlock	PRESENT-80	LED-80	SIMON 96/96	AES-128
Počet LUT	142	311	358	435	318

7 NÁROKY IP BLOKU A JEHO VYUŽITELNOST

V tabulce 3 lze vidět jednotlivá rozložení prostředků pro modul axi_stream_wrapper podle funkcionálních kategorií po syntéze ve Vivado design Suite.

			2	1	1	11	
Název	LUT6	LUT5	LUT4	FDRE	I/O	LUT celkově	Flips-flops celkově
Počet	5	4	6	9	36	15	9

Tabulka 3: Využití prostředků pro AXI stream wrapper.

Z tabulky je patrné, že se jedná o návrh s nízkými požadavky na vyhledávací tabulky. IP blok je navržen pro vývojovou desku ZYBO Z7-20, která má 53 200 dostupných vyhledávacích tabulek. To znamená využití přibližně 1% celkové kapacity. V případě vstupů a výstupů lze tento návrh považovat za náročný.

Momentální cílová frekvence pro hodiny ACLK je 125MHz, kdy probíhá testovaní z možností navýšení frekvence. Celková spotřeba návrhu na čipu by se dle kompilátoru měla pohybovat na úrovni 0,106W. Aktuální návrh by měl být použitelný na většině FPGA, které splňují podmínky dostatečného počtu I/O portů.

8 ZÁVĚR

V článku byla přiblížena problematika lehké kryptografie na FPGA a implementace modulu šifry LBlock a jejího IP bloku, který je určen pro šifrovaní na vývojové desce ZYBO Z7-20. Funkčnost IP bloku popsaného v článku je již implementována a otestována. Dle testů se implementace IP bloku prokázala jako funkční. IP blok není finální a dále bude upravována maximální frekvence k nalezení optima s dostupnými prostředky.

- [1] POSCHMANN, Axel. LIGHTWEIGHT CRYPTOGRAPHY: Cryptographic Engineering for a Pervasive World [online]. Bochum, 2009, [cit. 5. 3. 2021]. Dostupné z URL: <https: //www.ei.ruhr-uni-bochum.de/media/crypto/veroeffentlichungen/ 2011/09/16/thesisp.pdf> Disertační práce. Ruhr-University Bochum, Germany, Faculty of Electrical Engineering and Information Technology.
- [2] NEKUŽA, Karel, Odlehčená kryptografie pro embedded zařízení [online]. Brno, 2016, [cit. 5. 3. 2021]. Dostupné z URL: <https://www.vutbr.cz/studenti/zav-prace? zp_id=93665.> Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce Ing. Zdeněk Martinásek, Ph.D.
- [3] YALLA, Panasayya a KAPS, Jens-Peter, Lightweight Cryptography for FPGAs [online]. 2009 International Conference on Reconfigurable Computing and FPGAs, Cancun, Mexico, 2009, s. 225-230 [cit. 5. 3. 2021]. Dostupné z: doi: <https://doi.org/10.1109/ReConFig.2009.54>
- [4] WENLING, Wu a ZHANG, Lei, LBlock: A Lightweight Block Cipher. Applied Cryptography and Network Security [online]. Germany: Springer, Berlin, Heidelberg, 2011, s. 327-344 [cit. 5. 3. 2021]. ISBN 978-3-642-21554-4. Dostupné z: doi: https://doi.org/10.1007/978-3-642-21554-4_19>
- [5] DIEHL, William, FARAHMAND, Farnoud, YALLA, Panasayya, KAPS, Jens-Peter, GAJ, Kris Comparison of hardware and software implementations of selected lightweight block ciphers [online]. 2017 27th International Conference on Field Programmable Logic and Applications (FPL), Ghent, Belgium, 2017, [cit. 23. 3. 2021]. Dostupné z: doi: <https://doi.org/10. 23919/FPL.2017.8056808>

GRAPHIC USER INTERFACE FOR PASIVE OPTICAL NETWORK CONFIGURATION

Dávid Kováč

Bachelor Degree Programme (3),FEEC BUT E-mail: xkovac52@stud.feec.vutbr.cz

Supervised by: Tomáš Horváth

E-mail: horvath@feec.vutbr.cz

Abstract: This paper presents a Graphic user interface for the configuration of the optical line terminal in passive optical networks. The aim of the GUI is to create an app that makes the configuration of the OLT easier. The GUI connects to the OLT with the Secure Shell protocol. The GUI not only allows the configuration of the OLT but also has the ability to show the status of the connected optical network units and show the activity on the OLT.

Keywords: GUI, PON, Java, SSH

1 ÚVOD

Článok sa venuje vytvoreniu grafického rozhrania na konfiguráciu riadiacej jednotky pasívnej optickej siete. Grafické rozhranie emuluje písanie príkazov do príkazového riadku a zjednoduší proces konfigurácie pre užívateľa, ktorému zaniká potreba znalosti konfiguračných príkazov. Grafické rozhranie poskytuje výpis statusov koncových jednotiek, uchováva udalosti a vykonané zmeny v riadiacej jednotke.

2 PROGRAMOVANIE GRAFICKÉHO ROZHRANIA

Grafické rozhranie je naprogramované v jazyku Java s rozšíreniami SpringWEB[2], Thymeleaf[1] a s knižnicou Jsch[3]. Jazyk Java umožňuje grafickému rozhraniu fungovanie bez ohľadu na operačnom systéme na ktorom bol spustený. SpringWEB umožňuje zobraziť grafické rozhranie vo webovom prehliadači na adrese localhost:8080/. Thymeleaf rozširuje funkcionalitu jazyka HTML, napríklad umožňuje používanie podmienok v jazyku HTML a výmenu dát medzi HTML a Java kódom. Knižnica Jsch poskytuje jazyku Java možnosť pracovať s protokolom SecureShell.

3 PRINCÍP GRAFICKÉHO ROZHRANIA

Princíp grafického rozhrania je založený na spojení užívateľom zadaných dát s konštantnými časťami príkazu. Správnosť užívateľom zadaných dát sa kontroluje na správnosť, napríklad či užívateľ do číselného poľa nezadal písmeno. Správne údaje sa spoja s konštantnou časťou príkazu a odošlú sa riadiacej jednotke na spracovanie. Grafické rozhranie má funkcionalitu prijímania spätnej väzby od riadiacej jednotky. Proces kontroly a spájanie prebieha na pozadí a užívateľ ho nevidí.

4 PRIHLÁSENIE DO ZARIADENIA

Na prihlásenie do zariadenia užívateľ musí na webovej stránke localhost:8080/index zadať prihlasovacie meno, heslo, IP adresu zariadenia a port na ktorom zariadenie pracuje, zvyčajne to je port 22, alebo 2222. Po vyplnení údajov užívateľ stlačí tlačidlo "prihlásiť" a tým odošle dáta z webovej stránky na spracovanie do programu písanom v Jave. V Jave sa tieto údaje spoja s príkazmi z knižnice Jsch a nastane pokus o prihlásenie do zariadenia. Po úspešnom prihlásení je užívateľ presmerovaný na localhost:8080/DBA, kde môže nastaviť DBA profil, alebo pomocou tlačidiel vybrať inú sekciu konfigurácie. Ak je prihlásenie neúspešné je vyhlásená chyba a užívateľ sa môže znova pokúsiť o prihlásenie.

Aby sa grafické rozhranie zobrazilo na webovej stránke je potrebné v Jave vytvoriť metódu, ktorá vracia názov HTML súboru, pred ktorou je príkaz @GetMapping("/adresa"). Týmto sa webová stránka a kód písaný v Jave prepojí. Textové polia do ktorých sa zadávajú údaje sú zapuzdrené v HTML tagu <form></form>. Tieto údaje sa posielajú, ako objekt s názvom post do metódy, ktorá má pred sebou @RequestMapping(value = "/adresa", method = RequestMethod.POST). Táto metóda pri prvom spustení prihlasuje užívateľa do zariadenia a v následných cykloch slúži na konfiguráciu DBA profilu. Pri odoslaní údajov sa jednotlivé časti spoja do jedného slova, ktoré je oddelené čiarkami (napr. a, b, c, d) čo vyvoláva nutnosť rozdelenia pomocou príkazu .split(post), ktorý slovo rozdelí a uloží do poľa typu String. Na pokus o prihlásenie sa využije funkcia session, kde pomocou getSession() sa pridávajú prihlasovacie meno, IP adresa a port a pomocou .setPassword() sa priradí heslo. S funkciou .connect nastane pokus o prihlásenie.

Login:							
Pasword:							
Ip address:							
Port:							
_	Prihlasiť						
@GetMapping(<pre>@GetMapping("/index")</pre>						
public Strin	g index(Model model){						
Post pos	t = new Post();						
firstRur							
model.ad	<pre>model.addAttribute("post",post);</pre>						
return '							
}							

Obrázek 1: Prihlásenie do zariadenia

5 KONFIGURÁCIA

Ako už bolo spomenuté, konfigurácia prebieha po úspešnom prihlásení do zariadenia, napríklad na adrese localhost:8080/DBA, kde užívateľ zadá meno, typ a požadovanú rýchlosť od DBA profilu. Údaje sa odovzdajú do triedy SSH, ktorá ich ďalej odovzdáva do triedy CommandCompiler. Tá zadané údaje spojí so konštantnou časťou príkazu a vráti triede SSH na odoslanie do zariadenia. Ak nastala chyba so zadanými údajmi tak sa vypíše chybové hlásenie v grafickom rozhraní.

Na prepojenie HTML súboru s Javou je použitý príkaz @RequestMapping(value = "/adresa", method = RequestMethod.POST), ktorý umožňuje pracovať s údajmi poslanými z webovej stránky. Údaje sa odovzdajú do triedy SSH, ktorá pomocou príkazu channel vytvorí komunikačný kanál so zariadením. Pomocou .setCommand("") sa nastaví text na odoslanie. Pre .setCommand("") je potrebné poslať naraz všetky príkazy na konfiguráciu danej časť, príkazy je nutné oddeliť pomocou \n. SSH pred odoslaním príkazu odovzdáva údaje triede CommandCompiler, ktorá má za úlohu údaje spojiť so statickou časťou príkazu. Na rozoznanie o aký príkaz sa jedná, slúži funkcia switch a identifikátor konfigurovanej sekcie odovzdávaný z metódy @RequestMapping.

6 VÝPIS UDALOSTÍ

Výpis udalostí na zariadení je riešený pomocou triedy Changelog, ktorá od jednotlivých metód webových stránok dostáva údaje, čo sa zmenilo. Changelog zapisuje do súboru kto sa prihlásil a čo nakonfiguroval na zariadení. Ďalej zapisuje dátum udalosti a IP adresu užívateľa. Aby sa udalosti vypisovali od najnovšej udalosti k najstaršej, bolo nutné zapisovať do dvoch súborov. Najskôr do dočasného, kde sa zapíše nová udalosť. Následne sa údaje prenesú z trvalého súboru do dočasného súboru. Trvalý súbor sa vyčistí a prekopírujú sa do neho údaje z dočasného súboru a ten sa nakoniec vymaže.

7 VÝPIS STATUSU KONCOVÝCH JEDNOTIEK

Pre výpis statusu koncových jednotiek sa spätná väzba od zariadenia zapíše do súboru. Novovytvorený súbor je prehľadávaný pomocou cyklu while, až kým sa nenájde text "display ont info", ktorý označuje začiatok výpisu a (config)\#, ktorý označuje jeho koniec. Cyklus po nájdení začiatku začne zapisovať jednotlivé riadky do objektu Arraylist, umožňujúci uchovávanie viacerých údajov jedného typu, až do nájdenia konca. Arraylist sa odošle na webové rozhranie, kde sa pomocou cyklu foreach vypíše.

8 ZÁVER

Cieľom projektu bolo vytvorenie grafickej nadstavby pre konfiguráciu riadiacej jednotky a na uľahčenie práce s konfiguráciou užívateľovi. Testovanie grafického rozhrania prebiehalo na lokálnej sieti so serverom SSH na mobile prostredníctvom aplikácie SSHDroid a na vzdialenej sieti cez VPN s SSH pripojením na riadiacu jednotku optickej siete. V oboch prípadoch sa grafické rozhranie vedelo pripojiť bez problémov cez SSH protokol a príkazy boli odosielané rýchlo a spoľahlivo. Výsledkom je komplexné riešenie grafického konfiguračného prostredia, ktoré značne používateľovi uľahčuje prácu s konfiguráciou riadiacej jednotky optickej siete. Toto riešenie je možné po určitých úpravách kódu implementovať aj na iné zariadenia podporujúce konfiguračné nastavenia prostredníctvom protokolu SSH.



Obrázek 2: Výpis z riadiacej jednotky po úspešnom prihlásení

- [1] Thymeleaf [online]. [cit. 2021-3-10]. Dostupné z: <u>https://www.thymeleaf.org</u>
- [2] SpringWEB [online]. [cit. 2021-3-10]. Dostupné z <u>https://spring.io/guides/gs/serving-webcontent/</u>
- [3] JavaSecureChannel JSch [online]. [cit. 2021-3-10]. Dostupné z: http://www.jcraft.com/jsch/

COMPUTER NETWORKS BASED ON VYOS AND CISCO IOS OPERATING SYSTEMS

Denis Jalovecky

Bachelor Degree Programme (3), FEEC BUT E-mail: xjalov05@vutbr.cz

Supervised by: Aneta Koláčková E-mail: xkolac15@stud.feec.vutbr.cz

Abstract: This thesis focuses on configurations and issues of network protocols in operating systems of Cisco IOS and VyOS routers. One of the main goals include choosing the most appropriate emulator and installation of this emulator in virtual environment. The thesis also describes the properties and differences of protocols and their application in the network operating systems mentioned above.

After learning about possibilities of emulations, two scenarios, in which the protocols were applied and analyzed, were created. These scenarios meet standard of a lab excersise which takes around one and a half hour. The most frequent issues and their solutions are mentioned as well.

Keywords: Cisco, VyOS, networking, internet protocols, firewall

1 ÚVOD

Cieľ om tejto práce je poukázať na jednotlivé funkcie smerovačov s operačnými systémami Cisco IOS-XR a VyOS. Poukázať na rozdiely medzi nimi z hľadiska funkčnosti a konfigurácie. Taktiež sa bude zaoberať možnosť ami emulácie týchto operačných systémov na hardvéri a vo virtuálnom prostredí pomocou emulačných programov.

Ďalšou časť ou bude využitie sieť ových protokolov ako OSPF (Open Shortest Path First), BFD (Bidirectional Forwarding Detection), IPv4 (Internet Protocol version 4), IPv6 (Internet Protocol version 6) na fungovanie komunikácie medzi Ciscom IOS-XR (Internetworking Operating System-XR), VyOS a užívateľ mi. V dnešnej dobe zohráva významnú úlohu aj bezpečnosť a tak sa práca zaoberá aj zabezpečením sieť ovej komunikácie.

Navrhnuté sú dva rôzne scenáre, ktoré sa venujú analýze a problematike použitých sieť ových protokolov. Obidva scenáre by mali zodpovedať dĺžke laboratórneho cvičenia o dĺžke cca 1,5 hodiny.

Prvý scenár obsahuje nastavenie IPv6 adries na portoch smerovačov a na užívateľ ov. Následne konfiguráciu protokolu OSPFv3 (Open Shortest Path First version 3) s priradením smerovačov a užívateľ ov do oblastí. Posledná časť sa zameriava na protokol BFD, ktorý rýchlejšie zistí, keď spojenie zlyhá.

Druhý scenár obsahuje konfiguráciu IPv4 adries pomocou protokolu DHCP (Dynamic Host Configuration Protocol), nastavenie OSPF a konfiguráciu firewallu na VyOS smerovači a filtrovanie paketov na smerovači Cisco IOS.

2 MOŽNOSTI EMULÁCIE

Existuje viacero možností emulácie sieť ových prvkov. Jedným zo spôsobov je nahranie obrazu na ľubovoľný hardvér, ktorý podporuje virtualizáciu (stolný počítač, notebook). Ďalším spôsobom je nahranie obrazu smerovačov priamo do programov, ktoré podporujú virtualizácie, t.j. VMWare alebo VirtualBox. Poslednou možnosť ou je vytvorenie servera v programe VMware alebo Virtualbox a

pomocou emulačných programov akými sú GNS3 (Graphical Network Simulator-3) alebo EVE-NG (Emulated Virtual Environment-Next Generation), je možné nahrať obrazy smerovačov a následne ich konfigurovať. [1]

Táto práca sa realizuje podľa poslednej možnosti. Najprv je potrebné stiahnuť z oficiálnych stránok program VMware Workstation Pro, ktorý sprostredkúva vytvorenie servera. Následne aj emulátor GNS3, ktorý slúži priamo na virtualizáciu. Na stránkach GNS3 sa nachádza aj server pre VMware, ktorý je taktiež potrebné stiahnuť a otvoriť priamo vo VMware. Nastavenie parametrov je následovné: 4GB RAM a 20GB pamäť na disku. Serveru je automaticky priradená IP adresa 192.168.106.128. Ak je server správne nainštalovaný, po spustení programu GNS3 by sa mala zobraziť pri IP adrese, vyššie spomenutej, zelená bodka. Na obrázku 1 je zobrazený emulátor GNS3 so smerovačmi a užívateľ mi na ľavej strane. Na pravej strane sa nachádza server GNS3 VM v programe VMware.



Obrázek 1: GNS - Prostredie

3 SCENÁRE PRE KONFIGURÁCIU

Práca obsahuje 2 scenáre pre laboratórne cvičenia, v ktorých sa využívajú rôzne sieť ové protokoly pre prepojenie, zabezpečenie a plynulú komunikáciu. Týmto sa ukáže rozdiel konfigurácií a nastavení medzi smerovačmi Cisco IOS XR a VyOS.

3.1 PRVÝ SCENÁR

V prvom scenári sa najprv konfigurujú IPv6 adresy pomocou vhodne vytvorenej adresnej tabuľ ky.Priradenie IPv6 adries vidieť na obrázku 2. Medzi smerovačmi Cisco a VyOS je nastavený subinterface, a až na ten sa konfiguruje IPv6 adresa. Pre VyOS router je nutné nastaviť VLAN, ktorá funguje ako subinterface.Tu je možné vidieť zásadný rozdiel medzi konfiguráciami týchto smerovačov.

Ku každému smerovaču sú pripojení dvaja užívatelia (PC). Ďalším krokom je konfigurácia OSPFv3 protokolu, práve táto verzia podporuje IPv6. Na obrázku 2 sú zobrazené oblasti pre konfiguráciu protokolu OSPF, kde je možné vidieť, že Cisco je hraničný smerovač, ktorý prepája 3 vytvorené oblasti. Oblasti v tejto práci sú simulované z dôvodu, že keď dôjde k zmene v sieti, tak OSPF rozposiela LSU (Link State Update) pakety len v rámci jednej oblasti, kde nastala zmena. Týmto sa znižuje zať aženie siete. [2]



Obrázek 2: OSPF oblasti a IPv6 adresy

Overenie konfigurácie protokolu OSPF na smerovačoch je možné napríklad pomocou príkazu ping. Na obrázku 3 sa overovalo spojenie medzi PC1 a PC4, ktoré bolo úspešné.

PC1>	2001:db8:2	2ab::c43/64		
2001: 2001: 2001: 2001: 2001: 2001:	db8:2234::1 db8:2234::1 db8:2234::1 db8:2234::1 db8:2234::1	<pre>icmp6_seq=1 icmp6_seq=2 icmp6_seq=3 icmp6_seq=4 icmp6_seq=5</pre>	ttl=62 ttl=62 ttl=62 ttl=62 ttl=62 ttl=62	time=69.949 ms time=6.158 ms time=5.966 ms time=6.752 ms time=6.223 ms

Obrázek 3: Overenie konfigurácie

Poslednou experimentálnou súčasť ou scenáru je aplikovanie BFD protokolu na protokol OSPF. BFD protokol primárne slúži na veľmi rýchlu detekciu výpadku linky, ale monitoruje aj parametre ako: stratovosť paketov, jitter a latenciu. Informácie posiela ď alším protokolom.

3.2 DRUHÝ SCENÁR

Druhý scenár bude komunikovať pomocou IPv4 adries. Tentokrát sa nevytvára adresný plán manuálne, ale pomocou protokolu DHCP, ktorý bude na oboch smerovačoch. Podobne ako v prvom scenári sa aj tu konfiguruje protokol OSPF, ale verzia 2. Práve táto verzia podporuje IPv4 adresy. Po konfigurácií a pred ďalším nastavovaním je potrebné overiť spojenie medzi smerovačmi a koncovými užívateľmi.

Ďalším krokom je nastavenie firewallu na smerovači VyOS. Firewall slúži napríklad pre povolenie alebo zakázanie protokolu ICMP (Internet Control Message Protocol), do ktorého spadá príkaz ping. Taktiež je možné obmedziť komunikáciu z určitých IP adries a portov. [3]

Cisco smerovač nemá firewall, obmedzenie komunikácie je možné len cez ACL (Access Control Lists-Prístupové zoznamy). Pomocou ACL sa zakáže komunikácia s jedným užívateľ om, tu si treba dať pozor, pretože v prípade zmeny IP adresy je potrebné ACL vymazať a nastaviť odznova. Výber práve týchto protokolov použitých v druhom scenári je z dôvodu, že pri ich konfigurácií a nastavovaní je vidieť rozdiely medzi smerovačmi. Po úspešnej konfigurácií sa simuluje troubleshooting (riešenie chýb) pri nastavovaní OSPF alebo DHCP hlavne medzi smerovačmi. Jedná sa o najčastejšie chyby a ich riešenia, pričom sa porovnajú jednotlivé konfigurácie smerovačov.

4 ZÁVER

V tejto práci bol vypracovaný spôsob, kde sa používali programy GNS3 a VMWare, pomocou ktorých sa vytváralo virtuálne prostredie pre smerovače Cisco a VyOS. Program VMware slúžil na vytvorenie virtuálneho servera a GNS3 priamo na emuláciu. Pri inštalácií daných programov si treba dávať pozor na verzie programov a verziu servera na VMware. Pri odlišných verziách smerovače nefungujú ako by mali. Napríklad pri vydaní novej verzie GNS3 nebola vydaná nová verzia práve pre server, VyOS smerovač mal problémy s ukladaním konfigurácie. Pri reštarte smerovača sa celá konfigurácia stratila. Ďalšie možné problémy a ich riešenia sú popísané v bakalárskej práci.

V prvom scenári bolo zahrnuté vytvorenie adresného plánu IPv6 adries a následné nastavenie na rozhrania smerovača, kde medzi smerovačmi bol použitý subinterface. Protokol OSPFv3 bol natavený pre komunikáciu koncových užívateľ ov a protokol BFD pre rýchlu detekciu zlyhania spojenia a monitorovania parametrov siete.

V druhom scenári bol využitý protokol IPv4. Tentokrát sa nevytváral manuálne adresný plán, ale vytvoril ho protokol DHCP aplikovaný na oboch smerovačoch. V tomto scenári sa použil OSPFv2 protokol, ktorý funguje výhradne len pri IPv4 adresách. Po overení fungujúceho spojenia medzi užívateľmi sa konfiguroval firewall na smerovači VyOS a ACL na smerovači Cisco. Po konfigurácií sa simuloval troubleshooting spojený s protokolmi IPv4, OSPF, ACL a Firewallu.

Pri realizovaní vytvorených scenárov bolo vidieť aké sú konfigurácie smerovačov rozličné. Či už pri nastavovaní IP adries, kde na Cisco smerovači sa najprv treba dostať na rozhranie a až potom nastaviť IP adresu, u VyOS je to možné jedným príkazom. Podobné rozdiely boli aj pri nastavovaní OSPF. Ďalším dôležitým rozdielom bol práve subinterface a jeho konfigurácia. U smerovača Cisco bolo nastavenie jednoduchšie. Porty routra Cisco treba zapínať po konfigurácií, VyOS ich zapne automaticky. VyOS má funkciu zabezpečenia systému (firewall), pričom na Ciscu je možný len ACL. Okrem tejto funkcionality s firewallom som nenarazil na zásadný rozdiel, ktorý by jeden z operačných systémov uprednostňoval pred iným. Je tým myslená napríklad nefunkčnosť protokolov (OSPF, BFD) pri virtualizácií OS v emulátoroch. V konečnom dôsledku to zavisí len na rozhodnutí užívateľ a a jeho preferencií, ktorý sieť ový operačný systém si vyberie.

- [1] CBT Nuggets. CBT Nuggets [online]. Copyright ©. [cit. 11.3.2021] Dostupné z URL: https://www.cbtnuggets.com/blog/career/career-progression/5-best-network-simulators-for-cisco-exams-ccna-ccnp-and-ccie>.
- [2] OSPF protokol [online]. Copyright © [cit. 9.03.2021]. Dostupné z URL: http://www.cs.vsb.cz/grygarek/SPS/lect/OSPF/ospf.html.
- [3] VyOS 1.4.x (sagitta) documentation. [online]. Copyright © [cit. 10.03.2021]. Dostupné z URL: https://docs.vyos.io/en/latest/configuration/firewall/index.html DE, b-Quadrat, 2004, s. 131-145, ISBN 3-933609-02-X.

AUTOMATIC ACQUISITION OF VALUES FROM MEA-SUREMENT DEVICES WITHOUT COMMUNICATION IN-TERFACE

Martin Dohnálek

Bachelor Degree Programme (3), FEEC BUT E-mail: xdohna46@stud.feec.vutbr.cz

Supervised by: Jan Kunz

E-mail: kunzj@feec.vutbr.cz

Abstract: This paper deals with optical character recognition (OCR) of measured values from displays of measuring instruments without communication interface. Proposed algorithm functions as a bridging between a screen of the instrument and a measuring software. It enables the acquisition of the displayed value from a snapped picture from the input camera stream, so that it is comprehensible for computers. The execution is, after necessary initialization done by the user, fully automated. Supported camera connection interfaces are USB and WiFi, meaning that either standard office webcam, or smartphone with third party app running is usable. The final algorithm recognized both 99 % valid values and its speed was exceeding the refresh rate of most common instruments (as fast as 34 ms per iteration). This means that it is not bottlenecking the measurement itself.

Keywords: Computer vision, optical character recognition, automated measurement systems, LabVIEW, preprocessing, character segmentation

1 ÚVOD

Automatizované měřicí systémy jsou nedílnou součástí novodobé laboratorní práce. Snižují nejenom časovou vytíženost experimentátorů při pokusech, ale také eliminují chybovost odečtení hodnoty z přístroje. Automatizované měření je však možné pouze s použitím takových přístrojů, které disponují možností přenášet naměřená data do obslužného softwaru spuštěného například na počítači. Vzhledem k ceně kvalitního laboratorního vybavení je obměna přístrojů zdlouhavý proces a ladem tak často leží kvalitní přístroje, které nejsou využívány právě kvůli absenci komunikačního rozhraní. Řešením této problematiky se zabývá následující text. Popisuje aplikaci optického rozpoznávání znaků, anglicky Optical Character Recognition (OCR), které je podoblastí rozpoznávání struktur v počítačovém vidění [1], k nahrazení experimentátora v procesu odečítání hodnot z přístrojů. Znakům nasnímaným webkamerou je algoritmicky přiřazen počítači srozumitelný význam. Po zralé úvaze bylo rozhodnuto, že k vypracování této problematiky bude vhodné využít programovacího jazyka LabVIEW, předně kvůli robustním nástrojům sloužícím k práci s obrazem, ale také díky faktu, že ve většině laboratoří je LabVIEW pro účely měření využíváno a bude tak možné implementovat vytvořený kód do již existujících programů.

2 OPTICKÉ ROZPOZNÁVÁNÍ ZNAKŮ

Základní strukturu aplikace pro realizaci OCR lze rozdělit do čtyř základních částí: získání obrazu, preprocessing, segmentace znaků a samotné rozpoznání znaků s případným postprocessingem [2]. S tímto dělením korespondují také dílčí kroky, které musely být algoritmem splněny. Převod z grafické podoby na význam zajišťuje dataset, což je soubor obsahující podobu znaků seřazených do skupin podle jejich významu. Rozpoznávání potom spočívá ve vyhodnocování podobnosti na-

snímaného znaku se skupinami známých znaků, přičemž hledá se největší shoda. Jedná se tedy o OCR princip srovnávání se vzorem.

3 ALGORITMUS PRO TVORBU DATASETU

Na obrázku 1 je zobrazeno blokové schéma navrženého algoritmu pro vytváření datasetu pro OCR aplikaci. Výchozím bodem celého procesu jsou fotografie displeje měřidla, pro nějž je dataset vytvářen. Je klíčové, aby na nich byly rovnoměrně zastoupeny všechny číslice, jednotky, znaménko minus a desetinná tečka, tedy všechny znaky, které budou později rozpoznávány. Při testování se osvědčilo zastoupení každého znaku alespoň 10x.



Obrázek 1: Blokové schéma tvorby datasetu pro OCR algoritmus

Snímky jsou po jednom předzpracovány, což spočívá v extrakci zvolené barevné složky. Vhodná složka je doporučena algoritmicky na základě srovnání míry zašumění obrazu pro jednotlivé složky, přičemž se hledá ta obsahující nejméně šumu, avšak finální volba zůstává uživateli. Tím dojde k přechodu do stupňů šedi. Dále je z obrazu vyříznuta uživatelem zvolená oblast zájmu, na níž se nachází číslice. V neposlední řadě je snímek podroben binarizaci, kdy číslice jsou reprezentovány logickou 1 a pozadí 0.

Binarizovaná oblast zájmu je dále rozdělena pomocí navrženého stavového automatu na řádky. Oblast je nejprve převedena na pole binárních hodnot, kterým se iteruje po řádcích (řádek pole odpovídá řádku pixelů obrazu). Výchozím stavem je SPACE, reprezentující pozadí znaků. V každém dalším řádku se sleduje četnost výskytu černých pixelů. Pokud je vyšší než zvolený threshold (při tvorbě datasetu byl nastaven na 5 pixelů), uloží se číslo řádku jako hranice a stavový automat přechází do stavu LINE. V něm algoritmus setrvává, dokud iteruje řádky s vyšším počtem černých pixelů, než je zvolený threshold. Zde bylo zapotřebí ošetřit možnost, kdy se v sedmisegmentové číslici může nacházet horizontální mezera, i když znak ještě nebyl ukončen (např. zobrazený údaj "07.100"). Algoritmus v takovém případě čeká s rozhodnutím o ukončení řádku a několik iterací setrvává ve stavu MAYBE SPACE, což pokryje možnost mezery v údaji. Pokud nebyla splněna podmínka návratu do stavu LINE, uloží se druhá hranice řádku a přechází se do stavu SPACE. Iterování probíhá, dokud nebyly prověřeny všechny řádky pixelů v oblasti zájmu.

Jednotlivé nalezené řádky jsou následně rovněž pomocí stavového automatu rozděleny na jednotlivé znaky (číslice). Jsou vytvořeny oblasti zájmu odpovídající hranicím nalezených řádků, které jsou postupně předkládány k vyhodnocení. Celý proces je analogický jako v případě segmentace řádků, kromě toho, že iterace probíhá sloupci pixelů nalezených řádků. Liší se také zvolené pixelové konstanty pro přechod mezi stavy. Ve stavu CHAR (analogický k LINE) je navíc ověřováno, zda hraniční černé pixely v právě zpracovávaném sloupci nejsou vzdálenější, než aktuální uložená hranice znaku (horní a spodní hrana ohraničujícího obdélníku), a pokud ano, dojde k její aktualizaci.

Testování odhalilo problém algoritmu s kurzívou znaků displeje, kdy se překrývaly hranice znaků a nedocházelo tak k segmentaci desetinné tečky. Toto bylo vyřešeno vyrovnáním oblasti zájmu o počet pixelů zjištěný hranovou transformací provedenou metodou First Edge Rake a následným proložením úsečky nalezenou hranou. Výchozí snímek pro multimetr Voltcraft VC940 a výsledek předzpracování a segmentace je na obrázku 2. Samotná tvorba datasetu by nebyla možná bez přispění uživatele. Jsou mu postupně zobrazovány jednotlivé segmentované znaky, k nimž je zapotřebí zadat jejich význam. Tedy, zobrazí-li se například číslice "3", uživatel musí z klávesnice zadat právě "3".



Obrázek 2: Výchozí a výsledný snímek pro multimetr Voltcraft VC940

4 ALGORITMUS PRO ROZPOZNÁVÁNÍ ZNAKŮ

Druhý vytvořený algoritmus slouží již pro samotné rozpoznávání znaků z displejů měřicích přístrojů. K tomuto je využíváno obrazu buď z webkamery připojené přes rozhraní USB, ale také smartphonu připojeného přes WiFi díky aplikaci DroidCAM umožňující bezdrátový přenos obrazu z kamery telefonu do počítače bez výrazného zpoždění. Algoritmus je realizován jako stavový automat. Jeho klíčové části lze shrnout do následujícího řetězce: inicializace, zachycení snímku, předzpracování, rozeznání údaje, rozhodnutí o validnosti, uložení hodnot.

V prvním kroku uživatel nastaví polohu kamery vůči měřícímu přístroji, případně upraví osvětlení podle situace a dále zvolí oblast zájmu, ve které budou rozpoznávány číslice. Snímek je pak v určitých časových intervalech s uživatelem nastavenou periodou zachycen a je podroben obdobnému předzpracování popsanému v předchozí kapitole. K rozeznání údaje je použit dataset vytvořený v ní popsaným postupem, přičemž principiálně se jedná o metodu porovnávání se vzorem. Znaku je tedy přiřazen význam jemu nejvíce podobné skupiny vzorů z datasetu. Výstupem rozeznání je řetězec znaků (číslic, desetinné tečky a znaménka), přičemž následně je vyhodnocena jeho validnost dle následujících kritérií: nesmí obsahovat znak "?", který byl zvolen jako zástupný symbol při nerozpoznání znaku a dále musí obsahovat právě jednu desetinnou tečku. Tím je určena smysluplnost údaje a dle uživatelských preferencí se rozhodne o jeho zařazení do pole výsledků, které bude na konci běhu programu uloženo do výsledného souboru. Ten obsahuje nejenom hodnotu, ale i jednotku a časovou značku, kdy byl údaj kamerou zachycen. Měření probíhalo na dvou multimetrech, Metex M-3850D a Voltcraft VC940 a měřené napětí bylo generováno laboratorním zdrojem Tesla BK 127. K zachycení obrazu byla použita kancelářská webkamera Creative Live! Cam VF0770 a smartphone OnePlus 7. Sledována byla nejenom validnost údajů, ale také rychlost běhu algoritmu, která může být klíčovým parametrem u jistých typů měření rychle se měnících veličin.

4.1 VÝSLEDKY TESTŮ VALIDNOSTI

V tabulce 1 jsou uvedeny výsledky měření pro oba multimetry a obě kamery, přičemž pro každou kombinaci kamera-multimetr bylo provedeno právě 5005 měření. Kritéria, dle kterých bylo rozhodováno o validnosti údaje, jsou uvedena v kapitole 4.

Kamera	Multimetr	Celkem měření	Validní	Nevalidní	Úspěšnost [%]
Creative	Voltcraft VC940	5005	4953	52	98,96
Creative	Metex M-3850D	5005	4034	971	80,60
OnePlus	Voltcraft VC940	5005	4949	56	98,88
OnePlus.	Metex M-3850D	5005	4502	503	86,95

Tabulka 1:Úspěšnost rozpoznávání znaků z displejů Voltcraft a Metex s webkamerou Creative a
smartphonem OnePlus

Z výsledků je patrné, že zatímco pro multimetr Voltcraft je dosahovaná validnost takřka 99 %, pro multimetr Metex je tato hodnota pouhých 80 %, respektive 87 %. Důvodem tohoto rozdílu je rozdílná perioda překreslování displeje obou měřidel, kdy Voltcraft aktualizuje hodnoty frekvencí 3 Hz [2] a Metex 10 Hz [3]. Rozdílná úspěšnost rozpoznávání je tedy vysvětlena statisticky častějším zaznamenáním neúplně vykresleného údaje v případě multimetru Metex.

4.2 VÝSLEDKY TESTŮ RYCHLOSTI

V tabulce 2 jsou zobrazeny délky naměřených časových intervalů potřebných k vykonání jednotlivých kroků. Testování bylo prováděno na notebooku Acer Aspire E5-575G-5660 disponujícím dvoujádrovým procesorem Intel Core i5-7200U (frekvence jádra 2,5 GHz) a 16 GB paměti RAM. Je vidět, že největší část doby iterace zabírá zachycení, přičemž toto se liší mezi USB a WiFi o třetinu. Z časového rozložení jednotlivých kroků plyne dostatečná optimalizace algoritmu.

	ITR	ACQ	PRP	OCR	VAL	SAV
Creative [ms]	100,693	94,121	4,410	1,600	0,006	0,556
OnePlus [ms]	34,401	29,435	3,075	1,641	0,006	0,244

Tabulka 2:Úspěšnost rozpoznávání znaků z displejů multimetrů Voltcraft a Metex se smartpho-
nem OnePlus (Použité zkratky: ITR = celá iterace, ACQ = zachycení snímku, PRP = předzpracování,
OCR = rozpoznání znaků, VAL = validnost a SAV = uložení snímku.)

5 ZÁVĚR

Vyčítání údajů z displejů měřících přístrojů bez komunikačního rozhraní je možné díky využití OCR. Pro účely rozeznávání údaje z displeje byl vytvořen algoritmus, jehož výstupem jsou měřené údaje společně s časovou značkou (časem, kdy byla hodnota zachycena). Pro jeho funkčnost je klíčové využití datasetu, jehož tvorba je pro libovolné měřidlo možná díky druhému vytvořenému algoritmu. Testování algoritmu bylo prováděno na dvou multimetrech (Voltcraft VC940 a Metex M-3850D) za použití dvou kamer (webkamera Creative Live! Cam VF0770 a smartphone OnePlus 7), přičemž byla sledována validnost rozeznaných údajů a rychlost rozpoznávání. Dosažená maximální validnost je takřka 99 % (viz tabulka 1). Rychlost běhu programu se liší dle zvoleného obrazového vstupu, kdy za předpokladu využití WiFi připojení dosahuje až 34,401 ms/iteraci (iterací se rozumí provedení celého řetězce OCR pro daný snímek) (viz tabulka 2).

- [1] EIKVIL, Line. *OCR Optical Character Recognition* [online]. Oslo, 1993 [cit. 2020-10-17]. Dostupné z: https://www.nr.no/~eikvil/OCR.pdf
- [2] ZHANG, Tao, Yanqiu CUI a Yaning YANG. Design of a Character Recognition System Based on LabVIEW. Proceedings of the 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering 2015 [online]. Paris, France: Atlantis Press, 2015, , - [cit. 2020-10-10]. ISBN 978-94-6252-133-9. Dostupné z: doi:10.2991/icmmcce-15.2015.72
- [3] VOLTCRAFT VC940. Voltcraft® [online]. Version 01/13. Hirschau [cit. 2021-02-26]. Dostupné z: https://asset.conrad.com/media10/add/160267/c1/-/%23%23/000123297ML04/000123297ML04.pdf
- [4] METEX M-3850D. *Metex*® [online]. [cit. 2021-02-26]. Dostupné z: https://www.manualslib.com/manual/1369505/Metex-M-3850d.html#manual

UAV SWARM COMMUNICATION AND MISSION PLANNING

Josef Kolísek

Bachelor Programme (3.), FEEC BUT E-mail: xkolis02@stud.feec.vutbr.cz

Supervised by: Jiří Janoušek

E-mail: xjanou09@stud.feec.vutbr.cz

Abstract: Project deals with the issue of swarm UAV flying and uses. The connection is realized by a Wi-Fi network which secure communication with drones and helps to send MAVLink commands through the ground control station via MAVProxy console. The operator is able to control all drones together or each drone separately.

Keywords: Autonomous drone, MAVLink, MAVProxy, Pixhawk, Swarm

1 ÚVOD

Problematika rojového létání se v dnešní době dostává do popředí více a více. S novými inovacemi v elektrotechnickém odvětví a se zvyšováním nároků na technologická zařízení je kladen důraz na automatizaci strojů. Nejsou opomíjeny ani letové jednotky a jejich využití v budoucnosti. Autonomní řízení UAV (Unmanned Aerial Vehicle), jakožto bezpilotního letounu řízeného na dálku, se stává zajímavějším jak pro širokou společnost, tak i pro soukromé organizace. Rojové létaní disponuje širokou škálou využití. Lze jej využít pro zábavu společnosti např.: ekologické ohňostroje. Desítky, ba dokonce stovky dronů jsou schopny pomocí precizních instrukcí a přesných GPS souřadnic rozzářit oblohu. Také mohou sloužit k zabezpečení oblasti před vniknutím cizího objektu. Roje dronů mohou pomoci i při pátracích akcích v rizikových oblastech s častým výskytem požárů či povodní.

V tomto projektu jsem se zabýval řízením a vytvořením komunikačního kanálu mezi pozemní řídící stanicí a rojem dronů. Ve své práci používám komunikační protokoly MAVLink pro zajištění bezpečného letu dronů za pomoci jedné řídící stanice.

2 REALIZACE ROJOVÉHO LÉTÁNÍ

Náplň této práce spočívá v realizaci zapojení a komunikace roje dronů s pozemní řídící stanicí. Komunikace s drony probíhá vzájemným připojením pozemní řídící stanice a dronu. Pro spojení slouží připojené Wi-Fi moduly pracující na frekvenci 2,4 GHz, moduly jsou k řídící stanici a dronu připojené sériovou linkou.

Komunikaci mezi drony a pozemní řídící stanicí zajišťuje již zmíněný protokol MAVLink. Řídící pokyny jsou realizovány softwarovou aplikací v jazyce Python, která odesílá MAVLink rámce z pozemní stanice frekvencí 2,4 GHz pomocí UDP protokolu do modulu připojeného k dronům, rámce dat zpracovává řídící jednotka dronu. Připojeným komunikačním kanálem se pomocí MAVLink rámců nepřenáší jenom řídící signály, ale také se zpětně přenáší telemetrická data o stavových parametrech dronů. Tato data jsou následně vypisována pozemní řídící stanicí a slouží jako kontrolní informace o stavech dronů.

Propojení pozemní řídící jednotky a dronů, vytvoření komunikační Wi-Fi sítě pro přenos MAVLink rámců je znázorněno na **Obrázek 1**. Každý jednotlivý dron je opatřen GPS modulem pro určení souřadnic a senzory pro stabilní let.



Obrázek 1: Schéma zapojení a komunikace dronů s pozemní řídící stanicí

3 MAVPROXY A MISSION PLANNER

Knihovna MAVProxy slouží k vytvoření plně funkční pozemní řídící stanice pro bezpilotní letouny podporující MAVLink protokol. Hlavním účelem je možnost vytváření síťové infrastruktury pro přenos protokolu MAVLink, vytváření paketů na síťové vrstvě a možnost přesměrování na libovolné porty. Software může být rozšířen o přídavné moduly nebo může být doplněný o další software jako je např.: Mission Planner. Obsahuje možnost směrování informací z pozemní řídící stanice do dronu přes TPC nebo UDP protokol. Používá se také ke směrování vytvořených testovacích dronů s pomocí SITL (Software In The Loop) [1].

4 DRONEKIT- SITL

Dalším způsobem připojení pozemní řídící stanice a dronu je s pomocí knihovny DroneKit. Knihovna byla využita pro vytvoření aplikace, která běží na řídícím počítači komunikujícím s řídícími letovými jednotkami Pixhawk se open source firmwarem ArduPilot. Komunikace s řídící stanicí může probíhat v reálném čase s desítkami bezpilotních letounů, pro obsluhu komunikace se data posílají protokolem MAVLink. Použitá knihovna byla využita jako rozhraní pro čtení telemetrie, k získání aktuálních informací o stavu letounu, k síťovým parametrům o připojení dronů a využívá se také pro řízení mise i přímou kontrolu nad pohybem dronu.

Připojení k dronu přes pozemní stanici k dronu je nastaveno ve skriptu v jazyce Python. Po připojení bezpilotních letounů musí operátor nastavit jejich letové parametry. První parametr určuje cílovou

adresu, tedy adresu zpětné smyčky pro port UDP. Druhý parametr wait_ready se používá k určení, zda se počká, jestli se k dronu operátor připojí ihned nebo se počká na naplnění parametrů letounu. Ve většině případů se čeká na naplnění parametrů.

Pro ověření softwaru provádím před každým testovacím letem jeho simulaci. Simulace je důležitá k ověření letových instrukcí a zabránění kolizi jednotlivých dronů při jejich letu. Jedná se tak o bezpečnostní prvek, který je důležitým krokem před reálnými lety. Pro simulování jednotlivých dronů používám Software In The Loop, zkráceně SITL, který umožňuje nasimulovat letovou jednotku založenou na bázi ArduPilot v jakémkoliv reálném prostředí a komunikovat s ní pomocí MAVLink protokolu na síťové vrstvě. Obslužný skript využívá jednotlivých instancí knihovny DroneKit-SITL pro co nejrychlejší způsob spuštění simulace bezpilotních letounů. Nainstaluje se z pythonského pip nástroje a poskytuje několik jednoduchých příkazů pro spuštění předem vytvořených binárních souborů, které jsou vhodné pro daný hostitelský operační systém. Tento software používám k testování aplikací DroneKit a nových zdrojových kódů [2][3].

5 PROPOJENÍ ŘÍDÍCÍCH JEDNOTEK A POZEMNNÍ STANICE

Komunikace mezi drony a pozemní řídící stanicí je realizována Wi-Fi sítí komunikující na frekvenci 2,4 GHz. Na letové jednotce jsou připojeny senzory pro správnou stabilizaci dronů ve vzduchu. GPS modul na dronu je připojen k letové jednotce pomocí UART (Universal Asynchronous Receiver-Transmitter) portu, což je sériový port na letové jednotce. Wi-Fi modul s řídící jednotkou komunikuje adresované síťové spojení pomocí UDP protokolu. UDP protokol byl v této práci upřednostněn před protokolem TCP kvůli jeho nízké latenci.

Připojení bylo realizováno v programovacím jazyku Python. Navázání spojení bylo naprogramováno s použitím knihovny Dronekit. Prvním krokem je přidání UDP portu do MAVProxy a následná definice typu připojení UDP

Pro přenos telemetrických údajů, nastavení parametrů dronů a letových stavů bezpilotních letounů lze využít i jiné způsoby přenosu dat, jakožto s pomocí rádiové telemetrické komunikaci použitím frekvence 433 MHz nebo s použitím komunikační frekvence 868 MHz, což je alternativa k vysílací frekvenci 900 MHz, která je ale v Evropě zakázána. Dalším způsobem je využití vysílání a přijímání údajů potřebných k řízení dronu prostřednictvím mobilních datových kanálů. Bohužel tyto kanály nejsou nastaveny na stálé připojení v reálném čase, ale na krátkodobé přenášení paketů dat např.: načítání webové stránky. V důsledku toho by mohlo dojít k častému vypadávání spojení mezi dronem a pozemní stanicí i mimo města, ve kterých by bylo ovládání ještě více ztíženo přetížením sítě. V dnešní době s přibývající technologií se nezavřelo ani na Bluetooth komunikaci a s technologií Bluetooth 5 má operátor větší toleranci ke vzdálenosti letu dronu.

```
Connecting...
Connected vehicle1
Connected vehicle2
Armuji, vzlétám do 5m
Armování
ARM , probíhá vzlet
>> Výška = 0.0 m
>> Výška = 0.0 m
>> Výška = 0.2 m
>> Výška = 1.5 m
>> Výška = 3.4 m
>> Výška = 4.5 m
výška dosažena
```

Obrázek 2: Odpověď dronů během provádění letových příkazů

6 ZÁVĚR

Během práce jsem se potýkal s problematikou komunikace mezi jednotlivými UAV. Vytvořil jsem komunikační Wi-Fi síť, zajišťující předávání informací z jednotlivých dronů do pozemní jednotky a naopak. Zajištění řízení mého programu je realizováno v jazyce Python. Pro přenos příkazů a telemetrie vytvářím pakety MAVLink, které slouží k plynulému letu roje dronů. Výsledkem této práce je roj dronů, komunikující s pozemní řídící stanicí skrze vytvořenou Wi-Fi pracující na protokolu UDP. Tímto projektem lze dosáhnout řízení více UAV z jedné pozemní řídící stanice místo řízení jednotek zvlášť. UAV lze do budoucna opatřit kamerou, aby bylo možné jej využít k zajištění bezpečnosti perimetru či při pátracích akcích.

PODĚKOVÁNÍ

Chtěl bych poděkovat škole VUT za možnost projektu řízení roje dronů a předně bych chtěl poděkovat panu Ing. Jiřímu Janouškovi za odbornou přípravu, za výborné vedení a pomoc na celém projektu.

- [1] Síťové protokoly (VII. část), Protokol UDP [online]. [cit. 2021-02-20]. Dostupné z: https://www.banan.cz/serialy/sitove-protokoly/Sitove-protokoly-VII-cast-Protokol-UDP
- [2] Setting up a Simulated Vehicle (SITL) [online]. [cit. 2021-02-20]. Dostupné z: https://dronekit-python.readthedocs.io/en/latest/develop/sitl_setup.html
- [3] About DroneKit [online]. [cit. 2021-02-20]. Dostupné z: https://dronekitpython.readthedocs.io/en/latest/about/overview.html

WIRELESS TELEMETRIC DATALOGGER

Matúš Halgoš

Bachelor Degree Programme (3), FEEC BUT E-mail: xhalgo00@ vutbr.cz

Supervised by: Tomáš Benešl

E-mail: xbenes23@stud.feec.vutbr.cz

Abstract: This paper describes how to effectively use telemetry commonly used in motorsport, but is rarely seen used by ordinary drivers. The purpose of the telemetry datalogger is to provide the driver with as much information about his vehicle as he wants, while leaving the vehicle without mechanical intervention. The datalogger consists of sensors that send values and a datalogger that processes the data and plots it into graphs.

Keywords: telemetry, datalogger, sensor, temperature, motobike

1 ÚVOD

V dnešnej dobe keď sme čoraz viac obklopení elektronikou, nosíme inteligentné hodinky, ktoré merjaú počet našich krokov, aktivitu, spánok máme tendenciu z týchto dát vychádzať a na ich základe zlepšiť napríklad svoj bežecký výkon. Podobne je to aj z vozidlami. Moderné monoposty F1 sú vybavené až 300 senzormi ktoré sa starajú o bezpečnosť, stabilitu vozidla, iné pomáhajú pilotovi a inžinierom k dosiahnutiu kratšiemu času na jedno kolo čo i len o desatinu sekundy. V bežnej premávke samozrejme nieje potrebné vedieť také množstvo údajov, ale v prípade že sa jedná o vodiča najmä motocyklu ktorý má záujem analyzovať svoje jazdy, bol by schopný z niektorých dát ako sú napríklad tlak, teplota pneumatík alebo náklon motocykla a pôsobiace postranné sily, vyčítať dôležité informácie o správaní sa vozidla a na ich základe prispôsobiť štýl jazdy.

2 POPIS FUNKCIE

Jednotlivé senzory sú rozmiestnené tak aby boli schopné čo najpresnejšie získavať merané dáta. Datakoncentrátor, nachádzajúci sa na motocykli komunikuje so senzormi a prijíma namerané hodnoty ktoré odosielajú senzory každých 10 sekúnd cez 2.4 GHz.



Obrázok 1: Rozloženie senzorov na motocykli [5]

Senzory sú navrhnuté tak aby boli schopné odolať vysokým vibráciám spôsobenými jazdou po ceste, nenarušovali dizajnové spracovanie vozidla no zároveň boli dostatočne presné a odosielali relevantné údaje. Všetky komponenty sú pre zminimalizovanie konštrukčných zásahov napájané dobíjateľnou batériou. Senzory sa po zapnutí automaticky spárujú s dataloggerom a začnú odosielať dáta. V dataloggeri sú dáta uložené na pamäťovú kartu, pripravené na spracovanie. Po strate kontaktu s dataloggerom sa prepnú senzory do režimu spánku.

Parametre senzoru – Rozsah meratel'nej teploty: -70 – 380°C

Rozsah okolitej teploty: -40 – 85°C

Maximálna vzdialenosť od meraného objektu: 15cm

Presnosť merania: 0,5°C

Vysielací dosah modulu: max 100m[1,2]

3 MODUL TEPLOTNÉHO SENZORU

Senzor teploty pneumatík je jedným zo senzorických modulov na ktorom bude vysvetlený princíp funkcie. Jedná sa o funkčný celok, ktorý obsahuje všetky komponenty potrebné k tomu aby spĺňal vyššie popísané funkcie.

3.1 VÝBER KOMPONENTOV

Pri výbere komponentov bol kladený vysoký dôraz na to, aby jednotlivé súčiastky odoberali z batérie čo najmenší možný výkon a to aj v čase keď modul neodosiela žiadne dáta. Ďalšími kritériami boli ich veľkosť, možnosť využitia v prostredí s vysokými vibráciami a cena.

Najdôležitejším komponentom je senzor. Pre tieto účely bol vybratý bezdotykový teplotný senzor od firmy Melexis typ MLX90614-ESF-BAA ktorý je napájaný 3.3 V s vnútorne implementovaným komunikačným protokolom I2C. Tento senzor je vybavený sleep módom v ktorom jeho prúdový odber klesne až na 2.5 μ A. [1]



Obrázok 2: Senzor Melexis MLX90614 [1]

Na spracovanie vstupno-výstupných signálov, prevod teploty a odosielanie dát do vysielača bol vybratý 32-bitový Arm procesor STM31L031K6T6 od firmy ST. Jedná sa nízko-výkonový procesor ktorého prúdový odber v sleep móde môže klesnúť až na 0,23 µA. [3]

Ako vysielač/prijímač bol vybratý nRF24L01 ktorý bude použitý aj pri ostatných moduloch vrátane dataloggeru. nRF24L01 je jedno-čipový nízko-výkonový bezdrôtový čip, ktorý slúži ako aj na vysielanie tak aj prijímanie dát na frekvencii 2.4 GHz rýchlosťou až 2 Mb za sekundu s dosahom cca 100 m. Registrová mapa je dostupná procesoru cez SPI. Má integrovaný protokol Enhanced Shoc-kBrust[™] ktorý automaticky zaobchádza s predanými hodnotami v registroch. [2]

3.2 PRINCÍP FUNKCIE

Senzor zaznamená teplotu pneumatiky a odosiela ju cez protokol I2C na spracovanie procesorom

MCU komunikuje so senzorom z ktorého dostáva hodnoty, prepočítava ich na teplotu a pomocou SPI posiela hodnoty do vysielača odkiaľ sú poslané na datalogger. MCU, odoslaním signálu do odpájacieho obvodu, sa tiež stará o to aby pri poklese napätia na batérii bolo možné zariadenie bezpečne vypnúť a zabrániť tak zničeniu Li-pol batérie. Batériu je možné dobíjať 5 V "inudstry standard" konektorom USB-C.

Vysielač nRF24 – zhromažďuje dáta na odoslanie a po privedení impulzu ich cez filter a anténu odošle na datalogger. Při návrhu boli uvažované 2 varianty: použiť komunikačný modul s anténou a pripojiť ho k procesoru ako shield, alebo navrhnúť vysielač do spoločnej schémy. Po zvážení oboch možností, bola v rámci šetrenia priestoru a udržania vizuálnej celistvosti výsledného návrhu, zvolená druhá možnosť.

Blok regulátoru konvertuje napätie dodávané batériou z rozsahu 3-4.2 V na stabilných 3.3 V na výstupe ktoré je distribuované do všetkých komponentov. [1, 2, 3]



Obrázok 3: Bloková schéma senzoru

3.3 DPS



Obrázok 4: 3D model návrhu DPS

Doska plošných spojov vychádza z návrhu schémy. Pri tvorbe dps bol kladený dôraz na usporiadanie komponentov v okolí NRF24 aby sa čo najviac zhodovali s overeným zapojením navrhnutým priamo výrobcom. Hrúbka ciest je 0.254 mm. Dve tlačidlá sú použité na párovanie s dataloggerom, zapnutie a vypnutie. Svetelná dióda signalizuje nabíjanie batérie druhá je programovateľná. Piny 1-5 sú programovacie piny pre procesor. Veľkosť DPS je 44 x 30 mm

3.4 ENERGETICKÁ BILANCIA

Keďže jednou zo základných požiadaviek na modul je jeho nízka spotreba tak všetky použité súčiastky boli vyberané tak, aby odoberali čo najnižší výkon v prevádzkovom čase ale aj v čase keď sú prepnuté do sleep módu. Pri predpokladanom výpočte spotreby boli uvažované prúdové odbery v prevádzkovom režime 20 mA a čase spánku 70 μ A. Uvažovaný čas jedného meracieho a odosielacieho cyklu bol 100 ms, 6 – krát za minútu. To z hodiny činí 36 sekúnd kedy zariadenie pracuje a 3564 sekúnd kedy je prepnuté do režimu spánku. Priemerná spotreba elektrickej energie bola vypočítaná na základe vzorca:

$$I_{average} = \frac{I_{run} \cdot t_{run} + I_{sleep} \cdot t_{sleep}}{3600} = \frac{20 \cdot 36 + 0.07 \cdot 3564}{3600} = 0.269 \ mA \tag{1}$$

Dosadením výsledku z rovnice 1 do vzorca pre výpočet výdrže batérie spolu s kapacitou batérie 150 mA

$$t = \frac{kapacita \ bat{\acute{e}rie}}{I_{aeverage}} = \frac{150}{0.269} = 577.6 \ h \tag{2}$$

bude získaný približná doba prevádzky za ideálnych podmienok na jedno nabitie.

4 ZÁVER

Práca bola vyhotovená na základe semestrálnej práce ktorej výsledkom bolo uvedenie čitateľa do problematiky zbierania jazdných dát, vysvetlenie princípu fungovania dataloggeru a ukážka spracovania konkrétneho modulu. Pri návrhu senzorického modulu boli dodržané hlavné požiadavky na spracovanie a to: možnosť použitia v podmienkach s vysokými vibráciami a veľký dôraz bol kladený na malú veľkosť zariadenia. Obe podmienky sa podarilo objektívne dosiahnuť a preto pri finálnej montáži bude možné zariadenie bezproblémovo používať, pričom nebude narúšať samotný dizajn vozidla. Výpočet spotreby batérie ukázal že zariadenie bude schopné na jedno nabitie pracovať približne 578 hodín čo zodpovedá 24 dňom. Je nutné dodať že sa jedná o výpočet ideálnej batérie za ideálnych podmienok. V reálnej prevádzke môže byť tento údaj menší aj o 1/5.Tento čas je možné predĺžiť manuálnym vypnutím po ukončení jazdy. Na semestrálnu prácu bude nadväzovať bakalárska práca ktorej náplňou bude vývoj software. Software bude navrhnutý v prostredí STM a nahraný do zariadenia pomocou programovacích pinov 1-5.

- Melexis: MLX90614 family Single and Dual Zone Infra Red Thermometer in TO-39 (datasheet) [PDF] [online].06/2015 Dostupné z URL: https://www.mouser.sk/datasheet/2/734/MLX90614-Datasheet-Melexis-953298.pdf
- [2] NORDIC Semiconductor: Single chip 2.4 GHz Transceiver nRF24L01
 [PDF][online].03/2006 [cit. 29. 11. 2020]. Dostupné z URL: https://www.sparkfun.com/datasheets/Components/nRF24L01_prelim_prod_spec_1_2.pdf
- [3] STMicroelectronics: STM32L031x4 STM32L031x6, Access line ultra-low-power 32-bit MCU (datsheet) [PDF] [online]. ©2018 Dostupné z URL: https://www.st.com/resource/en/datasheet/stm32l031k6.pdf
- [4] Obrázok 1: https://www.pngaaa.com/detail/51816

SOFTWARE FRAMEWORK FOR PLC PROGRAM STANDARDIZATION

Otto Hrubý

Bachelor Degree Programme (3), FEEC BUT E-mail: xhruby27@stud.feec.vutbr.cz

Supervised by: Miroslav Jirgl

E-mail: jirgl@feec.vutbr.cz

Abstract: This article deals with standardization of PLC software used in company. The aim is to design software framework for single-purpose machines. The program uses elements from IEC 61131-3 to make program portable among different integrated development environments. The software was designed to be easily reusable and modifiable for different machines.

Keywords: PLC, IEC 61131-3, framework, single-purpose machine, finite state machine

1 ÚVOD

Tato práce vznikla za účelem standardizace návrhu softwaru pro jednoúčelová zařízení ovládaná programovatelným logickým automatem (PLC) ve firmě ALPS Electric Czech, s.r.o. pomocí frameworku, který se skládá z programů, funkčních bloků (FB), funkcí a dokumentace.

Některé části frameworku jsou předem vytvořené a definují pro různá zařízení stejnou softwarovou architekturu. Použitím těchto částí zrychlíme zavádění nových zařízení, neboť nemusíme začít s vývojem softwaru od začátku. Jednotná struktura usnadní také budoucí údržbu softwaru.

Další části jsou pro různá zařízení specifická. Jejich tvorbu můžeme urychlit pomocí vytvoření šablon včetně ukázkových příkladů, které jsou obsažené v referenční dokumentaci.

2 EXISTUJÍCÍ KNIHOVNY

Organizace PLCopen se zabývá standardizací aplikací pro řízení motorů, komunikaci a bezpečnost pomocí vytvoření knihoven obsahující FB. Zdrojové kódy k nim však nejsou dostupné. Nelze je tak lehce upravovat a přenášet do libovolných vývojových prostředí.

Volně přístupnou knihovnu poskytuje komunita Open Source Community for Automation Technology (OSCAT) obsahující funkce a FB například pro komunikaci, měření, řízení. Zdrojové kódy jsou napsané v textovém jazyku Structured Text (ST). Použití grafických jazyků není efektivní pro tvorbu snadno přenositelného kódu.

Pro FB obsažené v obou knihovnách platí, že obsahují výstupy informující o jejich aktuálním stavu. Vlastní aplikace se vytváří pomocí opakovaného použití již vytvořených a otestovaných bloků s podobným rozhraním.

3 ANALÝZA

Před vývojem byla provedena analýza programů používaných ve firmě. Skládají se z inicializace (počáteční nastavení proměnných), Origin (uvedení zařízení do počátečního stavu) a konečného stavového automatu (KSA).

V každém programu se používají u kontroly vstupů PLC časové limity. Pokud by došlo k poruše hardwaru, tak by bez jejich použití program neustále čekal na úspěšné vykonání kontroly. Zařízení

by však bylo v nečinnosti. Navíc by obsluha zařízení nebyla o poruše informována. Po vypršení časového limitu zařízení nahlásí chybu s určitým kódem. Je výhodné, když kód obsahuje přesnou informaci, proč porucha nastala. Sníží se tak ztráty během odstávky.

Dalším používaným prvkem je softwarové ošetření zákmitů vstupů PLC. Pro některé aplikace může být ošetření zákmitů výhodné, poněvadž se snižuje riziko, že vlivem zákmitu dojde nedopatřením k přechodu KSA do jiného stavu a nastane chyba. Nevýhodou je, že se snižuje doba odezvy zařízení.

KSA tvoří hlavní část programu. V jednotlivých stavech se ve většině případů využívá mnoho operací jako je ovládání výstupů, kontrola vstupů s časovým limitem, s ošetřením zákmitů a s tím související generování chybových kódů. Je výhodnější, když se kód rozdělí na menší části, neboť se lépe opakovatelně používají pro podobná zařízení a lépe se testují v simulátoru.

V případě používání stejných operací v rámci různých stavů, jako je například ovládání stejného pneumatického válce včetně provádění kontrol a generování chybových kódu, není vhodné tyto operace mezi stavy kopírovat. Pokud je potřebujeme změnit je nutné procházet větší část kódu, aby bylo zajištěno, že změny byly provedeny ve všech kopiích.

4 HLAVNÍ PROGRAM

Od programu se požaduje, aby byl dobře přenositelný mezi různými výrobci a vývojovými prostředími podporující normu IEC 61131-3 a aby se skládal z vícero menších částí, což umožní snadnější testování v simulátoru a rychlejší provádění změn. Na základě analýzy a požadavků byla navržena architektura hlavního programu obsahující 3 vrstvy, z nichž každá zapouzdřuje odlišnou funkcionalitu (viz Obrázek 1).

V hlavním programu probíhá inicializace, Origin, kontrola, jestli může zařízení bezpečně pracovat, tzn. že se nevyskytla žádná chyba, a výběr mezi manuálním a automatickým módem.

4.1 MANUÁLNÍ A AUTOMATICKÝ MÓD

Manuální mód obsahuje příkazy pro manuální obsluhu zařízení zadávaných přes operátorský panel. Automatický mód je realizován pomocí KSA. V každém stavu se postupně volá vybraný úkol. Po jeho dokončení nastane přechod do dalšího stavu. Když se úkol nepovede provést, tak se nahlásí chyba. Po resetu chyby se pokusí znovu provést poslední metoda.

4.2 ÚKOLY

Úkoly slouží k rozdělení KSA na vícero menších částí. V případě pneumatického lisu mohou být úkoly například proveď Origin, počkej na vložení výrobku do přepravníku a přesuň ho pod pneumatický lis, slisuj výrobek, dovez zpátky výrobek a počkej na jeho vyjmutí. Úkoly se vykonávají pomocí postupného volání metod. Každý úkol se tvoří podle šablony a nachází se ve zvláštním FB.

4.3 METODY

Pro snadnější kontrolu a nastavování proměnných byly vytvořeny 2 uživatelské funkce, ze kterých jsou tyto instrukce volány. Změny tak stačí provádět pouze v těchto funkcích.

V první funkci se definují metody pro nastavování proměnných (například: vysunout/zasunout píst válce). Druhá funkce obsahuje metody pro kontrolu proměnných (například: zkontroluj, zdali se píst válce vysunul/zasunul). Při neúspěchu vrátí funkce chybu obsahující informaci, proč kontrola neproběhla úspěšně. U kontroly je potřeba použít časovače pro časový limit a pro ošetření zákmitů, což funkce neumožňuje, protože nemá paměť, proto bude funkce volána z předdefinovaného FB. Pro univerzální použití FB umožňuje výběr, které časovače mají být spuštěny.



Obrázek 1: Blokový diagram hlavního programu

5 KOMUNIKACE

Jednoúčelová zařízení mohou být použita ve větších výrobních systémech, ve kterých jich může být instalován větší počet. Dalším z požadavků na framework byl, že zařízení budou pracovat v Master-Slave systému. Role slave je přidělena jednoúčelovým zařízením, poněvadž pouze přijímají příkazy od nadřazeného zařízení (master).

Komunikaci zajišťuje další program pracující paralelně s hlavním programem. V programu jsou implementovány příkazy pro ovládání zařízení (viz Obrázek 2). Příkazy mohou být přijaty pouze pokud to umožňuje aktuální stav zařízení. Příjem a odeslání dat zajišťuje FB volaný z tohoto programu, jehož implementace závisí na použitém komunikačním protokolu a na výrobci či modelu PLC. Pro každý protokol musí být vytvořen FB se stejným rozhraním. Výhodou tohoto řešení je, že při změně protokolu stačí pouze měnit FB. V programu se tak nemusí již nic upravovat.



Obrázek 2: Blokový diagram komunikačního programu

6 ZÁVĚR

Na základě získaných poznatků a požadavků byl vytvořen hlavní program pro jednoúčelová zařízení. Program umožňuje rychlejší zavádění nových zařízení a rychlejší provádění změn programu. Dále byl vytvořen komunikační program pracující nezávisle na použité komunikačním protokolu. Tento program byl prozatím otestován na vytvořeném protokolu přes TCP/IP Socket ve vývojovém prostředí CODESYS. Nyní probíhá testování programu na průmyslovém protokolu EtherNet/IP. Po otestování bude k programům vytvořena referenční dokumentace. Na závěr bude řešení ověřeno na reálném zařízení.

- [1] PLC Controls with Structured Text (ST). 3rd ed. Randers, Denmark: Books on Demand, 2020. ISBN 8743015549.
- [2] Leverage object-oriented industrial programming. Control Engineering. 2019, 2019(7), 17-21.
- [3] OSCAT Neuigkeiten [online]. Wörth Donau: OSCAT, c2015 [cit. 2021-03-27]. Dostupné z: http://www.oscat.de/
- [4] Creating Reusable, Hardware Independent Motion Control Applications via IEC 61131-3 and PLCopen Motion Control Profile. In: IEEE/ASME (AIM) International Conference on Advanced Intelligent Mechatronics [online]. Como, Italy: IEEE, 2001, s. 6 [cit. 2021-03-27]. ISBN 0-7803-6736-7. Dostupné z: https://ieeexplore.ieee.org/document/936763

MICRO HYDRO POWER PLANT

František Rusnák

Bachelor (3), FEEC BUT E-mail: xrusna06@stud.feec.vutbr.cz

Supervised by: Tomáš Benešl

E-mail: xbenes23@stud.feec.vutbr.cz

Abstract: This paper deals with realization of functional model of control unit for micro hydro power plant. Main task of this work is realization rectifier with active power factor correction, which is used instead of commonly used frequency converter. Control unit is designed for asynchronous motor in generator operation with self-excitation. The propeller turbine is used as propulsion. To convert rectified voltage from PFC rectifier to mains will be used solar inverter.

Keywords: Hydroelectric power, Propeller turbine, Power Factor Correction, PFC, CCM, Solar Inverter

1 ÚVOD

Projekt je vyvíjen ve spolupráci se společností ELZACO spol. s r.o. Společnost ELZACO spol. s r.o. je specializovaná v oblasti řízení technologických procesů, návrhů a výroby jednoúčelových strojů a malých vodních elektráren.

Funkční model je navrhován tak, aby z něj mohl vycházet návrh prototypu a následně celkové zařízení, které může být aplikováno v provozních podmínkách. Pozornost je věnována především sestavě PFC měniče, která je v rámci této práce navržena a sestavena. Ostatní komponenty jsou pořizovány jako hotové výrobky.



Obrázek 1: Blokové schéma řídící jednotky MVE

Řídící jednotka zajišťuje transformaci elektrické energie dodávané generátorem do rozvodné sítě. Jako generátor je použit asynchronní motor v zapojení se samobuzením. Toto řešení zjednodušuje běžně používané zapojení třífázového asynchronního motoru řízeného frekvenčním měničem a je vhodné pro provoz při výkonech v řádu kilowatů. Celkové zapojení se skládá z generátoru, usměrňovače s korekcí účiníku, fotovoltaického měniče a standardní energetické ochrany. Jako generátor bude použit přípravek s asynchronními motory, kde je jeden motor použit jako pohon a druhý je v zapojení se samobuzením a plní zde funkci generátoru. Tyto motory mají pevně spojené hřídele.

2 POŽADAVKY NA ŘÍDÍCÍ JEDNOTKU

Byly stanoveny následující požadavky na řídící jednotku:

- Turbína/generátor s maximálním výkonem: 2,2 kW
- Rozsah provozních otáček turbíny: 800 1600 min⁻¹
- Typ generátoru, jedna z možností:
 - asynchronní samobuzený ve 3-fázovém, nebo 1-fázovém zapojení
 - synchronní s permanentními magnety 3-fázové zapojení
- Výstup z řídicí jednotky do 1 f. sítě AC 230 V/ 50 Hz
- Ochrany: standardní energetické ochrany

3 TEORIE PFC

Řízené spínání výkonových prvků umožňuje zachování účiníku přes daleko větší rozsah odebíraného výkonu z měniče, než je to možné u zdroje s pasivní korekcí účiníku. "Korekce účiníku upravuje tvar průběhu vstupního proudu napájecích zdrojů, aby se zvýšil skutečný výkon dostupný ze sítě. V ideálním případě by měl elektrický spotřebič představovat zátěž, která napodobuje ideální odpor. V takovém případě je jalový výkon odebíraný zařízením nulový." [1]

Na trhu je dostupné veliké množství měničů i řídicích čipů, které umožňují aktivní PFC regulaci. Výkony dostupných zdrojů se pohybují v rozmezí 100 W až jednotky KW. Výkonné měniče jsou schopny transformovat výkon s velikou účinností, která se u některých měničů pohybuje až do 98 %. Účinnost je nejvíce závislá na odebíraném výkonu a napájecím napětí. Obecně platí, že při velice nízkém zatížení měniče je jeho účinnost malá. Je to kvůli výkonu, který spotřebovává řídící jednotka měniče a okolní součástky. Také musíme uvažovat ztráty, které vznikají v důsledku spínání výkonových prvků. Spínáním dochází k velikým napěťovým přechodům se strmou hranou. V reálném obvodu vždy musíme počítat s parazitní kapacitou součástek i plošného spoje. Opakovaným vybíjením a nabíjením těchto parazitních kapacit vznikají tepelné ztráty.

4 MÓD ŘÍZENÍ NEPŘERUŠOVANÉHO PROUDU (CCM)

Mód řízení nepřerušovaného proudu je vhodným řešením pro výkonné měniče. Řízení CCM nachází uplatnění v řadě aplikací jako jsou zdroje LCD monitorů, zdroje počítačů, nabíječe baterií, průmyslové napájecí zdroje, univerzální PFC usměrňovače a jiné. Toto řízení nabízí řadu výhod. Spínání výkonového tranzistoru má konstantní periodu. To umožňuje snadno realizovat filtrační prvky pro odrušení očekávané frekvence. Díky spínání s nižší periodou, než je čas potřebný k úplnému vyproudění cívky, proud cívkou neklesá na nulu. Díky vyšší frekvenci spínání je možné dosáhnout podstatně menšího zvlnění proudu než u jiných módů řízení s proudem klesajícím na nulovou hodnotu. Při návrhu výkonných měničů se musí cívka navrhnout s ohledem na špičkový proud. Díky udržování proudu cívkou klesá špičková hodnota proudu a s frekvencí spínání se přibližuje ke střední hodnotě proudu. Důležitou vlastností CCM módu řízení je nízká hodnota špičkového proudu cívkou, která má zásadní vliv při návrhu vhodných parametrů cívky. S nižší maximální proudovou zatížitelností cívky klesá cena a prostorový objem cívky. Pozornost je třeba věnovat době spínání tranzistoru. Sepnutí tranzistoru nastává v době, kdy cívkou a diodou stále protéká proud. Pak je nutné brát ohled na "dobu zpětného zotavení diody" [2], kvůli kterému mohou vznikat veliké proudové špičky při špatně nastavené době sepnutí.

V návrhu je použit obvod UC3854, který umožňuje přesnou regulaci proudu, který je úměrný aktuální hodnotě vstupního napětí. Klasické vnitřní zapojení řídících obvodů pro řízení nepřerušovaného proudu obsahuje složitější násobičku. Tyto obvody jsou složité a výroba analogové násobičky je také finančně nákladnější. Hlavní benefit, který takové zapojení nabízí, je konstantní odezva regulátoru na změnu výstupního napětí bez závislosti na vstupním napětí. V této aplikaci napětí na vstupu PFC měniče není stabilní a je žádoucí, aby regulátor PFC měniče pracoval stabilně i přes možné výkyvy vstupního napětí.

5 SESTAVENÍ PFC MĚNIČE

Pro danou aplikaci byl zvolen typ zapojení PFC měniče podle topologie Boost. Hlavní výhodou takového zapojení je jednoduchost a malý počet komponentů, s čímž souvisí ztrátový výkon.

V návrhu byly použity výkonové součástky tranzistor Q1 IPW60R099P6[4], dioda D1 FFH30S60STU[5] a Cívka L DEHF-42/0.15/28DLA[3]. Výběr součástek tranzistoru a diody byl proveden na základě schématu dokumentace [6]. Společné zapojení tranzistorů je dimenzováno na maximální proud I_{MAX}=48A při teplotě 100°C, maximální napětí U_{DS}=650V a výkon P_{TOT}=220W při 100°C. Společné zapojení diod je dimenzováno na průměrný proud v propustném směru I_F=64A při teplotě 100°C a maximální závěrné napětí U_R=600V. Pro výpočet maximální střední hodnoty proudu cívkou I_{PK} je použit vzorec (1). Zvolená cívka má výrobcem udávaný maximální proud I_{MAX}=28A.

$$I_{L_PK} = \frac{\sqrt{2} \cdot P_0}{\eta \cdot U_{IN \ MIN}} + \frac{\left(1 - \frac{\sqrt{2} \cdot U_{IN} MIN}{U_0}\right) \cdot \frac{1}{f_{SW}} \cdot U_{IN} MIN \cdot \sqrt{2}}{2 \cdot L} = 23,1 \text{ A}$$
(1)



Obrázek 2: Principiální schéma topologie Boost

Sestava PFC měniče se skládá ze tří desek plošných spojů. Tato sestava byla navržena jako funkční model pro zkušební provoz v laboratorních podmínkách. Sestava obsahuje vstupní filtr pro odstranění rušení, výkonovou část PFC měniče a řídící jednotku.

Řídící jednotka PFC měniče je navržena jako samostatná DPS. Obvod je založen na analogové části, která plní funkci regulátoru řízení proudu PFC měniče. Analogové řízení je ovládáno signály z mikropočítače, který zajišťuje řízení celé sestavy PFC měniče. Díky použití mikropočítače, je možné výrazně zvýšit funkčnost řídící DPS a zdroj lze také uživatelsky ovládat pomocí sériové komunikace.


Obrázek 3: Sestava PFC měniče

6 ZÁVĚR

Cílem projektu je navrhnout a sestavit funkční model řídící jednotky malé vodní elektrárny. Součástí sestavy je aktivní PFC měnič, který byl navrhnut a sestaven v rámci této práce. Zdroj byl úspěšně oživen a otestován při výkonu 1500 W v zapojení s přípravkem generátoru se samobuzením. Práce bude dále pokračovat výběrem vhodných součástí pro sestavu řídící jednotku MVE.

PODĚKOVÁNÍ

Děkuji vedoucímu projektu Ing. Tomášovi Zatloukalovi a Ing. Romanovi Kubíčkovi za pedagogickou a odbornou pomoc v průběhu mé práce na projektu. Rovněž bych chtěl poděkovat jednateli společnosti ELZACO spol. s r.o. Jiřímu Vénosovi za umožnění realizace projektu v prostorách firmy a vedoucímu semestrální práce Ing. Tomášovi Benešlovi za cenné rady při zpracování mé práce.

REFERENCE

- [1] Power Factor Correction (PFC) Handbook [online]. ON SEMICONDUCTOR. 5. Duben 2014 [cit. 2021-03-02]. Dostupné z: <u>https://www.onsemi.com/pub/Collateral/HBD853-D.PDF</u>.
- [2] Prof. Ing. Jaroslav Boušek, CSc., Ing. Petr Kosina, Ph.D. a Ing. Barbora Mojrová. Elektronické součástky. 2015.
- [3] DEHF Hi-Flux high current inductors [online]. [cit. 2021-03-28]. Dostupné z: https://www.tme.eu/Document/1fe237522db87fe387a223c95dca0a5e/dehf.pdf
- [4] 3600V CoolMOS[™] P6 Power Transistor [online]. [cit. 2021-03-28]. Dostupné z: https://www.tme.eu/Document/1fe237522db87fe387a223c95dca0a5e/dehf.pdf
- [5] Stealth II Rectifier30 A, 600 VFFH30S60S [online]. [cit. 2021-03-28]. Dostupné z: https://cz.mouser.com/datasheet/2/308/FFH30S60S_D-1808667.pdf
- [6] 230-V, 3.5-kW PFC With>98% Efficiency, Optimized for BOM and Size Reference Design. [Online] Říjen 2017. [Citace: 26. Prosinec 2020.] <u>https://www.ti.com/lit/ug/tidube1c/tidube1c.pdf?ts=1609332644353&ref_url=https%253A%252F%252Fwww.google.com%252F.</u>

Bakalářské projekty

Silnoproudá elektrotechnika a elektroenergetika, Teoretická elektrotechnika, Fyzika a matematika

THERMAL ANALYSIS OF SOLID ROTOR COATED WITH COPPER LAYER

Jakub Súkeník

Bachelor Degree Programme (3), FEEC BUT E-mail: xsuken00@stud.feec.vutbr.cz

> Supervised by: Marek Toman E-mail: marek.toman@vutbr.cz

Abstract:

This paper deals with thermal calculations of rotor of a high-speed machine. The first part describes thermal network and the basic relationships for the calculation of thermal resistances. In the second part, the rotor thermal network for a high speed machine is designed. Next, this thermal network of the rotor is compared with the thermal model of a rotor created in Ansys software.

Keywords: Thermal network, high-speed induction motor, solid rotor

1 ÚVOD

Elektrické stroje se významně podílejí na spotřebě elektrické energie, podle [1] a [2] elektrické motory spotřebovávají přibližně 45 % celosvětové vyrobené energie. Díky tomu je potřeba zvyšovat účinnost elektrických strojů pro minimalizaci energetických ztrát.

V posledních dekádách se vysokootáčkové elektrické stroje dostávají do popředí díky rozmachu výkonové elektroniky a frekvenčních měničů. Vysokootáčkové stroje pracují s vysokými provozními otáčkami. To má za důsledek značně vyšší ztráty než u klasických elektrických strojů. Jedná se především o ztráty v železe a mechanické ztráty. Na druhou stranu pracují s vyšší hustotou výkonu, což se příznivě projevuje na velikosti a hmotnosti elektrického stroje. Díky těmto vlastnostem je nezbytné tyto stroje konstruovat s adekvátním chlazením, které musí zajišť ovat dostatečný odvod tepla ze stroje. Chlazení může být realizováno buď proudícím vzduchem či héliem, nebo vodním chlazením. Pro vysokootáčkové elektrické stroje je důležitý z hlediska odvodu tepla také rotor. Při jeho dlouhodobém tepelném namáhání by se mohl materiál na povrchu poškodit, což by vedlo k nepříznivému chování celého stroje [3].

Tento článek se zabývá tepelnou sítí rotoru. Nejdříve jsou zde popsány vztahy pro výpočet tepelných odporů. Dále se tento článek zabývá popisem tepelné sítě a nakonec je tepelná sít² rotoru porovnána s tepelným modelem v programu Ansys.

2 METODA TEPELNÉ SÍTĚ

Tepelný výpočet stroje se dá realizovat metodou tepelné sítě, nebo-li metodou ekvivalentních tepelných obvodů. Cílem této metody je vytvořit schéma, které svým uspořádáním bude simulovat tepelné chování stroje. Tepelné schéma se skládá z jednotlivých prvků, jako jsou uzly, větve a nory. Uzel je charakterizován svou teplotou a může, či nemusí být zdrojem tepelného toku. Větev představuje cestu tepelného toku mezi jednotlivými uzly. Nor zprostředkovává přestup tepla do okolního prostředí. Tato metoda se používá na ustálené a tranzientní stavy [4].

2.1 VÝPOČET DŮLEŽITÝCH TEPELNÝCH ODPORŮ GEOMETRICKÝCH TĚLES

Pro výpočet tepelných odporů v tepelné síti rotoru je nejvíce využívaný tepelný odpor mezikružím a tepelný odpor ve válcové tyči s vývinem tepla. Tepelný odpor válcového mezikruží R_v je podle [5] ve tvaru

$$R_{\rm v} = \frac{\ln \frac{r_2}{r_1}}{2\pi l \lambda},\tag{1}$$

kde r_1 je vnitřní poloměr válce, r_2 je vnější poloměr válce, l je délka válce a λ je měrná tepelná vodivost. Rovnice pro výpočet vnitřního odporu vedením uvnitř tyče s vývinem tepla R_t je podle [5] vyjádřen

$$R_{\rm t} = \frac{1}{8\pi l\lambda},\tag{2}$$

kde *l* je délka tyče a λ je měrná tepelná vodivost tyče.

3 TEPELNÁ SÍŤ ROTORU

Geometrické rozměry a ztráty v rotoru vychází z práce [3]. Motor, ve kterém je tento rotor umístěn, má jmenovitý výstupní výkon na hřídeli $P_2 = 12$ kW při jmenovitých otáčkách $n = 45\,000$ ot/min a celkové účinnosti přibližně $\eta = 90$ %. Na obrázku 1 jsou zobrazeny jednotlivé geometrické rozměry rotoru. Tyto parametry byly nezbytné při výpočtech jednotlivých odporů pro tepelnou síť rotoru i modelu vytvořeném v programu Ansys. Modré přerušované čáry na obrázku 1 značí středy jednotlivých úseků rotoru.



Obrázek 1: Značení geometrických rozměrů rotoru.

Na obrázku 2 je rozmístěno 28 bodů, které znázorňují uzly, ve kterých jsou počítány střední teploty daných částí rotoru. Při samotné realizaci tepelné sítě v programu Matlab bylo zapotřebí tyto uzly mezi sebou propojit a vypočítat tepelné odpory pomocí vztahů uvedených v sekci 2. Rozmístění uzlů je provedeno univerzálně, aby bylo možné případně upravit geometrii pro lepší odvod tepla [1].



Obrázek 2: Rozmístění středních teplot v rotoru.

Návrh, uspořádání uzlů a následné očíslování těchto uzlů bylo vhodně provedeno s ohledem na geometrické rozměry rotoru, rozmístění objemových a mechanických ztrát rotoru. Proto bylo nutno rozdělit rotor na jednotlivé sekce, které jsou znázorněny přerušovanými čarami. Díky rotační symetrii není třeba řešit tepelnou síť rotoru trojrozměrně, ale postačuje ji řešit jako síť dvojrozměrnou.

4 TEPELNÝ MODEL ROTORU VYTVOŘENÝ V SOFTWARU ANSYS

Tepelný model vytvořený v softwaru Ansys slouží jako reference pro tepelnou síť rotoru v sekci 3. Na obrázku 3 je zobrazen tepelný model navržený v softwaru Ansys. Pro rychlejší řešení tepelného modelu byla navržená geometrie rotoru z původně plného válce upravena do podoby válcové výseče se středovým úhlem válcové výseče $\theta = 45^{\circ}$, protože je předpokládaná rotační symetrie rozložení teplot. Teploty spolu se součiniteli přestupu tepla byly zadány do modelu na obrázku 3 pomocí konvekce na povrch rotoru. Teploty v konvekcích na povrchu rotoru jsou zadány nesymetricky, protože se předpokládá proudění vzduchu ve vzduchové mezeře. To se projeví nesymetrickým rozložením teplot v rotoru. Teploty zadané do modelu jsou fiktivní, později se nahradí reálnými teplotami. Hodnoty součinitelů přestupu tepla byly určeny na základě obvodové rychlosti rotoru. Všechny rozměry zadané do tepelného modelu na obrázku 3 vycházejí z geometrických rozměrů na obrázku 1.



Obrázek 3: Tepelný model v softwaru Ansys.

Ansys	Matlab	Rozdíly teplot	
Teplota [°C]	Teplota [°C]	Teplota [°C]	
163,52	162,38	1,14	
163,46	164,48	1,02	
167,56	166,72	0,84	
172,04	166,26	5,78	
163,57	162,45	1,12	
163,49	164,52	1,03	
167,58	166,75	0,83	
172,08	166,31	5,77	
164,22	163,78	0,43	
170,31	166,72	3,59	
164,32	164,06	0,26	
170,28	166,76	3,51	
165,24	163,27	1,97	
	Ansys Teplota [°C] 163,52 163,46 167,56 172,04 163,57 163,49 167,58 172,08 164,22 170,31 164,32 170,28 165,24	AnsysMatlabTeplota [°C]Teplota [°C]163,52162,38163,46164,48167,56166,72172,04166,26163,57162,45163,49164,52167,58166,75172,08166,31164,22163,78170,31166,72164,32164,06170,28166,76165,24163,27	

5 POROVNÁNÍ VYPOČTENÝCH TEPLOT

 Tabulka 1:
 Porovnání středních teplot modelu v Ansysu a tepelné sítě vypočtené v Matlabu.

V tabulce 1 jsou uvedeny výsledné střední teploty vypočtené pomocí tepelné sítě a střední teploty tepelného modelu rotoru realizovaném v softwaru Ansys. Číslování uzlů odpovídá obrázku 2. Hodnoty teplot v posledním sloupci tabulky 1 jsou uvedeny v absolutní hodnotě a značí rozdíly mezi

tepelným modelem vytvořeným v softwaru Ansys a danou tepelnou sítí rotoru. Pro přehlednost byly některé uzly v tabulce 1 sloučeny, méně podstatné uzly nebyly uvedeny do tabulky. Název hřídel značí v tabulce 1 aritmetický průměr uzlů 23 až 28.

6 ZÁVĚR

Tento článek se zabývá návrhem a praktickým výpočtem tepelné sítě rotoru vysokootáčkového asynchronního stroje. Nejdříve jsou uvedeny rovnice pro výpočty tepelných odporů, ze kterých se tepelná sít skládá. Poté je zde uveden seznam geometrických rozměrů rotoru. Tyto rozměry slouží k určení tepelných odporů v tepelné síti vytvořené v programu Matlab a také jsou zadány do návrhu geometrie tepelného modelu vytvořeném v softwaru Ansys.

Teploty definované v okrajových podmínkách konvekcí byly zadány tak, aby přibližně odpovídaly reálným teplotám ve stroji. Bylo také snahou zadat okrajové podmínky tak, aby výsledné teplotní pole bylo v axiálním směru nesymetrické a ověřila se tak správná funkčnost navržené tepelné sítě. Tyto nastavené okrajové podmínky byly použity pro ověření výsledků ze sítě pomocí softwaru Ansys. Ve výsledné tepelné síti vysokootáčkového stroje se dají očekávat vyšší teploty v rotoru. V návaznosti na tuto tepelnou síť se v budoucnu vytvoří další tepelné sítě částí analyzovaného stroje (např. stator, kostra), které se následně sloučí do jedné tepelné sítě, která bude reprezentovat tepelné chování celého stroje.

Střední teploty uzlů tepelné sítě rotoru a střední teploty uzlů modelu vytvořeném v softwaru Ansys byly uvedeny do tabulky 1. Střední teploty sítě rotoru se v několika uzlech liší od středních teplot tepelného modelu v softwaru Ansys, ale jedná se o velmi malé odchylky.

PODĚKOVÁNÍ

Tento příspěvek byl vytvořen v rámci Centra výzkumu a využití obnovitelných zdrojů energie (CV-VOZE). Autor vděčně děkuje za finanční podporu z Ministerstva školství, mládí a sportu České republiky.

REFERENCE

- BÁRTA, Jan. Návrh elektrického stroje 6kW, 120 000 ot/min pro turbocirkulátor hélia. Brno, 2017. Disertační práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Vedoucí práce Čestmír Ondrůšek.
- [2] KIM, Youn-Hwan, Hee-Deuk JUN, Jae-Won MOON, Rae-Eun KIM, Se-Hyun RHYU a Sang-Young JUNG. *Motor Efficiency Determination of SynRM and Measurement Uncertainty* [online]. IEEE, 2019, 2019, , 233-239 [cit. 2020-12-28]. ISBN 978-1-5386-7687-5. Dostupné z: https: //ieeexplore.ieee.org/document/9007174/
- [3] VÍTEK, Ondřej. *Vysokootáčkové elektrické motory*. Brno, 2016. Habilitační práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií.
- [4] TOMAN, Marek. Vázané modelování asynchronního motoru metodou fyzikálního modelování. Brno, 2015. Diplomová práce. Vysoké učení technické v Brně, Fakulta strojního inženýrství. Vedoucí práce Radek Vlach.
- [5] CENGEL, Yunus A. a Afshin J. GHAJAR. *Heat and Mass Transfer: Fundamentals and Applications*. Vyd. 5. New York: McGraw-Hill Education, 2015. ISBN 978-0-07-339818-1.

POTENTIAL DEVELOPMENT OF FAST CHARGING STATI-ONS IN SELECTED AREA

David Hysek

Bachelor (3.ročník), FEEC BUT E-mail: xhysek05@stud.feec.vutbr.cz

Supervised by: Martin Paar

E-mail: paar@feec.vutbr.cz

Abstract: In this article, a selected locality was evaluated in terms of charging infrastructure and a scenario for three types of installations of AC and DC charging stations according to NAP were determined. Furthermore, the parameters of the charging stations used in the scenario were compared.

Keywords: charging station, fast charging, EV charging

1 ÚVOD

Článek je zaměřen na rozvoj nabíjecí infrastruktury ve vybrané lokalitě v Brně, jelikož rozvoj elektromobility je závislý na rozvoji infrastruktury nabíjecích stanic. Článek vychází z předpokladu znalosti rozdílu mezi AC a DC nabíjecími stanicemi a rozdílem mezi rychlým a pomalým nabíjením. V článku budou řešeny parametry vybraných nabíjecích stanic a plánovaný odhad počtu nabíjecích stanic ve vybrané lokalitě podle NAP.

2 TYPY VYBRANÝCH NABÍJECÍCH STANIC

Nabíjecí stanice byly vybrány na základě již existujících instalací a podle odhadu rozvoje EV v dané oblasti.

Název stanice Typ nabíjení	Jmenovitý	Výstupní	Výstupní proud [A]	Příkon	Vstupní	Vstupní	
	výkon ¹	napětí		stanice	napětí	proud	
	[kW]	[V]		[kVA]	[V]	[A]	
Terra AC		11	230	16	11	230	16
wallbox	AC	11	230	10	11	230	10
Terra 54 CT	DC	50	200-500	125	77	400	112
Terra 184 CC	DC	180	150-920	200	192	400	280
Terra HP 350	DC	350	150-920	200 (CHAdeMO) 500 (CCS-1)	384	600	370

Tabulka 1 Přehled parametrů vybraných nabíjecích stanic [1] [2] [3] [4]

2.1 TERRA AC WALLBOX TAC-W11-G5-R-0

Jedná se o nabíjecí AC stanici se zásuvkou Mennekes Type-2, která disponuje maximálním jmenovitým výkonem 11 kW. Nabíječka je záměrně vybrána tak, aby disponovala menším výkonem. Jelikož většina palubních nabíječek elektromobilů není schopna vyšší výkon

¹ Jedné se o maximálně dosažitelnou hodnotu. Reálně bude nabíjecí výkon nižší, jelikož je řízen palubní nabíječkou elektromobilu.

dlouhodobě využít. Za předpokladu, že řidič denně nespotřebuje celou kapacitu baterie, je nabíjecí výkon 11 kW dostačující k opětovnému nabití elektromobilu během nočního nabíjení. [1]

2.2 ABB TERRA 54 CT

Nabíjecí DC stanice Terra 54 CT je vybavena jedním CCS2 konektorem a zásuvkou na AC nabíjení. Tato konfigurace byla vybrána na základě trendu standardizace nabíjecích konektorů, podporovaných ze strany Evropské unie. Konektor CCS2 disponuje maximálním nabíjecím výkonem 50 kW a AC konektor Mennekes Type-2 22 kW. [2]

2.3 ABB TERRA 184 CC

Nabíjecí stanice Terra 184 CC disponuje rychlým DC nabíjením o maximálním výkonu 180 kW. Konfigurace CC umožňuje nabíjení přes dva konektory CCS. Při stejnosměrném nabíjení dvou aut zároveň je odebíraný výkon 2x90 kW. [3]

2.4 ABB TERRA HP 350

Jedná se o ultra rychlou nabíjecí DC stanici, vybavenou maximálním výkonem 350 kW a stálým výkonem 320 kW. Výstupní proud je rozdílný, podle použitého konektoru, kdy CHAdeMO disponuje maximálně 200 A DC a CCS-1 až 500 A DC. Stanice je rozdělena na dvě části nabíjecí stojan a dvě silové části (power cabinet). Nabíjecí stojan slouží pouze k chlazení a samotné předávce energie s elektromobilem. V silové části dochází k usměrnění proudu, který se tedy do stojanu dostává již usměrněný. Silová část obsahuje i transformátory kvůli galvanickému oddělení. Při instalaci dvou výdejních stojanů a dvou silových částí lze provozovat tento nabíjecí bod jako rychlonabíjecí DC stanici o maximálním výkonu 350 kW (při připojení jednoho elektromobilu). Ovšem pokud budou připojeny elektromobily na obou stojanech, tak se silové části rozpojí a každá zásobuje jeden stojan a ve výsledku tedy dostaneme dvě nabíjecí stanice, každou o maximálním výkonu 175 kW. [4]

3 SOUČASNÝ STAV LOKALITY

Vybraná oblast se nachází v Brně, ovšem nebude blíže specifikována, jelikož se nejedná o oficiální studii a zadavatel si žádá utajení oblasti. Ve vybrané lokalitě se nachází dva obchodní domy, čtyři prodejny automobilového průmyslu, firmy stavebního průmyslu a velká bytová jednotka. Elektrizační síť této oblasti je převážně řešena silovým kabelem 22-AXEKVCEY 1x240/25. V oblasti vedení se nachází 52 transformátorů. Oblast je napájena z transformovny 110/22 kV a disponuje jedním transformátorem 25 MVA a jedním 40 MVA. [5]

Ve vybrané oblasti se v současné době nachází tři pomalé střídavé dobíjecí stanice jedna s výkonem AC 11 kW, dvě s výkonem AC 2x22 kW. Je zde instalována jedna rychlodobíjecí stanice DC, která disponuje 50 kW při připojení konektorem CHAdeMO, a 150 kW při připojení přes konektor CCS2 i AC dobíjením o výkonu 22 kW. V oblasti napájené ze stejné rozvodny se dále nachází tři DC nabíjecí stanice s výkony 50 kW (s konektory CCS2 i CHAdeMO), které jsou vybaveny i AC nabíjením 22 kW přes konektor Mennekes Type-2 a dvě nabíjecí stanice AC 2x22 kW. Zhodnocení stávající infrastruktury nabíjecích stanic v této lokalitě proběhlo pomocí mapy nabíjecích stanic fDrive.cz. [6]

4 SCÉNÁŘ INSTALACE STANIC

Podle cíle NAP CM 2019 by mělo být v roce 2030 k dispozici 19 000-35 000 dobíjecích stanic, aby infrastruktura dokázala pokrýt nároky 220 000-500 000 elektromobilů a 800-1 200 elektrobusů, které by měly na českých silnicích dle NAP jezdit právě do roku 2030. [7] Jelikož se v lednu roku 2020 nacházelo v České republice 750 dobíjecích stanic, jedná se o pětadvaceti až padesátinásobný nárůst. Dnes se ve vybrané lokalitě nachází devět nabíjecích stanic, do roku 2030 by to tedy mohlo být 225-

450. Samozřejmě tento odhad není úplně přesný, kvůli tomu, že nárůst nabíjecích stanic nebude rovnoměrný, ovšem vybraná lokalita disponuje dobrým potenciálem k podobnému nárůstu stanic.

Scénář uvádí instalace AC i DC nabíjecích stanic (viz tabulka 1) ve třech oblastech rozvoje nabíjecích stanic: bytová jednotka, průmyslové objekty, čerpací stanice.



Obrázek 2 Schéma vedení vybrané lokality

4.1 INSTALACE 100 AC NABÍJECÍCH STANIC TERRA AC WALLBOX 11 KW A 4 DC NABÍJECÍ STANICE ABB TERRA 54 U BYTOVÉ JEDNOTKY

Tato instalace by byla realizovatelná například u velké bytové jednotky nacházející ve vybrané lokalitě. Nabíjecí stanice by byly instalovány na parkovišti před bytovou jednotkou. Celkem by zde bylo k dispozici sto míst pro připojení elektromobilu na AC nabíjecí stojany 11 kW a osm míst na připojení k DC nabíjecí stanici, z nichž čtyři by disponovaly nabíjením DC 50 kW a zbylé čtyři AC 22 kW. Celkový instalovaný příkon nově instalovaných nabíjecích stanic by tedy činil 1 408 kVA. (100 x 11 kVA + 4 x 77 kVA = 1 408 kVA)

4.2 INSTALACE 8 RYCHLÝCH DC NABÍJECÍCH STANIC TERRA 184 A 8 ULTRARYCHLÝCH NABÍJECÍCH STANIC TERRA HP 350 S DVĚMA NABÍJECÍMI STOJANY V PRŮMYSLOVÝCH OBJEK-TECH

Jednalo by se o instalaci rychlonabíjecích stanic v několika průmyslových objektech, například: stavební firmy, automobilové salóny a prodejna nákladních automobilů. Jelikož se jedná o automobilku, která již projevila velký zájem v rozvoji elektromobility mezi nákladní dopravou, je velmi pravděpodobné, že se i v této pobočce časem budou prodávat nákladní elektromobily. Ty budou potřebovat silnější nabíječky, jelikož budou mít baterie o větších kapacitách. Celkový instalovaný příkon všech zařízení by při této instalaci činil 3 008 kVA. (8 x 184 kVA + 4 x 384 kVA= 3 008 kVA)

4.3 INSTALACE 6 RYCHLÝCH DC NABÍJECÍCH STANIC TERRA 184 A 2 ULTRARYCHLÝCH NABÍJECÍCH STANIC TERRA HP 350 S DVĚMA NABÍJECÍMI STOJANY NA ČERPACÍ STANICI

U této varianty by bylo instalováno šest rychlonabíjecích DC stanic s nabíjecím výkonem 180 kW a dvě ultrarychlé nabíjecí stanice s maximálním nabíjecím výkonem 350 kW. Byly by umístěny na čerpací stanici, která si již v dané lokalitě nachází. Výhodou této lokality je, že se nachází u sjezdu

z dálnice, tudíž by tento nabíjecí bod mohly využít i elektromobily, které by pouze sjely z dálnice a po nabití se na ni opět vrátily. Čerpací stanice, která se v současné době ve vybrané lokalitě nachází, disponuje šesti čerpacími stojany. Pokud by například jedno čerpání pohonné hmoty trvalo i s platbou do pěti minut a nabití Tesly Model S s kapacitou baterie 75 kWh z 10 % na 80 % trvá do dvaceti minut při využití nabíječky Terra 184 a do deseti minut při využití nabíječky Terra HP 350, byl by počet osmi stanic pro kompletní nahrazení všech čerpacích stojanů dostatečný. Celkový instalovaný příkon všech zařízení by tedy činil 1 920 kVA. (6 x 192 kVA + 2 x 384 kVA = 1 920 kVA)

5 ZÁVĚR

V článku byl na základě odhadu počtu nabíjecích stanic v České republice do roku 2030 z NAP čisté mobility stanoven scénář instalace vybraných AC i DC nabíjecích stanic. První instalace popisuje zavedení AC nabíjecích stanic a 50 kW DC stanic u bytové jednotky, kde je předpoklad dlouhodobějšího parkování EV. Druhá instalace popisuje vybudování rychlonabíjecích stanic v průmyslových objektech, kde je předpoklad potřeby vyšších nabíjecích výkonů. Třetí instalace uvádí vystavění rychlých a ultra rychlých nabíjecích stanic na čerpací stanici v blízkosti dálnice, kde je požadavek na co nejrychlejší nabití EV a na minimální tvoření front. V následující práci dojde k uskutečnění výpočtů a možnému rozšíření počtu nabíjecích stanic v jednotlivých instalacích.

LITERATURA

- [1] E-mobilita pro každého. ABB Brozura Terra AC Wallbox [online]. Praha: ABB, 2020 [cit. 2021-03-25]. Dostupné z: https://search.abb.com/library/Download.aspx?DocumentID=2CHC420005B4601&Languag eCode=cs&DocumentPartId=&Action=Launch
- [2] Terra 54. ABB [online]. Praha: ABB, 2021 [cit. 2021-03-11]. Dostupné z: https://search.abb.com/library/Download.aspx?DocumentID=2CHC420001L4601&Languag eCode=cs&DocumentPartId=&Action=Launch
- [3] Nabijeci stanice Terra DC. ABB [online]. Praha: ABB, 2020 [cit. 2021-03-11]. Dostupné z: https://search.abb.com/library/Download.aspx?DocumentID=2CHC420016B4601&Languag eCode=cs&DocumentPartId=&Action=Launch
- [4] Electric Vehicle Infrastructure. ABB [online]. Praha: ABB, 2020 [cit. 2021-03-11]. Dostupné z: https://search.abb.com/library/Download.aspx?DocumentID=4EVC700601-LFUS&LanguageCode=en&DocumentPartId=&Action=Launch
- [5] Návrh rekonstrukce distribuční sítě 0,4 kV E.ON v zadané oblasti [online]. Brno, 2012 [cit. 2021-03-11]. Dostupné z: https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=51956. Diplomová práce. Vysoké učení technické v Brně.
- [6] MAPA NABÍJECÍCH STANIC. FDrice.cz [online]. Praha: 24net s.r.o., 2021 [cit. 2021-03-11]. Dostupné z: https://fdrive.cz/mapa-nabijecich-stanic/
- [7] Rozvoj dopravní infrastruktury do roku 2050. MDCR [online]. Praha: Ministerstvo dopravy, 2020 [cit. 2021-01-12]. Dostupné z: https://www.mdcr.cz/Dokumenty/Strategie/Rozvojdopravni-infrastruktury-do-roku-2050
- [8] Nabíjecí stojan EVBox AC BUSINESS 2x22 kW. COMelectric [online]. Praha: COMelectric, 2018 [cit. 2021-03-11]. Dostupné z: https://www.comelectric.cz/wpcontent/uploads/2019/07/letak-A4-EVbox-AC-Business-2x22kW.pdf

A SELECTED WATER AREA FOR FLOATING PV POWER INSTALLATION-BASIC EVALUATION

Enas Al Halabi

Bachelor's degree Programme (3.) FEEC BUT E-mail: xalhal00@stud.feec.vutbr.cz

Supervised by: Martin Paar

E-mail: paar@feec.vutbr.cz

Abstract: The paper deals with floating-PV technology and conditions that have to be considered for chosen location for the FPV installation. The basic evaluation of placement for FPV installation is described. As evaluated waterbody the Nové Mlýny reservoir was selected.

Keywords: FPV, Floating PV system, site identification, waterbody.

1 INTRODUCTION

Although floating photovoltaic (FPV) technology is still at a nascent stage, it has been attractive for many countries and its market has grew up dramatically. According to the World Bank report, there is a potential of 204 GWp just in Europe, if 10% of the man-made freshwater reservoirs has been used for floating-PV power systems [5]. the paper selects recommendable site in the Czech Republic on the Nové Mlýny reservoir for PV floating installation.

2 OVERVEW OF THE FLOATING PV SYSTEM

Floating-PV is a new application of a solar technology, in which PV system is placed on massive water bodies. As a rule, the layout of a floating array is the same as is in a land-based PV power plant, but the difference is that in FPV system PV modules, and often the inverters and the transformers, are mounted on floating platforms [5].



Obrázek 1: Conceptual structure - floating solar power plant [5].

Compared to ground-mounted plants, the big advantage of floating-PV technology is that no agricultural land is used, therefor FPV technology as a renewable source energy is an attractive choice for small, high populated countries with lots of water bodies. It also offers unique benefits, like easier installation, the water-cooling effect improves the efficiency of the panels, reducing water evaporation, improving water quality by reducing algae growth and lower operation and maintenance costs [5]. The floating PV technology has still the challenges that include:

The floating PV plant is more expensive than ground-mounted PV but costs are constantly dropping due to technical improvement. Challenges in the marine environment due to high waves, corrosive caused by water salty and the heavy wind loads [3]. Uncertainty about environmental impact and the technical complexity of designing, constructing, and operating on and in water especially electrical safety, anchoring and mooring issues. Safe energy production and transportation on shore and to the grid [5].

3 SITE IDENTIFICATION OF FLOATING 6PV POWER PLANT

One of the most important prerequisites for the successful project of an FPV power plant is the propriate site selection. The essential considerations which estimate site suitability for the FPV construction include: the solar radiation, shading, climate conditions, wind speed, area and shape of water surface, bathymetry, level of water and waves size, condition of the subsurface soil, pollute, environmental regard and other site considerations. The following essential aspects must be considered when selecting a water body for an FPV system [3]:

Type of water body:

Artificial reservoirs; industrial water bodies like wastewater treatment basins and cooling ponds; dams; flooded coal mines; irrigation ponds; and there is a lower priority for natural lakes and attractive or recreational locations [3].

Specific considerations:

Considering that the depth of a reservoir should be shallow up to 15 m (high depth leads to complexity of anchoring and high construction costs). The shape and details of a waterbed - boundaries shape - the reservoir must have a regular shape like rectangular or square. Surrounded by pure land or land with a few trees, not narrow reservoir between mountains to avoid the horizon shading from trees or from the close mountains (which caused hot spots in solar panels). No presence of islands or barriers in the reservoir centre, the ground must be hard for anchoring. Water should be fresh with as possible lower pollute, calm water. The area with frequent bird activity is not preferred to avoid the trouble of the higher-than-average bird droppings on the FPV construction. In the case of water bodies with significant water-level variations, or which have short-term variations owing to waves or tides, the system will have to be designed for additional water-level variations [3].

Populated regions; near manufacturing utilities or ports; electrical infrastructure available. Unsuitable areas are with the risk of seasonal flooding.

Recommended climate: calm wind and high solar irradiation, not too cold region, exist a special designed floating installation to suit heavy ice or freezing environment [3].

4 THE EXPECTED SUITABLE WATER BODY FOR FPV SYSTEM IN THE CZECH RE-PUBLIC AT NOVÉ MLÝNY RESERVOIR

The Nové Mlýny reservoir is a cascade of three dam valley reservoirs on the river Dyje in southern Moravia, on the Brno-countryside, its three parts are: the lower reservoir (Novomlýnská), the middle reservoir (Věstonická), and the upper reservoir (Mušovská). Its average depth is very shallow, around 2 m, and the deepest point is about 7 m [1]. An area of 3,226 ha with floodplain forests and meadows is flooded. The main reason for its construction was to prevent annual floods. The upper reservoir is used for recreation like leisure activities - swimming, windsurfing, and boating [2].

The middle reservoir is a nature reserve and considers an ornithological locality with islands for birds nesting and has a big different group of birds. The lower reservoir (Novomlýnská) is used for irrigation in a very limited way and for recreation as yachting, surfing, and fishing. It is also used

for electricity generation by the small hydropower plant Nové Mlýny. The dam is 2,484 m long [1] and its area is about 1 668 ha, the Yacht club Dyje is located on its south bank [2].

5 PARAMETERS OF THE DESIGNED FPV PLANT AT RESERVOIR NOVÉ MLÝNY

In general, installation of 1 MWp floating PV system requires an area of around 1 hectare for the floating platform and 1.7 hectares of water area after taking into consideration anchoring [3]. Consequently, to installed FPV with 2 MWp, a water area of 3.4 ha would be needed. This area is equal to about 0.2% of the area of the chosen reservoir which equals 1 668 ha. This small percent of the total area will affect as a minimum on the purposes of the reservoir as a recreation area, the environmental effect would be lower as well [3].

Because the reservoir is used for recreation especially water sports and yachts, it is suitable to build the solar installation somewhere near the shore and considering that the south bank near the yacht club and other tourists' facilities is not the good option for installation.

Southern Moravia region is considered as one of the areas with the best intensive solar radiation in Czechia, its annual solar radiation is about 3.2 kWh/m² per day [4].

5.1 **TYP ANCHORING OF THE SYSTEM**

In general, geotechnical, civil engineering, and structural experts determine the anchoring type after applicate soil tests to identify soil composition, toughness, and chemical properties. The anchoring types to the bottom or to the bank are basically influenced by subsurface soil conditions of the reservoir [3].



Obrázek 2: Selected site on the map of Nové Mlýny reservoirChyba! Nenalezen zdroj odkazů.

6 THE BASIC ELECTRICAL SYSTEM PARAMETERS

General information about electrical part and used components, description of connection: The 2 MWp plant consists of 2 similar parts connected by a 1.5 m access way with string inverters. The number of solar modules is 4464.Every 1 MW floating PV plant consists of structure, in which 2223 solar panels with dimensions of 2094x1038 mm per panel are installed, every 72 modules are connected to one inverter then 2x31 inverters are required for the 2 MWp system.

24 modules in one row are connected in serial then the final current is the same for every module and equal to 10.85 A, the final voltage will be $41.5x \ 24 = 996$ V, then three rows connected together in parallel to one inverter which its parameters must be at least U= 996 V, I= 3x10.85 = 32.55 A. It has been chosen plastic solar floating mounting system made of material HDPE.

Assembly of the constructor is done at the location on the bank, the single units are linked together. After a few units are linked, the entire row is pushed to the water body, it is important to ensure that DC cables and connectors stay above the water surface at all the time, DC cables touching the water could deteriorate the cables' properties and increase the risk of corrosion. All the DC cables/conduits should be secured with proper cable ties or clamps to avoid water contact. The transformer, and other electrical components should be dry at small room in the bank [3].

PV panels: Typ: monocrystalline 9BB perc half-cell, Pmax= 450 W, Voltage at Maximum Power Vmp = 41.5 V, Current at maximum Power Imp = 10.85 A, Module efficiency = 20.7 %, 144 cells, size: 2094x1038x35 mm, Company: BLUESUN, quantity: 4464 pcs.

String DC/AC Converters: Typ: Growatt30000TL3, Max. DC power: 34500 W, Max. DC voltage:1000 V, Max. input current: 35 A, Max. efficiency: 98.2 %, quantity: 2x 31 pcs.

Transformer: Typ: earthing dry transformer, Power: 2 MVA, Voltage: 0.69/36 KV, quantity:1 pc, **Connection to MV network**: the system is connected to MV network 22 kV line.

7 CONCLUSION

The selected water body in South Moravia is considered for basic evaluation, in future work the deep evaluation is planned. The selected location matches lots of the key elements of the site considerations like solar irradiation; the presence of electrical infrastructure; the shallow water depth; the fresh clean water; the large area of waterbody; populated location; the suitable climate but the challenges still that is not a calm water and being it the attractive area for recreation and sports, the installations has to be safe for surrounding and should not limit the tourism and other activities to achieve the aim of multiuse of the waterbody. this study is an analysis for a bachelor thesis and isn't a study for a real project. another sites in the Czech Republic were selected to agree as far as possible with the site conditions according to above mentioned considerations: the Rozkoš reservoir; the Křižanovice reservoir; Mácha Lake.

REFERENCES

- [1] *Vodní nádrž Nové Mlýny* [online]. MÍSTOPISNÝ PRŮVODCE PO ČESKÉ REPUBLICE. Available at: <u>https://www.mistopisy.cz/pruvodce/body-zajmu/179/vodni-nadrz-nove-mlyny/</u>
- [2] *Novomlýnské nádrže* [online]. Město Břeclav. Available at: <u>https://breclav.eu/kultura/cestovni-ruch/novomlynske-nadrze</u>
- [3] Where Sun Meets Water: Floating Solar Handbook for Practitioners (English). Washington, D.C.: World Bank Group. <u>http://documents.worldbank.org/curated/en/418961572293438109/Where-Sun-Meets-Water-Floating-Solar-Handbook-for-Practitioners</u>
- [4] *Solar resource maps of Czech Republic* [online]. SOLARGIS. Available at: <u>https://solargis.com/maps-and-gis-data/download/czech-republic</u>
- [5] World Bank Group; ESMAP; SERIS.2018.
 Where Sun Meets Water: Floating Solar Market Report Executive Summary (English).
 Washington, D.C.: World Bank Group. Available at: http://documents.worldbank.org/curated/en/579941540407455831/Floating-Solar-Market-Report-Executive-Summary
- [6] *Mapy.CZ* [online]. Available at: <u>https://en.mapy.cz/s/guvanaraku</u>
- [7] Earthing transformers. *Hitachi ABB* [online]. Available at: <u>https://www.hitachiabb-powergrids.com</u>
- [8] Solar panel. *BlueSun* [online]. Available at: <u>https://www.bluesunpv.com/bluesun-144cell-</u> solar-panel-9bb-perc-half-cell-440w-450w-455w-mono-pv-module_p382.html
- [9] inverter. *BlueSun* [online]. Available at: <u>https://www.bluesunpv.com/growatt-30000w-40000w-grid-tie-solar-inverter_p19.html</u>

DESIGN AND CONSTRUCTION OF MEASURING APPARATUS FOR LIGHTNING RESEARCH

Ivan Kirsanov

Bachelor Degree Programme (3), FEEC BUT E-mail: xkirsa00@stud.feec.vutbr.cz

Supervised by: Zdeněk Roubal

E-mail: roubalz@feec.vutbr.cz

Abstract: The present work is concerned with the problems of measuring the electric field intensity in the earth's atmosphere. Within the article theoretical principles of circulating charges in the atmosphere leading to the increasing of the electric field strength are discussed. The minimum mathematical apparatus for calculating the required quantity is given. The principle of operation of the electrostatic flux meter type device is explained. Special attention has been given to subjects as follows: sensor noise clipping and an example of the project of an active filter for determining required measurement frequency.

Keywords: intensity, synchronous detection, atmosphere, field mill.

1 ÚVOD

Cílem práce je návrh měřicí aparatury pro výzkum elektrického pole. Experimentální výzkumy elektrického pole v atmosféře byly donedávna omezeny na měření blízko povrchu Země. Tato měření byla prováděna na několika stanicích v průběhu desetiletí. Trvalé průzkumy umožnily objasnit řadu charakteristických rysů chování elektrického pole v povrchové vrstvě a vysvětlit jejich vztah k dalším prvkům atmosférické elektřiny, jako je vodivost atmosféry, proud vzduch– země atd. Bylo zjištěno, že za příznivých povětrnostních podmínek elektrické pole obvykle směřuje tak, jako kdyby Země měla záporný náboj a ionosféra kladný. Střední intenzita pole na úrovni země je asi 130 Vm⁻¹. Během srážek a zejména během bouřek může pole změnit směr a dosáhnout přibližně 10 000 Vm⁻¹. [1]

2 TEORIE

Výsledky průzkumů ukázaly také, že intenzita atmosférického elektrického pole prochází změnami, které probíhají současně po celé planetě. Měření v různých výškách nad zemským povrchem zároveň ukázalo, že vodivost vzduchu stoupá s nadmořskou výškou, přičemž změna probíhá dle exponenciálního zákona. Intenzita pole se mění také v závislosti na zeměpisné poloze. Střední hodnoty intenzity jsou minimální na rovníku a směrem k mírným zeměpisným šířkám vzrůstají. Tyto skutečnosti a řada dalších jevů byly vysvětleny pomocí teoretických modelů (teorie kulového kondenzátoru). Tato teorie obvykle slouží k vysvětlení chování atmosférické elektřiny. Podle této teorie elektrické pole atmosféry existuje díky skutečnosti, že záporný náboj Q– je koncentrován na zemi a kladný náboj Q+ v horní vrstvě atmosféry. V daném případě zemský povrch a atmosféra slouží jako kondenzátorové desky. Nabité částice vytvářejí určitý rozdíl potenciálů mezi "deskami", v důsledku čehož v atmosféře vzniká elektrické pole. Vzhledem k atmosférické vodivosti mezi zemským povrchem a horními vrstvami atmosféry zde proudí elektrický proud *I*, který má tendenci vybíjet kondenzátor ionosféry a zemského povrchu. Hustota tohoto proudu je konstantní a v závislosti na výšce se rovná

$$I_d = E_h * \lambda_h \tag{1}$$

kde E_h je pole a λ_h je vodivost vzduchu ve výšce *h*. Atmosférické pole by se mělo zmenšovat s nadmořskou výškou podle exponenciálního zákona, protože vodivost, jak je uvedeno výše, se s nadmořskou výškou zvyšuje dle exponenciálního zákona. Například vodivost ve výšce 6 km je přibližně 10krát vyšší než na zemském povrchu, kde elektrické pole bude pouze 10%. V obyčejný den nad rovnou pouštní zemí nebo nad mořem se ve směru nahoru od povrchu země zvyšuje elektrický potenciál přibližně o 100 voltů na metr. Existuje tedy vertikální elektrické pole $E = 100 \text{ Vm}^{-1}$. Znamení pole odpovídá zápornému náboji na zemském povrchu. [2]

V zařízení typu elektrostatického fluxmetru se síla elektrického pole převádí na elektrický proud pomocí rotačního elektrostatického generátoru, který je založen na jevu elektrostatické indukce. Tok elektrostatické indukce měřeného pole indukuje elektrický náboj na měřicí desce. Modulátor – stínicí deska – pravidelně stíní měřicí desku v elektrickém poli, díky čemuž se hodnota indukovaného náboje periodicky mění. Náboj proudící do desky a ven z desky vytváří proud v zatěžovacím obvodu. Amplituda tohoto proudu je úměrná síle měřeného elektrického pole, frekvenci otáčení modulační desky a ploše měřicí desky, fáze je určena směrem elektrického pole na povrchu měřicí desky. [3]

3 NÁVRH

Pro výpočet konstanty byl použit následující vzorec: [4]

$$K = \frac{n * S * \varepsilon_0}{d} * (E * d)$$
⁽²⁾

kde *n* je počet lopatek [-], *S* je plocha jedné lopatky $[m^2]$, ε_0 je permitivita vakua $[F * m^{-1}]$, *d* je vzdálenost mezi elektrodami a *E* je intenzita elektrického pole $[V * m^{-1}]$.

Ze vzorce (2) lze odvodit potřebnou plochu jedné lopatky:

$$S = \frac{K}{n * \varepsilon_0 * E} \tag{3}$$



Obrázek 1: Měřicí sonda

Měřicí sonda je založena na rovnoměrném rozložení elektrického pole mezi dvěma rovnoběžnými deskami s otvory ve tvaru ventilátoru. Dopadající elektrické pole na okraji elektrody však může způsobit zkreslení, což má za následek nerovnoměrné rozdělení. Všechny "field mill" nástroje mají vícebřitý rotor rotující před vícebřitou deskou snímače, která je stacionární (tato deska senzoru je stator). Jak se rotor točí, je každá lopatka statoru střídavě vystavena elektrickému poli okolí, nebo je před ním chráněna. Po vystavení elektrickému poli je na každé lopatce statoru indukován náboj a je vyloučen (odváděn), když list rotoru zakrývá lopatku statoru. Signál indukovaného proudu je snímán na statoru a je veden do zesilovače. Při zvětšení šířky mezery indukovaný elektrický náboj klesá a množství změny indukovaného elektrického náboje také klesá. Výzkum ukazuje, že průměrný indukovaný proud klesá s rostoucí mezerovou vzdáleností. Mezera by měla být co

nejmenší, aby se zlepšila citlivost. V praxi může umístění snímacích elektrod v blízkosti rotoru způsobit kolize a zvýšit obtížnost procesu montáže.

Pro snímání rychlosti otáček byl zvolen fotoelektrický snímač, který bude umístěn na statické části sondy. Pod rotorem bude nalepen kousek fólie pro světelný odraz, aby snímač mohl detekovat otáčky. [5]

Důležitou částí měřiče je operační zesilovač, který zesílí slabý vstupní signál. Pro tento účel byl zvolen operační zesilovač OPA131. Řada OPA131 nabízí vynikající univerzální výkon včetně nízkého offsetového napětí, driftu a dobrých dynamických charakteristik. OPA131 je k dispozici v balíčcích DIP a SO. Výkonnostní třídy zahrnují komerční a průmyslové teplotní rozsahy. OPA131 má malou napěťovou nesymetrii $\pm 0,2$ mV, velmi malý vstupní klidový proud 5 μ A, nízký šum a nízkou nelinearitu 0,005 %. Za těchto podmínek byl tento zesilovač zvolen jako nejvhodnější ze všech uvedených výše. [6]

Pro digitalizaci analogového signálu byl zvolen běžný A/Č převodník ADS1115, který má kompatibilitu s Arduinem MEGA.

Obrázek 2 ukazuje blokové schéma přístroje. Z měřicí sondy postupuje signál na vstup operačního zesilovače. Obvod AD630, který je spojen se snímačem otáček na sondě a kontrolérem motoru dohromady, představuje synchronní detekci pro odstranění vlastního šumu sondy. Dolní propust potlačuje vysoké kmitočty. A/Č převodník ze spojitého signálu vytváří odpovídající výstupní diskrétní číselnou hodnotu, kterou zpracuje mikrokontrolér Arduino MEGA, a pak uloží data na mikro SD kartu. DC motor má za úkol roztočit rotor, aby vznikl proud iontů.



Obrázek 2: Blokové schéma

Celý měřicí systém byl navržen pro napájení 5 V. Na obrázku 3 je schéma napájecího systému.



Obrázek 3: Napájecí systém

Napájecí systém se skládá z elektrického akumulátoru na bázi Li-Ion článků a obvodu ICL7660 pro bipolární napájení operačního zesilovače a aktivního filtru.

Pro odstranění vlastního šumu sondy byla implementována synchronní detekce. Tato detekce je realizována pomocí obvodu AD630, zpětné vazby ze snímače otáček a filtru typu dolní propust.

Minimální rychlost ventilátoru pro měření je 2 m/s. [3]

$$v = r * 2 * \pi * f \tag{4}$$

kde r je poloměr [m], v je rychlost [m/s] a f je frekvence [Hz].

Ze vzorce (4) lze odvodit minimální frekvenci

$$f = \frac{v}{2*\pi*r} \tag{5}$$

Při r = 0,04 m je frekvence f = 7,96 Hz, z čehož plyne, že šířku pásma pro aktivní filtr typu dolní propust lze zvolit do f = 16 Hz. Pomocí programu NAF byl navržen aktivní filtr 5. řádu. Na obrázcích 4 a 5 je příklad filtru a jeho frekvenční charakteristika.



Obrázek 4: Aktivní dolní propust 5. řádu



Obrázek 5: Průběh závislosti útlumu na frekvenci v programu NAF

ZÁVĚR

V rámci článku byla probrána teorie atmosférické elektřiny a vypracován teoretický návrh měřiče intenzity elektrického pole. Důležitou částí návrhu je odstranění šumu a potlačení vysokých frekvencí, které mohou zkreslit měření. Pro tento účel bylo navrženo použití synchronní detekce a k ní doplněn následný aktivní filtr 5. řádu. Pro zesílení slabého vstupního signálu byl zvolen operační zesilovač OPA131 s nízkým offsetem 0,2 mV.

PODĚKOVÁNÍ

Chci poděkovat vedoucímu bakalářské práce Ing. Zdeňku Roubalovi, Ph.D.

REFERENCE

[1] IMYANITOV I., CHUBARINA E. Electricity of the free atmosphere. Results of Measurements Carried Out During the International Geophysical Year and Year of International Geophysical Collaboration, Gidronieteorologicheskoe Izdatel'stvo Leningrad 1965, Israel Program for Scientific Translations Jerusalem 1967, 218s

- [2] FEYNMAN R. Lectures on physics. Basic Books, 2010, 566s. ISBN: 978-0-465-07998-8
- [3] SMIRNOV E. Změny v elektrickém poli Země v seismických regionech jako ukazatele silných zemětřesení a erupčních jevů na Slunci. Disertační práce. Federální státní rozpočtový ústav vědního ústavu pro kosmofyzický výzkum a distribuci rádiovln, dálněvýchodní odvětví Ruské akademie věd Petropavlovsk Kamčatskij, 2018, 259s
- [4] MARINOV T., POPOVA P., MITEV M. Atmospheric electrostatic field acquisition system with charge sensing, Texas Instruments European Analog Design Contest 2013 Project Report. Technical University – Sofia Bulgaria, 2013, 10s
- [5] YONG Cui, HAIWEN Yuan, XIAO Song , LUXING Zhao, YUMENG Liu, and LIWEI LIN. Model, design, and testing of field mill sensors for measuring electric fields under highvoltage direct-current power lines, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, 2018 VOL. 65, NO. 1, s 608-615
- [6] OPA131 datasheet General-Purpose FET-INPUT OPERATIONAL AMPLIFIERS, Texas Instruments, 2020

Magisterské projekty

Biomedicínské inženýrství a bioinformatika

A SYSTEM FOR CONTROLLING A COMPUTER PRESENTATION USING GESTURES

Radek Němec

Master Degree Programme (2), FEEC BUT E-mail: xnemec81@ vutbr.cz

Supervised by: Vratislav Čmiel

E-mail: cmiel@ feec.vutbr.cz

Abstract: This paper focuses on the design of a device and gestures for contactless control of the computer presentation. The key elements for this work are the development board ESP32 and IMU sensor BNO055. The goal is to connect the board, sensor and power supply on a printed circuit board and develop an appropriate software in Arduino IDE containing the Bluetooth communication between the microcontroller and the computer. In this paper, the general overview of the solution is presented as well as the communication and two gestures which has been already implemented. The results are evaluated and future improvements are discussed.

Keywords: Contactless control, gesture, presentation, sensor, ESP32, BNO055

1 INTRODUCTION

Nowadays, in the world full of modern technology, it is possible to encounter a large number of smart devices for contactless control of computers, robots or any other software. In recent years, the devices, such as camera or different types of sensors are mainly used for recognizing the gestures of individual limbs, face or straight the whole body, which is then converted into signals or is image-processed and then sent to a computer as an instruction. Companies around the world are also developing various types of gloves or other devices with a gyroscope, accelerometer or similar types of sensors which make it no longer necessary to use cameras, but the data is read directly from the moving limb, sent wirelessly to computers and subsequently processed. One of the device which this work was inspired by, is the DigiTouch system represented by a QWERTY keyboard implemented to the glove, so the user can easily tap the fingers, whereas each phalanx represents one key on the keyboard [1]. Another system from biomedical sphere using similar components as this work is [2], which uses sensors detecting the bend of the fingers in combination with gyroscope, accelerometer and Arduino Nano board as the automatic translator of the sign language.

Earlier, the user had to perform various types of non-natural movements with the above-mentioned devices. In case of longer operation of the device, this type of control could be very inconvenient fot the user. Although many gestures are intuitive and simple, even the best devices still contain many bugs and limitations from the external environment that need to be considered during development and not omitted.

Undoubtedly the most used computer accessory is a mouse and keyboard. However, during the presentation, the user may need to control the computer even from its immediate vicinity. That could be solved by using a presenter or a device that will send instructions to the computer based on the hand gestures of the user in the form of signals. With such a device, the user would be able to control the presentation remotely without the need for any other accessories, just a hand and the device itself.

The aim of this work is to design a device that will allow the user to wirelessly control the mouse cursor on the screen, open a presentation and use appropriately designed gestures to switch individual slides in the presentation.

2 REALIZATION OF THE DEVICE

The device uses ESP32-DevKitC V4 development board in combination with the Adafruit BNO055 IMU sensor. The power bank that is used as the power supply is connected to the ESP32 via a mini USB port and provides an output voltage of 5 V. However, this implementation is used for development purpose only and Li-Ion or Li-Po battery will be used for the final version of the device. The BNO055 sensor is then powered from the ESP32 development board via its 3v3 pin with output voltage of 3.3 V. The development board with the IMU sensor, which provides input data from the gyroscope and the accelerometer, communicate with each other via the I2C bus, specifically via SDA and SCL serial data lines. This position data are represented by signals obtained from hand movement with the attached sensor. The signals are then algorithmically processed in ESP32 and transefered to an instruction which is sent to a computer using a Bluetooth module.

The ESP32 has a dual-core Xtensa LX6 processor running at 160 MHz. The main advantage of ESP32 is the integrated Bluetooth module version 4.2 with the BLE (Bluetooth Low Energy) function. ESP32 also includes a WiFi module with support for WPA and WPA2 security in the 40 MHz band. The baud rate of 115200 is used in this work. The BNO055 sensor has a sampling frequency of up to 100 Hz and 9 degrees of freedom (9-DOF). These degrees indicate all possible movements of the sensor in 3D space. Thus, three degrees of freedom can be specified for linear motion, three for rotational motion and the last three are used to collect magnetometric data.

The basic scheme of the device as well as the scheme illustrating the processing of the signal during gesture detection can be seen in Figure 1.



Figure 1: Basic scheme of the device (left) and signal processing (right)

3 GESTURES DESIGN

For intuitive and simple control of the computer and presentation it is necessary to design appropriate gestures that will be easily detectable. However, the algorithm should not swap these gestures with random hand movements. Therefore, during the design of detection logic, great emphasis is placed on the speed and originality of the given movement. Gestures should also be comfortable for the user.

This device has five functions whereas a different type of movement is designed for each functionality. The instruction representing the emulation of keyboard or mouse is send immediately to the computer, so the gestures can be performed immediately one after each other with no time delay. An overview of the proposed gestures and their corresponding instructions can be seen in Table 1.

Gesture	Instruction
Move the hand to the sides	Move the mouse cursor
Shake the hand to the sides	Left double-click the mouse
Raise the forearm	Start the presentation (key F5)
Wave the hand from right to left	Move the presentation forward (key N)
Tilt the hand to the right and back	Move the presentation backward (key P)
Pull up the hand to the breast	Take/drop the laser (key Ctrl+L)
Hold the hand at the back	Close the presentation (key Esc)

Table 1: Overview of suggested gestures and instructions

For the gestures mentioned in the Table 1, there are used either raw accelerometric data or data containing the position and tilt of the IMU sensor in the space for x, y and z axis.

As the example and more realistic idea of the proposed gestures there is shown a gesture for starting the presentation on the left side of Figure 2, where the hand movement of the particular gesture is indicated. The starting point at which the gesture must be started is indicated by a white circle. The orientation of the movement is then indicated by arrows. For correct motion detection, it is necessary to perform the movement in full range, whereas the average length of one movement should be in range within 0.8 - 1.2 s. On the right side of the Figure 2 it can be seen the graphic visualization of data showing the position of sensor in time. The gesture itself is visualized between the red vertical lines.



Figure 2: Indication of hand movement (left side) and position data (right side) of the gesture

In order to detect the start of any gesture, the key is the detection of hand movement, which in Figure 2 represents the signal starting at 0.4 s. This detection is based on accelerometric data. Once the hand starts its movement, the buffer begins to fill.

The detection of the particular gesture in Figure 2 involves maintaining the values for x and y axes in experimentally determined band. According to Figure 2 there should be no position deviations greater than 20° in these two axes. In the z axis, the key will be detect the forearm lift represented by a deflection of the sensor by about 90° and the subsequent return to the initial position.

Each proposed gesture has its own original detection logic which was designed based on exploration and data analysis of captured data when designing and performing the individual gestures.

4 RESULTS AND FUTURE IMPROVEMENTS

At this time the functional connection between the microcontroller and sensor with its power supply is designed, as well as the Bluetooth communication between the microcontroller and PC and two of the gestures – moving the mouse cursor and left double-click with the mouse. Although the device occasionally encounters the Bluetooth signal dropouts, the gesture for moving the cursor works very well with 100 % accuracy. The gesture for left double-click works with accuracy of 95 %. Both the gestures were tested by 2 users for 20 attempts in total.

The proposal prototype of the device placed on the hand, which is being used for testing and developing purposes can be seen in Figure 3.



Figure 3: Prototype of the device

The final device should be smaller and lighter. One of the considered components is TinyPICO ESP32 development board, which is 2 times smaller than the ESP32-DevKitC V4 shown in the Figure 3. The aim will be also to design a printed circuit board for connecting the microcontroller and sensor without the need to solder the pins together via cables. Instead of the big and heavy powerbank the small Li-Ion or Li-Po battery will be used. All the mentioned components will be placed in the case, which can be printed on 3D printer.

5 CONCLUSION

The purpose of this paper was to give an idea about developing a new device, which can be used for contactless control of the computer presentation. The components were already successfully connected together and the communication with the computer via Bluetooth module was designed and tested. Five gestures were designed and its belonging signals were processed and explored. Two of the gestures were already successfully implemented.

The aim of the master thesis will be implementation of the remaining three gestures and verification of their functionality in practice. Next step will be designing the printed circuit board with the power supply and finally placing all the components to the case, which can be mounted to a hand.

REFERENCES

- [1] WHITMIRE, Eric, Mohit JAIN, Divye JAIN, Greg NELSON, Ravi KARKAR, Shwetak PATEL a Mayank GOEL. DigiTouch. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* [online]. 2017, **1**(3), 1-21 [cit. 2020-10-12]. ISSN 2474-9567. Dostupné z: doi:10.1145/3130978
- [2] YUDHANA, A., J. RAHMAWAN a C. U.P. NEGARA. Flex sensors and MPU6050 sensors responses on smart glove for sign language translation. *IOP Conference Series: Materials Science and Engineering* [online]. 2018, 403(1). [cit. 2020-11-01]. ISSN 1757899X. Dostupné z: doi:10.1088/1757-899X/403/1/012032

RESPIRATORY RATE ESTIMATION FROM ECG AND PPG SIGNALS VIA CONTINUOUS WAVELET TRANSFORM

Pavlina Sikorova

Master Degree Programme (1), FEEC BUT E-mail: xsikor19@stud.feec.vutbr.cz

Supervised by: Lukas Smital E-mail: smital@feec.vutbr.cz

Abstract: In this paper, a novel approach to estimate the respiratory rate from ECG and PPG signals is proposed. The respiration signal is extracted from both signals using the continuous wavelet transform and the rate is estimated in time as well as frequency domain. Subsequent fusion of different approaches provides the result more robust. Designed algorithm predicts the breathing frequency with the mean absolute error of 0.96 breaths per minute.

Keywords: respiratory rate, ECG, PPG, continuous wavelet transform, BIDMC

1 INTRODUCTION

The analysis of respiration signal is an important part of medical diagnostics and health condition assessment. The basic information that can be derived from the signal is the respiratory rate (RR), which is one of vital signs. The unit of RR is the number of breaths per minute (bpm). The normal value for an adult is 12-20 bpm [1]. The RR out of this range is considered abnormal and indicates possible cardiopulmonary disorders [2].

Usually, the respiratory rate can be measured by manual counting of chest wall movements per minute (which is highly inefficient, inaccurate, and non-objective), impedance pneumography or by using other additional sensors (which is disadvantageous for unnecessary expenses and extra stress for patients). [3] However, this information can be also obtained from routinely measured signals, whose purpose is not primarily the RR estimation - those include electrocardiogram (ECG) and photopletysmogram (PPG) signals. They are both strongly amplitude and frequency modulated by the respiration and hence make it possible to extract the respiration signal. [3] Breathing influences the beat morphology by changes in thoracic impedance and changes in the orientation of the electrical axis of the heart [3] and the tissue blood volume by transmitted changes in intrathoracic pressure and vasoconstriction of arteries during inhalation [4]. Heart rate increases during exhalation [2].

There are two different approaches how to extract the respiration signal - filter-based and featurebased techniques [5]. This work is focused on the first one, namely on the continuous wavelet transform (CWT). It is used to decompose the ECG/PPG signal into frequency bands, using consecutive bandpass filters, which split the original information among those bands [6]. Some of those bands contain information about patient's breathing where the respiration signal can be derived [7].

2 DATASET CHARACTERISTICS

The algorithm was tested on the BIDMC PPG and Respiration Dataset [8], which is available from the PhysioNet [9]. The data was acquired from 53 patients during the hospital care.

Every recording contained PPG, ECG, and impedance respiration signals (sampled at 125 Hz), which were all 8 minutes long. Besides the signals, there were also some physiological and fixed parameters

included (such as heart rate or directly measured RR, age, etc). Database also included annotated individual breaths (by two human annotators) in the impedance respiration signal, which were used to calculate a respiratory rate. For the final reference of RR, the information from directly measured RR and manual annotations were combined. Reference RRs are defined for each minute of all records [8].

3 RESPIRATORY RATE ESTIMATION

Proposed algorithm can be split into three main parts, each focused on a different step of the process. The block scheme representing the algorithm used for the respiration signal extraction and following RR estimation can be seen in Figure 1. Algorithm is designed and tested to work with one minute time frames but could be redesigned accordingly.



Figure 1: The block scheme of the algorithm

Both PPG and ECG signals were decomposed by the continuous wavelet transform. For this reason was used the analytical Morlet wavelet, which is suitable for the localization of transients in a signal [6]. The interesting frequency band ranged between 0.20 and 0.40 Hz - the choice of this frequency range was based on the typical physiological breathing frequency of adults [1].

The CWT was followed by fusion of the decomposed signals, which fell within the established frequency range. Those frequency bands were averaged, which resulted into single signal. The respiration signal was obtained by the inverse CWT.

The RR was then estimated from the extracted signals. There were two methods used:

1. In the time-domain (Figure 2): detected peaks of the respiration signal correspond with the inhalation. Minimal peak distance of 2.1 seconds was set to avoid false positive detection. Estimation of the RR is based on position of detected peaks.



Figure 2: Peak detection in the time-domain, (BIDMC signal ID: 25)

2. In the frequency-domain (Figure 3): the dominant frequency of the power spectrum was found, as it corresponds with the breathing frequency, from which the final RR could have been estimated.



Figure 3: Dominant frequency of the power spectrum, (BIDMC signal ID: 25)

The median fusion was executed to calculate final respiratory rate. Both methods described before provided each 1-minute-long record with several RR estimations. The median value was presented as the most accurate and plausible respiratory rate estimation.

4 **RESULTS**

The described algorithm was evaluated by mean absolute error (MAE). Estimated respiratory rate was compared to the references and the difference between them was used to calculate the MAE. Based on this quality assessment, the algorithm predicts the respiratory rate with MAE of 0.96 bpm (see Figure 4). Using just one signal leads to different MAEs - for PPG signal is the MAE equal to 1.01 bpm and for ECG signal it is 1.17 bpm. Therefore, the combination of both signals is considered more accurate.



Figure 4: Mean Absolute Error of RR estimations for every record

As can be seen in Figure 4, the records 5 and 27 have MAEs considerably bigger than the rest of the records. Their breathing frequencies (according to the reference provided by the BIDMC database) were significantly lower and thus did not fall within the physiological range. Therefore, the correct respiratory signal could not have been derived. Extending the frequency band would decrease the accuracy of the estimation in general, thus the introduced algorithm has been retained.

5 CONCLUSION

In this study, the possibility of estimating the respiratory rate from non-respiratory signals was introduced, which might be beneficial in medicine. This was achieved through respiratory signal extraction from ECG and PPG, using continuous wavelet transform, and time and frequency-domain analysis. The accuracy for physiological respiratory signals was very high.

Future work could improve the RR estimation from significantly pathological respiratory signals, extracted from non-respiratory signals.

REFERENCES

- [1] GANONG, William F. Prehled lekarske fyziologie: dvacate vydani. Praha: Galen, c2005. ISBN 80-726-2311-7.
- [2] LARSEN, P.D., Y.C. TZENG, P.Y.W. SIN a D.C. GALLETLY. Respiratory sinus arrhythmia in conscious humans during spontaneous respiration [online]. 2010, 174(1-2), 111-118 [cit. 2021-03-12]. ISSN 15699048. DOI: 10.1016/j.resp.2010.04.021
- [3] CHARLTON, Peter H, Timothy BONNICI, Lionel TARASSENKO, Jordi ALASTRUEY, David A CLIFTON, Richard BEALE a Peter J WATKINSON. Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants. *Physiological Measurement* [online]. 2017, **38**(5), 669-690 [cit. 2021-03-12]. ISSN 0967-3334. DOI: 10.1088/1361-6579/aa670e
- [4] NITZAN, Meir, Igor FAIB, Haim FRIEDMAN a D.C. GALLETLY. Respiration-induced changes in tissue blood volume distal to occluded artery, measured by photoplethysmography. *Journal* of Biomedical Optics [online]. 2006, 11(4), 111-118 [cit. 2021-03-12]. ISSN 10833668. DOI: 10.1117/1.2236285
- [5] CHARLTON, Peter H., Drew A. BIRRENKOTT, Timothy BONNICI, et al. Breathing Rate Estimation From the Electrocardiogram and Photoplethysmogram: A Review. *IEEE Reviews in Biomedical Engineering* [online]. 2018, **11**, 2-20 [cit. 2021-03-12]. ISSN 1937-3333. DOI: 10.1109/RBME.2017.2763681
- [6] THE MATHWORKS, INC. CWT-Based Time-Frequency Analysis. The MathWorks, Inc. [online]. 2020 [cit. 2020-12-07]. https://www.mathworks.com/help/wavelet/ug/ cwt-based-time-frequency-analysis.html
- [7] ESPIRITU SANTO, A., C. CARBAJAL, Timothy BONNICI, et al. Respiration rate extraction from ECG signal via discrete wavelet transform: A Review. 2010 2nd Circuits and Systems for Medical and Environmental Applications Workshop (CASME) [online]. IEEE, 2010, 2010, 11, 1-4 [cit. 2021-03-12]. ISBN 978-1-4244-9994-6. ISSN 1937-3333. DOI: 10.1109/CASME.2010.5706679
- [8] GOLDBERG, A., L. AMARAL, J. HAUSDORFF, P. C. IVANOV, R. MARK a H. E. STANLEY. Physiobank, Physiotoolkit, And Physionet. Components of a new research resource for complex physiologic signals [online]. 101(23), pp e215-e220. Circulation, 2000 [cit. 2020-12-07]. ISSN 0009-7322.
- [9] PhysioNet: The Research Resource for Complex Physiologic Signals [online]. Circulation Electronic Pages, 1999 [cit. 2021-03-11]. https://physionet.org/

ATRIAL FIBRILLATION MODEL

Richard Ředina

Master Programme (2nd year), FEEC BUT E-mail: xredin00@stud.feec.vutbr.cz

Supervised by: Marina Ronzhina

E-mail: ronzhina@vut.cz

Abstract: The aim of this thesis is to create a basic 2D electroanatomic model of a heart atria and perform an atrial fibrillation on this model. Model development and simulations are performed in COMSOL Multiphysics. Depolarization wave propagation is simulated by using FitzHugh-Nagumo model. A representation of the pathology is obtained by parameter modulation of this model. The achieved results can be used later to predict the locations of different pathologies in heart which should improve diagnostics and hopefully shorten the duration of operations.

Keywords: Atrial fibrillation simulation, heart modelling, FitzHugh-Nagumo model, computational modelling, depolarization wave propagation

1 ÚVOD

Ve své práci se zabývám vytvořením 2D elektroanatomického modelu srdce za účelem modelování atriálních fibrilací. Tato srdeční patologie je jednou z nejčastěji se vyskytujících arytmií u starší populace [1]. I přes to, že sama o sobě nemusí ohrožovat pacienta na životě, předznamenává tato porucha možnost výskytu závažnějších onemocnění, jejichž projevy mohou vést ke smrti (infarkt, komorové tachykardie a fibrilace nebo cévní mozkové příhody). Ze simulací *in silico* mohou být získány umělé signály EKG, které v porovnání se skutečným EKG pacienta mohou ukázat na patologii. To pomůže zrychlit diagnózu a spolu s tím i urychlit léčbu pacienta, oddálit fatální komplikace a zlepšit kvalitu života. Další z výhod simulací je i jejich efektivní použití bez účasti pacienta. To může být výhodou v počátcích zkoumání nějaké patologie nebo obecně, když není dostatek reálných dat od pacientů. Díky neúčasti pacientů také odpadají etické otázky, na které bývá v medicíně kladen velký důraz.

Simulování šíření depolarizační vlny myokardem je v posledních letech rozvíjejícím se odvětvím. Nejvíce je dáváno do souvislosti s katétrovými ablacemi. Takto vytvořené modely zkoumají možné dopady operací na propagaci depolarizace. Díky tomu je možné lépe naplánovat, kam bude aplikována radiofrekvenční energie, aby bylo docíleno zamezení vzniku fibrilací [2].

2 ANATOMICKO-ELEKTRICKÝ MODEL SRDEČNÍCH SÍNÍ

Srdce se stahuje na popud změny membránového napětí. Jeho velikost je měněna díky střídavému otevírání a zavírání specifických iontových kanálů. Srdce musí pracovat vysoce organizovaně, aby byly jeho stahy efektivní a zvládlo udržovat tlak ve velkém krevním oběhu. Při fibrilaci síní se tato organizovanost ztrácí na úrovni atrií a okrsky svaloviny se chaoticky depolarizují. Pro popis průběhu membránového napětí byly v minulosti vyvinuty matematické modely.

FitzHughův-Nagumův model (FHNM) byl zvolen pro svou jednoduchost a relativní nenáročnost. Je odvozen z obecného Hodgkinova-Huxleyho modelu, který byl vytvořen pro šíření nervového vzruchu na axonu sépie. Původní model počítal s konkrétními membránovými iontovými proudy a sestával se ze čtyř diferenciálních rovnic [3]. Takovéto řešení je sice z biologického hlediska přesné, ale zároveň je výpočetně náročné. Hodgkinův-Huxleyho model se hodí pro sledování dějů, které se odehrávají přímo na buněčné membráně (např. účinky léků na iontové kanály). FHNM pracuje pouze se dvěma diferenciálními rovnicemi pro popis membránového napětí a uměle vytvořené proměnné, která slouží k udržení cyklicity [4]. Sledování průběhu membránového napětí v čase je postačující pro sledování šíření depolarizačních vln v srdci. Rovnice obsahují řadu parametrů, díky kterým lze různě modulovat časový průběh napětí nebo rychlost šíření v prostoru s cílem nasimulovat změny elektrické aktivity síní odpovídající fibrilaci síní. Rovnice májí následující tvar:

 $dV/dt = \nabla(\sigma\nabla V) + kc_1(V - B)(((V - B)/A) - a)(1 - ((V - B)/A)) - kc_2 \upsilon$ (1)

$$d\upsilon/dt = ke(((V - B)/A) - \upsilon),$$
⁽²⁾

kde V je membránové napětí, v je pomocná veličina pro chod modelu, parametry A a B popisují amplitudu napětí (respektive jeho klidovou hodnotu) parametr σ vodivost myokardu a zbylé parametry (*k*, *a*, *c*₁, *c*₂, *e*) jsou charakteristiky membrány jako takové (v literatuře často bez bližšího určení).

Při vytváření anatomicko-elektrického modelu bylo vycházeno z anatomických podkladů [5], které byly přiměřeně zjednodušeny (pro efektivní implementaci) se zachováním značné míry věrohodnosti. Srdeční síně jsou reprezentovány eliptickou úsečí, která je rozdělena na poloviny síňovým septem (viz Obr. 1). Ve stěně pravé síně se nachází sinoatriální uzel (SA – primární pacemaker). Ve stěně levé síně jsou naznačena ústí dvou plicních žil. Síně jsou uzavřeny síňokomorovým septem.



Obrázek 1: Geometrie modelu srdečních síní s SA uzlem a naznačením dvou plicních žil

Model byl následně proložen sítí uzlových bodů, v nichž probíhají výpočty řídících rovnic. Tyto "lokální" jsou následně aproximovány do zbytku plochy.

3 SIMULACE FIBRILACE SÍNÍ

Fibrilace síní mají mnoho různých spouštěčů. V této práci proběhla simulace nejčastějšího spouštěče – ektopické aktivity z ústí plicních žil. Svalovina srdce neostře přechází ve svalovinu cévy. Tyto buňky se časem mohou stát fokusy rychlých depolarizací, které se pak šíří dál na srdeční síně [6]. Ložisko defektních buněk bylo vytvořeno přidáním nové komponenty a její parametry byly pozměněny tak, aby se ložisko chovalo jako zdroj depolarizačních vln.

Je-li srdce zdravé, tak zdrojem depolarizačních impulzů je pouze SA uzel, který udává tepovou frekvenci. Všechny zbylé části mu funkčně podléhají. Membránové napětí buněk v SA uzlu se pomalu zvyšuje až překoná kritickou hodnotu, kdy skokově vzroste. Tím vznikne dostatečný napěťový impulz, který se šíří dál po myokardu (viz Obr. 2 – zelená křivka). Buňky stěny myokardu se samovolně nedepolarizují, a proto následuje jejich depolarizace vždy až v návaznosti na předešlou depolarizaci v SA uzlu (viz Obr. 2 – modrá křivka).

Zapojením ektopického ložiska, které bylo umístěno do jedné z plicních žil bude křivka membránového napětí svaloviny síní desynchronizována oproti SA uzlu (viz Obr. 3 – modrá křivka). Tento rychlý a nepravidelný rytmus není hemodynamicky účinný, a tudíž nedochází k efektivnímu přečerpávání krve ze síní do komor.

Ektopické ložisko bylo původně nastaveno podobně jako SA uzel. Následně byl změněn parametr k (5x zvýšena hodnota), který zvýšil frekvenci impulzů vycházejících z ložiska. Toto zrychlení samo o sobě zajistilo i lehkou nepravidelnost impulzů, protože tkáň v okolí ložiska neměla dost času na to se dostatečně repolarizovat a být tak připravená na další příchozí pulz.



Obrázek 2: Průběh membránového napětí u buněk zdravého srdce (vlevo) a u srdce s ektopickou aktivitou z plicních žil (vpravo)

(zelená - SA uzel; modrá - svalovina síní)



Obrázek 3: Šíření depolarizace ve zdravém srdci, kdy vzruch vychází z SA uzlu (vpravo) a šíření depolarizace z ektopického ložiska (vlevo; barevná škála udávána ve V; t = 1.01 s)

Signály byly snímány ze SA uzlu, respektive ze stěny atrioventrikulárního septa pravé síně. Při implementaci byly zvoleny Neumanovy (no-flux) okrajové podmínky. Z obrázku 2 je patrné, že průběh membránového napětí v SA uzlu se vlivem fibrilace nemění. Je zde aplikována zjednodušená představa o struktuře SA uzlu, kdy je celý reprezentován svými vnitřními buňkami, které jsou od rychlých změn odstíněny buňkami vnějšími [7].

4 ZÁVĚR

Byl vytvořen jednoduchý, ale názorný anatomicko-elektrický model srdečních síní, který byl řízen rovnicemi FHN modelu. Vhodným nastavením parametrů modelu byla ukázána fyziologická aktivita svaloviny síní. Změnou parametrů a uvažováním ektopického zdroje vzruchů pak byla namodelována fibrilace síní charakteristická rychlostí a nepravidelností stahů, které vedou k celkové neefektivitě síní jako pumpy. Limitací metody může být její časová náročnost, která silně závisí na vzorkování modelu sítí konečných prvků. Její výhodou je, že probíhá bez jakékoli účasti pacienta nebo složité manipulace s buněčnými kulturami. Dalším rozvíjením práce by bylo vytvoření 3D modelu, který by ještě o něco lépe vystihoval realitu a následně extrakce uměle vytvořeného EKG, které by sloužilo k porovnání se skutečným EKG pacientů s fibrilací síní. Tímto porovnáním by mohla být predikována lokalizace patologie u pacienta.

ACKNOWLEDGEMENT

I would like to thank Ing. Marina Ronzhina Ph.D. for her advices. I am also very grateful for her helpfulness and patience. I would also like to thank my schoolmates and family for their support during this pandemic era.

REFERENCE

- [1] VÍTOVEC, Jiří, Jindřich ŠPINAR, Lenka ŠPINAROVÁ a Ondřej LUDKA. Léčba kardiovaskulárních onemocnění. Praha: Grada Publishing, 2018. ISBN 978-80-271-0624-0.
- [2] Shim Jaemin, Hwang Minki, Song Jun-Seop, Lim Byounghyun, Kim Tae-Hoon, Joung Boyoung, Kim Sung-Hwan, Oh Yong-Seog, Nam Gi-Byung, On Young Keun, Oh Seil, Kim Young-Hoon, Pak Hui-Nam. Virtual In-Silico Modeling Guided Catheter Ablation Predicts Effective Linear Ablation Lesion Set for Longstanding Persistent Atrial Fibrillation: Multicenter Prospective Randomized Study. Frontiers in Physiology, 2017, DOI=10.3389/fphys.2017.00792
- [3] Zhu Zhenyu, Wang Rubin, Zhu Fengyun. The Energy Coding of a Structural Neural Network Based on the Hodgkin–Huxley Model. Frontiers in Neuroscience, 2018, DOI=10.3389/fnins.2018.00122
- [4] NIEDERER, S.A., LUMENS, J., TRAYANOVA, N.A. Computational models in cardiology. Nat Rev Cardiol, 2019, vol. 16, pp 100-111.
- [5] PAULSEN, Frierich., WASCHKE, Jens., Sobotta Atlas of Human Anatomy 15th ed., Munich: Elsevier 2011, ISBN 978-0-7234-3732-1
- [6] LUKL, Jan. Fibrilace síní. Praha: Grada, 2009. ISBN 978-80-247-2768-4.
- KIRCHHOF Charles J.H.J. ALLESSIE Maurits A. Sinus Node Automaticity During Atrial Fibrillation in Isolated Rabbit Hearts. Circulation. 1992; 86:263–271. DOI=10.1161/01.CIR.86.1.263

Magisterské projekty

Elektronika a komunikace

LORA-BASED INDOOR LOCALIZATION

Marek Šimka

Master Degree Programme (2), FEEC BUT E-mail: xsimka01@vutbr.cz

Supervised by: Ladislav Polák E-mail: polakl@feec.vutbr.cz

Abstract: This paper focuses on possible utilization of LoRa (Long Range) technology for indoor localization purposes. The SK-iM282A starter kit is used to create a LoRa-based wireless link in the 2.4 GHz ISM band. The rightness of the proposed measurement setup and methodology is verified and compared with a state-of-the-art concept. Preliminary results show that the accuracy of LoRa-based indoor localization is highly depending on the used LoRa signal configuration.

Keywords: LPWAN, LoRa, indoor localization, RSSI, 2.4 GHz ISM band

1 ÚVOD

LoRa[®] (Long Range) představuje technologii pro realizace bezdrátového spojení na dlouhé vzdálenosti (až 30 km), s nízkou rychlostí přenosu dat (0,3 kbps až 50 kbps) při zachování nízkých vysílacích výkonů [1]. Využívá modulace s technikou rozprostření spektra a pro Evropskou Unii je poskytována především v pásmu sub-GHz (863–870 MHz). V roce 2017 společnost Semtech představila na trh moduly, které umožňují provozovat technologii LoRa také v ISM (Industrial, Scientific and Medical) pásmu 2,4 GHz [2]. Díky zmíněným vlastnostem se naskytuje více způsobů aplikace této technologie a jednou z nich je lokalizace zařízení ve vnitřním prostředí nebo ve volném prostoru. V případě vnitřní lokalizace ([3]) se v praxi nabízí užitečné možnosti, například nalezení polohy zboží či materiálu uvnitř velkých skladových prostor, trasování jednotlivých dílů v rámci výrobních linek průmyslových společností apod. Cílem této práce bylo prozkoumat a ověřit potenciál technologie LoRa[®] pro lokalizaci uvnitř budov s komunikací v pásmu 2,4 GHz.

2 VNITŘNÍ LOKALIZACE POMOCÍ TECHNOLOGIE LORA®

Měření bylo realizováno s využitím vývojového kitu SK-iM282A od firmy Semtech a společně s ním také softwarového prostředí WiMOD LR Studio pro řízení rádiové sítě [4]. Jedná se o víceúčelovou sadu dvou rádiových vysílačů/přijímačů komunikujících v bezlicenčním pásmu 2,4 GHz. Jeden modul pracoval ve funkci vysílače (TX), připojen USB konektorem k PC. Druhý modul plnil funkci přijímače (RX) jako vzdálené zařízení napájené bateriemi, jenž na základě přijatého LoRa[®] signálu zasílal nazpět zprávu s naměřenými údaji rádiových parametrů. V rámci experimentální lokalizace byly využity tři vysílače. Pro každou pozici bylo na jednotlivých vysílačích naměřeno 30 vzorků (celkem tedy 90 pro jednu pozici) a z nich byla vypočtena průměrná hodnota RSSI (Received Signal Strength Indication).

2.1 PRINCIP LOKALIZACE

Modul SK-iM282A sice umožňuje zaznamenávat časy odeslání zprávy z vysílače a přijetí zpětné odpovědi od přijímače, avšak tyto časové údaje jsou vyjádřeny v řádu sekund, což pro výpočet doby šíření v rámci metody ToA (Time of Arrival) představuje nedostatečné rozlišení hodnot. Z těchto důvodů byla zvolena metoda RSSI, založená na principu měření síly přijímaného signálu.

Pro zjištění hledané vzdálenosti na základě hodnoty RSSI se využívají modely útlumu signálu, neboli path loss modely [5].

Za účelem získání path loss modelu bylo provedeno měření pro vzdálenosti v rozsahu 0.1 až 7 m. Konkrétně v rozmezí 0,1 - 0,5 m s jemným krokem 0,1 m a v intervalu 0,5 m až 7 m s krokem 0,5 m. Závislost útlumu signálu v prostoru má logaritmický charakter (viz Obr. 1), tudíž při vzdálenosti do 0,5 m dochází k nejprudšímu poklesu RSSI. Z toho důvodu byl zvolen jemnější krok a tím dosaženo přesnějšího modelu. Pro každý bod vzdálenosti bylo naměřeno 10 hodnot a z nich následně vypočten medián, čímž se odstranil vliv skokově se měnících hodnot. Pro potřeby lokalizace byl zvolen typ modelu, který vychází z měřené hodnoty RSSI při referenční vzdálenosti 1 m a nikoliv z teoretického vztahu ($20 \cdot log_{10}(f) - 28$) na základě pracovní frekvence. Model je definován následující rovnicí [5]:

$$RSSI \ [dBm] = -10 \cdot n \cdot log_{10}(\frac{d}{d_0}) + A + X_{\sigma}, \tag{1}$$

kde *d* je vzdálenost mezi vysílačem a přijímačem, *n* je koeficient ztrát šířením v daném prostředí, *A* značí hodnotu RSSI naměřenou při vzdálenosti $d_0 = 1$ m, X_{σ} představuje Gaussovo normální rozdělení s nulovou střední hodnotou a k tomu příslušná směrodatná odchylka σ .

Pro referenční vzdálenost 1 m byla naměřena hodnota A = -42,54 dBm. Měřením na různých vzdálenostech došlo k nalezení průměrné hodnoty parametru *n*, která činila 1,5669. Obr. 1 ukazuje měřený i teoretický průběh závislosti RSSI na vzdálenosti. Teoretická závislost pro vzdálenosti 0-7 m byla vypočtena dosazením nalezených hodnot *A* a *n* do rovnice 1. Přesnost odhadu path loss modelu byl posuzován podle hodnoty koeficientu determinace (Rsquare), který nabývá hodnoty od 0 do 1, kde hodnota 1 znamená přesnou shodu teoretického modelu s reálným. V našem případě, při vyjádření v procentech, se vypočtený model shoduje s měřenou závislostí na 88 %.



Obrázek 1: Závislost hodnoty RSSI na vzdálenosti mezi TX a RX v přenosovém prostředí.

Následně s využitím předem získaného path loss modelu byly k daným velikostem RSSI nalezeny odpovídající hodnoty vzdáleností mezi vysílači a přijímačem (d_a , d_b , d_c). Výpočet lokalizované pozice proběhl za pomoci matematického principu trilaterace, což je technika založená na nalezení průsečíku tří kruhů, jejichž poloměrům odpovídají zmíněné vzdálenosti. Tento princip lokalizace v daném prostředí včetně konkrétního rozmístění vysílačů A, B, C ukazuje Obr. 2.



Obrázek 2: Princip lokalizace hledané polohy.

Předpokládáme-li, že každý vysílač vůči přijímači lze popsat rovnicí, potom dostaneme soustavu tří rovnic (2, 3, 4). Stanovení lokalizovaných pozic bylo provedeno na základě řešení zmíněné soustavy o dvou neznámých, které představují hledané souřadnice X a Y charakterizující danou pozici.

$$(X - X_a)^2 + (Y - Y_a)^2 = d_a^2$$
⁽²⁾

$$(X - X_b)^2 + (Y - Y_b)^2 = d_b^2$$
(3)

$$(X - X_c)^2 + (Y - Y_c)^2 = d_c^2$$
(4)

2.2 KONCEPT MĚŘENÍ

Pro experimentální ověření lokalizace pomocí LoRa[®] technologie a navrženého konceptu bylo vybráno prostředí nacházející se v rodinném bytě. Jednalo se o místnost s půdorysem ve tvaru písmene T a rozměry 7,4 × 4,25 m. V okolí místnosti se nacházelo 11 Wi-Fi sítí pracujících také v pásmu 2,4 GHz. Analyzováním těchto sítí však bylo zjištěno, že žádná nepracuje na frekvenci bližší než $\Delta f = 24$ MHz, tudíž by nemělo dojít k výraznému ovlivnění LoRa[®] signálu. Přenosové prostředí umožňovalo komunikaci s přímou viditelností (LOS), jelikož se v ose přenosové cesty nenacházely žádné překážky. Vysílač i přijímač byly umístěny na volné desce psacího stolu, což přibližně odpovídá výšce pasu člověka a tím emuluje reálnou situaci, kdy se koncové zařízení nachází v kapse osoby. Robustnost LoRa[®] signálu souvisí především s hodnotou činitele rozprostření (Spreading Factor - SF) a velikosti šířky pásma (Bandwidth - BW) [4]. Měření bylo provedeno ve dvou variantách. V 1. variantě se jednalo o konfiguraci LoRa[®] signálu s nízkým datovým tokem, ale především robustnějšího vůči rušení (kombinace BW = 800 kHz a SF = 12). Druhé měření proběhlo s odlišným nastavením LoRa[®] signálu, kde 2. varianta byla naopak zvolena v nejméně odolné konfiguraci a zároveň s nejvyšším datovým tokem (kombinace BW = 1600 kHz a SF = 5). Další parametry signálu během měření byly nastaveny následovně: frekvence 2,449 GHz, kódový poměr 4/6, vysílací výkon 8 dBm.

3 VÝSLEDKY MĚŘENÍ

Měření proběhla pro dvě varianty konfigurace rádiových parametrů LoRa[®] signálu, přičemž pro tato měření platily stejné přenosové podmínky, princip a postup lokalizace. Ve vybraném prostředí byla provedena měření celkem pro šest různých pozic s náhodně vybranými souřadnicemi za pomoci referenčních vysílačů A, B a C, které se nacházely na souřadnicích A [2,20 ; 3,65], B [3,30 ; 0,40], C [3,90 ; 2,55]. Grafické znázornění veškerých výsledků lokalizace pro obě varianty měření nabízí Obr. 3. Zde jsou v naznačeném půdorysu místnosti vykresleny jednak polohy vysílačů (ve čtvercích označené A, B, C), ale také vypočtené a skutečné polohy lokalizovaných pozic. Každá poloha je vyznačena symbolem křížku s příslušným názvem pozice. Polohy skutečné jsou pro orientaci zvýrazněny zelenou barvou a vypočtené barvou červenou.

V případě měření při 1. variantě konfigurace LoRa[®] se odchylky pohybovaly v rozmezí 0,245 m až 2,82 m, přičemž k nejvyšší odchylce (2,82 m) došlo v případě pozice P6. Naopak nejpřesněji lokalizovaná pozice byla P5 s odchylkou 0,245 m. Z pohledu celého měření činí hodnota průměrné odchylky 1,062 m. V 2. variantě konfigurace byl signál méně robustní než v 1. případě, což se projevilo zvýšenými hodnotami výsledných odchylek, konkrétně byl rozsah odchylek 0,914 m až 4,625 m. Jedinou pozicí s odchylkou nižší než 1 m byla P5. V souhrnu všech pozic činila průměrná odchylka 2,242 m. Z hlediska přesnosti lokalizace je tedy důležitá robustnost signálu oproti vysoké přenosové rychlosti.

Přesnost lokalizace lze porovnat například s článkem [3], kde provedli lokalizaci ve dvou odlišných prostředích do vzdálenosti 5 m. V tomto případě činila průměrná odchylka měření 0,846 m pro 1. prostředí (kancelářská místnost) a 1,534 m pro 2. prostředí (školní učebna). Ze získaných výsledků a těchto údajů vyplývá, že přesnost lokalizace lze také ovlivnit vhodným nastavením systémových parametrů LoRa[®].


Obrázek 3: Polohy lokalizovaných pozic (× – skutečná poloha, × – vypočtená poloha).

4 ZÁVĚR

Pro ověření využitelnosti LoRa[®] za účelem vnitřní lokalizace a správnosti konceptu bylo experimentálně lokalizováno celkem šest pozic. Z důvodu přetrvávajících protiepidemiologických opatření byl experiment proveden v domácích podmínkách. Pro metodu RSSI byl aplikován path loss model s koeficientem determinace 88 %. Bylo provedeno měření ve dvou verzích, lišících se konfigurací parametrů LoRa[®] (BW a SF). Následně byly vypočteny souřadnice jednotlivých pozic a k nim odpovídající odchylky, jenž vyjadřují vzdálenost mezi skutečnou a vypočtenou polohou. Dosažené výsledky ukazují, že LoRa[®] v pásmu 2,4 GHz představuje vhodnou technologii pro lokalizaci uvnitř budov. Téma bude rozvíjeno v diplomové práci, ve které budou provedena měření v šiřším rozsahu. Měřící scénáře budou realizovány pro více přenosových prostředí (LOS i NLOS), ale také především pro vícero konfigurací LoRa[®] signálu.

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory projektu interního grantu Vysoké Učení Technické v Brně FEKT-S-20-6325.

REFERENCE

- [1] LoRa Alliance, Inc. *LoRaWAN™ What is it?: A technical overview of LoRa*® *and LoRaWAN™* [online]. Semtech Corporation, November 2015, San Ramon, s. 20, [cit. 2021-02-12].
- [2] MOUSER ELECTRONICS. Semtech's Low-Power SX128x 2.4GHz Transceivers, Available from Mouser, Deliver Integrated Long-Range RF for IoT [online]. Mouser Press Room, July 2017, [cit. 2021-02-11].
- [3] SADOWSKI, S., SPACHOS, P. *RSSI-Based Indoor Localization with the Internet of Things* [online]. IEEE, 2018, ISSN 2169-3536, [cit. 2021-02-12]. DOI: 10.1109/ACCESS.2018.2843325.
- [4] IMST GMBH. *SK-iM282A LoRa*® *Starter Kit with iM282A Radio Module* [online]. Wireless Solutions, [cit. 2021-02-12].
- [5] LI, G., GENG, E., YE, Z., XU, Y., LIN, J., PANG, Y. Indoor Positioning Algorithm Based on the Improved RSSI Distance Model [online]. Sensors 2018, 18, 2820, [cit. 2021-02-02]. DOI: 10.3390/s18092820.

QUANTIFICATION OF TURBULENCE BY THE EQUIVALENT TEMPERATURE GRADIENT

Alžbeta Kovaľová

Master Degree Programme (2), FEEC BUT E-mail: xkoval13@stud.feec.vutbr.cz

Supervised by: Lucie Hudcová

E-mail: hudcova@feec.vutbr.cz

Abstract: This work gives an introduction to the analysis of optical beam transmission through a turbulent atmosphere using matrix optics. A model of a cascade of turbulent cells is shown and a modified method of equivalent temperature gradient is introduced for the final calculation of the turbulence intensity.

Keywords: Atmospheric turbulence, equivalent temperature gradient, optical beam, matrix optics, cascade of turbulent cells.

1 ÚVOD

Stála prítomnosť turbulencie v atmosférickom prenosovom kanáli je často limitujúcim faktorom spoľahlivej funkcie bezdrôtového optického komunikačného spojenia. Turbulencie majú vždy za následok zvýšenú mieru bitovej chybovosti spôsobenú odklonením optického lúča, fluktuáciou optickej intenzity alebo zmenou uhla dopadu. Keďže sa jedná o jav, ktorý má významný vplyv na správnu činnosť komunikačného systému, je nutné ho vedieť správne kvantifikovať.

V súčasnej dobe existuje niekoľko metód so štatistickým prístupom ku kvantifikácii turbulencie [1], preto je táto práca zameraná na metódu spojenú s okamžitým prístupom k vyhodnoteniu turbulentnej aktivity, a to konkrétne na metódu ekvivalentného teplotného gradientu s využitím popisu prenosu optického lúča pomocou maticovej optiky.

2 ATMOSFÉRICKÉ PRENOSOVÉ PROSTREDIE

Atmosféra je považovaná za nestacionárne a nehomogénne prostredie, v ktorom majú kvalitatívne parametre prenosového kanálu náhodný charakter. Najdôležitejšie javy z hľadiska optických bezdrôtových spojov sa realizujú v najnižšej vrstve atmosféry, a to v troposfére, kde vplyv prenosového prostredia na kvalitu optickej bezdrôtovej komunikácie je reprezentovaný časovo, priestorovo a frekvenčne závislými parametrami: koeficientom extinkcie a indexom lomu prostredia [2].

Optická turbulencia je definovaná ako fluktuácia indexu lomu vyplývajúca z teplotných zmien medzi atmosférou a zemským povrchom. Spojením s mechanickým pohybom vzduchu dochádza k vzniku interagujúcich vírov, ktoré sú tiež označované ako turbulentné cely. S uvážením stochastického charakteru zmeny vetra, môžeme považovať turbulenciu za chaotický pohyb atmosféry, z čoho vyplýva, že ide o ťažko predvídateľný jav [3].

3 MATICOVÁ ANALÝZA PRENOSU OPTICKÉHO LÚČA

Táto práca je tvorená simuláciou turbulentného prostredia pomocou jednoduchých optických elementov, ktoré slúžia na aproximáciu veľkých turbulentných ciel. Technika prenosu lúčov je založená na dvoch referenčných rovinách nazývaných vstupná a výstupná rovina, ktoré sú kolmé

na optickú os systému. Existuje niekoľko optických elementov, ktoré sú reprezentované ABCD maticou spájajúcou parametre odklonu optického lúča na vstupnej rovine s parametrami odklonu na výstupnej rovine. Týmito parametrami sú vzdialenosti odchýlok zväzku od osi systému (x_1, x_2) a uhly, ktoré zvierajú s pomocnou osou systému (θ_1, θ_2) . Všeobecný vzťah pre výpočet parametrov vystupujúceho optického lúča je daný vzťahom [4]:

$$\begin{pmatrix} x_2 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix}.$$
 (1)

V tejto práci bola zvolená aproximácia turbulentnej cely hrubou šošovkou, ktorá pracuje s nestacionárnym prostredím, kde dochádza k zmene indexu lomu, čo bližšie popisuje reálne podmienky. ABCD matica pre tento optický element má tvar **Chyba! Nenalezen zdroj odkazů.**:

$$\begin{pmatrix} 1 & 0\\ \frac{n_2 - n_1}{R_2 n_1} & \frac{n_2}{n_1} \end{pmatrix} \begin{pmatrix} 1 & t\\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0\\ \frac{n_1 - n_2}{R_1 n_2} & \frac{n_1}{n_2} \end{pmatrix},$$
(2)

kde t predstavuje centrálnu hrúbku šošovky, n_1 značí index lomu mimo šošovku, n_2 index lomu vnútri šošovky, R_1 poukazuje na polomer krivosti prvej plochy a R_2 na polomer krivosti druhej plochy šošovky.



Obrázok 1: Technika prenosu optického lúča (vľavo). Parametre hrubej šošovky (vpravo).

4 URČENIE ODCHÝLKY OPTICKÉHO ZVÄZKU KASKÁDOU TURBULENTNÝCH CIEL

Táto časť práce slúži pre bližší popis turbulentného prostredia, ktorý je tvorený väčším počtom turbulentných ciel tvoriacich nerecipročnú optickú trasu. Náhradou turbulentných ciel hrubými šošovkami je možné zistiť pozíciu optického lúča na výstupnej rovine šošovky a smer, ktorým sa ďalej šíri. Taktiež tento postup môže slúžiť k predikcii správania optického zväzku, ktorý by sa šíril spätným kanálom a rekonštrukcii modelu približného turbulentného prostredia z nameraných odchýlok.



Obrázok 2: Kaskáda troch hrubých šošoviek reprezentujúcich tri turbulentné cely

Dané experimentálne riešenie je tvorené kaskádou troch hrubých šošoviek aproximujúcich prítomnosť troch turbulentných ciel. V prvej turbulentnej cele dôjde k odklonu paraxiálneho lúča, ktorý sa šíri skrz kaskádu turbulentných ciel, až s výslednou odchýlkou je detekovaný na prijímacej rovine detektora. Jednotlivé odchýlky na výstupných rovinách šošoviek je možné získať prostredníctvom matíc ABCD. Pomocou trojuholníkovej podobnosti a goniometrickej funkcie tangens je možné vyrátať odchýlku lúča v medziturbulentnom priestore a na prijímacej rovine detektora. Vo výsledku má hlavný vplyv na hodnotu odchýlky veľkosť uhla odklonu a dĺžka trasy, ktorú musí laserový lúč prejsť.

5 MODIFIKOVANÁ METÓDA EKVIVALENTNÉHO TEPLOTNÉHO GRADIENTU

Táto metóda vyhodnocuje ekvivalentný teplotný gradient (ETG) z polohy laserového lúča na tienidle nachádzajúcom sa v prijímacej rovine a z parametrov charakterizujúcich okolitú atmosféru. Využíva sa za prítomnosti homogénneho rozloženia turbulencie medzi vysielačom a prijímačom [6]. Je možné ju využiť len za predpokladu vychýlenia optického lúča spôsobeným prítomnosť ou veľkých turbulentných ciel. Vzťah pre výpočet ETG pomocou tejto metódy nadobúda tvar [6]:

$$\frac{dT}{dy} = -2 \cdot \frac{V_m}{R} \cdot \frac{n}{n-1} \cdot \frac{y_a}{L^2} \cdot P,$$
(3)

kde dT/dy predstavuje ETG, V_m molárny objem plynu, R univerzálnu plynovú konštantu, n označuje index lomu, ktorý je platný pre vlnovú dĺžku optického žiarenia na základe vlastností atmosférického prenosového prostredia. y_a poukazuje na odklon optického lúča od osi systému, L na vzdialenosť medzi vysielačom a prijímačom, ktorá zároveň zodpovedá veľkosti nehomogenity v trase optického zväzku a P označuje tlak vzduchu. Pomocou modifikovanej metódy ETG je možné dôjsť využitím metódy uhla dopadu z okamžitého popisu turbulencie pomocou dT/dy ku štatistickej hodnote štruktúrneho parametra indexu lomu C_n^2 , ktorý hodnotí silu turbulencie [6]:

$$C_n^2 = \frac{8}{3} \cdot \frac{1}{2,91} \cdot \frac{1}{L} \cdot D_{RXA}^{1/3} \cdot \frac{1}{m} \cdot \sum_{i=1}^m \left(\operatorname{arctg}\left(\frac{1}{-2 \cdot \frac{Vm}{R} \cdot \frac{n}{n-1} \cdot \frac{P}{L}} \cdot \frac{dT}{dy}\right) i - E\left(\operatorname{arctg}\left(\frac{1}{-2 \cdot \frac{Vm}{R} \cdot \frac{n}{n-1} \cdot \frac{P}{L}} \cdot \frac{dT}{dy}\right)\right) \right)^2.$$

$$(4)$$

 D_{RXA} značí priemer prijímacej apertúry, *m* je počet nadobudnutých hodnôt uhla dopadu a E(arg) poukazuje na strednú hodnotu argumentu.

6 SIMULÁCIA

V prvom bode tejto simulácie boli namodelované turbulentné cely tvoriace kaskádu hrubých šošoviek s náhodne generovanými rozmermi v rozmedzí od 1 *cm* do 3 *cm*. Následne boli definované vzdialenosti medzi jednotlivými turbulentnými celami, ktoré boli rovnako náhodne generované v rozmedzí od 1 *cm* do 3,5 *cm*. Sumou jednotlivých medziturbulentných vzdialeností a centrálnych hrúbok šošoviek bola daná celková vzdialenosť medzi vysielačom a prijímačom. Počiatočný uhol odchýlenia laserového lúča bol rovný $\theta_1 = 0,4^\circ$ tak, aby spĺňal podmienku prítomnosti paraxiálnych lúčov. Využitím maticovej analýzy optického zväzku a goniometrie bolo možné určiť odchýlky optického zväzku za každou šošovkou.

V tomto modeli je index lomu konštantný v okolí turbulentných ciel, počítaný pre vlnovú dĺžku $\lambda = 632,8 nm$. K jeho zmene dochádza vo vnútri každej cely, a to až na šiestom desatinnom mieste. Pre výpočet ETG sa ráta s molárnym objemom plynu $V_m = 22,4136 \cdot 10^{-3} m^3 \cdot mol^{-1}$, univerzálnou plynovou konštantou $R = 8,315 J \cdot mol^{-1} \cdot K^{-1}$ a tlakom $P = 102\,850\,Pa$.

Turbulentnou kaskádou bol vytvorený model popisujúci správanie optického lúča, prostredníctvom ktorého sa došlo nielen k výsledným odchýlkam zväzku y_a na detektore, ale aj k odchýlkam zväzku po prechode jednotlivými turbulentnými celami. Zo znalosti výsledných odchýlok bolo následne možné vyrátať hľadanú hodnotu ETG využitím modifikovanej metódy pre 200 iterácií. Využitím týchto hodnôt a s predpokladom priemeru apertury detektora $D_{RXA} = 2 cm$ bolo možné vyrátať hodnotu štruktúrneho parametra indexu lomu C_n^2 v medziturbulentných priestoroch a na konci kaskády. Odsimulované výsledky z piatich pokusov sú zobrazené v nasledujúcej tabuľke ako priemery odchýlok zväzku a hodnôt ETG z 200 iterácií spolu s výsledným ukazovateľom intenzity turbulencie, ktorý poukazuje na prítomnosť veľmi silných turbulencií, čo vyplýva z geometrie simulovaného prostredia. (Vzdialenosti krajných rovín od vysielača a prijímača sa pohybujú v rozmedzí od 3 *cm* do 5 *cm*, preto je hodnota priemeru apertury detektora malá.)

	$\overline{x_3} \ [mm] \ \overline{z_3}$	$\overline{x_5}$ [mm]	$\overline{y_a} \ [mm]$	$\overline{dT/dy} \ [K/m]$			$C_n^2 \left[m^{-2/3} \right]$		
ι				<i>x</i> ₃	<i>x</i> ₅	Уа	<i>x</i> ₃	<i>x</i> ₅	Уа
1	0,342842	0,338847	0,209231	-578984,09	-129495,96	-45516,94	6,42E-06	2,43E-07	5,31E-08
2	0,347796	0,350604	0,207242	-574258,01	-134379,23	-49265,64	1,79E-05	2,13E-06	3,03E-07
3	0,343749	0,337940	0,200160	-517915,23	-117563,46	-44478,79	1,56E-08	4,55E-09	7,15E-09
4	0,344045	0,343086	0,201538	-523590,95	-126559,44	-45409,97	4,50E-06	2,01E-07	5,14E-08
5	0,347150	0,343068	0,207417	-577916,95	-133044,15	-49076,08	5,61E-07	1,00E-07	6,15E-10

Tabul'ka 1: Priemerné hodnoty odchýlok zväzku, ekvivalentného teplotného gradientu a výsledný štruktúrny parameter indexu lomu. x₃ predstavuje odchýlku detekovanú po prechode prvou turbulentnou celou, x₅ predstavuje odchýlku detekovanú po prechode druhou turbulentnou celou.

7 ZÁVER

Táto práca predstavuje teoretický základ k okamžitej kvantifikácii turbulencie podporený matematickým modelom turbulentnej atmosféry tvorenej optickými elementami charakterizujúcimi turbulentné víry. Ďalšie kroky práce budú zamerané na bližšiu analýzu zmien v hodnotách ETG a vo výchylkách optického zväzku pomocou spomínaných metód. Tento matematický experiment môže slúžiť k analýze nerecipročných turbulentných systémov, ktorými sa šíri laserový zväzok.

REFERENCIE

- Hudcová, L., Wilfert, O.: Quantification of the atmospheric turbulence by the method of the equivalent temperature gradient. In: 2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA) [online]. IEEE, 2018, s. 1-4. Dostupné z: doi:10.1109/RADIOELEK.2018.8376401
- [2] Wilfert, O.: Optoelektronika: přednášky. Brno: VUT FEKT, 2002, 121 s. ISBN 80-214-2264-5.
- [3] Andrews, L. C., Phillips, R. L.: Laser beam propagation trough random media. Bellingham: SPIE Optical Engineering Press, 1998, 433 s. ISBN 0-8194-2787-X.
- [4] Kogelnik, H., Li, T.: Laser beams and resonators. Proceedings of the IEEE [online]. IEEE, 1966, 54(10), 1312-1329. ISSN 0018-9219. Dostupné z: doi:10.1109/PROC.1966.5119
- [5] Gerrard, A., Burch, J. M.: Introduction to matrix methods in optics. New York: Dover, 1994, xi, 355 s. ISBN 0-486-68044-4.
- [6] Hudcová, L., Wilfert, O.: Determination of the Atmospheric Turbulence by the Analysis of the Optical Beam Deflection. In: 2020 30th International Conference Radioelektronika (RADIOELEKTRONIKA) [online]. IEEE, 2020, s. 1-5. Dostupné z: doi:10.1109/RADIOELEKTRONIKA49387.2020.9092387

THE OCDM SYSTEM MODEL

Tomáš Sekanina

Master Degree Programme (2), FEEC BUT E-mail: xsekan11@vutbr.cz

Supervised by: Roman Maršálek E-mail: marsaler@feec.vutbr.cz

Abstract: This paper focuses on the multi-carrier modulation technique called Orthogonal Chirp Division Multiplex (OCDM). The first part of the paper contains theoretical and mathematical description of the OCDM system in general. Then digital model of the OCDM modulator and demodulator is realized in programming language MATLAB. Realization of the model represents two possible mathematical implementations of the Fresnel transform, whose discrete form is kernel of the system. In the digital model is performed simulation of information from transmitter to receiver in AWGN channel. Results are evaluated from the viewpoint of the system resilience against errors.

Keywords: OCDM, modulation, digital communication system, Fresnel transformation, MATLAB

1 ÚVOD

Realizace digitálního komunikačního systému pomocí *Orthogonal Chirp Division Multiplex* (OCDM) je relativně nový způsob, jehož princip byl poprvé publikován v článku [1]. K přenosu mezi vysílačem a přijímačem je využito tzv. chirp signálů, které nachází uplatnění zejména v radarových a sonarových systémech. OCDM pracuje s více nosnými vlnami a řadíme jej tedy mezi *Multi-Carrier* (MC) systémy. Známějším MC systémem je *Ortogonal Frequency Division Multiplex* (OFDM), který je rozšířený zejména díky možnosti realizace pomocí rychlé Fourierovy transformace. Oba zmíněné systémy pracují na stejných principech, ale jelikož systém OCDM využívá specifických chirp signálů, tak je pro zajištění maximální spektrální účinnosti v modulátoru i demodulátoru aplikována Fresnelova transformace [2], a to konkrétně její diskrétní forma. Cílem článku je systém OCDM nejprve uvést a následně realizovat jeho digitální model v prostředí MATLAB.

2 SYSTÉM OCDM

OCDM k přenosu signálu využívá více nosných vln. Jedná se tedy o MC modulační schéma, které na vysílací straně multiplexuje skupiny lineárních chirp signálů a na straně přijímací aplikuje inverzní kroky, aby bylo dosaženo původního signálu [1]. Systém OCDM pracuje na stejných principech jako systém OFDM, ale jelikož se využívá chirp signálu, tak jsou aplikovány matematicky odlišné postupy.

2.1 DISKRÉTNÍ FRESNELOVA TRANSFORMACE

Pro popis systému OCDM je nejprve nutné věnovat pozornost matematickým vlastnostem diskrétní Fresnelovy transformace, díky kterým jsme schopni transformovat bloky, již namodulovaných chirp signálů, na komplexní sériovou posloupnost vzorků. Sériový datový tok je poté možné vyslat přenosovým rádiovým kanálem směrem k přijímači. Po obdržení signálu na přijímací straně je opět využito transformace kvůli nutné rekonstrukci signálu do původní podoby. Inverzní diskrétní Fresnelova transformace (IDFnT) je v systému OCDM tedy využita v modulátoru a diskrétní Fresnelova transformace (DFnT) v demodulátoru.

Pro digitální implementaci OCDM systému lze využít popisu pomocí DFnT matice Φ řádu $N \times N$ s prvky (*m*,*n*). Pro jednotlivé prvky této matice platí rovnice 1 a 2 [2], [3]:

$$\Phi(m,n) = \frac{1}{\sqrt{N}} e^{-j\frac{\pi}{4}} e^{j\frac{\pi}{N}(m-n)^2} \quad \dots \quad \text{sud} \notin N,$$
(1)

$$\Phi(m,n) = \frac{1}{\sqrt{N}} e^{-j\frac{\pi}{4}} e^{j\frac{\pi}{N}(m+\frac{1}{2}-n)^2} \cdots \text{ lich} N.$$
(2)

3 DIGITÁLNÍ MODEL SYSTÉMU OCDM

Digitální model OCDM systému se skládá z vysílací části (modulátor), přenosového kanálu a přijímací části (demodulátor). Schématicky znázorněno na obrázku 1.

Do vysílací části OCDM systému vstupují sériové řazená binární data. Binární datová posloupnost je za pomocí sériově/paralelního (S/P) převodníku převedena na paralelní posloupnosti, které jsou v jednotlivých větvích mapovány pomocí modulačního schématu QPSK nebo M-QAM. Na namapované symboly x(k) je poté aplikována IDFnT a získáváme diskrétní signál s(n). Datový tok je poté převeden do sériové symbolové posloupnosti a vstupuje do přenosového kanálu. V přijímací části jsou digitální sériová data pomocí S/P převodníku převedena zpět do původních paralelních větví, signál vyskytující se v tomto okamžiku označujeme r(n).



Obrázek 1: Základní digitální model OCDM systému.

Poté je využito DFnT a po transformaci získáváme paralelní symboly x'(m), které jsou dále demapovány a pomocí převodníku paralelně/sériového (P/S) převedeny na výstupní bitovou posloupnost. Na obrázku 1 je zanedbán vliv přenosového kanálu a jelikož nedochází k zašumění vyslaných symbolů, tak v tomto idealizovaném případě platí x(k) = x'(m). Za tohoto předpokladu se musí vstupní bitová posloupnost na vysílači rovnat té na výstupu přijímače.

3.1 MATEMATICKÁ IMPLEMENTACE IDFNT A DFNT

Jádrem OCDM systému je diskrétní forma Fresnelovy transformace, pomocí které jsou symboly vhodně transformovány pro další zpracování a přenos. Tuto transformaci lze implementovat v zásadě dvěma způsoby. První možností je implementovat IDFnT a DFnT do systému OCDM pomocí definičního vztahu této transformace, tedy pomocí akumulace součinu vstupního vektoru s vektorem báze vzorek po vzorku. Druhou možností je implementace maticová, která nepracuje s jednotlivými vzorky, ale s celým souborem vzorků současně [1], [4].

Maticová implementace pracuje se čtvercovou DFnT Φ maticí, která již byla vyjádřena v rovnicích 1 a 2. Ve vysílači OCDM lze signál v maticové formě zapsat jako:

$$\boldsymbol{s} = \boldsymbol{\Phi}^{\boldsymbol{H}} \boldsymbol{x} \tag{3}$$

kde $s = [s(0), s(1), ..., s(N-1)]^T$ je OCDM vektor, $\mathbf{x} = [x(0), x(1), ..., x(N-1)]^T$ je vektor symbolů a symbol Φ^H charakterizuje DFnT matici o velikosti $N \times N$, na kterou je aplikována komplexně konjugovaná transpozice. Neuvažujeme-li zašumění a úniky symbolů vlivem přenosu, tak platí rovnost r=s a pro zrekonstruované symboly \mathbf{x}' na přijímací straně platí rovnice 4:

$$\mathbf{x}' = \Phi \mathbf{s} = \mathbf{x}.\tag{4}$$

Pro lepší orientaci lze využít náhledu na obrázek 1. Zde jsou jednotlivé úseky systému OCDM popsány i maticovými symboly.

Pro srovnání časové náročnosti obou způsobů implementace byla v prostředí MATLAB provedena simulace základního modelu OCDM systému znázorněného na obrázku 1 oběma těmito implementacemi. Výsledky jsou zaznamenány v tabulce 1.

Maticový zápis klade větší nároky na paměť. Uvažujeme-li RAM paměť o velikosti 8GB v kombinaci s 64 bitovou verzí MATLABu,

e	Tabulka 1. Stovhall Casove hardenosti implementace ibi iff a D1 iff.					
		Doba simulace [s]				
1 -	Počet symbolů <i>x</i>	Implementace vzorek po vzorku	Maticová implementace			
3	1 024	11,3594	5,731			
_	10 000	55,5333	6,1291			
	102 400	4 638,2801	6,3643			

Tabulka 1: Srovnání časové náročnosti implementace IDFnT a DFnT.

tak lze pracovat s maticí Φ o maximální velikosti 4 000 × 4 000. Maticová implementace je i přes zmíněnou závislost na velikosti paměti RAM v digitální realizaci volbou číslo jedna, jelikož vykazuje mnohonásobně menší časovou náročnost, což je zřejmé i z tabulky 1.

4 VLIV AWGN KANÁLU NA OCDM MODULACI

Základní model OCDM systému (viz obrázek 1) neuvažuje odezvu přenosového kanálu a neřeší otázku omezení spektra. V této kapitole je uvažován přenosový kanál a jeho vliv na data. Je provedena simulace přenosu informace z vysílače do přijímače v *Additive White Gaussian Noise* (AWGN) modelu kanálu. Je tedy uvažován aditivní šum, který na přenášené symboly působí a signál po průchodu kanálem lze popsat následující rovnicí:

$$\boldsymbol{r} = \boldsymbol{s} + \boldsymbol{n} \tag{5}$$

kde *n* je šumový vektor AWGN.

Tvarování modulačních pulsů je realizováno pomocí Raised Cosine (RC) filtru. Jedná se o filtr typu dolní propust, který je určen k omezení spektra signálu a potlačení mezisymbolových interferencí (Inter Symbol Interference, ISI). RC filtr je realizován pomocí dvou samostatných Root Raised Cosine (RRC) filtrů. Schématicky je AWGN kanál společně s RRC filtry znázorněn na obrázku 2. Zbytek schématických bloků je naznačen třemi tečkami a je stejný jako u schématu 1, kde je znázorněn základní model OCDM. Zpravidla se nastavení filtru provádí tak, aby zejména u nižších hodnot činitele tvaru β nedocházelo k výrazné chybovosti. Vhodné nastavení parametrů RRC filtru je uvedeno v tabulce 2. Omezené spektrum pomocí RRC filtru umístěného ve vysílači a nastaveného dle tabulky 2 je znázorněno na obrázku 3.

Obrázek 2: AWGN kanál v OCDM.

| n

Tabulka 2: Volba parametrů pro RRC filtr.

Parametr filtru	Hodnota		
činitel tvaru β	0,3		
vzorků na symbol	2		
délka impulzní odezvy	50		



Obrázek 3: Omezené spektrum OCDM signálu pomocí RRC filtru.

Vliv AWGN kanálu na demodulované symboly v přijímači lze ukázat za pomocí konstelačních diagramů, viz obrázek 4. Konstelační diagramy namapovaných symbolů pomocí 4-QAM byly vygenerovány pro různé hodnoty *Signal to Noise Ratio* (SNR). Efektivnější metrikou pro popis odolnosti systému OCDM vůči chybám je však závislost BER na poměru $\frac{E_b}{N_0}$, jenž byla vygenerována pro modulační schéma 4-QAM, 16-QAM, 64-QAM a 256-QAM, viz obrázek 5.



Obrázek 4: Konstelační diagramy 4-QAM po průchodu AWGN kanálem.



Obrázek 5: Chybovost BER v systému OCDM pro kanál AWGN.

5 ZÁVĚR

V prostředí MATLAB byl vytvořen model modulátoru a demodulátoru OCDM. Byla provedena simulace základního modelu OCDM systému a tím i ověřena jeho základní funkčnost. Byly otestovány a porovnány dvě možné matematické implementace diskrétní Fresnelovy transformace, tvořící jádro OCDM modemu. Na základě získaných poznatků byla maticová implementace pro digitální realizaci a další simulace v systému MATLAB upřednostněna a veškeré navazující simulace využívají právě ji. Závěr článku se věnuje popisu simulace přenosu informace z vysílače do přijímače v AWGN modelu kanálu. Pro tvarování signálu byl aplikován RRC filtr, díky němuž bylo omezeno spektrum vysílaného OCDM signálu. Vliv AWGN kanálu byl zobrazen za pomocí konstelačních diagramů a zároveň byla vynesena závislost BER na poměru $\frac{E_b}{N_0}$. Křivky dosažené simulací odpovídaly těm teoretickým.

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory projektu specifického výzkumu Vysokého Učení Technického v Brně, *Mobilní komunikační systémy 5. a vyšších generací* (FEKT-S-20-6325).

REFERENCE

- OUYANG, Xing a Jian ZHAO. Orthogonal Chirp Division Multiplexing. IEEE Transactions on Communications. 2016, 64(9), ISSN 0090-6778. DOI: 10.1109/TCOMM.2016.2594792.
 Dostupné také z: http://ieeexplore.ieee.org/document/7523229/
- [2] F. V. JAMES, Daniel a Girish S. AGARWAL. The generalized Fresnel transform and its application to optics. Optics Communications. 1996. ISSN 0030-4018. Dostupné také z: http: //www.sciencedirect.com/science/article/pii/0030401895007083
- [3] OUYANG, Xing, Cleitus ANTONY, Fatima GUNNING a Hongyu ZHANG. Discrete Fresnel Transform and Its Circular Convolution. ArXiv:1510.00574, 2015. Dostupné také z: https: //arxiv.org/abs/1510.00574
- [4] OUYANG, Xing a Jian ZHAO. Orthogonal Chirp Division Multiplexing for Coherent Optical Fiber Communications. Journal of Lightwave Technology. 2016, 34(18), 4376 - 4386. ISSN 0733-8724. Dostupné také z: https://ieeexplore.ieee.org/document/7536125

2×2 PATCH ANTENNA ARRAY WITH SIW-BASED BEAM-SWITCHING NETWORK

Dominik Krejčíř

Master Degree Programme (2), FEEC BUT E-mail: xkrejc62@stud.feec.vutbr.cz

Supervised by: Tomáš Mikulášek

E-mail: mikulasekt@feec.vutbr.cz

Abstract: The paper deals with a design of a 2×2 patch antenna array with an SIW-based beamswitching network. This antenna operates in the ISM band with the central frequency of 5,8 GHz. A feeding network based on Butler matrix provides required amplitude and phase distribution for beam control in four directions. The simulations were performed using the CST Microwave Studio 2020.

Keywords: Beamsteering, Butler matrix, Phase array, Patch antenna

1 ÚVOD

Metody vychylování svazku jsou dnes častějším tématem v bezdrátových komunikačních technologiích. Jejich využití můžeme nalézt i v audiotechnice, radarové komunikaci či optických systémech. V rádiových a radarových systémech lze řízení svazku dosáhnout změnou fází vysokofrekvenčních signálů. V akustice se řízení svazku používá pro směrování zvuku z reproduktorů na konkrétní místo. V optických systémech lze řízení paprsku dosáhnout pomocí změny indexu lomu média, kterým je signál přenášen. Nejnovější využití řízení svazku je však v nové technologii 5G.

Cílem této práce bylo navrhnout anténu s řiditelným svazkem, která pracuje na frekvenci 5,8 GHz. Pro řízení vyzařovacího svazku se využívá různých metod. Při návrhu této antény byla použita analogová metoda vychylování svazku pomocí fázového pole pro řízení svazku do čtyř směrů.

2 ANALOGOVÉ ŘÍZENÍ

Pro analogové řízení vyzařovacího paprsku se používá fázovací pole, kde jsou jednotlivé příspěvky s daným fázovým posunem sečteny a následně je vytvořen výsledný paprsek. Tato metoda tvarování svazku je relativně levná a nároky na spotřebu jsou nízké. Nicméně je zde nevýhoda pro aplikace, kde je potřebná velká šířka pásma, jelikož se mění fáze a ta je frekvenčně závislá. To omezuje použití na frekvencích milimetrových vln, kde je zásadní velká šířka pásma [1]. Jako fázové pole, kde se sčítají jednotlivé příspěvky elementů, se v tomhle případě používá pevně nastavený obvod, který má stejný počet vstupů jako výstupů. Vychylování svazku se provádí přepínáním mezi jednotlivými vstupy. Mezi nejznámější fázovací obvody patří Blassova a Butlerova matice [2].

2.1 BUTLEROVA MATICE

Matice se skládá z několika hybridních členů a fázovacích článků. Tento obvod vyžaduje stejný počet vstupů a výstupů. Chceme-li svazek přepínat do více směrů, je nutné zvýšit počet vstupních a výstupních portů a tím se zvětšuje celkový rozměr matice. Vý-



Obrázek 1: Butlerova matice

hodou může být relativní jednoduchost návrhu a teoretický bezeztrátový přenos. Koncepce Butlerovy matice je zobrazena Obrázek 1.

3 NÁVRH ANTÉNY

Anténní systém je vytvořen pomocí napájecí sítě a anténního pole. Napájecí síť pro řízení svazku do čtyř směrů byla vytvořena jako Butlerova matice v provedení vlnovodu integrovaného do substrátu (SIW). Pro vyzařování bylo navrženo 2×2 pole flíčkových antén s aperturovým napájením. Obrázek 2 znázorňuje model Butlerovy matice.



3.1 NAPÁJECÍ SÍŤ

Obrázek 2: Model Butlerovy matice

Pro splnění podmínky řízení svazku do čtyř směrů bylo potřeba využít fázovací pole, které bylo provedeno jako planární pole 2×2. Pro napájení takového pole bylo potřeba Butlerovu matici patřičně upravit. Návrh byl řešen postupně po dílčích částech. Bylo potřeba určit fázové posuvy, které budou zajišťovat určité nasměrování vyzařovacího paprsku. K dosažení daných fázových posuvů bylo zapotřebí minimálně dvou vazebních článků. Ty byly navrženy a simulovány samostatně. Poté bylo potřeba daný signál z jednoho vazebního členu dostat do druhého, aby se daný signál dostal na všechny výstupní porty. Za tímto účelem byl vytvořen propojovací vlnovod. Pro celou napájecí síť byl použit substrát Arlon 25N s relativní permitivitou ε_r = 3,38 mm, tloušťkou *h* = 1,52 mm a ztrátovým činitelem tg δ = 0,0001. Tyto parametry jsou platné pro kmitočet 5,8 GHz. Výsledky simulací při buzení portu 1 jsou zobrazeny na Obrázek 3, kde parametry S51, S61, S71 a S81 jsou znázorněny plnou čarou a platí pro ně vertikální vedlejší osa (vpravo). Pro S-parametry S11, S21, S31 a S41 platí vertikální hlavní osa (vlevo) v grafickém znázornění.



Obrázek 3: Závislost fází jednotlivých výstupů napájecí sítě na frekvenci – Butlerova matice při buzení portu 1, Závislost S-parametrů na frekvenci – Butlerova matice při buzení portu 1

3.2 ANTÉNA

Jako anténní element byla navržena obdélníková flíčková anténa. Anténní prvek je umístěn na dielektrickém substrátu o tloušťce *h* s relativní permitivitou ε_r . Na spodní straně substrátu je nanesena vodivá vrstva, která se chová jako zemnicí plocha. Napájet takové antény je možné pomocí koaxiálního kabelu, mikropáskového vedení, nebo apertury. Jelikož bylo uvažováno napájecí pole pomocí SIW vlnovodu, tak se přistoupilo k řešení s flíčkovou anténou, která je napájená pomocí apertury (viz Obrázek 4). Flíček byl umístěn na substrátu Cuclad 233 s relativní permitivitou $\varepsilon_r = 2,33$ mm, tloušťkou *h* = 1,52 mm a ztrátovým činitelem tg δ = 0,0001. Tato anténa pracuje na kmitočtu 5,8 GHz.



Obrázek 4: Model flíčkové antény s aperturovým napájením, Závislost činitele odrazu na kmitočtu

3.3 CELKOVÝ ANTÉNNÍ SYSTÉM

Celkový anténní systém se skládá ze dvou vrstev. Na spodní vrstvě se nachází napájecí síť, která je vytvořená pomocí Butlerovy matice. Vstupní signál je zde přiveden na vstupy jednotlivých vazebních článků. Díky spojovacímu vlnovodu je budící signál přiveden na další fázovací články. Za fázovacími členy jsou umístěny štěrbiny, které slouží pro napájení pole flíčkových antén. Anténní systéme je zobrazen na Obrázek 5.





Díky fázovým posuvům je možné vždy patřičně vychýlit vyzařovací svazek. Pro každý vstupní port je nastaven vždy jeden směr vyzařování. Výsledný zisk celé antény se pohyboval v rozmezí 10,1 – 10,6 dBi. Elevační úhly pro jednotlivé kvadranty byly následovné 31 ° pro první kvadrant (2 port), 23° pro druhý kvadrant (4 port), 25° pro třetí kvadrant (3. port), 28° pro čtvrtý kvadrant (1. port). Princip vychylování svazku do čtyř kvadrantů je znázorněn na Obrázek 6.



Obrázek 6: Vychylování svazku do čtyř kvadrantů

4 ZÁVĚR

V této práci byla navržena anténa s řízením svazku do čtyř kvadrantů. Nejprve byla navrhnuta a simulována napájecí síť pro dané vychylování pomocí analogové metody. Práce se dále zabývá návrhem a simulací planárního pole flíčkových antén. Celý anténní systém bude následně zrealizován. Následně budou simulace srovnány se skutečnými naměřenými hodnotami, které budou změřeny v bezodrazové komoře.

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory projektu č. FEKT-S-20-6526 Interní grantové agentury Vysokého učení technického v Brně.

REFERENCE

- UCHENDU, I., KELLY, J. Survey of Beam Steering Techniques Available for Millimeter Wave Applications: Progress In Electromagnetics Research B [online]. In: . 2016, s. 35-54 [cit. 2020-11-16]. Dostupné z: http://www.jpier.org/PIERB/pierb68/03.16030703.pdf
- [2] KHATTAK, M. Kamran, Sungtek KAHNG, M. Salman KHATTAK, A REHMAN, C LEE a D HAN. A low profile, wideband and high gain beam-steering antenna for 5G mobile communication. In: 2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting [online]. IEEE, 2017, s. 2575-2576 [cit. 2020-12-10]. Dostupné z: doi:10.1109/APUSNCURSINRSM.2017.8073330

RF TRANSMITTER CLASSIFICATION BASED ON FRONT-END IMPAIRMENTS

Kristina Youssefová

Master's degree, program Electronics and Communication Technologies (2), FEEC BUT

E-mail: xyouss02@vutbr.cz

Supervised by: Roman Marsalek

E-mail: marsaler@feec.vutbr.cz

Abstract: This paper focuses on classifying Radio Frequency transmitters depending on their Radiofrequency imperfections using a machine learning algorithm. The theoretical part of the paper can be divided into two branches. In the first branch, possible imperfections in the radio frequency transmitters are presented. In the second one, the support vector machine algorithm is explained. The practical part deals with the implementation of support vector machines in the MATLAB program and the evaluation of results.

Keywords: I/Q imbalance, CFO, classification, machine learning, SVM, ROC, Matlab

1 INTRODUCTION

The wireless devices classification based on the transmitter imperfections is one of the promising ways to provide additional security to future wireless communication networks [1]. These imperfections (such as IQ modulator imbalance, DC offset, carrier frequency offset) can be used to classify the data of a specific transmitter amongst the others representatives. Classifying radiofrequency (RF) transmitters depending on their imperfections can be applied by using a selected machine learning algorithm. In this paper, the support vector machine algorithm is used as a classifier showing good performance in this task.

2 RADIO FREQUENCY TRANSMITTER IMPERFECTIONS

These imperfections exist because of the nonlinear behavior of the transmitter components, limiting the quality of the signal and degrading it. Examples of such imperfections can be explained as follows:

• I/Q imbalance

I/Q imbalance can be divided into gain imbalance and phase imbalance. The gain imbalance occurs because of the gain difference between I and Q branches while the phase imbalance occurs when the phase shift between I and Q branches differs from the theoretical value of 90°.

• Carrier frequency offset

The Carrier Frequency offset (CFO) exists because of the frequency mismatch between transmitter and receiver local oscillators. As a result, the received signal will be slightly shifted in frequency.

• Constellation origin offset

The signal from the local oscillator leaks to the input of the power amplifier, in which it is amplified. At the receiver side, this results in shifting the center of the constellation diagram of the origin, on Inphase and Quadrature axes.

3 SUPPORT VECTOR MACHINES

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification [2]. This algorithm searches for the optimal hyperplane that best separates the data into the classes.

3.1 **BINARY CLASSIFICATION**

Let's assume we have two types of data presented in a two-dimension space. The blue data points represent the first class with label 1 and the pink data points represent the second class with label -1. The optimal hyperplane (a line for a two-dimensional space) can be achieved by choosing the line that has the maximum distance from the closest data points of both classes. These closest data points are called the support vectors. A distance between the hyperplane and the support vectors is called a Margin. The data points can be linearly separable (known as a hard margin case, see Figure.3-1-(a)) or non-linearly separable (known as a soft margin case, see Figure.3-1-(b)).



Figure 3-1: Binary classification: (a) Hard margin. (b) Soft margin

3.2 MULTI-CLASS CLASSIFICATION SVM

The most commonly used SVM algorithm for multiclass classification is a one-vs-One [3]. The one-vs-one algorithm divides the multiclassification module into $\frac{l.(l-1)}{2}$ binary SVM models. So, $\frac{l.(l-1)}{2}$ hyperplanes should be found, with each of these hyperplanes separating the data points between two individual classes [2] (see Fig.3-2).



Figure 3-2: One-vs-one SVM classification

4 RESULTS

A set of measurements was performed on a sample of nine different USRP N200 transmitters by Martin Pospišil [1]. In my paper, I use the measured data of transmitters RF1 and RF4, RF1 and RF2, for binary classification, and the data of the transmitters RF1, RF4, and RF7 for the multiclass classification. The classification is done in Matlab software with the use of *Statistics and Machine learning Toolbox*. The outputs of applying the classification by SVM algorithm are discussed as shown in Fig. 4-1 to Figure.4-3. In all figures, the measurements for the first, second, and third transmitters are denoted by different markers colored according to the measured transmitter temperature. The points enclosed in the pink circles are the support vectors we have found. In the case of binary classification, only two features at once for each transmitter of RF1 and RF4 have been chosen. The first case is the gain imbalance and the quadrature error (see Figure.4-1-(a)). It is evident that the data are linearly separable. In this case, the SVM classifier best separates the data into two classes and thus there will be no classification errors. The second case is the gain imbalance and the CFO imbalance (see Figure.4-1-(b)). It can be noticed that the data is not entirely separable. This case introduces a soft margin where some amount of data points are strayed over the line into the margin.



Figure.4-1: Binary classification for measured data of RF1 and RF4: (a) linearly separable case. (b) linearly non-separable case.

Figure.4-2-(a), and Fig.4-2-(b) present various Receiver Operating Characteristic (ROC) curves for classification between RF1/ RF4 and RF1/RF2, respectively. These curves are used to evaluate the performance of the SVM classifier. It is possible to see that the worst ROC curve is between RF1 and RF2 transmitters.



Figure.4-2: ROC curve for SVM classification : (a) between RF1 and RF4. (b) between RF1 and RF2. The CFO and the gain imbalance have been used as the SVM features

The last Figure in this section – Fig 4-3, illustrates the principle of multi-class SVM classification according to the one-vs-one algorithm. Various hyperplanes separating between individual pairs of transmitters are denoted in different colors



Figure.4-3: Multi-class classification for real data of RF1, RF4, and RF7

5 CONCLUSIONS

In this paper, I have presented my experimental work on Support Vector Machines method for RF transmitter classification. Several front-end impairments have been used as SVM features. In future work, I will concentrate on a selection of robust multi-dimensional classifiers and their comparison with other classification approaches including a neural network.

REFERENCES

- [1] POSPÍŠIL, M., MARŠÁLEK, R., GOTTHANS, T., Wireless device classification through transmitter imperfections — Evaluation of performance degradation due to the chip heating, In Proceedings of *IEEE Radio and Wireless Symposium (RWS) 2017*, Phoenix, AZ, USA, 2017, pp. 169-172, doi: 10.1109/RWS.2017.7885978.
- [2] AWAD, M., KHANNA, R., *Efficient Learning Machines*, Apress, Berkeley, CA, doi: 10.1007/978-1-4302-5990-9_3.
- [3] KUMAR, M.A., GOPAL, M., A comparison study on multiple binary-class SVM methods for unilabeltext categorization, Pattern Recognition Letters, 31 (2010) 1437–1444, doi: 10.1016/j.patrec.2010.02.015

OMNI-DIRECTIONAL ANTENNA FOR 60 GHZ BAND CHANNEL MEASURMENT

Kristián Levocký

Master Degree Programme (2), FEEC BUT E-mail: xlevoc00@stud.feec.vutbr.cz

Supervised by: Tomáš Mikulášek

E-mail: mikulasekt@feec.vutbr.cz

Abstract: This paper describes a design of an omni-directional antenna for a 60 GHz band. A monocone antenna is integrated with a mm-wave coaxial connector. Such a structure is small in size and provides an omnidirectional radiation pattern with a wide vertical angle. The paper pre-sents a comparison of simulated and measured results of a fabricated prototype. The antenna is in-tended to be used for mm-wave channel measurement.

Keywords: 60GHz band, omni-directional antenna, design

1 ÚVOD

Táto práca sa zaoberá na realizáciu všesmerovej antény v pásme milimetrových vĺn (konkrétne 55 – 65 GHz), návrhom vybranej realizácie pomocou programového prostredia CST Microwave Studio a výrobe navrhnutej antény.

60 GHz pásmo spadá pod pásmo EHF (extrémne vysoké frekvencie). V porovnaní s pásmami o nižších frekvenciách je toto pásmo vysoko ovplyvňované okolitým prostredím, za ich najväčším útlmom stojí absorpcia atmosférickými plynmi. Ďalším významným zdrojom útlmu sú vplyvy počasia ako je dážď alebo v prípade púštnych oblastí to môže byť aj samotná veterná búrka. Práve kvôli týmto znevýhodneniam sa často spája s automotive a komunikáciami na krátke vzdialenosti [1].

Navrhovaná anténa ma slúžiť práve na merania prenosových kanálov na tomto pásme a z toho budeme vychádzať práve aj pri požadovaných, sledovaných parametroch.

2 NÁVR A SIMULÁCIE

Pri výbere konkrétnej realizácie navrhovanej antény boli zohľadnené hlavne praktické a dostupné riešenia. Z tohoto pohľadu bola vybraná realizácia monokónickej antény ktorá by mohla byť pripevnená na koaxiálny konektor a nemuseli by sa tak riešiť prechody medzi koaxialom a vlnovodom aký by sme riešili pri iných realizáciách. Takisto je vyhovujúca očakávaná vyžarovacia charakteristika danej antény ktorá má symetrickú vyžarovaciu charakteristiku okolo vertikálnej osy, čo je jedna z vyžadujúcich vlastností pre meraciu anténu [2][3].

Na Obrázku 1 môžeme vidieť navrhnutý základný model vybraného typu antény v prostredí CST Microwave Studio z ktorého sa vychádzalo. U prvotného modelu sa jednalo len o jednoduchý kónus pripevnený na kuse koaxiálneho vedenia. Model na obrázku už počíta s vybraným koaxiálnym konektorom a jeho pájecím pinom.



Obrázek 1: Rez základného modelu antény

Tento model bol následne veľkostne prispôsobený pre naše pásmo s konečnými rozmermi kónusu : 2 mm na výšku a 4,8 mm v širšom priemere. Osvedčilo sa aj rozšírenie zeme ktoré v novom modeli tvorí protikónus. V dôsledku miniatúrnych rozmerov a potenciálnej krehkosti pájecího pinu musela byť pre anténu navrhnutá mechanická podpora ktorá zabezpečí pevnosť a uľahčí tak manipuláciu s anténou zároveň boli do modelu pridané zmeny ktoré boli spojené s komplikáciami pri výrobe. Ako vhodný spôsob sa javil 3D tlačený obal ktorý by bol dostatočne pevnou a jednoduchou voľbou.

Na základe vlastnosti materiálu by bol vhodný PLA materiál avšak touto technológiou nie sme schopný tlačiť s dostatočnou presnosťou a tak sa pri ďaľších simuláciach počíta s fotopolymérovým materiálom u ktorého vieme tlačiť s presnosťou v desiatkach mikrometrov. Tento materiál má podľa meraní tieto parametre : Relatívna permitivita – 2,58 Merná vodivosť – 0,27 S/m Strátovy činiteľ – 0,023. Tieto parametre boli prebrané síce z iného 3D Resinu avšak po parametrickej analýze bolo vidieť že rozdiely týchto parametrov v desatinách jednotiek nemajú na parametre antény veľký vplyv.

Tento model môžeme vidieť na Obrázku 2.



Obrázek 2: Rez upraveného modelu antény

S výsledným modelom sa nám podarilo dosiahnuť nasledujúce parametre antény (zobrazené na nasledujúcich Obrázkoch 3 a 4) :

S11 – (-)6dB a menej na celej šírke pásma

Vyžarovací uhol - 150 stupňov na 55 GHz až 100 stupňov na 65 GHz



Obrázek 3: S11 navrhovanej antény



Obrázek 4: Smerové charakteristiky navrhovanej antény zľava pre 55 GHz, 60 GHz a 65 GHz

Pre dosiahnutie týchto výsledkov boli parametricky menené a upravované parametre na ktoré bola anténa najcitlivejšia a to spodný priemer kónusu, vzdialenosť nasadeného kónusu od pajecího pinu. Dôležitým parametrom boli aj parametre obalu, konkrétne veľkosť vzduchového priestoru ktorý ovplyvňoval S11 antény a samotnú hrúbku steny obalu ktorá pri veľkej hrúbke znižovala vyžarovací uhol antény.

3 VÝROBA ANTÉNY

Výroba týchto miniatúrnych časti je samozrejme veľmi zložitá avšak po konzultáciách a pomoci dielni na ústave UREL boli vyrobené oba kónusy z mosadze s presnosťou +-2% a zároveň upravený kúpený konektor do potrebnej podoby. Na skonštruovanej anténe z Obrázku 5 sa nám už podarili spraviť prvé merania z Obrázku 3.



Obrázek 5: Zhotovená meraná anténa a zasadenie do vyrobeného obalu

4 ZÁVĚR

Navrhnutý model antény sa zo simulácií javí ako vhodný typ pre našu aplikáciu ktorou je meracia anténa pre pásmo 60 GHz nakoľko dosahuje solídny vyžarovací uhol 100 a viac stupňov pri prijateľnom S11 na celej šírke pásma. Nevýhodou je však zložitosť výroby v dôsledku miniatúrnych rozmerov a zároveň potrebnej presnosti ktorej bude potreba dosiahnuť, o čom sme sa presvedčili pri prvom meraní skonštruovaného modelu. Zhotovená anténa na ktorej sa meralo mali lepšie parametre ako zmenený model antény v simulačnom prostredí. Ďalším krokom tak bude nielen upraviť dostať parametre vyrobenej antény na ešte lepšie hodnoty ale zároveň doladiť model, tak aby odpovedal skutočnosti. Vďaka prvotným meraniam a konštrukcii boli spozorované prvky na ktoré bola anténa najcitlivejšia a teda na ktoré bude najprioritnejšie sa zamerať. Jedná sa predovšetkým o vzdialenosť nasadeného kónusu od pájecího pinu, zbavenia sa vyosenia kónusu voči osi konektoru ale aj potreba zbaviť sa prebytočného cínu ktorý zvýšil výšku rozšírenej zeme a spôsobil nerovnomernosť v tomto bode. Väčšina nedostatkov by sa mohla odstrániť a presnosť výroby by sa mohla zvýšiť výrobou formy do ktorej by sa konektor a kónus zasadili pri pájení. Táto forma by mohla byť takisto vyrobená na fotopolymérovej 3D tlačiarni nakoľko materiál vydrží kontakt s pájkou o teplote 360 °C. Ďalším prínosom by mal mať presnejší kónus v užšej časti ktorý by mohol byť v blízkej dobe vyrobený. Pri doladení výrobného procesu sa počíta so zlepšením parametrov.

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory projektu č. FEKT-S-20-6526 Interní grantové agentury Vysokého učení technického v Brně.

REFERENCE

- [1] BALANIS, C.A. Antenna theory: Analysis and design. Wiley & Sons, 1996
- [2] W. S. Yeoh and Wayne S. T. Rowe, An UWB Conical Monopole Antenna for Multiservice Wireless Applications. IEEE Antennas and Wireless Propagation Letters, January 2015
- [3] Francisco Estêvão Simão Pereira, Maurício Henrique Costa Dias, "On the Design of Conical Antennas for Broadband Impedance Matching Performance", International Journal of Antennas and Propagation, vol. 2017, Article ID 1691580, 13 pages, 2017. https://doi.org/10.1155/2017/1691580

Magisterské projekty

Komunikační technologie a informační bezpečnost

POLYGON SCENARIOS OF THE CZECH DISTRIBUTION NETWORK

Antonín Bohačík

Master Degree Programme (2), FEEC BUT E-mail: xbohac10@vutbr.cz

> Supervised by: Petr Blažek E-mail: blazekpetr@feec.vutbr.cz

Abstract:

This work is focused on the design and implementation of communication scenarios to created polygon that simulates the data communication of the transmission network of the Czech Republic. Scenarios represent usual but also unusual situations in which real stations of the transmission system may be. Furthermore, the assumptions are described here to simulate the created scenarios or the criteria and ranges that the station must cover during the simulation.

Keywords: communication, distribution network, IEC 60870-5-104, industry, polygon, SCADA, scenarios, simulation

1 ÚVOD

S rozvojem moderní technologie se začal rozvíjet trend dálkového řízení. K tomuto účelu byly navrženy systémy a standardy jako Supervisory Control And Data Acquisition (SCADA), které informace nejen shromažďují, ale také podporují vzdálenou regulaci a řízení. Za účelem testování nových zařízení, protokolů nebo technologií obecně vznikají simulátory či celé polygony, které představují bezpečný způsob pro jejich testování. Další využiti těchto polygonů nastává při definování sekvence kroků tzv. scénářů, díky nimž můžeme simulovat chování zařízení při určitých událostech.

Cílem této práce je na základě znalostí chování datové části přenosové soustavy (PS) navrhnout a implementovat komunikační scénáře, včetně samotného ovládacího a kontrolního rozhraní pro vzdálenou správu simulovaných stanic. Tyto scénáře reprezentují stavy reálných rozvoden a elektráren PS, a to jak obvyklé, tak i neobvyklé provozní stavy, jako např. výpadek či podpětí soustavy. V neposlední řadě budou vytvořeny scénáře reprezentující útoky na infrastrukturu polygonu.

V současnosti existuje mnoho polygonů implementující protokol IEC 60870-5-104, ovšem většina z nich neobsahuje všechna důležitá kritéria jako např. možnost zapojení reálných zařízení, obsah dohledového systému (SCADA), možnost propojení s dalšími systémy, či zda jsou zaměřeny přímo na PS. Výčet obdobných polygonů a porovnání zmíněných vlastností shrnuje Tabulka 1.

Název	Realizace	Reálné prvky	SCADA	Další systémy	PS
Polygon PS	Emulace	Ano	Ano	Ano	Ano
Jarmakiewicz et al.	Emulace	Ne	Ano	Ano	Ano
Maynard SCADA	Virtualizace / Emulace	Ano	Ano	Ne	Ne
SNL Testbed	Virtualizace	Ne	Ano	Ano	Ne

Tabulka 1: Porovnání polygonů implementující IEC 60870-5-104 [1].

2 POLYGON DATOVÉ KOMUNIKACE PS ČR

Na ústavu telekomunikací byl vytvořen polygon PS pro výukové, simulační a testovací účely, viz Obrázek 1. Polygon simuluje datový provoz blížící se reálnému provozu mezi jednotlivými stanicemi

a dohledovým střediskem, který je založen na open-source softwaru OpenMUC¹. Stanice reprezentující hraniční komunikační brány rozvoden či elektráren jsou realizovány pomocí jednodeskového počítače Raspberry Pi 3B+ (RPi). Propojení všech prvků zajišť ují přepínače od společnosti Mikrotik a samotná simulovaná komunikace je založena na open-source knihovně *lib60870-2.2.0*, která implementuje standard IEC 60870-5-104, včetně části zabezpečení TLS, dle normy [2].

Stanice polygonu přijímají příkazy/dotazy od dohledového střediska a odesílají na ně odpovědi. Dále pak tyto stanice zastávají funkci simulátoru hodnot fyzikálních veličin, mezi které patří např. napětí (U), proud (I), frekvence (f) a teplota (T). Tyto hodnoty jsou generovány pomocí implementované funkce v rozsahu stanoveném konfiguračním souborem odpovídajícím standardním rozsahům reálných stanic definovaných souborem "Kodex přenosové soustavy ČR" [3]. Hodnoty činného (P) a jalového (Q) výkonu jsou oproti ostatním hodnotám dopočítávány, viz níže. Pro zobrazení nejdůležitějších dat mají všechny stanice připojen informační displej IIC I2C OLED, viz Obrázek 1. V neposlední řadě je zde vytvořen logovací systém pro účel zpětného testování a analýzy přenesených dat.

Celý polygon obsahuje více než padesát prvků, které je nutné kontinuálně kontrolovat a nastavovat pro správnou funkci. Z tohoto důvodu je v polygonu zavedena další oddělaná síť, díky které je možné komunikovat s jednotlivými stanicemi polygonu bez přímého zásahu do datové komunikace, popřípadě měnit jejich parametry. Správu nad sítí zajišť uje webová aplikace, která je založena na open-source webovém frameworku Django². Schéma zapojení stanic polygonu, včetně vytvořené webové aplikace pro management stanic či dohledového střediska, je zobrazeno na Obrázku 2.





Obrázek 1: Polygon přenosové soustavy ČR.



3 SCÉNÁŘE STAVŮ POLYGONU

Pro polygon energetické PS byly vytvořeny a realizovány scénáře, které simulují stavy, ve kterých se můžou stanice PS nacházet. Konkrétně byly vytvořeny scénáře standardního provozu, podpětí, přepětí, výpadku, dále pak scénář útoku vyřazení lokální kontroly stanice a scénář útoku na elektrárny.

3.1 SCÉNÁŘ STANDARDNÍHO PROVOZU

Dle kodexu PS vytvořené ČEPSem musí všechny stanice PS splňovat stanovené rozsahy po dobu 99% času provozu (ovšem existují i výjimky). Jedná se tedy o nejběžnější stav stanice PS, ve kterém se nachází a pro který byla navržena a uzpůsobena. Ovšem celá PS a toky elektrického proudu v ní jsou závislé na mnoha ovlivnitelných, ale i neovlivnitelných faktorech. Tyto faktory, mezi které patří např. fyzikální zákony, konstrukce zařízení či vliv okolních připojených soustav, není možné zcela přesně předpovídat či určovat jejich stálost a neměnnost. Proto kromě standardního provozu musíme předpokládat i stavy nestandardní, jako je např. výpadek stanice. Jednotlivé elektrárny či rozvodny PS mohou obsahovat více než jeden výkonový transformátor a stejně je tomu i u simulovaných stanic

¹Jedná se o java framework pro práci se SCADA systémy, dostupný na https://www.openmuc.org.

²Jeho charakteristickým rysem je architektura Model-view-controller, dostupný na https://www.djangoproject.com.

polygonu. Z tohoto důvodu jsou níže popsané hodnoty stanoveny pro jeden transformátor stanice. Stanovené rozsahy přímo vychází z kodexu PS.

Transformátor stanice polygonu při scénáři standardního provozu simuluje data vstupních a výstupních napětí (U_a, U_b, U_c) v rozsahu 360–420 kV pro napěť ovou úroveň 400 kV, 198–246 kV pro úroveň 220 kV a 99–121 kV pro úroveň 110 kV. Dále pak velikost vstupních a výstupních proudů (I_a, I_b, I_c) pohybujících se ve stovkách až tisících ampérech podle typu rozvodny/elektrárny. Velikosti hodnot činného výkonu (P_a, P_b, P_c) jsou dopočítávány pomocí vzorce:

$$P = U \cdot I \cdot \cos(\varphi) \quad [W]. \tag{1}$$

Velikosti hodnot jalového výkonu (Q_a, Q_b, Q_c) jsou dopočítávány pomocí vzorce:

$$Q = U \cdot I \cdot sin(\varphi) \quad [var], \tag{2}$$

kde hodnota φ představuje fázový posun a je nastavena na hodnotu 0,525 *rad* (přibližně 30°, představující efektivitu 85%). Dále jsou simulovány hodnoty frekvence v rozsahu 49,5–50,5 *Hz* pro jmenovitou frekvenci 50 *Hz*. Pro teplotu transformátoru jsou simulovány hodnoty 120 ±20 °C a hodnoty reprezentující vnější podmínky 20 ±5 °C v souladu s normou [4].

Při vytvoření statistiky zatížení polygonu všemi stanicemi pro scénář standardního provozu dostáváme simulované zatížení polygonu zobrazeného na Obrázku 3 (oranžová křivka). Jedná se o simulované zatížení v rozsahu od 8 GW do 11 GW v závislosti na denní době. Toto simulované zatížení usiluje o napodobení reálného zatížení PS. Pro porovnání jsou v grafu zobrazena i reálná data³ měřená dne 3. 3. 2021 (modrá křivka) získaná ze statistiky vedené ČEPSem.



Obrázek 3: Porovnání simulovaného a reálné zatížení PS ze dne 3.3.2021.

3.2 SCÉNÁŘ ÚTOKU VYŘAZENÍ LOKÁLNÍ KONTROLY STANICE

Podle Desarnaud [5] je energetický sektor ve většině vyspělých zemí součástí kritické infrastruktury a kritické informační infrastruktury. Představuje tedy soustavu velkého významu pro fungování daného státu, která by měla odolávat kybernetickým hrozbám. Za účelem existence možnosti zkoumání různých útoků v bezpečném (simulovaném) prostředí byly v rámci polygonu navrženy a implementovány scénáře útoků, které mají za cíl narušit správné fungování soustavy.

Scénář útoku vyřazení lokální kontroly stanice byl realizován na základě analýzy malwaru "Stuxnet", objeveného roku 2010 běloruskou bezpečnostní společností VirtusBlokAda. Dle Shakariana [6] je považován za nejsofistikovanější malware, který měl za cíl íránský nukleární program. Jednalo se o úspěšný útok zpomalující proces obohacování uranu a výroby nukleárních zbraní. Jedná se o první malware cílený přímo na industriální zónu, jehož součástí byl i programovatelný rootkit, který představuje techniku používanou při falzifikaci informací o škodlivém kódu s cílem skrýt jeho přítomnost.

Tento scénář cílí na změnu odesílaných dat bez vědomí dohledové kontroly. Skládá se ze čtyř základních částí: (i) získání přístupu do stanice, (ii) zachytávání generovaných hodnot, (iii) vytvoření

³Tato data jsou volně dostupná na webu https://www.ceps.cz/cs/data.

zobrazovací smyčky, (iv) vzdálené přetěžování soustavy, viz Obrázek 4. Při tomto scénáři útočník získá přístup do sítě dané stanice. Následně sesbírá dostatečný vzorek dat pro zacyklení veškerých zobrazovacích prostředků. Poté začne útočník vzdáleně přetěžovat síť PS. Dohledové středisko bude od stanice dostávat nereálné údaje o sledovaných hodnotách, které nebudou odpovídat hodnotám skutečným.



Obrázek 4: Diagram útoku na stanici polygonu.

Za účelem simulace útoku je do infrastruktury připojeno další zařízení představující útočníka. Ten pomocí zjištěného přihlašovacího jména a hesla, získaného např. jako při útoku Stuxnet pomocí sociálního inženýrství, získá přístup do stanice polygonu. Útočník spustí skript, který zapříčiní zobrazování zacyklených dat podobných reálně snímaným datům stanice. Následně útočník upraví konfigurační soubory stanice k přetěžování přenosové soustavy. Na základě toho bude dohledové středisko zaznamenávat neobvyklý přebytek energie v celé soustavě, ovšem data přijatá od této stanice nebudou vykazovat známky výkyvů od stanovených rozsahů. Vzhledem k faktu, že útočník má nad stanicí plnou kontrolu, může kromě zmíněného dlouhodobého přetěžování soustavy také odpojit stanici, přeposílat reálná data, či odstranit bezpečnostní prvky stanice.

4 ZÁVĚR

Problematika kyberbezpečnosti se díky stále se rozšiřující automatizaci projevuje do stále více oborů. Jedním z nich je i odvětví energetiky, které je zároveň i kritickou infrastrukturou, a vzniklé poruchy tedy mohou mít dopad na úrovni státu. V tomto článku je rozebrán polygon přenosové soustavy České republiky, který obsahuje všechny klíčové prvky pro simulační, testovací a výukové účely. Tento polygon simuluje datovou část přenosové soustavy s možností simulace kritických stavů, které není možné provádět v reálném prostředí. Článek dále popisuje dva scénáře simulující jak standardní datový provoz přenosové soustavy, tak i scénář útoku na integritu rozvodné sítě. Navazující práce na polygonu bude zaměřena na rozšíření scénářů a dosažení přesnějších simulovaných hodnot.

PODĚKOVÁNÍ

Popsaná práce je součástí grantového projektu registrovaného pod č. VI20192022132, který je financován Ministerstvem vnitra České republiky.

REFERENCE

- [1] CONTI, M; DONADEL, D; TURRIN, F. A Survey on Industrial Control System Testbeds and Datasets for Security Research [cit. 23.03.2021]. Dostupné z: https://arxiv.org/pdf/2102.05631.pdf
- [2] PNE 18 4310:2010. Standardizované informační soubory dispečerských řídicích systémů. Praha: ÚJV Řež, a.s., 2010, 61 s [cit. 15.02.2021].
- [3] Kodex přenosové soustavy. Část V. Bezpečnost provozu a kvalita na úrovni PS. 2020-06-22. ČEPS, a.s., 2020, 17 s [cit. 15.02.2021]. Dostupné z: https://www.ceps.cz/cs/kodex-ps
- [4] IEC 60076-7:2018. Power transformers Part 7: Loading guide for mineral-oil-immersed power transformers. 2018-01-12, 58 s [cit. 15.02.2021].
- [5] DESARNAUD, G. *Cyber attacks and energy infrastructures*. [online] Copyright © 2017 [cit. 15.02.2021]. Dostupné z: https://lurl.cz/DzxmO
- [6] SHAKARIAN, P. Stuxnet: Cyberwar Revolution in Miliraty affairs. [online] Copyright © 2011 [cit. 15.02.2021]. Dostupné z: https://apps.dtic.mil/sti/pdfs/ADA546439.pdf

PROPOSAL OF CYBER THREAT DETECTOR USING RASPBERRY PI

David Hirš

Master Degree Programme (2), FEEC BUT E-mail: xhirsd00@stud.feec.vutbr.cz

Supervised by: Zdeněk Martinásek E-mail: martinasek@feec.vutbr.cz

Abstract: Nowadays, the number of discovered vulnerabilities increases rapidly. In 2019, the 20, 362 vulnerabilities were discovered. Therefore, the probability of cyber-attacks realization and their economic impact are real. Currently, it is necessary to propose and implement automated and low-cost Intrusion Prevention Systems (IPS) that is applicable for home use or small corporate networks. The main goal of the system is to mitigate cyber-attack impact as fast as possible. In this article, we propose IPS based on Raspberry Pi that can detect and prevent many various cyber-attacks.

Keywords: Raspberry Pi, IDS, IPS, Suricata, Kismet, arpwatch

1 ÚVOD

Přínos výpočetní techniky pro společnost je bezpochyby nepopiratelný, avšak přináší i rizika v podobě kybernetických útoků. Tyto útoky mohou cílit na odepření služeb, odposlouchávání dat, jejich úprava, nebo i ovládnutí uživatelova zařízení. Tato skutečnost představuje velkou hrozbu pro zařízení komunikující uvnitř sítě. Ať už se jedná o domácí, pracovní či veřejnou síť. Žádoucím je ochránit svá aktiva před zneužitím. Současná situace zaznamenala růst rizik nejen ze zdravotního hlediska, ale i toho kybernetického [1]. Vzrůstá tedy potřeba chránit svá aktiva, avšak použití běžných prostředků je v některých případech nedostačující. Při pohledu na automatizované spuštění kybernetických útoků je vhodné přikročit i k systémům automatizované detekce hrozeb.

Článek neobsahuje popis útoků a možnosti jejich mitigace, avšak o této problematice pojednává publikace [2]. Hlavním přínosem článku je teoretický návrh nízkonákladového detektoru kybernetických útoků. Zařízení Raspberry Pi 4 typ B bylo vybráno jako vhodné pro implementaci detektoru. Zaměření detektoru je závislé na použitých nástrojích a je téměř všestranné. Vhodnými nástroji pro realizaci detektoru jsou IDS (Intrusion Detection System) a IPS (Intrusion Prevention System) zařízení [3]. Detekce kybernetických útoků probíhá na základě anomálií či signatur ze zachyceného síť ového provozu [4]. Čtenář získá náhled na problematiku tvorby detektoru zaměřeného na útoky řazené do vrstev L2 až L3 modelu OSI/ISO. Detektor musí poskytnout dostatečnou ochranu před možnými hrozbami tak, jak si uživatel zvolí a nastaví.

2 VÝZKUM DOSTUPNOSTI NÁSTROJŮ PRO NÍZKO NÁKLADOVOU SONDU

Tato sekce popisuje IDS/IPS systémy, které byly otestovány a vybrány jako vhodné pro splnění cílů nízkonákladového detektoru. Jedná se o nástroje: **Arpwatch** - sledování ARP dotazů a zjištění nových zařízení v síti, **Kismet** - IDS systém pro bezdrátové komunikace a **Suricata/Snort** - systémy nabízí funkce IDS i IPS pro přímo připojená zařízení.

Systémy Suricata a Snort byly testovány na detekci útoku záplavou ICMP paketů. Testování hardwarového vytížení systémy Suricata a Snort probíhalo při monitorování sítě a zachycení útočníkem generovaného útoku. Oba systémy pracovaly v režimu IPS. Využitá pravidla pro monitorování sítě byla volně poskytována samotnými organizacemi. Na základě těchto výsledků a subjektivního ohodnocení uživatelské přívětivosti byl vybrán systém Suricata verze 6.0.0.0. Následně byl testován IDS systém Kismet při monitorování bezdrátové sítě a oznamování zjištěných změn. Dále byl testován nástroj Arpwatch. Tento nástroj oznámil připojení nových zařízení do sítě a odhalil probíhající útok ARP spoofing. Nelze opomenout možnost užití vlastních skriptů psaných v programovacím jazyce Python. Programovací jazyk Python byl vybrán na základě obliby širokou veřejností a pozitivního ohlasu při jeho užití. Vlastní skripty představují nutný doplněk celého detektoru. Nedílnou součást představuje menu, které uživatele provede kompletním nastavením, více v kapitole4. Během testování použitelnosti jednotlivých systémů byly zaznamenány nároky na procesor (CPU) a operační paměť (RAM) zařízení Raspberry Pi. Výsledky měření těchto dvou systémů obsahuje Tabulka 1, kde jsou zapsaný hodnoty i zbylých vybraných systémů. Hodnoty v tabulce představují průměr z naměřených hodnot linuxovým softwarem *top*.

		J
Systém a verze	Vytížení CPU [%]	Vytížení RAM [%]
Suricata 6.0.0.0	23,7	1,2
Snort 2.9.7.0 GRE	28,3	2,7
arpwatch 2.1a15	0,7	1,3
KISMET 2020-00-GIT	4,1	0,5
KISMET chromium web browser	42,9	5,4

Table 1: Hardwarové nároky vybraných systémů

3 REALIZACE EXPERIMENTÁLNÍHO PRACOVIŠTĚ

Nízkonákladový detektor bude realizován na zařízení Raspberry Pi 4 typ B. Jedná se o zařízení drobných rozměrů na které budou instalovány vybrané IDS/IPS systémy. Monitorovaný provoz zde bude vyhodnocován a na základě vyhodnocení budou provedeny příslušné akce. Uvnitř topologie se také nachází přístupový bod Mikrotik, ze kterého bude zrcadlený veškerý provoz na detektor. Zrcadlení provozu je zde nastaveno z důvodu, kdy bude sám Mikrotik vystaven útokům a i tato situace musí být monitorována. Diagram experimentálního pracoviště zobrazuje Obrázek 1. Zde je naznačen směr komunikace mezi jednotlivými uzly realizované topologie a jejich IP adresy. V diagramu jsou stanice útočníka i uživatele vyobrazeny jako dva odlišné stroje, avšak jedná se o dva virtuální stroje uvnitř jednoho fyzického zařízení. Raspberry Pi disponuje Wi-Fi přijímačem, který dovoluje bezdrátové připojení k zařízení Mikrotik. Doplněním popisu realizovaného experimentálního pracoviště je fotka reálného vzhledu pracoviště, viz Obrázek 1. Jak lze z fotky určit, Raspberry Pi je rozšířeno o USB síť ový adaptér, skrze který je spojený jedním metalickým kabelem k přístupovému bodu Mikrotik. Rozšíření bylo přidáno z důvodu, kdy by bylo nutné připojit Raspberry Pi zároveň skrze metalický kabel i Wi-Fi. Mikrotik je dále spojený dalším metalickým kabelem k osobnímu počítači, na kterém je spuštěný systém útočníka Kali Linux, spolu s uživatelem Ubuntu Linux.

4 NÁVRH PROGRAMOVÉHO VYBAVENÍ DETEKTORU

Navržený detektor umožní uživateli vybrat použité systémy, funkce a zabezpečovanou síť. Významnou roli bude zastupovat IPS systém Surikata, který bude zaměřen na metalické spojení v síti. Uživatel by měl být schopný vybrat soubor pravidel, který Surikata bude následovat. Samozřejmostí je výběr rozhraní, na kterém bude Surikata naslouchat a definice dalších parametrů, se kterými systém Surikata může být spuštěn. Pro bezdrátové sítě bude sloužit jako detekční mechanismus systém Kismet. Umožněný musí být i výběr pravidel, které nesmí být porušeny komunikujícími zařízeními. Pro komunikaci na druhé vrstvě ISO/OSI bude sloužit nástroj Arpwatch. Konkrétně bude zaměřený na ARP protokol a jeho užití v síti. Situaci, kdy nebudou využité systémy dostatečné, bude pokrývat



Figure 1: Realizované experimentální pracoviště.

tvorba vlastního skriptu, který bude zaměřený na daný incident. V případě většího počtu skriptů je nutné umožnit jejich výběr. Očekávaným výsledkem tedy bude jeden program, který bude ovládat zmíněné systémy. Reakce může probíhat na základě výpisu do logu od IDS systémů či výpisu do konzole. Grafické rozhraní není předpokládáno, což znamená ovládání jen z příkazové řádky. Tato skutečnost však nesmí záporně ovlivnit uživatelskou přívětivost programu. Veškerá menu a tištěné texty musí uživatele provést celým nastavením tak, aby výsledkem byl IPS systém dle jeho představ a požadavků. Po spuštění programu dojde k vytištění uvítacího textu a následně úvodního menu.

Toto menu obsahuje list dostupných systémů a skriptů, které uživatel může použít. Tištění textu do terminálu je rozdělené do dvou procesů kvůli situacím, které vrátí uživatele na začátek programu, ale již není žádoucí tisknout uvítání. Program čeká na interakci uživatele, který nyní může ukončit program, zvolit systém/skript a jejich nastavení, výpis aktivních systémů a skriptů, uvést do režimu Stand-By a nechat detektor pracovat. Uvedení programu do Stand-By vypíše aktivní systémy a skripty, následně čeká na ukončení celého programu od uživatele. Program v první fázi vypne jednotlivé systémy či skripty a následně ukončí sám sebe. Volba konkrétního systému či skriptu dále vede k vytištění informací o očekávaných parametrech a jejich hodnotách. Volba parametru následně poskytuje prostor pro zadání hodnoty vybranému parametru. Po zadání hodnoty dojde k opětovnému vytištění dostupných parametrů. Tímto způsobem uživatel definuje tolik parametrů, kolik bude potřebovat. Pokud jsou některé parametry vyžadované či chybně zadané, bude o této skutečnosti informován. Tato konfigurace bude obdobná každému z dostupných systémů i skriptů. Bude-li uživatel se spuštěnými systémy/skripty spokojený, musí se vrátit o krok zpět do úvodního menu. V tomto menu vybere uvedení programu do režimu Stand-By a nechá program pracovat tak dlouho, dokud uzná za vhodné. Vhodnou funkcionalitou se jeví možnost importu a exportu vybraných systémů, spolu s jejich nastavením. Grafický návrh programu pro nízkonákladový detektor představuje Obrázek 2.

5 ZÁVĚR

Efektivní obranu proti kybernetickým útokům nelze realizovat manuálně. Záměrem tohoto článku bylo popsat programové vybavení vhodné pro realizaci detektoru kybernetických hrozeb. Detektor není nutné vybudovat na vysoce výkonném zařízení. Dostačující je i rozměrově menší a výpočetně omezené zařízení, jakým je například Raspberry Pi 4 B. Takto realizovaný detektor je možné snadno přenášet a umístit jej takřka kamkoliv. Hlavním přínosem však je návrh samotného programu, který automatizuje proces nastavení a spuštění všech prvků detektoru. Praktická realizace navrženého programu pro detektor je do budoucna předpokládána, avšak nyní se jedná pouze o teoretický návrh.



Figure 2: Diagram výsledného programu pro nízkonákladový detektor.

PODĚKOVÁNÍ

Výzkum byl podpořen projektem MVČR s reg.č. VI20192022149.

REFERENCES

- LALLIE, H. S., SHEPHERD, L. A., NURSE, J. R.C., EROLA, A., EPIPHANIOU, G., MAPLE, C., BELLEKENS, X.: Cyber Security in the Age of COVID-19: A Timeline and Analysis of Cyber-Crime and Cyber-Attacks during the Pandemic. Computers & Security, 2021.
- [2] HIRŠ, D., MARTINÁSEK, Z.: Přehled kybernetických útoků na linkové a transportnívrstvě. Elektrorevue - Internetový časopis (http://www.elektrorevue.cz), 2020, roč. 22, č. 1, s. 1-15. ISSN: 1213-1539
- [3] SCARFONE, K., MELL, P.: Guide to Intrusion Detection and Prevention Systems (IDPS). In: NIST Special Publication (SP), 2012, 800-94 [online]. Gaithersburg, MD: National Institute of Standards and Technology, p. 1-111.
- [4] NASEER, S., SALEEM, Y., KHALID, S., BASHIR M., HAN J., IQBAL, M., HAN, K.: Enhanced Network Anomaly Detection Based on Deep Neural Networks. IEEE Access [online]. 2018. 6, 48231-48246.

SUITABILITY OF THE LORAWAN TECHNOLOGY FOR MASSIVE MACHINE-TYPE COMMUNICATION

Petr Mašek

Master Degree Programme (2nd year), FEEC BUT E-mail: xmasek17@vutbr.cz

> Supervised by: Pavel Mašek E-mail: masekpavel@vutbr.cz

Abstract: A remarkable progress in the LPWAN (Low Power Wide Area Network) technologies over the recent years opens new opportunities for developing versatile massive IoT (Internet of Things) applications. In this paper, we focus on one of the most popular LPWAN solutions operating in the license-exempt frequency bands, named LoRaWAN (Long Range Wide Area Network). The key contribution of this work is the implementation of the LoRaWAN technology within the Network Simulator 3. The collected simulation results reveal the expected spectrum utilization within the unlicensed band as well as show the transmission success rate for different communication distances between end-devices and LoRaWAN gateways.

Keywords: Industry 4.0, Internet of Things, LPWAN, LoRaWAN, massive MTC

1 ÚVOD

Princip IoT (Internet of Things) je založen na komunikaci M2M (Machine to Machine), která spočívá v předávání dat mezi stroji. V rámci Průmyslu 4.0 zavádějí průmyslové podniky z ekonomických důvodů automatizované postupy a uvádějí do provozu stoje, které spolu vzájemně komunikují. Může se jednat o stovky až tisíce připojených zařízení. Takový způsob komunikace se označuje jako mMTC (Massive Machine-Type Communication). Praktickým příkladem jsou zařízení zajišť ující dálkový odečet energií z domácností (elektřina, voda, plyn) [1].

Z důvodů finanční nákladnosti uvedení velkého množství "chytrých" zařízení do provozu se nejprve přistupuje k použití repliky sítě ve virtuálním prostředí, která zohledňuje propagaci signálu prostředím (zastavěná či nezastavěná lokalita). Tato replika bývá označována jako digitální dvojče a simulační výsledky bývají používány jako vstupy pro reálné nasazení. Pro co nejpřesnější predikci chování sítě s mnoha připojenými zařízeními lze využít vhodné simulační scénáře, které implementují komunikační parametry uvažovaných síť ových technologií. Očekává se, že do roku 2023 bude v provozu 1,9 miliardy připojených zařízení využívající technologie LPWAN (Low Power Wide Area Network) [2, 3, 4].

Na základě predikce rostoucího počtu "chytrých" zařízení se tento článek zaměřuje na komunikační technologii LoRaWAN a vyhodnocení simulačních scénářů pro mMTC.

2 KOMUNIKAČNÍ TECHNOLOGIE LORAWAN PRO IOT

Pro reálné nasazení IoT zařízení do provozu je vhodné využít technologie LPWAN. Jedná se o sítě s rozšířeným komunikačním dosahem a nízkou spotřebou energie, což je vhodné pro bateriově napájené zařízení. Mezi zástupce LPWAN technologií, které jsou dostupné i v České republice, se řadí technologie LoRaWAN (Long Range Wide Area Network), Sigfox a NB-IoT (Narrowband IoT) [5]. Každá z těchto technologií využívá různé komunikační parametry a je vhodná pro odlišné komunikační scénáře. Infrastruktura sítě LoRaWAN se skládá z: (i) *koncových zařízení*, která mohou být bateriově napájená (očekává životnost baterie je 5 – 10 let v závislosti na komunikačním scénáři), (ii) *bran*, se kterými komunikují koncová zařízení pomocí bezdrátové LoRa (Long Range) modulace, která využívá techniku rozprostřeného spektra CSS (Chirp Spread Spectrum), (iii) *síť ového serveru*, který přijímá informace od bran pomocí libovolného komunikačního protokolu (například Ethernet nebo 4G) a (iv) *aplikačního serveru*, který poskytuje uživateli požadovaná data. V České republice technologii LoRaWAN provozuje společnost České radiokomunikace [3, 6].

Mezi hlavní výhody technologie LoRaWAN lze zařadit nízkou cenu modulů $(5-12 \text{ EUR}^1)$ nebo bezplatné využití nelicencovaného spektra (úrovně pod 1 GHz). Oproti ostatním dostupným LPWAN technologiím v České republice disponuje také možností vytvoření vlastních privátních LoRaWAN sítí, které jsou tak nezávislé na třetí straně. Toho lze využít v průmyslu pro pokrytí konkrétního území, které je pod plnou kontrolou provozovatele [7]. Nevýhody plynoucí z využití nelicencovaného spektra jsou například omezení vysílacího výkonu (14 dBm = 25 mW) a dodržení střídy (1%) dle požadavků Českého telekomunikačního úřadu² pro ISM (Industrial, Scientific, Medical) pásmo 868 MHz nebo omezení maximální velikosti přenášených zpráv (243 B). Stručné porovnání technologií LoRaWAN, Sigfox a NB-IoT je uvedeno v Tabulce 1 [8].

Tabulka 1: Klíčové parametry LPWAN technologií LoRaWAN (Specification v1.1), Sigfox a NB-IoT
(3GPP Release 13) pro komunikaci ve směru Uplink [8].

	LoRaWAN	Sigfox	NB-IoT (NB1)
Spektrum	Nelicencované	Nelicencované	Licencované (LTE)
Střída (duty cycle)	1 %	1 %	Není omezeno
Frekvenční pásmo	ISM 868 MHz	ISM 868 MHz	LTE 800 MHz
Maximální vysílací výkon (ERP)	$14 \mathrm{dBm} = 25 \mathrm{mW}$	$14 \mathrm{dBm} = 25 \mathrm{mW}$	$23 \mathrm{dBm} = 200 \mathrm{mW}$
Šířka pásma	125 kHz, 250 kHz	100 Hz	200 kHz
Přenosová rychlost	0,3-50 kb/s	<1 kb/s	0,3-32,25 kb/s
Maximální velikost zprávy	243 B	12 B	1600 B
Pokrytí (MCL)	157 dB	162 dB	164 dB
Privátní infrastruktura	Ano	Ne	Ne

3 SIMULAČNÍ SCÉNÁŘ PRO KOMUNIKACI V LORAWAN SÍTI

Pro provedení simulací bylo využito simulační prostření NS-3 (Network Simulator 3) ve verzi 3.30.1. Do tohoto prostředí byl integrován LoRaWAN modul *signetlabdei*³ a byl vytvořen komunikační scénář, jehož topologie je zobrazena na Obrázku 1. Pro spuštění simulace lze volit: (i) počet koncových zařízení, (ii) počet bran, (iii) poloměr komunikační oblasti, (iv) čas simulace, (v) periodu zasílání zpráv a (vi) velikost užitečných dat. Dále je možné nastavit zařízením náhodnou nebo konkrétní polohu v simulované oblasti. Aby koncová zařízení dodržela požadavky na velikost střídy, je nutné nastavit vhodnou velikost užitečných dat a periodu zasílání zpráv.

Po provedení simulace lze v terminálovém okně vyčíst pomocí logování konkrétní komunikační parametry, které byly v simulaci použity. Lze tak vyčíst: (i) faktor rozprostření, (ii) použitou šířku pásma, (iii) nosnou frekvenci, (iv) přenosovou rychlost, (v) dobu přenosu zprávy, (vi) vysílací výkon, (vii) přijímaný výkon, (viii) vzdálenost koncových zařízení k nejbližší bráně nebo (ix) počet vyslaných a v pořádku doručených zpráv. Zaznamenaná data z provedené simulace pro jedno koncové zařízení jsou uvedena v Tabulce 2.

¹Cena modulu Microchip RN2483A k březnu 2021 je přibližně 11 EUR. Viz: https://bit.ly/RN2483A.

²Požadavky Českého telekomunikačního úřadu pro nelicencované spektrum: https://bit.ly/CTUomezeni. ³Modul dostupný z: https://github.com/signetlabdei/lorawan.

Obrázek 1: Rozmístění 100 koncových zařízení LoRaWAN, jedné komunikační brány a síťového serveru v komunikační oblasti s přímou viditelností LOS (Line of Sight) mezi jednotlivými uzly. **Tabulka 2:** Výsledky simulace pro koncové zařízení č. 1.



4 DOSAŽENÉ VÝSLEDKY V ZÁVISLOSTI NA VYUŽITÍ FREKVENČNÍHO SPEKTRA

V rámci simulací byla analyzována závislost úspěšnosti přenosu na základě rozdílnosti vybraných komunikačních parametrů. Pro každý z případů bylo realizováno 10 simulací a uvedené výsledky jsou tak průměrem hodnot z provedených simulací, viz Tabulka 3.

V simulacích je uvažováno se 100 "chytrými" koncovými zařízeními, která vysílají vždy zprávu s užitečnou zátěží (user data) o velikosti 23 B. Závislost byla zkoumána pro komunikační oblasti ve tvaru kruhu s poloměrem 1 km, 7 km a 10 km, kdy byla koncová zařízení rozmístěna náhodně po celé komunikační oblasti. Počet bran v komunikační oblasti se měnil od jedné do pěti. Pokud byla brána pouze jedna, byla umístěna ve středu komunikační oblasti. Pro případy s dvěma a více branami byly tyto brány rozmístěny pravidelně na kružnici s polovičním poloměrem, než je poloměr komunikační oblasti (0,5 km, 3,5 km a 5 km).

Perioda vysílání [s] Počet bran		Úspěšnost při poloměru 1 km	Úspěšnost při poloměru 7 km	Úspěšnost při poloměru 10 km	
	1	84,9 %	45,3 %	24,3 %	
20	3	95,1 %	84,0 %	59,7 %	
	5	97,5 %	97,5 %	81,3 %	
	1	98,3 %	88,3 %	62,7 %	
100	3	99,0 %	99,0 %	96,0 %	
	5	99,7 %	100,0 %	100,0 %	
	1	100,0 %	96,2 %	69,2 %	
165	3	100,0 %	100,0 %	98,4 %	
	5	100,0 %	100,0 %	100,0 %	

Tabulka 3: Úspěšnost přenosu pro zvolené periody vysílání a počet bran. Hodnoty s úspěšností přenosu pod 70 % jsou zvýrazněny (hranice na základě spolupráce s výrobci LoRaWAN zařízení).

Simulace byla provedena pro tři různé hodnoty periody vysílání: (i) 20 s, (ii) 100 s a (iii) 165 s. Pouze v případě použití periody vysílání s hodnotou 165 s a vyšší lze garantovat, že všechna koncová zařízení při vysílání zprávy o užitečné zátěži 23 B dodrží hodnotu střídy nepřevyšující 1 %, což je vyžadováno telekomunikačními předpisy. V opačném případě mohou koncová zařízení s nastaveným vyšším faktorem rozprostření tyto požadavky porušovat. Výsledky naměřených úspěšností přenosu jsou uvedeny v Tabulce 3.

5 ZÁVĚR

V tomto článku byla popsána problematika technologie LoRaWAN pro využitelnost v IoT sítích, kdy byly zmíněny její hlavní komunikační parametry a byly uvedeny hlavní výhody a nevýhody. V praktické části byl do simulačního prostředí NS-3 integrován LoRaWAN modul a následně byly provedeny simulace zabývající se problematikou úspěšnosti přenosu v síti LoRaWAN v závislosti na použitých komunikačních parametrech.

Naměřená úspěšnost přenosu roste se zvětšující se periodou vysílání, s vyšším počtem bran a s menším poloměrem komunikační oblasti. Pro dosažení lepších výsledků je vhodné použít v LoRaWAN síti větší počet vhodně rozmístěných bran, což způsobí menší vzdálenosti mezi koncovými zařízeními a branami. Koncová zařízení tak budou smět komunikovat s nižším faktorem rozprostření (z rozsahu 7-12), což umožní i rychlejší komunikaci. Pro dodržení požadavků na velikost střídy je vhodné použít větší periodu vysílání, což lze však kompenzovat použitím menší velikosti vysílané zprávy. V simulaci byla použita fixní velikost zprávy o velikosti 32 B, pro ostatní scénáře je však možné velikost zprávy navýšit až na 243 B ve směru Uplink.

Naměřené výsledky je možné do budoucna porovnat s hodnotami získanými měřením v reálně nasazených sítích (veřejných či privátních) a následně zhodnotit přesnost výsledků z implementovaných simulačních scénářů.

REFERENCE

- A. Lavric and V. Popa, "Performance Evaluation of LoRaWAN Communication Scalability in Large-Scale Wireless Sensor Networks," *Wireless Communications and Mobile Computing*, vol. 2018, p. 6730719, Jun 2018. [Online]. Available: https://doi.org/10.1155/2018/6730719
- [2] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges," *IEEE access*, vol. 6, pp. 3619–3647, 2017.
- [3] D. Magrin, M. Capuzzo, and A. Zanella, "A Thorough Study of LoRaWAN Performance Under Different Parameter Settings," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 116–127, 2020.
- [4] Cisco, "Cisco Annual Internet Report (2018–2023)," White Paper, 2020.
- [5] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, and M. Pettissalo, "On the Coverage of LPWANs: Range Evaluation and Channel Attenuation Model for LoRa Technology," in 2015 14th International Conference on ITS Telecommunications (ITST). IEEE, 2015, pp. 55–59.
- [6] R. El Chall, S. Lahoud, and M. El Helou, "LoRaWAN Network: Radio Propagation Models and Performance Evaluation in Various Environments in Lebanon," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2366–2378, 2019.
- [7] P. Masek, M. Stusek, K. Zeman, J. Hosek, K. Mikhaylov, S. Andreev, Y. Koucheryavy, O. Zeman, J. Votapek, and M. Roubicek, "Tailoring NB-IoT for Mass Market Applications: A Mobile Operator's Perspective," in 2018 IEEE Globecom Workshops (GC Wkshps), 2018, pp. 1–7.
- [8] M. Stusek, D. Moltchanov, P. Masek, K. Mikhaylov, O. Zeman, M. Roubicek, Y. Koucheryavy, and J. Hosek, "Accuracy Assessment and Cross-Validation of LPWAN Propagation Models in Urban Scenarios," *IEEE Access*, vol. 8, pp. 154 625–154 636, 2020.

IMPACT OF CYBER SECURITY ON DATA VOLUME IN SMART METERING

David Kohout

Master Degree Programme (2), FEEC BUT E-mail: xkohou16@vutbr.cz

> Supervised by: Petr Mlýnek E-mail: mlynek@feec.vutbr.cz

Abstract: Smart metering is a new way of monitoring power grid and it can make billing of electricity consumption much easier. These systems can be marked as elements of critical infrastructure, so it needs to be protected. For cost-effective communication with smart meters, we need to find compromise between frequency of reading, data volume and cyber security.

Keywords: AMM, Communication, DLMS, Security, Smart Grid, Smart Metering

1 ÚVOD

S příchodem 5G (sítě páté generace) a IoT (internet věcí) dochází k stále většímu propojení zařízení do jedné sítě a společně tvoří chytrou infrastrukturu. S tímto trendem se nezaostává ani v případě měření energií. To lze označit pojmem Smart Metering, který využívá chytrých meterů k měření a k dálkovému sběru měřených dat. Mezi energie, které lze tímto způsobem měřit, patří především elektřina, avšak je možné použít tyto technologie i pro měření spotřeby vody, plynu, topení atd.

Nutnost nasazovat chytré elektroměry nad určitou roční spotřebu je již vázána legislativou. Právě díky této povinnosti bude implementace elektroměrů selektivní a tím se jeví bezdrátové přenosové technologie jako ideální, jedná se především o NB-IoT (Narrow Band IoT – úzkopásmová technologie) či LTE Cat. M1 (Long Term Evolution – širokopásmová technologie). Nově jsou také přílohou č. 4 vyhlášky č. 359/2020 Sb. určeny bezpečnostní požadavky na chytré elektroměry. Kombinace zmíněných přenosových technologií a daných bezpečnostních požadavků přináší nové výzvy a výzkumné oblasti v nalezení kompromisu mezi objemem dat, četností přenosu, limitů přenosových technologií, finančními náklady a bezpečnostními požadavky. Článek se zaměřuje na tuto problematiku a popisuje výzkum a vývoj DLMS generátoru s bezpečnostními algoritmy dle DLMS Security Suite pro možnosti testování a hledaní kompromisů a limitů.

2 SMART METERING

Systém inteligentního měření (Smart Metering, SM) a řízení odběru elektrické energie je soustava vzájemně propojených zařízení, která vzájemnou spoluprací umožňují dálkové odečty odběru elektrické energie, dálkové vypínání a zapínání odběru, dálkové řízení kritických stavů v rozvodné síti NN (nízkého napětí), či dálkové a inteligentní řízení spotřeby a dodávky v odběrném místě.

Mezi technologie pro SM patří i protokol DLMS (Device Language Messege Specification), který má do budoucna největší potenciál k hromadnému využití v Smart Grid sítích. SM může sloužit také zákazníkovi, který má možnost zjistit informace o vlastní spotřebě v reálném čase, a tak může upravit svůj styl spotřeby pro efektivnější využití energie (například maximalizace spotřeby v nízkém tarifu). Vzhledem k možnostem technologií pro SM lze považovat chytré elektroměry jako prvky kritické infrastruktury (přepínání tarifů a odpojování zákazníků na dálku) a z tohoto důvodu je nutné zabezpečit jejich provoz. Dále se na přenášené datové odečty od zákazníků může vztahovat GDPR.

2.1 SMART METERING V ČR

V České Republice existuje již několik projektů, kde dochází k pilotnímu nasazování Smart Metering technologií. Mezi tyto projekty patří: Smart region Vrchlabí [1], SMARAGD [2]. Rozšíření zavádění SM technologií má také napomoci vyhláška č. 359/2020 Sb. [3], která nařizuje distributorům od roku 2024 nasazovat chytré elektroměry na všechna odběrná místa se spotřebou vyšší než 6 MWh. Do této spotřeby patří zejména větší odběratelé, ale také domácnosti vytápěné elektřinou. Přibližně se tak jedná o 1 milion odběrných míst. Pro přenos dat z chytrých elektroměrů se v ČR uvažuje nasazení protokolu DLMS, který dle DLMS Security Suite naplňuje zmíněnou vyhláškou požadované bezpečnostní požadavky.

3 DLMS A MOŽNOSTI ZABEZPEČENÍ

DLMS je specifikace určená pro použití u chytrých meterů. Popisuje model rozhraní a komunikační protokoly pro výměnu dat mezi měřícími zařízeními. Specifikace je rozdělena do 4 barevně označených knih, které se označují jako Coloured Books. Nejdůležitější je zelená [4], která obsahuje základ DLMS, zabezpečení a provázání s přenosovými technologiemi. Modrá kniha [5] popisuje tvorbu objektů pro ukládání a zasílání dat. Zbylé knihy jsou žlutá (testování) a bílá (slovník pojmů).

Autentizace

V průběhu navázání spojení v DLMS může docházet ke 3 úrovním autentizace. První úroveň je bez zabezpečení, v druhé dochází pouze k zaslání hesla v otevřené podobě a třetí, tedy nejvyšší autentizační úroveň se označuje jako High Level Security (HLS). Její princip je založen na modelu výzva-odpověď, kde k úspěšnému navázání spojení dojde k zaslání 4 zpráv. Tato úroveň lze použít v kombinaci s mechanizmy, jako jsou MD5, SHA-1, SHA-2, GMAC, SHA-256 a EC-DSA.

Šifrování

Navázání spojení je v otevřené podobě, poté je již možné využít šifrovaného přenosu datové části. Důvěrnost dat je zabezpečena pomocí algoritmu AES (Advanced Encryption Standard). Dohodnutí šifrovaného spojení, případně výměna klíčů probíhá společně s autentizací. Veškeré použité algoritmy a velikosti klíčů jsou specifikované v sadách, které se označují jako Security Suites. Aktuálně jsou definovány pouze první tři z celkově 8 možných sad. Všechny parametry sad jsou uvedeny v následující tabulce 1.

		·					
	ID	Označení	Šifrování	Digitální podpis	Dohodnutí klíče*	Hash	Komprese
	0	AES-GCM-128	AES-GCM-128	-	-	-	-
1	1	ECDH-ECDSA-AES-		ECDSA	ECDH	SHA-	V 44
	T	GCM-128-SHA-256	AES-GCIVI-128	s P-256	s P-256	256	V.44
2	n	ECDH-ECDSA-AES-	AES COM DES	ECDSA	ECDH	SHA-	V 44
	2	GCM-256-SHA-384		s P-384	s P-256	384	V.44
	3-7	Rezervováno pro bu	doucí použití		*Klíč je přenesen po	omocí A	ES-128/256

Tabulka 1: Security suites [4]

4 MĚŘENÍ VLIVU ZABEZPEČENÍ NA DATOVÝ OBJEM

Vzhledem k množství chytrých elektroměrů, které se do budoucna budou instalovat, je nutné určit, jak velký nárůst datového objemu nastane při využití vyšších úrovní zabezpečení. Zároveň lze předpokládat, že použité přenosové technologie, budou mít omezenou datovou propustnost a stabilitu. S určením celkového datového objemu souvisejí také náklady za přenesená data.

4.1 VUT DLMS TESTER

Pro testování datových objemů je na VUT vyvíjena aplikace pro testování DLMS [6]. Tato aplikace se skládá ze dvou částí – klient (aplikace pro odečítání dat) a server (emulovaný elektroměr). Díky vlastní aplikaci lze spolehlivěji kontrolovat a nastavovat přenos, dále aplikace není vázána na určité
technologie a tak není omezující ani na výpočetní výkon. Pro ukázku obrázek 1 zobrazuje hlavní okno aplikace a obrázek 2 ukazuje manuálního klienta pro ruční vyčítání požadovaných objektů. Aplikace je stále ve vývoji a využívá se na testování DLMS komunikace s reálnými elektroměry.



Obrázek 1: Hlavní okno aplikace

Obrázek 2: Manuální klient

4.2 NASTAVENÍ MĚŘENÍ

Pro měření byla použita aplikace VUT DLMS Tester, viz kapitola 4.1. Pro stranu elektroměru bylo využito emulovaného elektroměru za pomoci serveru, který je vyvíjen společně s klientskou aplikací. Měření probíhalo s využitím tří různých úrovní přístupu. První úroveň je nezabezpečená, druhá používá HLS autentizaci a poslední navíc přidává šifrování dat s využitím SecuritySuite 0. Vyčítané objekty odpovídají požadovaným objektům, které jsou distributory plánované pro dálkový sběr v rámci AMM (Automatic Meter Management). Jedná se o 18 objektů typu Register (3 atributy) a o 3 objekty typu DemandRegister (9 atributů). Celkově 21 objektů s 81 atributy.

4.3 POROVNÁNÍ OBJEMŮ

V následujících grafech jsou vždy porovnány části komunikací s využitím nezabezpečeného spojení (None), s vysokou autentizací HLS (High) a s šifrováním pomocí Security Suite 0 (Encrypted). Získaná data pochází ze zachycené komunikace programem Wireshark, který umožňuje zjistit velikost celé TCP relace a režie DLMS protokolu.

V grafu na obrázku č. 3 je porovnán datový objem v bajtech, kde na levé straně jsou uvedeny velikosti pouze z navázání a ukončení spojení. Sloupce na pravé straně uvádějí datovou velikost pouze ze čtení 1 objektu typu Register. Z grafu je patrné, že největší nárůst dat je v samotném navázání spojení, kde šifrované spojení má přibližně 45% nárůst datového objemu. Při samotném čtení je 15% nárůst.

Na grafu v obrázku č. 4 jsou uvedeny datové objemy z celého vyčítání všech 21 objektů. Celý objem je rozdělen na jednotlivé složky spojení (navázání a ukončení spojení, datová část a data vycházející ze šifrování). Data navíc u šifrování tvoří 12 % z celého objemu dat a jedná se především o padding algoritmu AES a údaje zasílané navíc (označení odesílajícího zařízení – System-Title).

Poslední graf č. 5 uvádí další metody, kterými lze data z meteru vyčítat. První sloupec znázorňuje čtení všech atributů, kde každý atribut znamená jednu žádost a jednu odpověď. Další sloupec uvádí opět čtení všech dat, ale jejich čtení se provádí za pomoci listu. V jednom listu se nachází 10 atributů. Celé čtení 81 atributů je tak zmenšeno do 9 žádostí a odpovědí. Poslední 2 sloupce uvádí opět čtení zvlášť a za pomoci listu. Zde však nedochází ke čtení všech atributů, ale vyčítány jsou pouze důležité části. Celkově se tak jedná pouze o 42 atributů.



Obrázek 3: Porovnání objemů pro části spojení

Obrázek 4: Počet bajtů na celé čtení



Obrázek 5: Porovnání objemů při rozdílném typu čtení

5 ZÁVĚR

S novou vyhláškou a s nutností zavádět prvky Smart Meteringu, roste i potřeba zajištění kyberbezpečnosti pro chytré elektroměry a s tím související infrastruktury. Dále vzhledem k množství instalací narůstají datové nároky na přenosové technologie. V tomto článku je rozebrán protokol pro Smart Metering označovaný jako DLMS včetně jeho možností zabezpečení. S využitím vyvíjené aplikace bylo možné otestovat různé způsoby zabezpečení a tím určit datové nároky na bezpečnost. Ze získaných dat je patrné, že největší nároky má vytváření samotného spojení. Při dostatečně velkém čtení, případně při ponechání navázaného spojení, bude datový nárůst při použití SS0 přibližně 15 % oproti nešifrovanému přenosu. Navazující práce při vývoji aplikace pro testování DLMS bude zaměřena na testování pokročilejších bezpečnostních mechanizmů, tedy na sady Security Suite 1 a 2, které právě díky vyvinutému VUT DLMS Testeru budou jednoduše analyzovány vzhledem k datovým objemům.

REFERENCE

- [1] MOLEK, Tomáš. Smart region Vrchlabí. *oEnergetice* [online]. [cit. 2021-03-01]. Dostupné z: http://oenergetice.cz/elektrina/smart-region-vrchlabi-prvni-ceska-chytra-sit/
- [2] Projekt SMARAGD. EG.D [online]. [cit. 2021-03-01]. Dostupné z: https://www.egd.cz/smaragd
- [3] Vyhláška č. 359/2020 Sb. Dostupné také z: https://www.zakonyprolidi.cz/cs/2020-359
- [4] Green Book: DLMS/COSEM Architecture and Protocols. Ed10. DLMS User Association, 2020.
- [5] Blue Book: COSEM Interface Classes and OBIS. Ed14. DLMS User Association, 2020.
- [6] KOHOUT, David. *Zátěžový generátor zpráv DLMS/COSEM*. Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií, 2019.

PRIVACY-PRESERVING INSTANT MESSAGING APPLICATION

Štefan Krajanec

Master, FEKT BUT E-mail: xkraja00@feec.vutbr.cz

Supervised by: Petr Dzuenda E-mail: dzurenda@vutbr.cz

Abstract: Instant messaging applications noted significant grow especially in the last decade. In fact, the Internet communication is cheap and convenient way how to communicate with distant people. However, this grow of user communication and data exchange through the online world impacts user security and privacy, as it was also shown recently by WhatsApp privacy issues. Firstly this article evaluates security and privacy issues of current mobile messaging applications. Secondly, we design our basic open source solution with the focus on security, privacy, and user centric features. Furthermore, we provide proof-of-concept implementation of our system.

Keywords: Privacy, Security, Instant Messaging, Android, Metadata

1 ÚVOD

Mobilné aplikácie sú neodmyslitelnou súčasť ou života ľudí po celom svete. Pred pár rokmi bol odštartovaný trend IM (Instant Messaging) aplikácii, ktoré sú v súčastnosti masovo používané. Avšak, dodnes sa nekladie dostatočný dôraz na súkromie ich užívateľ ov. Súkromie ale nie je prioritou, spoločnosti hromadia a spracúvajú užívateľ ské dáta skôr za účelom vytvorenia profitu. Napríklad Facebook, dokáže na každom aktívnom užívateľ ovi zarobiť každý štvrť rok minimálne 10 \$ na cielených reklamách a táto suma sa zväčšuje každým štvrť trokom o 10 % [1].

Väčšina IM aplikácii v skutočnosti nejakým spôsobom šifruje správy užívateľov (často ale nie Endto-End), ale aktívne pracuje s ich metádatami. Tie sú často odosielané buď s obsahom správy alebo v pravidelných intervaloch [3]. Treba brať na vedomie fakt, že hoci v metadátach nie je prenášaný samotný obsah správy, ale ich vystavenie môže mať značný vplyv na súkromie užívateľa [4]. Napríklad, ak zamestnanec väčšiej organizácie (vláda, veľká spoločnosť) komunikuje s novinárom za účelom odhalenia určitej pofidérnej činnosti, je možné, na základe metadát túto komunikáciu odhaliť.

Tento článok predstavuje návrh IM systému zameraný na zaistenie ochrany súkromia jeho uživatelov a nemožnosti profilovania či sledovania. Navrhnutý systém je uživatelsky centralizovaný, uživatel má všetky svoje dáta pod kontrolou, tj. môže mazať svoje správy, nastavovať ich životnosť či obmedziť množstvo informácii odhalovaných posktovatelom služby. Celý projekt je cielený na open source riešenie, to znamená, že IM systém môže byť prevádzkovaný samotnými uživatelmi, ktorí sa nemusia spoliehať na nadkorporátne společnosti. Napríklad nie je tak dávno, kedy sa zaktualizovali podmienky používania služby Whatsapp od Facebooku a užívatelia v miliónoch túto sociálnu sieť začali opúšťať [2].

1.1 BEZPEČNOSTNÉ POŽIADAVKY NA IM APLIKACE

Porovnanie bezpečnosti našej aplikace Grobardo s najpoužívanejšou konkurenciou je prehladne naznačané v tabul'ke 1. Navrhnutý systém podporuje mnoho d'alších vlastností, ktoré umožňujú dosiahnuť vyššieho súkromia užívateľ ov a dávaju užívateľ om väčšiu kontolu nad ich dátami, napríklad: 1) nastavenie životnosti správ po ktorej sa úplne odstránia zo systému aj keď je kľúč platný, 2) limitovanie informácii poskytnutých oponentovi (napr. komunikovanie pod dočasným pseudonymom), 3) žiadosť o povolenie vykonať screenshot obrazovky, 4) nastavenie životnosti všetkým typom správ (počet otvorení, možnosť otvorenia na celú obrazovku), 5) odoslanie informácie o prečítaní správy oponentom. Navrhnuté riešenie umožňuje užívateľ om využívať vlastný server so správami a réžiou miestnosti, na ktorý sa aplikácia pripojí len jednoduchou zmenou endpointov. Všetky servery okrem autorizačného je možné naklonovať a rozbehnúť vlastný server aplikácie. Vď aka tomu je možné pri prvom spustení aplikácie určiť, ku ktorému serveru sa bude zariadenie pripájať, de facto na ktorom serveri budú uložené užívateľ ove kľúče. Ten, kto si stiahne, nakonfiguruje a spustí servre vytvorené v tejto práci, si po prihlásení do aplikácie vpíše cestu k svojim servrom.

	Whatsapp	Threema	Facebook	Signal	Telegram	Grobardo
Poskytovanie dát	000	nio	000	nia	ano	nio
tretím stranám	ano	IIIC	allo	me	ano	me
Uchovávanie a zber	020	nio	000	nio	000	nio
užívateľ ských dát	ano	Inc	allo	me	allo	me
Zálohovanie dešifrovacieho	áno	nio	áno	nie	áno	nie
kľ úča na serveri	ano	me				
Logovanie času a IP	nie	nie	áno	áno	nie	nie
Samozničujúce správy	áno	nie	áno	áno	áno	áno
Užívateľ ská	nio	nio	nio	000	nio	000
réžia kľúčov	IIIC	me	IIIC	ano	me	allo
Úplné odstránenie	nia	nia	nia		nia	0.00
odoslaných správ	me	me	me	allo	ille	ano
Zneprístupnenie	ána	nia	ána		nia	000
zariadení	ano	me	ano	ano	ille	ano
Réžia práv oponenta	nie	nie	nie	nie	nie	áno
Open source	nie	nie	nie	ano	ano	ano
Klonovateľ ný kód	nio	nio	nio	nia	nio	0 n 0
serverov s plnou réžiou	me	me	me	me	me	ano

 Tabulka 1:
 Porovnanie bezpečnosti aktuálne najvýznamnejších IM app.

2 ARCHITEKTURA NAVRHNUTÉHO SYSTÉMU

Navrhnutý IM systém je založený na inovatívnom návrhu databázových tabuliek. Základom je tabuľka miestností, v ktorej je najdôležitejším identifikátorom ID miestnosti a dalšie konfiguračné dáta. Ďalšia tabuľka obsahuje informáciu, aký užívateľ (userId) sa nachádza v ktorej miestosti (roomId). Posledná najdôležitejšia tabuľka obsahuje správy, kde každý riadok obsahuje šifrovanú správu, identifikátor miestností do ktorej správa patrí a ID užívateľa, ktorý správu odoslal plus ďalšie konfiguračné dáta. Základná architektúra tabuľkového systému a ich závisostí je zobrazená na obrázku 1. Návrh je multifunkčný, ale aktuálne je implementovaný len na android OS.

Systém je implementovaný pomocou troch klonovatelných open source serverov, ktoré zastrešujú správnu distribúciu kľúčov a ukladanie správ (kryptogramov). Medzi neklonovatelné servery patria tie, ktoré zabezpečujú autentizáciu užívateľ a a autorizáciu zariadení. Momentálne je používaný jeden volitelný server, vď aka ktorému sú na zariadenia správne doručované notifikácie v prípade novej správy alebo nového člena v miestnosti, rozširovaním funkcionalit v aplikácii bude narastať počet volitelných serverov.

Device keys							
userld	public key		deviceUnique				options
\downarrow		Cha	t me	ember	s table		
userld	room	roomId tempCon		onnector		RoomsTable	
				id	options	te	empConnector
Messages			_ √				
cryptogram	options	r	oom	ld			

Obrázek 1: Základná architekrúra tabuľ kového systému.

Vď aka návrhu distribúcie kľ účov, je každá správa šifrovaná symetrickým kľ účom a uložená na server do už spomínanej tabuľ ky. Správny kľ úč k danej správe majú len jej adresáti a odosielateľ. A do miestnosti sa nedostane užívateľ, ktorý do nej nebol pozvaný a ak by sa tam nejakým spôsobom dostal alebo by nejakým spôsobom dokázal zistiť, ktoré správy patria do miestnosti na ktorú útočí, tak bude vidieť len kryptogramy správ ku ktorým nemá kľ úč. Kľ úč má len zariadenie, ktoré bolo do miestnosti pozvané z vnútra. Princíp komunikácie medzi dvoma zariadeniami v miestnosti je zobrazený na obrázku 2.



Obrázek 2: Princip konvrzácie pomocou miestnosti.

2.1 IMPLEMENTÁCIA SYSTÉMU

Technológie, pomocou ktorých bol implementovaný proof-of-concept systém, ktorý obsahuje: Vývoj len na operačný systém Android s programovacím jazykom Kotlin Android 5.+, programovací jazyk na serveri Golang a príležitostne Python používajúc technológie GRPC - diaľkové volanie procedúr od Googlu, umožňuje reaktívny prístup komunikácie serveru s aplikáciou spolu s PostgreSQL - voľne šíriteľný objektovo-relačný databázový systém, komunikácia zariadení so serverom cez TOR sieť a kryptotechnologie: 1.) hashovacia funkcia - SHA-256 pre integritu dát, 2.) symetrická šifra - AES-256 pre šifrovanie dát pred serverom a 3.) RSA-2048 - asymetrická šifra - na distribúciu symetrických kľúčov. Obrázok 3 zobrazuje snímky obrazovky vytvorenej aplikácie pre instant messaging.



Obrázek 3: Screeny IM aplikácie (zľava): správy v miestnosti, správa kľúčov a dostupné chaty

3 ZÁVER

V článku sme predstavili návrh IM systému zvyšujúci ochranu súkromia užívateľov a ich kontrolu nad ich dátami. Podarilo se overiť funkčnosť teoretického návrhu systému a implementovať základnú funkcionalitu distribúcie kľúčov, odosielanie a prijímanie správ, vytvorenie a pripojenie do miestnosti, zobrazovanie a aktualizovanie správ, miestností a kľúčov. Aplikácia ešte neobsahuje všetky požadované funkcionality. Keď sa uzatvorí vývoj distribúcie kľúčov a šifrovania bude čas maximálne zredukovať citlivé dáta (id) v prepojených tabuľkách na šifrovanú prípadne zahashovanú podobu, tak aby nebolo možné bez kľúčových znalostí dopátrať adresátov správ. Ďalšími plánovanými krokmi je doimplementovať zvyšnú funkcionaitu ako odosielanie obrázkov, konfigurácia jednotlivých správ, miestností a užívateľov, a integrovanie ToR technológie pre anonymnú komunikáciu cez internet.

REFERENCE

- [1] *Facebook Q4 and full Y2020* https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx
- [2] *Is it time to leave WhatsApp* https://www.theguardian.com/technology/2021/jan/24/is-it-time-to-leave-whatsapp-and-is-signal-the-answer
- [3] 26th Symposium on Operating Systems Principles: A Distributed Metadata-Private Messaging System. s. 440. ISBN 978-1-4503-5085-3. https://dl.acm.org/citation.cfm?id=3132783
- [4] Stefan Schiffner. Privacy and Data Protection by Design from policy to engineering. *CoRR*. 2015, 79. Dostupné také z: http://arxiv.org/abs/1501.03726

TESTING POLYGON FOR QUANTUM KEY DISTRIBUTION

Jakub Širjov

Master Degree Programme (2), FEEC BUT E-mail: xsirjo00@stud.feec.vutbr.cz

Supervised by: Michal Látal E-mail: xlatal08@stud.feec.vutbr.cz

Abstract: The article explains issues in QKD and their sources. It next describes principle of QKD operation and its individual parts that are required for its function. Part of the article includes simulations from QKDNetsim and NS-3 that show possibilities of the simulation, that were used to create exemplary simulations for future check of QKD traffic on the test polygon.

Keywords: quantum, key, distribution, QKD, photon, simulation, NS-3, QKDNetsim

1 ÚVOD DO QKD

V dôsledku neustále prudko narastajúceho výpočtového výkonu sa museli aj šifrovacie metódy zlepšovať, aj keď sú založené na zložitých matematických operáciách a bolo ich už možné rozlúštiť v reálne krátkom čase. Kryptológia sa vyvíjala smermi, ako je šifrovanie chaosom, fraktálnym šifrovaním, kvantovými javmi alebo umelou inteligenciou. Práve kvantové javy sú využívané v kvantovej distribúcii kľúčov (Quantum Key Distribution, skrátene QKD), kde sa využívajú princípy kvantovej mechaniky. Práve QKD je technológia, o ktorej sa dá povedať, že je bezpečná voči útokom s neobmedzeným výpočtovým výkonom [1].

V QKD ako nositeľ informácie je používaná elementárna častica fotón. Technika pre generovanie kľúčov na kvantovej úrovni a ich distribúcia prostredníctvom fotónov sa označuje ako proces kvantovej distribúcie kľúčov. Ako je známe, existujú dve moderné základné metódy pre šifrovanie, a to symetrické a asymetrické. Práve QKD poskytuje prostriedky pre distribúciu symetrických kľúčov [2] [3] [4].

QKD je zobrazený na obr. 1. Alica pošle Bobovi sériu fotónov, kde každý je modulovaný náhodnou hodnotou (qubity). Príkladom je, že Alica pošle Bobovi kartu, kde je náhodne na jednej strane "1" a na druhej "0". Ak si Bob prečíta rovnakú stranu ako Alica, tak majú obaja rovnakú hodnotu, ale ak si Bob vyberie druhú stranu, tak si náhodne vyberie "0" alebo "1" (náhodou sa môže trafiť do čísla, ktoré Alica posielala, ale na to sa nemožno vždy spoliehať). Keď Alica odošle všetky fotóny a Bob ich všetky prečíta, tak vykonajú takzvanú "Sifting transaction" ako súčasť "Post processing", kde si navzájom oznámia len informáciu, ktoré strany karty čítali (nie aké hodnoty prečítali). Zahodia všetky karty, ktoré neboli vybrané správne a zostávajúce karty tvoria sled jednotiek a núl, ktoré sa použijú ako surový kľuč [5]. Pre bežnú implementáciu QKD zahrňuje tri základné komponenty.

- Kvantový kanál (quantum channel) buď optický kábel alebo bezdrôtový prenos slúži pre odoslanie kvantových stavov fotónov (qubity), v ktorých sa prenáša náhodná sekvencia bitov od odosielateľ a (Alica) k prijímateľ ovi (Bob). Tento kanál nemusí byť zabezpečený.
- Overený verejný komunikačný kanál (classical channel) medzi komunikačnými stranami. Jeho dôležitou úlohou je zabezpečiť synchronizáciu a výmenu dát medzi modulmi QKD, a mohli tieto strany uskutočniť "Post processing steps" a mohli vygenerovať správny a tajný kľuč.



Obr. 1: Základná schéma zostrojenia QKD medzi Alicou a Bobom

 Protokol pre výmenu kľúčov, ktorý využíva vlastnosti kvantovej mechaniky pre zaistenie bezpečnosti detegovaním odpočúvania alebo chýb a výpočtom množstva informácií, ktoré boli zachytené alebo stratené [6].

2 SIMULÁCIE

QKDNetsim je implementačný model, ktorý obsahuje sieť ový modul QKD, kľúče QKD, vyrovnávaciu pamäť QKD, sieť ové zariadenie QKD (QKD NetDevice) a "processing applications". QKD-Netsim bol vytvorený na Technickej univerzite v Ostrave [7].

Topológia simulácie sa skladá z 3 uzlov, kde uzol 0 je spojený s uzlom 1 a tento uzol je spojený s uzlom 2. Uzly komunikujú cez P2P spoje. Na tejto topológii boli zobrazené a otestované možnosti tohto simulačného prostredia. QKDNetsim nedokáže simulovať kvantovú úroveň QKD, teda nepoužíva sa na generovanie kľúča, ani pre žiadnu správu kvantového kanálu alebo odpočúvanie na kvantovej/fyzickej úrovni. Jeho činnosť je v prevažnej miere zameraná na použitie tajného kľúča a správu kľúčov na vyšších vrstvách.

Prvá simulácia zobrazuje vplyv rýchlosti tvorenia kľúčov a rýchlosť premávky na vyrovnávaciu pamäť kľúčov. Na začiatku simulácie boli už vo vyrovnávacej pamäti kľúče. Taktiež boli vybrané rôzne časy začiatku tvorby kľúčov tak, aby na grafoch bol viditeľný rozdiel.



Obr. 2: Priebeh zmeny množstva kľúčov vo vyrovnávacej pamäti medzi nultým a prvým uzlom.

Premávka tvorila celkové množstvo 3 000 000 bitov, jeden paket mal nastavenú veľkosť 60 bitov a ostatné parametre boli nastavené, viď. tab. 1. Z grafov je zjavné, že od definovaného začiatku tvorby kľúčov vždy prebehne 10 sekúnd a až potom sa odošle nastavené množstvo kľúčov a táto situácia sa opakuje.

Spojenie	QKD spoj 1	QKD spoj 2	Premávka				
Začiatok [s]	10	3	15				
Koniec [s]	50	50	50				
Keyrate [bit/s]	5 072 000	2 007 200	-				
Prenosová rýchlosť [Mb/s]	-	-	2				

Tabul'ka 1: Prenosové parametre prvej simulácie





Pri spustení premávky je zjavné znižovanie množstva kľ účov vo vyrovnávacej pamäti. V prípade keď množstvo kľ účov klesne pod prahovú úroveň (treshold), a táto pamäť sa dostane do stavu nebezpečne nízkeho množstva kľ účov (prezentované žltou častou grafu).

V **druhej simulácií** bolo sledované, ako sa bude sieť správať, ak bude nastavená príliš vysoká premávka, nedostatočné generovanie kľúčov a počiatočná hodnota bude nastavená pod prahovou hodnotou, viď obr. 4. Z grafu je zjavné, že v prípade ak v pamäti nie je dostatočné množstvo kľúčov, je na začiatku v stave dopĺňanie a po čase sa dostala do stavu pripravená a po spustení premávky množstvo kľúčov klesne pod minimálnu hodnotu, komunikácia skolabuje a už sa neobnoví. Tieto úrovne je možné nastavovať, ako parametre minimálneho množstva kľúčov (pod touto hodnotou spoj skolabuje), prahová hodnota (pod touto hodnotou je v stave nebezpečenstva), maximálna hodnota (maximálne množstvo uložených kľúčov) a počiatočná hodnota (množstvo kľúčov na začiatku simulácie).

Zdrojová IP adresa	ι	10.1.1.1			
Cieľ ová IP adresa		10.1.2	.2		
Odoslané [bit]	4000200	Prijaté [bit]	3112200		
Odoslané [paket]	6667	Prijaté [paket]	5187		
Pomer [bit]	0.778011	Pomer [paket]	0.778011		

Tabul'ka 2: Výstup z konzoly simulačného prostredia QKDNetsim pri simulácií vysokého toku dát, kde nebolo dostatočné množstvo kľúčov, ani dostatočné rýchla regenerácia kľúčov.



Obr. 4: Priebeh zmeny množstva kľúčov medzi nultým a prvým uzlom pri vysokom prenose dát.

3 ZÁVER

V článku bola vysvetlená problematika QKD a tiež princíp jeho fungovania. Následne boli zobrazené simulované vzorové simulácie dvoch scenárov. Vď aka simulátoru QKDNetsim sme schopní správne nastaviť jednotlivé parametre siete tak, aby sieť fungovala aj pri vysokej premávke, pri pevne danej pravidelnosti vytvárania kľ účov. Tieto parametre sú veľ mi dôležité, nakoľ ko pamäť kľ účov by bolo zbytočné mať príliš veľ kú, keď že každý kľ úč má svoju životnosť. Tento simulátor je vhodný na simuláciu sieť ovej vrstvy, kde nás zaujíma iba prenos kľ účov v sieti a nie samotná kvantová vrstva. Je možné si vopred nasimulovať rôzne prenosy a overiť, či by aktuálne nastavenie siete zvládlo daný kvantový prenos. Do budúcna je v pláne sa zaoberať navrhnutím a zostrojením testovacieho polygónu, kde bude možné pripojiť QKD a testovať rôzne druhy prenosov súbežne s prenosom kvantového signálu a sledovať ich vzájomný vplyv.

LITERATÚRA

- [1] PETROVSKÝ, Bc. Peter. *Formální analýza kryptografických protokolů*. Brno, 2015. Diplomová práca. VUT Brno. Vedoucí práce Ing. Vlastimil Člupek.
- [2] RUSSELL, J. Application of Quantum Key Distribution. *MILCOM 2008 2008 IEEE Military Communications Conference* [online]. 2008, 2008(1), 1-6 [cit. 2020-11-24]. Dostupné z: doi:10.1109/MILCOM.2008.4753169
- [3] KOTHARI, Abhishek. Qubit By Qubit. *Medium* [online]. Amerika: medium, 2018 [cit. 2020-11-24]. Dostupné z: https://medium.com/@abhishekkothari/qubit-by-qubit-104139024edc
- [4] Kvantový seriál díl 9. Kvantové sítě Současná situace [online]. Česko: Quantum Phi, 2020 [cit. 2020-11-24]. Dostupné z: https://qubits.cz/serialy/kvantovy-serial-dil-9-kvantove-sitesoucasna-situace/
- [5] ELLIOTT, C. Quantum cryptography. *IEEE Security Privacy* [online]. 2004, 2004(2), 57-61 [cit. 2020-11-24]. Dostupné z: doi:10.1109/MSP.2004.54
- [6] JAKUBÍČEK, Michal. Návrh zabezpečení systému dálkového měření kvality dodávky elektrické energie. Brno, 2013. Bakalárska práca. VUT Brno. Vedoucí práce Ing. Petr Mlýnek, Ph.D.
- [7] QKDNETSIM [online]. Česko, Bosnia and Herzegovina: QKDNetSim Team, 2020 [cit. 2020-12-10]. Dostupné z: https://www.qkdnetsim.info/

Magisterské projekty

Kybernetika a automatizace

CONTROL SYSTEM FOR SMALL INJECTION MOLDING MACHINE

Jan Novotný

Master Degree Programme (1), FEEC BUT E-mail: xnovot1i@vutbr.cz

Supervised by: Tomáš Benešl E-mail: xbenes23@stud.feec.vutbr.cz

Abstract: This paper deals with design and realization of control system in application for small injection molding machine. The small injection molding machine is intended to serve as a substitute for 3D printers in company, thus reducing producing time and improving mechanical properties of final product. The main goal was to design its own control system, which will be more cost-effective than using a PLC. An integral part was also the assembly, commissioning and programming of its own control system.

Keywords: control system, PLC, ARM, MCU, thermocoupler, injection molding machine, AT-SAME70, Atmel, Microchip

1 INTRODUCTION

3D printers are very popular and famous with DIY people (DIY - "Do it yourself"), because it's very low cost in 2021. 3D printers allow to do any shape (in technical possibilities) made of a plastic string, which have different properties depending on the material. The company, for which it this control system is designed, is printing small and specific components. FDM 3D printed part has inconsistent mechanical properties in many ways. The main differences are evident in the elasticity and tensile strength.

The main idea of the work is realize control system as one unit instead of control unit and lots of converters and control cards. The own control system can save place and be low cost.

This paper deals with the design and implementation of a control system for a small injection molding machine. The project is developed in cooperation with the company ELZACO spol. sr.o..

2 REQUIREMENTS FOR IMPLEMENTATION

The aim was to design and create a control system for a small injection molding machine at an affordable price and size. Because I am not the implementator of the control algorithm, all peripherals of control system must be implemented for the implementator who does not understand peripheral programming.

2.1 IMPORTANT PRINCIPLES OF INJECTION MOLDING MACHINE

For designing a specific control system it is important to know the main principle of injection molding machine to tailor-made all peripherals. On figure 1 there is a 3D model of small injection molding machine, which I get from a developer of mechanical design.

A simplified principle is the melting of the plastic granulate to the plasticization temperature. Melting is performed in a long tube with heaters. It is important to mention that there are several heaters that



Figure 1: 3D model of small injection machine.

ensure the correct course of the melting process. This means that we have to measure the temperature in several places. The time constant of the heater is in the order of tens of seconds and this is reason, why we don't need extra fast PWM outputs.

The injection of the plastic is realized by extruder feed screw. For rotating the extruder screw is used a stepper motor, which has a driver controlled by impulses and digital inputs for enable and direction. Plastic is injected into mold, where one half of mold is moving. The movement of the mold is performed by the lead screw and a stepper motor. [1]

Other important peripherals are digital inputs from sensors and buttons, or digital outputs for signaling.

3 CONTROL SYSTEM SELECTION

Chapter 2.1 and 4 describes the design of the control system, from which the requirements for a suitable control system (minimum requirements) are drawn.

3.1 PLC

The own control system will be compared with a PLC from the company Unitronics, UniStream, because this PLC offer an integrated display. Modular PLC with 7" display starts at 18000 *CZK* which need a processor for 8000 *CZK*. Other necessary items are **I/O** modules. In same count of peripherals, it was a modul of digital inputs and outputs, which can offer stepper motor signal (usually marked as PTO), analog inputs for current loop (ranging 0..20 *mA*) or voltage (0..10 *V*), thermalcoupler input (type K). The communication module for RS485 is sold separately. The price of the one module ranges between 3000 *CZK* and 8000 *CZK*.

The second option is to use a compact PLC, which offer an integrated display (5", 7" and 10.1") and several I/O. PLC with 5" diplay, 24 digital inputs, 2 analog inputs and 16 digital outputs is starting at 23000 *CZK*. For another module connection, it is necessary to buy a module extension and other modules can be added, like in the modular PLC.

All prices are only indicative and without VAT. The prices of PLC are taken from SCHMACHTL eshop[5].

3.2 OWN CONTROL SYSTEM

Given the requirement to minimize the cost, an alternative to a PLC control system was considered. The price of the own control system is estimated at 10000 CZK. This price is only for material per one piece. The final price is reduced by higher productions.

3.3 OTHER SOLUTION

During the market research, no product was found to be suitable as a cleaning system for injection molding machines. There are several companies that design control systems for large injection molding machines, which is an unsuitable solution in terms of price and size. At the time of the survey, there were some open-source projects for CNC, but it would be difficult to adapt or they were built with the idea of assembling arduino modules.

3.4 CONCLUSION

The analysis revealed that with the proposed control system, the price could be a third compared to a solution with a PLC. The savings do not only lie in the price of the control system, but also in the implementation of the specific peripherals like thermocouplers input into the control system itself. The price of a PLC can be justified by the simplicity and speed of implementation, for one manufactured piece. The own control system would be used in the production of multiple pieces, where the cost of development would be divided (the price of PLC will still be the same).

Mass-produced injection molding machines usually have their own control systems or designed to order.

4 DESIGN OF CONTROL SYSTEM

From the requirements described in Chapter 2.1 was created the block schematic on figure 2.



Figure 2: Block diagram of the control system.

4.1 **PERIPHERALS OF THE CONTROL SYSTEM**

I/O peripherals have been designed according to industrial requirements (similar to PLC). Thus, digital I/Os are designed with levels of 24 V and analog I/Os have ranges 0...10 V or 0...20 mA. For wide range use, the digitals peripherals and supply is designed to wide voltage range (10..32 V). And for safety all peripherals and supply is galvanic separated from processor.

The control system contains the following peripherals:

- **12 digital inputs** designed with common ground (SINK). An optocoupler is used for logic level conversion and galvanic isolation.
- 12 digital outputs, which are solved as an open collector using the circuit VNI4140. Subsequently, the outputs are equipped with a fuse on the power input.

- 4 fast digital outputs differ from normal digital output by speed and more option. All four digital output can use as PWM generator on high frequency. Two outputs can be use to generate specific number of pulses, that are using for control of drivers for stepper motors. Outputs is realized with circuit DRV8844 (Quad indipended ¹/₂ H-Bridge driver).
- 4 analog inputs are switchable between voltage and current. Voltage inputs are possible as $\pm 10V$ and current inputs in the range $\pm 20 mA$, where it is possible to convert range of 4...20 mA or any other in software. For measuring is used ADS1118, which is common for thermocoupler inputs.
- 4 thermocouplers inputs are designed to primary use with K-thermocoupler. For converting is used the analog-to-digital convertor (ADC) ADS1188. The law of cold junction compensation [2] is used for calculating the resulting temperature.
- **Display** that is, for example, capable of displaying measured data, statistics or settings. Display is monochrome graphicals. Display has the resolution of $240 \cdot 128$ pixels. In combination with an encoder and buttons it allows the setting of parameters.
- **USB** will allow the setting or showing advanced parameters using a PC application or show basic terminal. The USB-C connector was chosen as the USB connector for its recent expansion.
- **RS485** is for communication with other devices. Primary this peripheraly is for another project, where I want to use this control system as a diplay for modular control system. MODBUS/RTU will be used as the communication protocol, which offer communication with other mass-produced devices such as PLC. This means that this control system can be a device of decentralized production.
- Memory the control system have EEPROM (1 *Mbit* with M95M01 over SPI) and FLASH (32 *Mbit* with SST26VF032B over QSPI) memory to save statistics and parametrs.

4.2 **REALIZATION**

The control system (see figure 3), was designed as one four-layer printed circuit board. An ARM processor from Atmel (now is Microchip) ATSAME70Q21 [3] was chosen as the main processor.



(a) Top side with display.



(b) Bottom side.

Figure 3: Own control system

The HW was installed, revived, tested and SW equipment was created to operate all peripherals of the control system.

4.3 SOFTWARE

The programming environment Microchip Studio 7 (renamed from Atmel Studio) was used for the development of the SW. The control system contains a program for the operation of peripherals, simple GUI and example of their using.

For easier implementation of the resulting control algorithm by the end-programmer, RTOS (Real Time Operating System) was chosen. FreeRTOS was chosen from a wide range of RTOS. The main advantage of using RTOS versus superloop are the tasks and their apparent parallelism. Another very useful advantage is the resource manager, semaphore, mux, queues, events and timers. In addition to the kernel itself, FreeRTOS also offers a package called FreeRTOS+, which it is a library of function. The packages are optional and can offer TCP or UDP stack, CLI (Command Line Interface), IO manage, JSON decoder or application protocols (HTTP, MQTT). [4]

The ASF3 library (Atmel Software Framework), which uses an offline configurator, was used for peripheral programming. ASF4 versus ASF3 uses ATMEL START (configurator in a web browser). However, this library is not lightweight (for universally use) and some functions had to be rewritten. With regard to the library used, the C programming language was chosen.

5 CONCLUSION

I managed to design, revive and test our own control system for a small injection molding machine. It is estimated that the cost of the proposed control system could be a third compared to using a PLC solution. From SW FreeRTOS with CLI extension was implemented. The implementation of drivers for all peripherals is also a matter of course in the SW. In the next version of the control system development, the design of the control system will be reworked. The main change will be the partition into several sub-boards to minimaze size and use the of colored graphical display. The use of the control system does not have to be limited to a small injection molding machine, as it has been designed universally, but its use is limited only by the number of IOs.

ACKNOWLEDGMENT

I would like to thank the company ELZACO spol. s r.o., which offered me the background for development and the material support.

REFERENCES

- [1] MM Průmyslové spektrum. Konstruční provedení vstřikovacích lisů [online]. 2013 [cit. 19. 2. 2021]. Available from: https://www.mmspektrum.com/clanek/ konstrucni-provedeni-vstrikovacich-lisu.html
- [2] Maxim integrated. APPLICATION NOTE 4026: Cold Junction Compensation in Thermocouple [online]. 2013 [cit. 19. 2. 2021]. Available from: https://www.maximintegrated.com/ en/design/technical-documents/app-notes/4/4026.html
- [3] Microchip. ATMEL AT12874: Getting Started with SAM S70/E70 [online]. 2013 [cit. 19. 2. 2021]. Available from: http://wwl.microchip.com/downloads/en/DeviceDoc/ Atmel-42532-Getting-Started-with-SAM-S70-E70_ApplicationNote_ AT12874.pdf
- [4] FreeRTOS. Official site of FreeRTOS [online]. 2021 [cit. 19. 2. 2021]. Available from: https: //www.freertos.org/
- [5] SCHMACHTL. Online e-shop SCHMACHTL [online]. 2021 [cit. 22. 3. 2021]. Available from: https://www.schmachtl.cz/

POSITION MEASUREMENT OF MOVING OBJECTS USING A ROBOTIC TOTAL STATION

Tomáš Horeličan

Master Degree Programme (2nd year), FEEC BUT E-mail: xhorel00@stud.feec.vutbr.cz

> Supervised by: Tomáš Jílek E-mail: tomas.jilek@ceitec.vutbr.cz

Abstract: This paper attempts to evaluate an emerging use case for Robotic Total Stations (RTSs) as a tracking and guidance system for small Unmanned Aerial Vehicles (UAVs), especially applied to scenarios where conventional means of navigation (such as GNSS) are insufficient or completely unavailable. The focus is on indoors environments, where a constantly tracking total station could potentially be used to provide the UAV with a reference coordinate frame together with its immediate position within these coordinates. Preliminary experiments have been performed to evaluate the capabilities of two available RTSs, the Trimble S7 and S9 HP series, both controlled by a TSC7 controller. A Raspberry Pi mini-computer was used to timestamp specific events during the motion of a target tracked by the RTS. The Pi was synchronized with the RTS through Precision Time Protocol (PTP). This enabled direct timestamp comparison for time delay evaluation.

Keywords: robotic total station, kinematic tracking, unmanned aerial vehicle, navigation

1 INTRODUCTION AND MOTIVATION

Modern RTSs typically comprise two main measuring principles. An Electronic Distance Measurement (EDM) unit that usually utilizes a light source for Time of Flight (ToF) or phase shift measurements and a conventional theodolite system that performs extremely precise angle measurements using an electronically (or optically) read-out rotary encoder. The whole device is equipped with servo motors enabling automatic full 360 degree rotations in the horizontal as well as vertical plane. An onboard processor can then query measurements, control the servos, communicate with other devices and perform additional data processing and calculations.

In the early days, problems with internal component synchronization were occurring, however recent studies have shown a reduction of these issues into the sub-millisecond (millimeter) regime [1]. Over the past 10 years total station technologies have enjoyed a rapid development and new available feature sets. With the ability to continuously track a moving target and provide position measurements with high enough frequencies, new potential applications (apart from the standard geodetic or construction) have emerged [2, 3, 4]. One such use case with a demand for very high precision positioning is navigation of small UAVs. Nonetheless, correct overall synchronization is a large concern for near real-time applications and especially since RTSs still aren't primarily designed for these kinds of workflows, investigations into their parameters have to be conducted.

2 EXPERIMENT DESIGN AND METHODOLOGY

The performed experiment utilized these key notions:

- A reference measurement frame providing the real (true) time of an event.
- One objectively known position of a target (for all frames) during it's movement.

• Synchronization between reference and the observed time series.

2.1 TOTAL STATION CONFIGURATION

The RTS was locked onto a Trimble VX/S 360 prism, which was mounted on the end of a pendulum construction. A Trimble TSC7 controller running Windows 10 Pro was used to operate the RTS, which was set to track the prism continuously in Tracking (TRK) mode [5]. During tracking a timestamp (using local Windows time of the TSC7) for each measured point is provided by the RTS. A continuous output data stream was enabled through an on-board serial COM interface, which was connected to an external PC. A baud rate of 57 600 bits/s was set for the measurements, however a higher rate of up to 115 200 bits/s is available and should be more preferable [6]. The data stream was set as a custom format as is shown in Figure 1. A Trimble S9 HP [7] with 10 Hz tracking enabled was used for the experiment as the S7 does not have this feature in the present state. It has been observed that this rate is not stable and has a tendency to reach a maximum of 10 Hz rather than measuring at stable 10 Hz. This deficiency was also observed by Lienhart et al. with different RTSs [6].

Data output	
Continuous	GDM user defined 🗸
User defined record	
GDM label 1	GDM label 2
51 (Date)	▼ 52 (Time) ▼
GDM label 3	GDM label 4
7 (HA)	▼ 8 (VA) ▼
GDM label 5	GDM label 6
9 (SD)	▼ 10 (VD) ▼
GDM label 7	GDM label 8
11 (HD)	▼ 37 (N) ▼
GDM label 9	GDM label 10
38 (E)	▼ 39 (ELE) ▼
End of transmission character	
62 (">")	▼

Figure 1: RTS output data format.

2.2 EXPERIMENT WITH REFERENCE MEASUREMENT SETUP

The pendulum performed damped oscillations. A narrow tooth-like extension was attached to the end of the pendulum arm (with a radius of 129 cm) onto the prism. An Omron U-type photomicrosensor [8] connected to GPIO pins of a Raspberry Pi mini-computer running Raspberry Pi OS Lite was used to detect the rest position of the pendulum when the tooth crossed the sensor's light barrier. This position was also set as the zero azimuth backsight on the RTS, which was positioned at a 337 cm distance and such that the most significant movement was recorded in the Easting and Elevation axes and the Northing axis had minimal movement. Distances shorter than 300 cm caused the RTS to loose track of the moving prism completely. The RTS base station point was set as the origin of the coordinate system. Given the width of the tooth crossing the light barrier (7.5 mm), the exact moment when the sensor activates can differ between even and odd passes through the barrier.

The Pi OS was configured for maximum performance and tickless kernel mode was disabled to ensure minimal measurement delays. A custom measuring program, which can be run by a bash command was developed. Each time the pendulum crossed the rest position an ISR routine (using the PiGpio C library [9]) attached a timestamp to the event. The experiment layout can be seen in Figure 2, (a). The motion was always initiated from the right side as is shown in Figure 2, (b).

2.3 SYNCHRONIZATION PROCEDURE

Initially, correct local time was set on the Raspberry Pi from an NTP server and throughout the experiment local times of both devices (Pi and RTS) were being synchronized only through PTP. The Pi was set as the (grand) master clock and the TSC7 controller was a slave in the PTP hierarchy. Windows Time Service of the TSC7 was configured for high precision timing and SW-Timestamping





(a) Block diagram (green: RTS measurement,

blue: reference measurement, red: measured motion, purple: RTS data logging). (b) Motion initiation (RTS behind camera view).



mode and PTP [10] were all enabled by a custom developed PowerShell script. Presently, only the PTPd [11] implementation was compatible with Windows. Settings necessary to run PTPd with Windows were modified, otherwise all was left default.

3 OBTAINED RESULTS AND DISCUSSION

A thorough analysis of the results was conducted in the MATLAB environment. Since both devices were having their local times continually synchronized a direct comparison of their timestamps could be performed. Unix time values from the Raspberry Pi were converted into uniform nanosecond values. Data from the RTS first had to be parsed and filtered from corrupted or incomplete measurements. Then the proprietary time values could also be converted into uniform nanosecond values. Since positions of the RTS measured points did not necessarily collide with the precise pendulum rest point, the data had to be interpolated between each zero-bounding pair of points. This way timestamps corresponding to the exact zero position could be obtained. Linear interpolation is sufficient as position values with respect to time near the zero can be approximated linearly. Both final time series were shifted to a zero start according to the first Raspberry Pi value. Linear pendulum velocity when passing the rest point was approximated from the interpolated RTS positions and timestamps, from which a time error for even passes in the Pi data can be calculated.

Finally, a valid range from the whole oscillating pendulum motion had to be selected. The initial data from the RTS was too sparse due to a high prism velocity, therefore interpolation in this section could not have been trusted. Similarly, the end part contained low velocities that due to the asymmetrical sensor measurements would have created high potential time errors in the Pi data at even pendulum passes. The final velocity range was then $v \in [0.5,3]$ m/s, where the maximum calculated error was 20 ms at 0.5 m/s. A maximum error caused by the sensor's reaction area was 2.2 ms.

As can be seen in figures 3 and 4, timestamps from the RTS seem to have a delay clustering around 100 ms. Calculated corrections for the even points are shown, though no significant offset in the original data is visible. The reason might have been that the sensor setup was closer to being symmetrical than expected or different factors contributed to timing deviations more.



Figure 3: Time delays between Raspberry Pi and RTS measurements.



Figure 4: Histograms of measured time delays.

4 CONCLUSIONS AND FUTURE RESEARCH

Since the position of a moving target can change rapidly in time it is necessary to have correct temporal information of each measured position. This paper showed a delay of about 100 ms from the real position time when tracking a moving object with a Trimble S9 HP RTS. This result might be on the border of usability for small UAV navigation, since at reasonably higher velocities the position error would be comparable to the size of the aircraft itself.

Further experiments will be conducted to estimate more parameters of the RTS. A more precise sensor will be used for detection and different types of motion such as circular or straight could also be investigated. By connecting the RTS serial output directly to Raspberry Pi's UART pins further time analysis can be performed and data propagation delays can be estimated as well.

REFERENCES

- LENDA, G., A. UZNAŃSKI a M. STRACH. Influence of Time Delays of Robotic Total Stations Witch High Sampling Frequency on Accuracy of Measurements to Moving Prisms. *Archives of Civil Engineering* [online]. 2019, 65(1), 31-48 [ref. 2021-03-10]. ISSN 1230-2945. Available at: doi:10.2478/ace-2019-0003
- [2] LENDA, Grzegorz, Andrzej UZNANSKI a Michal STRACH. Comparison of Accuracy of Kinematic Methods for Localization of Mobile Targets. In: 2018 Baltic Geodetic Congress (BGC Geomatics) [online]. Olsztyn, Poland: IEEE, 2018, 2018, p. 138-144 [ref. 2021-03-10]. ISBN 978-1-5386-4898-8. Available at: doi:10.1109/BGC-Geomatics.2018.00032
- [3] ROBERTS, Craig a Peter BOORER. Kinematic positioning using a robotic total station as applied to small-scale UAVs. *Journal of Spatial Science* [online]. 2016, 61(1), 29-45 [ref. 2021-03-10]. ISSN 1449-8596. Available at: doi:10.1080/14498596.2015.1068232

- [4] PARAFOROS, Dimitris S., Marcus REUTEMANN, Galibjon SHARIPOV, Roland WERNER a Hans W. GRIEPENTROG. Total station data assessment using an industrial robotic arm for dynamic 3D in-field positioning with sub-centimetre accuracy. *Computers and Electronics in Agriculture* [online]. 2017, **136**(April 2017), 166-175 [ref. 2021-03-10]. ISSN 01681699. Available at: doi:10.1016/j.compag.2017.03.009
- [5] Trimble Access General Survey: HELP. Trimble: Trimble Geospatial help portal [online]. Sunnyvale, California, United States: Trimble, 2018 [ref. 2021-03-10]. Available at: https://help.trimblegeospatial.com/TALegacy/Help%20Files/ 2017_20_Help/English/General%20Survey%20Help%20v2017.20.pdf>
- [6] LIENHART, Werner, Matthias EHRHART a Magdalena GRICK. High frequent total station measurements for the monitoring of bridge vibrations. *Journal of Applied Geodesy* [online]. 2017, 11(1), 1-8 [ref. 2021-03-10]. ISSN 1862-9024. Available at: doi:10.1515/jag-2016-0028
- [7] Trimble S9/S9 HP: DATASHEET. Trimble: Geospatial [online]. Sunnyvale, California, United States: Trimble, c2015-2020, 2020 [ref. 2021-03-10]. Available at: https://geospatial.trimble.com/sites/geospatial.trimble.com/files/2020-04/022516-155H_TrimbleS9_DS_USL_0320_LRsec_0.pdf>
- [8] Photomicrosensor (Transmissive): EE-SX461-P11. TME [online]. Shiokoji Horikawa, Shimogyo-ku, Kyoto 600-8530, Japan: OMRON Corporation, c2007-2021, p. 146-147 [ref. 2021-03-10]. Available at: https://www.tme.eu/Document/7c3cde17f8178fdc71a91dbf142da7eb/ee-sx461-p11.pdf>
- [9] Pigpio: pigpio is a C library for the Raspberry which allows control of the General Purpose Input Outputs (GPIO). *GitHub: joan2937* [online]. San Francisco, California, United States: GitHub, 2021 [ref. 2021-03-10]. Available at: https://github.com/joan2937/pigpio
- [10] W32Time: This repo provides resources for high accuracy time on Windows. *GitHub: microsoft* [online]. San Francisco, California, United States: GitHub, 2020 [ref. 2021-03-10]. Available at: ">https://github.com/microsoft/W32Time>
- [11] Ptpd: PTPd official source master branch a.k.a. trunk. *GitHub: ptpd* [online]. San Francisco, California, United States: GitHub, 2019 [ref. 2021-03-10]. Available at: https://github.com/ptpd/ptpd

APPLIANCE FOR AUTOMATIC DISINFECTION OF INTERIORS

Alexander Korotynskiy

Master Degree Programme (2), FEEC BUT E-mail: xkorot01@stud.feec.vutbr.cz

Supervised by: Petr Petyovsky

E-mail: petyovsky@feec.vutbr.cz

Abstract: The work mainly deals with existing solutions for interior disinfection, their analysis and selection of a suitable situation. This article also deals with the appropriate selection of components for the implementation of mobile disinfection appliance. Subsequently, an analysis of the electrical connection of the driver for the DC motor is performed. Part of the article is also acquaintance with the programming language Rust, in which the firmware for the DC motor driver is created.

Keywords: Disinfection, Fogging, DC motor, Driver.

1 ÚVOD

V souvislosti s aktuální situaci okolo COVID-19 a dalších virů je hodně zvýšen požadavek na dezinfekci, a to včetně dezinfekci vnitřních prostor. V kancelářích, školách, úřadech, hotelích a jiných prostorách jsou vyžadovány každodenní dezinfekční procedury. Dezinfekce ve vnitřních prostorech nedává stoprocentní jistotu, že v prostoru se nenachází vir, ale je to jedna z cest, která obecně pomáhá v boji s nákazami.

V dnešní době dezinfekci vnitřních prostor většinou provádí člověk nebo zařízení, které člověk ovládá. Tyto zařízení nejsou plně automatizované. Tato práce se zabývá návrhem na plně automatizované mobilní zařízení pro dezinfekci vnitřních prostor.

Tato práce nejprve popisuje existující řešení pro dezinfekce vnitřních prostor a dále se zaměřuje na jejich analýzu a výběr vhodného řešení. Práce se také zabývá vhodným výběrem komponent pro realizaci mobilního dezinfekčního zařízení. Následně je uveden krátký popis elektrického zapojení driveru pro stejnosměrný motor, který bude sloužit jako aktuátor pro realizaci mobilní robotické platformy. V rámci práce jsem se také seznámil s programovacím jazykem Rust, ve kterém je tvořen firmware pro driver DC motoru.

2 NÁVRH DEZINFEKČNÍHO ZAŘÍZENÍ

2.1 ZPŮSOBY DEZINFEKCI

Dezinfekci vnitřních prostor lze provést pomocí několik metod, zejména pomocí ultrafialového záření, pomocí ozonu a také pomocí dezinfekčního aerosolu distribuovaného do prostoru formou studené nebo teplé mlhy. Dezinfekci pomocí ultrafialového záření byla odmítnuta, a to z důvodu, že vlny záření se nedostanou za překážku. Ozonová dezinfekce byla také zamítnuta z důvodu oxidačních vlastností ozonu – plyn je nebezpečný a jedovatý pro člověka a zvířata.

Princip dezinfekce studenou mlhou spočívá v tom, že z dezinfekční chemické kapaliny se udělá aerosol pomocí generátoru studené mlhy. Výhoda dezinfekce pomocí studené mlhy spočívá v tom, že na výstupu z generátoru dostaneme malé částice rozměrem 20-50 µm, které se dostanou všude v prostoru. Další výhodou je to, že studená mlha dezinfikuje vzduch a povrch. Také dezinfekce pomocí studené mlhy se může provádět za přítomnosti lidí, zvířat a rostlin za podmínky, pokud je vyfukována látka bezpečná. Nevýhodou je to, že po dezinfekci studenou mlhou se zvýší vlhkost v prostoru. Může také vzniknout kondenzát na stole, podlaze, což by znamenalo uklízení stolů od papírů, elektroniky a dalšího vybavení, které jsou citlivé na mlhu, před zapnutím dezinfekčního zařízení. Na rozdíl od UV a ozonové dezinfekce je potřeba více času po dezinfekci, až se všude látka usadí a projeví své dezinfekční schopnosti. V praxi by to znamenalo zapnutí zařízení v noci v pracovních místnostech a také v domácnosti, pokud člověk není doma. Vyfukování látky ve vnitřních prostorech se doporučuje provádět 1,5-2 metry nad podlahou pod úhlem směrem nahoru, a to z důvodu pokrytí dezinfekční kapalinou celého objemu prostoru. Pro neměnnost chemických vlastností dezinfekční kapaliny teplota vzduchu je doporučena v rozsahu 18-29 °C, což se má kontrolovat snímačem teploty. [1][2]

Obecný princip dezinfekci pomocí teplé mlhy je stejný jako u studené -z chemické kapaliny se udělá teplá mlha pomocí generátoru teplé mlhy, generátor je termomechanického typu. Dezinfekční kapalina se zahřívá v generátoru a stává se plynem, následně se vyfukuje plynová směs. Teplá mlha má lepší vlastností než studená mlha. Na výstupu z generátoru dostaneme menší částice, rozměrem 0,5-20 µm, které se dostanou všude a déle se drží ve vzduchu než částice studené mlhy. Na rozdíl od studené mlhy, teplá mlha může pokrývat nejen vodorovné povrchy ale i svislé. Nevýhodou je, že teplou mlhu není doporučeno používat za přítomnosti lidí, zvířat a rostlin v nezávislosti na dezinfekční kapalině. Další nevýhodou je to, že není doporučeno používat jakékoliv dezinfekční kapaliny, jelikož po zahřátí některé mohou ztrácet své dezinfekční schopnosti, dokonce mohou být i škodlivé. Po dezinfekci teplou mlhou se zvýší vlhkost a dochází ke kondenzátu stejně jako u studené mlhy. Po dezinfekci teplou mlhou, stejně jako studenou, je potřeba počkat několik hodin, v závislosti na rozměru prostoru, až látka projeví své dezinfekční schopnosti a následně vyvětrat. Generátory teplé mlhy jsou nebezpečné, jelikož může nastat požár nebo výbuch. Proto je doporučeno použití generátoru ve vnějších prostorách, ne ve vnitřních. Jelikož na výstupu z generátoru dostaneme plyn, tak nemusíme foukat pod úhlem nahoru, jako u studené mlhy, ale kondenzát může vznikat nejen na stole, ale i na stropě. Generátor teplé mlhy představuje zařízení s vyšší hmotností a vyššími pořizovacími náklady než generátor studené mlhy. [3][4]

Ve výsledku porovnání výhod a nevýhod dezinfekci pomocí studené a teplé mlhy bylo rozhodnuto pro dezinfekci studenou mlhou, a to z důvodu lepší bezpečnosti než u dezinfekci teplou mlhou, i když teplá mlha má lepší dezinfekční vlastností. Studenou mlhu je možné získat pomocí několika typů generátoru: pneumatického, diskového a ultrazvukového. Byl vybrán ultrazvukový generátor studené mlhy kvůli tomu, že má nejlepší vlastností, zejména je lehký, tichý a nejrychlejší a také je jednodušší v realizaci, než pneumatické a diskové generátory. Ultrazvukový generátor se umístí do nádoby s dezinfekční kapalinou a následně přes dvě trubky pod úhlem nahoru pomocí ventilátorů se studená mlha na výstupu vyfukuje do prostoru. Zpětná vazba o množství vyfukované kapaliny se dostává ze snímače výšky hladiny. Generátor má spotřebu 7 kg/h a napájí se 48 V DC a má výkon 250 W.

2.2 GENERÁTORY MLHY SPOLEČNOSTI IGEBA

Německá společnost IGEBA je orientována na vývoj a výrobou nových modelů generátoru teplé a studené mlhy. Většina jejich generátorů není vhodná pro vnitřní prostory, ale pro dezinfekci v průmyslu.

ULV (ultra-low volume) pneumatický generátor studené mlhy UNIPRO 40-T je nejnovější verzi generátoru společnosti IGEBA a je určen pro použití ve sklenících. Generátor má 4 trysky, které se dají manuálně výškově regulovat od 1,79 m do 4,3 m a také manuální otáčení o 360°. Má barvenou indikaci stavů, nastavení časovače pro dezinfekci a dálkový ovladač. Doporučený výkon generátoru je 42 l/h, maximální – 73 l/h. Studená mlha může být vyfukovaná dálkově až na 120 m. [5]

Generátor teplé mlhy TF 160 HD je nejnovějším modelem společnosti IGEBA. Takový generátor se dá použít stacionárně ve vnitřních prostorech nebo ve vnějších prostorech. Generátor TF 160 HD váží 65 kg, což pro ruční použití není vhodné. Objem benzinové nádrže je 10 litrů a nádrže s dezin-

fekční kapalinou až 60 litrů. Tento generátor se nabiji 12 V DC. Takové zařízení není nějakým způsobem automatizováno, má jenom tlačítka zapnout/vypnout na generátoru. [6]

2.3 VYBRANÉ KOMPONENTY

Hlavní řídicí jednotka bude zpracovávat signály ze snímačů a následně řídit pohyb celého zařízeni a vyfukování dezinfekční kapaliny. Jako hlavní řídicí jednotka celého zařízení byl vybrán jednode-skový počítač Raspberry Pi 4.

Pro vyfukování studené mlhy ze zařízení bylo rozhodnuto použít průmyslový ventilátor Sunon MagLev, který se umístí do trubky. Jeho rozměry jsou $80 \times 80 \times 25$ mm a pro jeho napájení je potřeba 24 V DC. Ventilátor má ložisko, které zaručuje nízkou hlučnost a je schopen vyfukovat průtok vzduchu 56.1 m³/h.

Pro mobilitu zařízení je třeba použít kolečka a byly vybrané čtyři průmyslové kolečka o průměru 100 mm a to z důvodu stability zařízení, jelikož doporučena výška při vyfukování mlhy je 150 cm. Celé zařízení bude poháněno dvěma motory, což znamená, že vepředu mají být použity 2 kolečka bez ložisek, které se upevní na motory a vzadu 2 kolečka s ložiskem.

Zařízení bude mít 7 litrů dezinfekční kapaliny a celá hmotnost zařízení bude s rezervou 15 kg. Pro pohanění takového zařízení na základě výpočtů byl vybrán stejnosměrný motor DCX32L s převodovkou GPX32HP a enkodérem ENX16 EASY od firmy Maxon. Motor DCX32L bude napájen 12 V DC. Krouticí nominální moment celé jednotky je 3.129 N.m při 12 V DC.



Obrázek 1: Motor DCX32L s převodovkou GPX32HP a enkodérem ENX16 EASY

Pro řízení stejnosměrného motoru se používá driver, který se vyvíjí na ústavu Automatizace a měřicí techniky. Komunikace mezi Raspberry Pi a driverem bude probíhat pomocí sériové komunikační sběrnice CAN. Řídicí jednotkou driveru je 32bitový ARM procesor STM32F091CBT6 z rodiny ARM Cortex-M0 s maximální frekvenci 48 MHz. Pro získání dat z enkodéru motoru na desce je využít integrovaný obvod MC3486D, na základě kterého lze sledovat aktuální úhel natočení motoru. Stejnosměrný motor se ovládá pomocí H-můstku. Pro měření motorového proudu v H-můstku je zapojen snímač proudu ACS712ELCTR-05B-T, který umožňuje AC a DC měření proudu v rozsahu -5...+5 A.



Obrázek 2: Driver pro stejnosměrný motor

3 ZÁVĚR

Během dosavadní práce byl proveden rozbor existujících řešení pro dezinfekci a jejích analýza. Bylo rozhodnuto pro dezinfekci chladnou mlhou pomocí ultrazvukového generátoru mlhy, a to z důvodu jeho nejlepších vlastností a jednoduchosti v realizaci. Následně byly vybrány vhodné komponenty pro realizaci mobilního dezinfekčního zařízení, kde součástí také byl výpočet vhodného motoru pro pohánění takového zařízení. Byl proveden rozbor existujícího elektrického zapojení driveru pro stejnosměrný motor. Proběhlo také seznámení s programovacím jazykem Rust, ve kterém bude pokračovat vývoj firmwaru pro driver DC motoru a jeho pomocnými nástroje, bez kterých se nelze obejit. Následně práce bude zaměřená na realizaci inovace firmwaru pro driver DC motoru, zejména: proudová a teplotní ochrana měniče, offline konfigurace parametrů měniče pomocí specializovaného souboru, runtime konfigurace parametrů pomocí komunikační sběrnice CAN, realizace regulace polohy.

REFERENCE

- MATHAI, C. V. Charged Fog Technology Part I: Theoretical Background and Instrumentation Development. *Journal of the Air Pollution Control Association* [online]. Pasadena, California,2012,,7[cit.2021-03-02]. Dostupné z: doi:https://doi.org/10.1080/00022470.1983.10465624
- [2] Обработка и дезинфекция холодным туманом. *DEZTRADE* [online]. [cit. 2021-03-02]. Dostupné z: https://deztrade.ru/info/obrabotka-i-dezinfektsiya-kholodnym-tumanom/
- [3] Обработка и дезинфекция горячим туманом. *DEZTRADE* [online]. [cit. 2021-03-02]. Dostupné z: https://deztrade.ru/info/obrabotka-i-dezinfektsiya-goryachim-tumanom/
- [4] Thermal Fog Application Efficiency to the last detail. *IGEBA* [online]. [cit. 2021-03-02]. Dostupné z: https://www.igeba.de/en/procedures/thermal-fogging/
- [5] UNIPRO 40 T. IGEBA [online]. [cit. 2021-03-02]. Dostupné z: https://www.igeba.de/en/products/ulv-aerosol-generators/unipro-40-t/
- [6] TF 160. IGEBA [online]. [cit. 2021-03-02]. Dostupné z: https://www.igeba.de/en/products/thermal-fog-generators/tf-160/

DATA CONCENTRATOR

Petr Dvorský

Master (5), FEEC BUT E-mail: xdvors11@stud.feec.vutbr.cz

Supervised by: Petr Fiedler

E-mail: fiedlerp@feec.vutbr.cz

Abstract: The topic of this paper is the design and realization of a modular Data concentrator for various types of measurements in diverse conditions. The device conception and basic principles are described herein. Also, the lightweight protocol design for radio communication is described. The electrical design of the device as well as the design of the printed circuit board was made using the Eagle Autodesk electronic design automation software (Eagle Autodesk EDA). Control software for a target microcontroller (ESP32-WROOM-32) is based on a FreeRTOS platform and ESP-IDF framework. An IDE for managing this software is Visual Studio Code with PlatformIO extension.

Keywords: data concentrator, eagle autodesk, visual studio code, platformio, cloud

1 ÚVOD

V mé magisterské práci se soustředím na vývoj Datového koncentrátoru pro autonomní sběr dat nejrůznějších fyzikálních veličin z přidružených inteligentních snímačů. Koncentrátor má za úkol komunikovat s jednotlivými připojenými snímači, koncentrovat z nich naměřená data, ty pak vhodně upravit a dále je ukládat ve vhodném formátu na jeho lokální uložiště. Takto získaná data bude koncentrátor exportovat do cloudového prostředí, kde bude docházet k jejich následné kontextualizaci a vizualizaci.

2 KONCEPCE DATOVÉHO KONCENTRÁTORU

Hlavním důvodem pro vytváření dále popisovaného Data koncentrátoru je reakce na ne zcela nový, avšak ne plně vyřešený požadavek vyvstávající v technických, ale i v různých dalších odvětvích, a to mít jednoduchou možnost měřit, ukládat a dále exportovat data v terénních a jiných neobvyklých podmínkách. Za tím účelem je třeba mít k dispozici jednoduché přenosné modulární zařízení, které dokáže data komunikovat s pomocnými senzorickými moduly určenými přímo pro danou měřenou aplikaci přes jednotné komunikační rozhraní a dále měřená data z těchto modulů extrahovat a schraňovat. Koncentrátor by měl takto získaná surová data nejen schraňovat na lokálním úložišti, ale i nahrávat extrahovaný obsah na cloud, kde lze data dále zpracovat, uvést je do širšího kontextu a vdechnout jim tak hlubší smysl. K takto zpracovaným datům by měl následně poskytnout v reálném čase vhodné vizuální vyjádření. Grafická podoba koncepce zařízení se nachází obrázku 1.



Obrázek 1: Grafické vyjádření koncepce Datového koncentrátoru

3 OBECNÝ PRINCIP ZAŘÍZENÍ

Princip zařízení je graficky vyjádřen níže na obrázku 2. Celé takovéto zařízení je tvořeno níže vyobrazenými moduly, které budou dále tvořeny vhodnou kombinací jednotlivých nižších hardwarových a softwarových modulů.

Zařízení se skládá ze základního výpočetního jádra, které vhodně komunikuje s jednotlivými podřízenými periferními moduly zařízení a vhodně ovládá jejich funkce. Jako konkrétní jádro koncentrátoru byl zvolen dobře známý MCU modul ESP32-WROOM-32, který obsahuje dvoujádrový 32bitový procesor Xtensa LX6, se zabudovanou 4 MB SPI flash pamětí a nabízí širokou řadu diverzitních komunikačních periferií, které jsou využity pro komunikaci s níže zmíněnými moduly. Další výhodou tohoto výpočetního modulu je nativně podpora komunikace v rámci Wifi a Bluetooth. [1]

Jedním z prvních, k jádru přidružených, periferních modulů je nízkoúrovňový komunikační modul, který má za úkol komunikovat s jednotlivými senzorickými jednotkami a vyměňovat s nimi základní surová data z jejich měření. Pro takovouto komunikaci byl taktéž navržen níže diskutovaný lightweight protokol. Tento nízkoúrovňový komunikační modul je v tomto případě zastoupen rádiovým vysílačem nRF24L01+, který komunikuje se senzory v rámci ISM pásma frekvencích v možném rozmezí od 2,4 GHz a do 2,525 GHz.

Dalším, pro takovýto koncentrátor naprosto nepostradatelným modulem, je modul lokálního uložiště dat. V tomto případě je v rámci zařízení lokální uložiště zajištěno pomocí SD karty ve formátu FAT32 s možností připojení k výpočetnímu jádru jak přes sběrnici SPI, tak přes nativní rozhraní SD/SDIO/MMC. Toto uložiště je důležité nejen z hlediska průběžného a konečného ukládání naměřených dat ve standartizovaném XML formátu v rámci každého probíhajícího měření koncentrátoru, avšak taktéž obsahuje konfigurační soubor se základními nastaveními a informacemi pro takováto měření Datového koncentrátoru.



Obrázek 2: Principiální blokové schéma Data koncentrátoru a jeho reálná podoba

Následně je taktéž nutné zmínit se jak o modulu vysokoúrovňové komunikace, tak o modulu geografické polohy. Oba tyto moduly v sobě integruje pro tento účel vybraná komponenta SIM808. Pro možnost komunikace dat do cloudového prostředí SIM808 zajišťuje přístup k internetu pomocí na koncentrátoru přítomné SIM karty a standartu GPRS. V rámci něj je umožněno naměřená data komunikovat pomocí HTTP protokolu.

Modul SIM808 taktéž disponuje i GPS modulem s obnovovací frekvencí 5 Hz a přesností 2,5 m CEP - 50 % (Circular error probable – Pravděpodobná kruhová chyba) pro potřeby zaznamenávání aktuální polohy koncentrátoru v případě mobilního venkovního měření. [2]

V neposlední řadě je nutné zmínit modul reálného času a synchronizační modul. Tyto moduly mají za úkol držet v rámci zařízení co nejpřesnější čas a ten vhodně distribuovat k jednotlivým senzorickým jednotkám. V rámci desky koncentrátoru jsou moduly tvořeny integrovaným obvodem DS3231SN v kombinaci s časovačem uvnitř výpočetního jádra, s korekcí na aktuální čas k dané geografické poloze pomocí modulu SIM808 a vstupně-výstupními piny.

Vstupně-výstupní periferie nemusíme nutně uvažovat jako nějaký modul, avšak jsou v podobě LED diody, tlačítka a vstupně-výstupních pinů přítomny pro základní ovládání a nastavení takovéhoto koncentrátoru.

4 LIGHTWEIGHT KOMUNIKAČNÍ PROTOKOL

V rámci nízkoúrovňové komunikace mezi jednotlivými senzorickými jednotkami a koncentrátorem je v rámci rádiového modulu nRF24L01+ již nativně implementován komunikační protokol Enhanced ShockBurst (ESB) pro výměnu dat mezi jednotlivými rádiovými moduly, včetně arbitrážní preambule, dynamické velikosti přenášených dat, či kontrolního součtu apod. Navržený lightweight protokol je však nadstavbou nad tímto zmíněným protokolem a definuje jak jednotnou strukturu přenášených dat, tak i celistvý průběh komunikace. [3]

Lightweight protocol - Datový rámec									
HE	EAD - Hlavič	ka	Payload - Vlastní přenášená data						
Délka	ID	CMD	Ser. č.	Protokol Ver.	Datový typ	Multiplikátor	Fyz. jednotka	Měř. hodnota	Čas. značka
1 Byte	1 Byte	1 Byte	1 Byte	1 Byte	1 Byte	1 Byte	1 Byte	4 Bytes	8 Bytes

Obrázek 3: Lightweight protocol – Datový rámec

Protokol je založen na komunikačním modelu Leader/Follower (dříve Master/Slave), kdy v roli Leadera – Vůdce je výše zmíněný datový koncentrátor a v roli Followerů – Následovníků jsou jednotlivé senzorické jednotky. Dále je v rámci protokolu definováno několik konfiguračních rámců pro možnost arbitráže v rámci komunikace a jeden datový rámec pro komunikaci již konkrétních měřených hodnot ze senzorických jednotek směrem k Datovému koncentrátoru.

Komunikaci vždy zahajuje vůdce, který nejdříve zjišťuje dostupnost jednotlivých následovníků na předem definovaných adresách. Pokud je daný následovník aktivní, odpoví definovaným potvrzovacím rámcem. S aktivními následovníky lze následně zahájit měření speciálním startovacím rámcem, který mimo hlavičku a další parametry obsahuje i přesný aktuální čas a čas, kdy bylo samotné měření zahájeno. Následovníci tento rámec standartně potvrdí a v rámci aktivního měření již komunikují směrem k vůdci pouze svá naměřená data zapouzdřená ve standartním datovém rámci. Vůdce přijetí těchto datových rámců již potvrzuje pouze v rámci výše zmíněné nižší komunikační vrstvy ESB. Ukončení celého měření a komunikace opět inicijuje vůdce definovaným ukončovacím rámcem, na který následovníci standartně odpovídají potvrzovacím rámcem.

Vzhledem ke skutečnosti, že se v našem případě v rámci komunikovaných dat jedná o měření nejrůznějších fyzikálních veličin, je třeba se samotnou změřenou číselnou hodnotou taktéž přenášet její fyzikální jednotku a její multiplikátor. Dále je taktéž nutné definovat datový typ přenášené číselné hodnoty pro její korektní uložení a zpracování a v neposlední řadě je taktéž nutné definovat časovou značku této hodnoty pro její evidenci vzhledem k časové oblasti. K tomu slouží datový rámec uvedený výše na obrázku 3.

Takto postavený protokol tedy pevně definuje strukturu a průběh komunikace a dále zajišťuje, že přenášená data budou vždy bezezbytku kompletně popsána a nemůže tak dojít k jejich dezinterpretaci a nevhodnému zpracování.

5 VÝVOJ A REALIZACE DATOVÉHO KONCENTRÁTORU

Prozatímní vývoj Datového koncentrátoru obsahoval vhodný výběr komponent pro zařízení, na základě takto vybraných komponent vytvoření návrhu elektrického schéma a následně návrhu výsledné desky plošných spojů ve EDA softwaru, kterým v tomto případě byl zvolen Eagle Autodesk.

Deska plošných spojů byla na zakázku realizována u externího výrobce JLCPCB a při dodání byla ručně osazena vybranými komponenty. V neposlední řadě bylo třeba takovéto zařízení oživit, ověřit jeho požadované chování a elektrické vlastnosti. Takto oživené a zprovozněné zařízení lze vidět na obrázku 2.

Řídicí software pro datový koncentrátor je spravován v prostředí Visual Studio Code v rámci doplňku PlatformIO, které je určeno právě pro vývoj takovýchto vestavěných zařízení. Samotný software je realizován v rámci frameworku ESP-IDF, což je nativní framework platforma pro výše vybraný MCU modul ESP32-WROOM-32 a využívá operační systém reálného času FreeRTOS.[4][5]

6 ZÁVĚR

V tomto článku jsme byli nejdříve seznámeni se samotnou koncepcí Datového koncentrátoru a důvody, které vedly k tvorbě, vývoji a realizaci takovéhoto zařízení.

Následně zde byl v rámci obrázku 2 vyobrazen obecný princip zařízení a dále podrobněji popsána funkčnost a výběr jednotlivých komponent, ze kterých se Datový koncentrátor skládá.

V neposlední řadě byl detailněji vysvětlen navržený lightweigth protokol pro komunikaci koncentrátoru s jednotlivými senzorickými členy a byl vysvětlen důvod vzniku takovéhoto protokolu, jeho princip a použití. Pro ukázku byl na obrázku 3 taktéž vyobrazen datový rámec tohoto protokolu.

Jako poslední byl v krátkosti zmíněn průběh samotného vývoje a realizace Datového koncentrátoru vzhledem k návrhu hardwaru a programového vybavení. Reálná podoba Datového koncentrátoru je taktéž uvedena na obrázku 2.

V budoucnu bude Datový koncentrátor rozšířen o ucelené programové vybavení, výběr a nastavení vhodné cloudové prostředí a vyzkoušení funkčnosti zařízení jako celku.

REFERENCE

- [1] ESP32-WROOM-32: Datasheet Version 3.0. Espressif.com [online]. Shanghai: Espressif Systems Co., 2021 [cit. 2021-03-12]. Dostupné z URL: <u>https://www.espressif.com/sites/default/files/documentation/esp32-wroom-32_datasheet_en.pdf</u>
- [2] SIM808 Hardware Design V1.03 [online]. Shanghai SIMCom Wireless Solutions, 2016 [cit. 2021-03-12]. Dostupné z URL:

https://simcom.ee/documents/SIM808/SIM808_Hardware%20Design_V1.03.pdf

[3] NRF24L01+ - Single chip 2.4Ghz Transceiver: Preliminary product specification v1.0 [online]. NORDIC Semiconductors [cit. 2021-03-12]. Dostupné z URL:

https://www.sparkfun.com/datasheets/Components/SMD/nRF24L01Pluss Preliminary Prod uct_Specification_v1_0.pdf

[4] ESP-IDF Programming Guide: Documentation for Espressif IoT Development Framework. Espressif.com [online]. Shanghai: Espressif Systems Co., 2021 [cit. 2021-03-12]. Dostupné z URL:

https://docs.espressif.com/projects/esp-idf/en/latest/esp32/

[5] BARRY, Richard. Mastering the FreeRTOS[™] Real Time Kernel: A Hands-On Tutorial Guide. Freertos.org [online]. Real Time Engineers, 2016 [cit. 2021-03-12]. Dostupné z URL:

https://www.freertos.org/fr-contentsrc/uploads/2018/07/161204_Mastering_the_FreeRTOS_Real_Time_Kernel-A_Hands-On_Tutorial_Guide.pdf

AUTOMATIC TABLE DISTILLATION COLUMN

Michal Ružička

Master Degree Programme (2), FEEC BUT E-mail: xruzic46@vutbr.cz

Supervised by: Petr Dejdar E-mail: xdejda00@stud.feec.vutbr.cz

Abstract: Main goal of this project was construction of small autonomous distillation column. It should serve as an example of potential uses of electrotechnics for general public. The column will have compact design based on Arduino board. It will contain everything necessary for its function, and will be operable even by layman. Part of the project is also closed cooling system based on Peltier modules.

Keywords: distillation, Arduino, cooling, peltier, automation

1 ÚVOD

Nápadov ako prilákať mladých ľudí k technike bolo už mnoho. Pre prezentačné účely je vytváraná malá stolná destilačná kolona plne riadená mikrokontrolérom. Pre jednoduché predstavenie a priblíženie problematiky je postavená na platforme Arduino, s ktorou je možné sa bežne stretnúť už v rámci výuky nielen na stredných, ale aj základných školách.

Bolo vytvorené prototypové zapojenie pre overenie teoretických poznatkov riadenia procesu destilácie. Vrámci tohto zapojenia je sledovaná teplota pár destilácie a prietok vzniknutého kondenzátu, na základe ktorých je aktívne riadený elektrický varič pre udržanie ideálnych parametrov destilácie.

Pre úplnú samostatnosť zariadenia je testovaný návrh chladiacej jednotky pomocou peltierových článkov. Takto bude zabezpečená funkčnosť aj v prípade nedostupnosti vodovodu. Ide o prietokový chladič výstupnej vody z chladiča, v ktorom kondenzujú pary.

2 NÁVRH KONŠTRUKCIE

K potrebám prezentácií bola zvolená malá nerezová destilačná kolona o objeme 101. Jej výhodou je integrovaný chladič na klobúku kotla. Takto je zabezpečená kompaktnosť riešenia a jej bezproblémové zostavenie.

Ovládanie prototypu je realizované štyrmi tlačidlami s výstupom na alfanumerickom displeji. Na ňom je možné zvoliť niektorý z predpripravených programov destilácie. Kedy pre vodu, víno a prvý destilát je potrebné udržiavať rôzne hodnoty teplôt a prietokov. Zobrazovaný aktuálny priebeh destilácie a aktuálne hodnoty sú následne zobrazované na displeji. Vo finálnom návrhu budú tlačidlá nahradené rotačným enkodrom a pridané bude aj Wi-Fi rozhranie pre možnosť jednoduchšieho nastavenia pomocou webového rozhrania a výpis parametrov v prehliadači.

Ako ohrev bol zvolený liatinový elektrický varič. Vďaka veľkej teplotnej kapacite styčnej plochy variču mohla byť jeho regulácia riešená pomocou SSR (Solid State Relay) spínaním napájacieho napätia. Teplota pár, ktorá indikuje v akej fáze sa proces destilácie nachádza je snímaná teplotným senzorom pripojeným sériovou jednolinkovou zbernicou, na ktorú môžu byť doplnené ďalšie teplotné senzory. Teplotu pár je potrebné snímať v najvyššom bode vedenia pred vstupom do chladiču pre získanie čo najpresnejších hodnôt. V konštrukcií destilačnej kolony bol nahradený pôvodný analógový

teplomer spomínaným teplotným senzorom v trubičkovom prevedení a upevnený v závite redukčnou maticou.



Obrázek 1: Bloková schéma zapojenia.

Prietok vytekajúceho kondenzátu je pre jeho malé hodnoty meraný nepriamo. Konkrétne vážením zbernej nádoby. Kedy zmena váhy v čase reprezentuje výsledný tok kondenzátu. Samotná váha je tvorená tenzometrom a ADC (Analógovo-Digitálnym Prevodníkom). Pre správne upevnenie a zabezpečenie spoľ ahlivej funkcie tenzometra bol vytvorený a vytlačený 3D model stojana s podložkou pre umiestnenie zbernej nádoby.

Obrázok 1 znázorňuje blokovú schému zapojenia destilačnej kolony s jej hlavnými funkčnými prvkami. Ďalej na obrázku 2 je možné si prehliadnuť funkčné prototypové zapojenie konštrukcie s jednotlivými prvkami zostrojenými v domácich podmienkach.



Obrázek 2: Prototyp destilačnej kolony.

3 REGULÁCIA DESTILÁCIE

Po zhotovení prototypového zapojenia riadenia na základe požiadavkov destilácie bolo potrebné oživiť celý návrh. Testovací firmware spracúva a vyhodnocuje aktuálne vstupy senzorov. Najmä hodnoty získané z ADC tenzometru snímajúceho hmotnosť, ktoré vyžadujú značné priemerovanie. Následne je potrebné prepočet váhy na prietok vhodne spriemerovať v čase a vylúčiť úseky, kedy je zberná nádoba vyprázdňovaná. Kalibračné údaje váhy sú uložené v pamäti mikrokontroléru a nie je tak potrebná kalibrácia po každom spustení, ale je možné ju vyvolať v menu na displeji.

Z takto získaných parametrov je následne vypočítavaný potrebný výkon variča, ktorým sa snaží regulátor dosiahnuť predom nastavený prietok. Vyparovanie obsahu kotla a následná kondenzácia je totiž závislá na privádzanej energii do systému varičom. [1]



Obrázek 3: Priebeh destilácie vína.

Na grafe znázornenom v obrázku 3 je možné si prehliadnuť praktickú funkciu regulácie. Postupne narastajúcu teplotu má za následok odparovanie alkoholu, ktorý má nižší bod varu ako voda. Teplota pár potom leží medzi týmito dvoma hodnotami v závislosti na koncentrácií zmesi vody a alkoholu. Nastavenie prietoku sa počas celého procesu destilácie nemenilo. Je tak možné vidieť prácu regulátoru, ktorý drží prietok v nastavených medziach. Zákmity vzniknuté medzi minútami 40 a 80 boli následne odladené zlepšením hodnôt regulátoru počas ď alších pokusov. Doposiaľ pokusov však stále nebolo dostatočné množstvo pre časovú náročnosť týchto meraní a úkonov s ním spojených.

4 TEST VÝKONNOSTI CHLADENIA

Pre možnosť fungovania destilačnej kolony aj v podmienkach mimo vodovodu je potrebné navrhnúť a zrealizovať chladiacu jednotku. Tá má za účel chladiť vytekajúcu vodu z chladiča pár, kde dosahuje jej výstup až 60 °C. Vstupujúcu vodu je potrebné chladiť na úroveň 10-15 °C. To je vo väčšine prípadov pod úrovňou okolitého vzduchu. Takže mimo bežné chladenie je potrebné dodatočne znížiť teplotu pretekajúcej vody. To je možné dosiahnuť peltierovými článkami.

V grafe na obrázku 4 je porovnanie účinnosti rôznych kombinácií článkov pri rôznych napájacích napätiach. Udávaná spotreba je odčítaná z grafu v datesheete [2], pričom reálny prúd pri danom napätí je pri chladnom článku zhruba o 0,5-1 A nižší. Graf znázorňuje priebeh teploty vody v objeme 21, kedy bola cirkulovaná jedným, alebo dvoma vodnými blokmi osadených dvojicou peltierových článkov.

Reálne to znamená, že pri obmedzenom prietoku vody v systéme spôsobí schladenie o cca 4,5 °C v prípade 4 článkov s napájacím napätím 7 V a zhruba o 8 °C pri napájaní 12 V, medzi vstupom a výstupom vodných blokov. Rozdiel v spotrebe energie je však natoľko veľký, že vyšší výkon nevyváži ich zníženú efektivitu. Pre finálne riešenie bude zvolené chladenie s ôsmimi článkami prevádzkovanými pri napätí cca 8 V a napájané regulovateľ ným zdrojom konštantného prúdu.



Obrázek 4: Meranie účinnosti chladenia.

5 ZÁVER

V článku bola popísaná konštrukcia malej destilačnej kolony spolu s predstavením fungovania jej regulácie. Pre fungovanie iba s napájaním pomocou elektrickej energie bolo vzhľadom na požadované malé rozmery zvolené vodné chladenie pár peltierovými článkami, ktorého funkčnosť a efektivita bola overená. Rovnako bude doplnené ovládanie pomocou webového rozhrania. V ňom bude umožnená úprava jednotlivých programov destilácie a vytváranie vlastných. Nasledujúcimi krokmi konštrukcie bude vytvorenie finálneho zapojenia a jeho osadenie do navrhnutých krabičiek pre jednoduché zostavenie a transport zariadenia.

REFERENCE

- SCHMICKEL, Helge a Bettina MALLE. *Domácí výroba lihovin*. Vyd. 2., rev. Praha: Beta, 2010. ISBN 978-80-7306-430-3.
- [2] *Thermoelectric Cooler: PM-40X40-89*. Dostupné také z: https://www.tme.eu/Document/ f90b783c096cf8ceee627b2a275a42d4/pm-40x40-89.pdf

AUTOMATION OF THE HORTICULTURAL SYSTEM

Martin Šimek

Master Degree Programme (2), FEEC BUT E-mail: xsimek29@stud.feec.vutbr.cz

Supervised by: Jakub Arm

E-mail: xarmja00@stud.feec.vutbr.cz

Abstract: This thesis deals with the issue of the Internet of Things and its use in the automation of gardening areas. The processes that are necessary to maintain the operation of gardening facilities are both physically demanding and time consuming. The aim of this thesis will be to design a system that will reduce the need for human intervention in performing these processes. The main prerequisites of the proposed system include the modularity of the whole concept. The automation of the gardening areas focuses mainly on the irrigation system, which will be controlled by the designed wireless modules. These are SHT-01 modules which have the task of measuring soil moisture, temperature and relative humidity in the area, BVU-01 modules which provide irrigation in the required area and last but not least FDU-01 modules with which it will be possible to enrich the water located in the drip pipe and used to irrigate the mentioned areas with mineral fertilizers.

Keywords: IoT, ESP8266, Raspberry Pi, MQTT, Node – RED, PI controller

1 ÚVOD

Osvojení zemědělství bylo bezpochyby jedním z nejvýznamnějších mezníku lidské civilizace. Umožnilo nám zejména přechod z lovu a sběru na udržitelný způsob získávání potravy, což vyústilo v rapidní růst lidské populace, který předtím nebyl možný. Mnoho lidí si od té doby kladlo za cíl zjednodušit či vylepšit dosavadní technologie a metody používané při pěstování rostlin a chovu zvířat. Od prvních vynálezů, které představovaly například motyky či jednoduché pluhy určené k obdělávání půdy, se nyní dostáváme až k pokročilým nejmodernějším pěstebním metodám, jako je aeroponie nebo hydroponie. Díky pokročilému rozvoji mikroelektroniky v posledních desetiletích jsou dnes k dispozici malé integrované obvody a různá cenově dostupná zařízení, díky kterým lze řadu procesů v oblasti pěstování rostlin zautomatizovat.

Předmětem této práce je tedy návrh systému, jenž by dokázal efektivně automatizovat dosavadní procesy, které musely být vykonávány manuálně. Jedná se zejména o řízení systémů závlah. Takovým řízením bude zajištěna optimální dávka vody a výživy pěstované komoditě. S ohledem na prostředky, které areál zahradnictví poskytuje, je nutné stanovit několik základních funkcí a požadavků, jež musí systém splňovat:

- Nezávislé řízení definovaných sekcí areálu
- Nadřazené řízení jednotlivých sekcí pomocí webové vizualizace
- Monitorování teploty, relativní vlhkosti a vlhkosti půdy v požadovaných oblastech sekce
- Možnost individuální závlahy požadované oblasti v sekci
- Efektivní a jednoduchý způsob hnojení při zavlažování

2 AUTOMATIZOVANÝ SYSTÉM

Areál zahradnictví byl vhodně rozčleněn do nezávislých sekcí, které budou individuálně spravovány. Rozdělením na sekce se zajistí modularita navrženého systému a jednoduchá správa jednotlivých sekcí. V případě potřebných úprav v dané sekci tak bude zachována plná funkčnost zbylých sekcí. Jak můžeme vidět z následujícího obrázku (obr. 2.1), pod dohled a řízení systému budou spadat 4 nezávislé sekce: Skleník, Fóliovník, Pole a Zahrada.



Obr. 2.1: Areál zahradnictví s vyznačenými sekcemi (převzato a upraveno z [1])

Na základě výše uvedených sekcí bylo vytvořeno blokové komunikační schéma systému (obr. 2.2). Jednotlivé sekce budou autonomní tzn. v případě výpadku serveru bude jejich proces řízen dle lokálně nastavených parametrů. Z toho důvodu byl pro jejich řízení zvolen modul ESP8266, který umožňuje jednoduchou komunikaci pomocí standardů IEEE 802.11, a zároveň poskytuje dostateč-ný výpočetní výkon. Pro komunikaci se serverem, jenž bude zprostředkovávat nadřazené řízení a dohled nad jednotlivými sekcemi, byl zvolen MQTT protokol a to hlavně díky jeho jednoduché implementaci a předpokládanému malému datovému toku mezi serverem a zařízeními. Server plnící úlohu MQTT brokeru poběží na jednodeskovém počítači Raspberry Pi 3B+. Na tomto počítači bude také umístěna vizualizace, která bude zprostředkována pomocí vývojového prostředí Node-RED. Raspberry Pi bude nadále připojeno do místní WiFi sítě. Přístup k vizualizaci bude tudíž umožněn všem účastníkům připojeným do uvedené WiFi sítě.



Obr. 2.2: Komunikační blokové schéma systému

Na základě požadavků formulovaných v úvodu práce budou navrženy a realizovány tři typy modulů. První z nich, modul SHT-01 (Soil moisture, Temperature, Humidity), má za úkol monitorovat lokální teplotu, relativní vlhkost vzduchu a vlhkost půdy v dané oblasti sekce. Tyto hodnoty bude zpracovávat základna sekce a dále bude vyhodnocovat, zda je potřeba danou oblast v sekci zavlažit. K ovládání distribuce vody do požadované oblasti pak bude sloužit modul BVU-01 (Bistable Valve Unit). Výše uvedené moduly budou napájeny Li-Ion bateriemi, jež budou dobíjeny pomocí fotovoltaického článku. Poslední realizovaný modul FDU-01 (Fertilizer Dosing Unit) se bude starat o automatické dávkování hnojiv do závlahového potrubí.



Obr. 2.3: Komunikační vazby inteligentního závlahového systému

2.1 MODUL FDU-01

Pro potřebu distribuce živin v podobě tekutých hnojiv bude sestaven modul obsahující řídicí systém, peristaltické čerpadlo, lopatkový průtokoměr a dva senzory konduktivity. Na základě údajů ze senzorů konduktivity a průtokoměru se bude pomocí přímého dávkování regulovat konduktivita závlahové vody. Kvůli velké spotřebě čerpadla je nutné modul konstantně napájet 5V. Komunikace se základnou bude probíhat pomocí protokolu ESP-Now. Technologické schéma modulu je následující:



Obr. 2.4: Technologické schéma modulu FDU-01

Na následujícím obrázku lze spatřit schéma popisující systém dávkovače hnojiv a jeho procesní veličiny.


Na základě rovnice směšování a údajů z funkčního schématu (obr. 2.5) byla odvozena diferenciální rovnice popisující chování tohoto nelineárního systému¹. Linearizací systému v okolí pracovního bodu, jenž je daný konkrétní hodnotou vstupního průtoku a konduktivitou dávkované kapaliny, získáme přenosovou funkci LTI systému 1. řádu s dopravním zpožděním:

$$F(p) = \frac{k}{T \cdot p + 1} \cdot e^{-Td \cdot p}$$
(2.1)

Kde je:

$$k = \frac{\kappa_d}{Q_i}, \ T = \frac{1}{Q_i}, \ T_d = 15000 \frac{l \cdot \pi \cdot d^2}{Q_i}$$

Z výše uvedeného přenosu vyplývá, že zesílení k, časová konstanta T a dopravní zpoždění systému T_d jsou závislé na velikosti vstupního průtoku (konduktivitu dávkované kapaliny lze považovat za konstantní a průtok dávkované kapaliny 0 až 100 *ml/min* je možné kvůli jeho malému přírůstku ke vstupnímu průtoku 1 až 15 *l/min* zanedbat). Dopravní zpoždění systému je způsobeno umístěním senzoru konduktivity za dávkovací jehlou. V reálné aplikaci je průměr potrubí a vzdálenost mezi jehlou a sensorem konstantní, tudíž jediná veličina, která může dopravní zpoždění během regulace ovlivnit, je vstupní průtok (konstanta 15000 vychází z úpravy jednotek při odvozování a přepočtu objemového průtoku na rychlostní). Díky průtokoměru umístěnému na vstupu systému lze monitorovat aktuální průtok a dle něho přizpůsobovat probíhající regulaci. Takové přizpůsobení je možné zajistit implementací metody gain-scheduling. Na základě velikosti vstupního průtoku se tedy mění pracovní bod soustavy a s ním i parametry lineárního regulátoru navržené pomocí metody Chien, Hrones a Reswick. Pro reálnou implementaci byl zvolen PI regulátor. S ohledem na požadavky regulace není rychlost dosažení žádané hodnoty zásadní. V reálné aplikaci se navíc předpokládá poměrně velký šum při měření konduktivity a soustava má rovněž znatelné dopravní zpoždění. Volba PI regulátoru je proto optimální cestou zajištující robustnost řízení a zároveň jeho jednoduchost.

3 ZÁVĚR

Domácí automatizace pěstování se díky dostupnosti široké škále různých komponent a zařízení stala trendem, který usnadňuje pěstitelům práci a zlepšuje podmínky pro jednotlivé pěstované komodity. V rámci chytrých domů se může jednat o inteligentní samozavlažovací květináče, kompaktní hydroponické systémy či senzory vlhkosti půdy a konduktivity, jež včas hlásí, kdy je potřeba danou rostlinu zavlažit nebo přihnojit. V případě zahrádkaření ve sklenících, fóliovnících nebo na polích je dostupnost chytrých zařízení umožňující automatizaci poměrně omezená a jejich instalace a uvedení do provozu bývá pro většinu uživatelů náročná. Cílem tohoto projektu je proto navrhnout a sestavit modulární systém inteligentních senzorů a akčních členů, který by bylo možné využít jak pro profesionální, tak pro komerční účely. Díky bezdrátové konektivitě modulů SHT-01 a BVU-01 lze jednoduše sestavit požadovaný systém závlah, jenž vyhovuje danému uživateli. Bezdrátový bistabilní ventil je možné použít pro většinu typů závlah (rozprašovací, kapénková, mlžicí). Ovšem v případě volby kapénkové závlahy je navíc zajištěno, že závlaha bude přiváděna vždy ke kořenové části rostliny, čímž lze razantně snížit spotřebu vody, a tím i ekologickou náročnost pěstování. Oproti rozprašovací závlaze je tak možné dosáhnout až 80% úspory vody [2]. U dávkovacích modulů hnojiv FDU-01 se nepředpokládá jejich nepřetržité používání. Je proto možné tento modul jednoduše připojovat pomocí rychlospojek a lze jej přemísťovat mezi jednotlivými sekcemi, u nichž je potřeba obohacovat závlahovou vodu hnojivem.

REFERENCE

- [1] Mapy.cz [online]. Seznam a.s. [cit. 27.03.2021]. Dostupné z: <u>https://www.mapy.cz/</u>
- [2] *How to Conserve Water: Drip Irrigation Systems* [online]. [cit. 28.03.2021]. Dostupné z: https://www.buildwithrise.com/stories/how-to-conserve-more-water-drip-irrigation-system

¹ Podrobné odvození rovnic je nad rámec rozsahu této práce

Magisterské projekty

Mikroelektronika a technologie

MATERIALS FOR BIODEGRADABLE BONES BASED ON Fe

Jan Hrabovský

Master Degree Programme (2.), FEEC BUT

E-mail: xhrabo12@vutbr.cz

Supervised by: Marie Sedlaříková

E-mail: sedlara@seznam.cz

Abstract: The thesis deals with biodegradable bone implants which must have good mechanical properties and also good compatibility with the human body. The thesis contains literary research in the area of bone physiology, biogenic materials and implants for the human body. The thesis focues on the selection of suitable materials for biogenic implants.

Keywords: Biodegradability, implants, bone, iron, corrosion, magnesium

1 ÚVOD

Poranění pohybového aparátu jsou často velmi omezující a jejich léčba bývá zdlouhavá. Zlomeniny ve velké míře postihují seniory, u kterých je léčba nejkritičtější a dlouhodobé upoutání na lůžko může mít fatální následky. Pro zlepšení léčby je snaha o hledání nových materiálů a možností, které by svými vlastnostmi předčily současné metody a mohly by tak zajistit lepší rekonvalescenci. Do dnešní doby proběhla celá řada výzkumů a dnes se hojně využívají nejrůznější kovové slitiny, nebo keramika.

V posledních letech roste zájem o biodegradabilní implantáty, které by mohly nahradit současné implantáty používané pro dočasné fixace. Biodegradabilní materiály mají funkci stejnou jako běžné implantáty, nicméně během působení tělních tekutin dochází k jejich pozvolnému rozkladu díky čemuž je eliminována nutnost sekundárního chirurgického zákroku pro vyjmutí fixátoru. Vhodné se zdají být kombinace různých kovů, jako jsou například železo, hořčík, zinek, mangan a jiné. Tyto látky by mohly mít dostatečné mechanické vlastnosti a zároveň také dobré korozní vlastnosti, které jsou nezbytné k degradaci. Důležité je také zajistit, aby degradace nebyla příliš rychlá nebo pomalá a aby nedocházelo k uvolňování toxických látek pro tělo.

Cílem této práce bylo pozorování dříve připravených vzorků, které byly vystaveny působení látek stimulující tělní tekutiny. V této práci byly také připraveny další vzorky, které obsahují odlišné složení a budou také podrobeny působení SBL (simulated body liquid).

2 PŘÍPRAVA VZORKŮ

Jako výchozí práškový kov pro přípravu vzorků bylo použito komerčně dostupné práškové železo. Byly vytvořeny také vzorky, ve kterých byla příměs nějakého dopantu – hořčík a zinek. Tyto kovy byly taktéž v práškové podobě. Vzhledem k požadavku porézní struktury, simulující spongiózní strukturu kosti, byla zvolena výroba pomocí replikace, při které vzniká vzorek pomocí replikace nosného materiálu. Jako nosný materiál byl zvolen polyuretan Bulpren S 28089 s počtem póru 60 PPI.

První sada vzorků byla připravena kombinací práškového kovu o hmotnosti 3 g a 4 ml destilované vody (případně i potřebné procento dopantu) za vzniku suspenze, do které byly polyuretanové vzorky namočeny. Po dostatečném obalení a nasáknutí byly přemístěny do sušičky (37 °C) na 24 hodin.

Druhá sada vzorků byla připravena s použitím 10% PVA pojiva, nebo 10% PS pojiva. Obdobně jako u první sady, i u této byly práškové kovy v potřebném zastoupení vmíchány do příslušného pojiva a

polyuretanové vzorky byly ve suspenzi ponořeny. Po dostatečném nasáknutí byly vytaženy a vloženy do sušičky po dobu 24 hodin při teplotě 37 °C.

Všechny vzorky byly následně slinovány, přičemž došlo k odstranění polymerní matrice a přeformování práškového kovu na pevnou formu. Slinování je realizováno difuzní cestou, z toho důvodu jsou rozhodujícími faktory teplota a čas.

Žíhání první sady bylo provedeno na Ústavu chemie materiálů Fakulty chemické VUTBR. Teplotní profil byl nastaven dle předchozích prací [3], tedy nejprve 2 hodiny na teplotě 450 °C a poté 1 hodina na teplotě 1120 °C. Pro snížení obsahu oxidů byla použita dusíková atmosféra. Vyžíhané vzorky jsou zobrazené na obrázku 1. Žíhání druhé sady vzorků bylo provedeno na Vysoké škole chemicko-technologické v Praze, přičemž pro různé složení vzorků byly použity odlišné teplotní profily. Seznam těchto vzorků je uveden v tabulce 1.

Č.	Složení	Atmosféra	Profil
1	FeSi15 + PVA	Dusík	450 °C (T = 2hod) -> 1120 °C (T = 1hod)
2	FeSi15 + PS	Dusík	450 °C (T = 2hod) -> 1120 °C (T = 1hod)
3	FeSi15 + H2O	Dusík	450 °C (T = 2hod) -> 1120 °C (T = 1hod)
4	FeSi15Zn15 + PVA	Dusík	400 °C (T = 3hod)
5	FeSi15Zn15 + PS	Dusík	400 °C (T = 3hod)
6	FeSi15Mg15 + PVA	Dusík	450 °C (T = 2hod) -> 600 °C (T = 1hod)
7	FeSi15Mg15 + PS	Dusík	450 °C (T = 2hod) -> 600 °C (T = 1hod)

Tabulka 1: Seznam vzorků včetně teplotních profilů z druhé sady



Obrázek 1: Vyžíhané vzorky na keramické podložce

3 EDS ANALÝZA A MAPOVÁNÍ PRVKŮ

Z důvodu vládních nařízení nebyla prozatím možná další práce se vzorky z druhé sady, nicméně vzorky z první sady byly po dobu několika měsíců uloženy ve fyziologických roztocích při teplotě 37 °C. U těchto vzorků byl v měsíčních cyklech měřen korozní potenciál, pH a vybrané vzorky byly také podrobeny EDS analýze před působením roztoků a po působení.

Vedlejší prvky se do vzorků mohly také dostat při manipulaci se vzorky v laboratoři nebo při výpalu. Při každém měření byla také zjištěna přítomnost uhlíku ve vzorku, který se nacházel ve struktuře polymeru. Důležitým prvkem je také kyslík, který hraje důležitou roli při oxidaci prvků, a tedy jeho degradaci [2], [3].

Vzorky v první sadě byly připraveny také v kombinaci Fe + Zn, ale prvková analýza ukázala, že po vypálení tyto vzorky zinek neobsahují. Vzhledem k tomu, že vzorky s obsahem zinku byly vypáleny stejným teplotním profilem jako ostatní vzorky, je možné, že při vyšších teplotách došlo k rozkladu zinku. Z tohoto důvodu byl pro vzorky z druhé sady obsahující Zn, resp. Mg zvolen odlišný teplotní profil jak u vzorků v první sadě.

U vzorků tvořených čistým železem byla po vypálení zjištěna přítomnost i jiných prvků než železa. Porovnání prvkové analýzy tohoto vzorku je v tabulce 2. Je patrné, že po působení SBL není ve vzorku zaznamenána síra a hořčík čili při degradaci došlo nejspíš k uvolnění a reakci s SBL. Naopak zaznamenán byl chlor, jehož původ je pravděpodobně z SBL. Uhlík před působením není zaznamenán, protože byl při analýze (manuálně) potlačen, nicméně lze předpokládat, že jeho obsah byl mírně vyšší než po působení SBL.

	Před působ. roztoků [wt. %]	Po působení SBL [wt. %]
Kyslík	30,04	31,72
Hořčík	6,41	-
Železo	57,85	58,12
Síra	1,22	_
Sodík	4,49	3,98
Chlor	-	2,67
Uhlík	-	3,51

Tabulka 2: Porovnání složení vzorku (Fe) před a po působení SBL

U vzorků s obsahem hořčíku (Fe + 2,5% Mg) bylo z prvkové analýzy zjištěno, že obsah hořčíku je menší než počáteční procentuální zastoupení (tabulka 3). Tento jev mohl být způsoben nehomogenním uspořádáním hořčíku ve vzorku. Pro zajištění lepší homogenity by bylo dobré využít míchacích zařízení, které by byly schopny dostatečně vmísit suspenzi práškových kovů do struktury polymeru.

	Před působ. roztoků [wt. %]	Po působení SBL [wt. %]
Kyslík	24,04	30,51
Hořčík	1,91	1,42
Železo	73,02	65,66
Sodík	-	1,42
Chlor	-	0,47
Uhlík	-	6,95
Vápník	1,03	-

Tabulka 3: Porovnání složení vzorku (Fe+5 % Mg) před a po působení SBL

V tabulce 3 je uvedeno procentuální zastoupení prvků naměřených pomocí EDS. Před vložením vzorku do SBL byl vzorek pravděpodobně znečištěn vápníkem, který se do vzorku mohl dostat manipulací. Je možné, že vápník byl špatně určen, protože po provedení EDS analýzy po namáhání v SBL nebyl identifikován ani ve stopovém množství. Naopak ve stopovém množství byly určeny prvky jako chlor a sodík, které byly nejspíš uvolněny z SBL.

4 SLEDOVÁNÍ KOROZNÍCH POTENCIÁLŮ

Jak již bylo zmíněno v kapitole 3, vzorky byly také podrobeny měření korozního potenciálu v měsíčních cyklech. Koroze, ať už její typ nebo rychlost, hrají důležitou roli při degradaci vzorku a ideální doba rozpadu vzorku by měla být 6-12 měsíců. Pro pochopení a regulaci koroze je důležité také pH, které ovlivňuje vznik pasivních vrstev – tuto závislost znázorňují Pourbaixovy diagramy. Korozní potenciál byl změřen před vložením vzorků do SBL a následovalo jejich uložení do SBL při teplotě 37 °C. Před měřením korozního potenciálu byly vzorky vyjmuty z SBL, očištěny destilovanou vodou a izopropylalkoholem tak, aby byl minimalizován vliv nečistot na měření. Některé vzorky byly také vystaveny působení roztoku NaCl, nicméně SBL se více podobá složení tělních tekutin, a proto lze říct, že naměřené hodnoty vzorků vystavených působení SBL jsou směrodatnější.

Měření bylo provedeno potenciostatickou metodou, která je založena na měření proudové hustoty v závislosti na potenciálu. Byl využit potenciostat AUTOLAB TYPE II v trojelektrodovém zapojení – pracovní (PIGE), pomocná (Pt elektroda) a referenční (Ag/AgCl) elektroda, přičemž elektrody byly ponořeny do 9% roztoku NaCl.



Graf 1: Naměřené korozní křivky pro vzorek 5 % Mg + Fe v SBL

U vzorku 5 % Mg + Fe uloženém v NaCl došlo při prvních měřeních k poklesu hodnot korozního potenciálu, což naznačuje zvýšenou tendenci ke korozi. Během působení roztoku NaCl na vzorek docházelo k tvorbě pasivních vrstev, které proces koroze postupně zpomalovaly a docházelo k poklesu korozního potenciálu. Po 375 dnech byl korozní potenciál u vzorku 5 % Mg + Fe větší než počáteční potenciál (-0,38 V vs -0,27 V). U vzorku v SBL došlo při druhém měření pravděpodobně vlivem fosforečnanových iontů k posunu ke kladnějším hodnotám. Posléze chloridové ionty způsobily snížení korozní odolnosti a poslední naměřený korozní potenciál byl -0,38 V.

U vzorku s čistým železem došlo k posunu z -0,54 V na hodnotu -0,33 V a z měření lze pozorovat trend růstu potenciálu do kladných hodnot. U vzorku s obsahem Mg (graf 1) byla během posledních měření zaznamenána stagnace kolem potenciálu -0,4 V a rozdíly mezi potenciály nebyly tak výrazné jako u vzorku s čistým železem (graf 2).



Graf 2: Naměřené korozní křivky pro vzorek Fe v SBL

5 ZÁVĚR

Cílem této práce byla příprava a zkoumání materiálů, které by byly vhodné pro výrobu biodegradabilních implantátů. Pro přípravu vzorků byla zvolena metoda replikace, u které bylo využito polyuretanové houby, která sloužila jako matrice pro suspenzi obsahující práškové kovy. Do připravené suspenze pojiva a práškových kovů byly vloženy vzorky z polyuretanové houby a po dostatečném nasátí byly vytaženy a vysušeny. Pro dokončení byly vzorky vyžíhány tak, aby došlo k vypálení polyuretanové matrice a zesintrování práškového kovu. Byly vytvořeny dvě sady vzorků, přičemž jedna sada byla vystavena dlouhodobému namáhání ve fyziologických roztocích a druhá sada, obsahující jiné složení, bude tomuto namáhání teprve vystavena. Pro stimulování tělního prostředí byly vzorky vloženy do fyziologických roztoků – klasický NaCl a simulated body liquid (SBL) - a uloženy do sušičky při teplotě 37 °C. V měsíčních cyklech byly měřeny korozní potenciály, pH, hmotnosti a u vybraných vzorků byla také provedena prvková analýza před a po působení fyziologických roztoků.

Dle naměřených hodnot jsou evidentní změny jak v hmotnosti vzorků, u kterých dochází ve většině k úbytku hmotnosti, tak také se mění korozní potenciály. U některých vzorků byla hmotnost stejná či vyšší než na začátku, což mohlo být způsobeno ulpěním korozních produktů ve vzorku. U korozních potenciálů dochází u většiny ze začátku k poklesu do záporných hodnot a posléze rychlost koroze zpomaluje, což je patrné z grafů, na kterých dochází k posunu korozních potenciálů ke kladným hodnotám. Důležitou roli hrají chloridové a fosforové ionty. Chloridové ionty narušují pasivní vrstvu, která vzniká při korozi, zatímco fosforové ionty ji naopak zlepšují.

PODĚKOVÁNÍ

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-20-6206.

REFERENCE

- HAVEROVÁ, ORIŇÁKOVÁ, ORIŇÁK, et al. An In Vitro Corrosion Study of Open Cell Iron Structures with PEG Coating for Bone Replacement Applications. Metals [online]. 28.6.2018, 2018(8) [cit. 2020-03-14]. DOI: 10.3390/met8070499. Dostupné z: www.mdpi.com/journal/metals
- [2] SEDLAŘÍKOVÁ, M.; VONDRÁK, J.; HÁVOVÁ, M.; KOŠÍČEK, A.; KADLEC, J. Preparation and corrosion of biodegradable iron based porous materials. In Advanced Batteries, Accumulators and Fuel Cells. ECS Transactions. Peddington USA: ECS Transaction, 2018. s. 1-6. ISSN: 1938-5862.
- [3] SEDLAŘÍKOVÁ, M.; VONDRÁK, J.; ČUDEK, V.; BINAR, T.; GALANOVÁ, Z. Chemical Corrosion of Porous Iron Alloys Prepared Pyrolytically. In *ECS Transaction. ECS Transactions*. Pennington USA: Electrochemical Society, 2018. p. 423-430. ISSN: 1938-5862.
- [4] SEDLAŘÍKOVÁ, M.; MIKA, M.; HRABOVSKÝ, J.; TKÁČOVÁ, T.; FAFILEK, G.; ČU-DEK, P. Iron magnesium materials for biodegradable implants Prepared by powder Metallurgy. In ECS transaction. ECS Transactions. 99. Pennington, USA: The Electrochemical Society, 2020. s. 249-253. ISSN: 1938-5862.

CUPROUS OXIDE AS A SEMICONDUCTOR PHOTOCATA-LYST USED FOR ORGANIC POLLUTANTS DEGRADATION

Filip Šmatlo

Master Degree Programme (2), FEEC BUT E-mail: xsmatl02@stud.feec.vutbr.cz

Supervised by: Petr Vanýsek

E-mail: petr.vanysek@gmail.com

Abstract: This paper contains discussion about photocatalytic water treatment. Contaminated water is purified by semiconductor photocatalytic material cuprous oxide exposed to light. Cuprous oxide is inexpensive and easy to synthesize and does not need any additional energy for its production. Cuprous oxide is synthesized purely chemically and the reaction requires commonly used chemicals as sodium hydroxide, cupric sulphate and ascorbic acid. The photocatalytic activity of cuprous oxide was examined on degradation of methyl orange dye that is used as a substitute for organic pollutants. The best achieved level of methyl orange degradation was 61.88 %. At the end of this paper the system for photocatalytic degradation was designed.

Keywords: cuprous oxide, Cu₂O, photocatalysis, water treatment, methyl orange

1 INTRODUCTION

In modern world the demand for innovation of reusable and renewable technologies is ascending more rapidly than ever. The use of fossil fuel as the main source of energy is not ecological; from this process a lot of toxic substances are generated. Photocatalytic materials have huge field of use in various environmental saving applications. These applications are CO_2 reduction, water splitting, self-cleaning surfaces or water purification.

Principle of CO_2 reduction methods is that the photocatalytic material can react with this greenhouse gas and produce harmless products from this reaction. Water splitting photocatalytic methods can produce hydrogen and oxygen from pure water. Hydrogen than can be used in fuel cells and oxygen produced by this method have huge window of use. The self-cleaning surfaces are mostly used in civil engineering. Facades of building in towns are being painted by photocatalytic material to stay clean as they were new. Last but not least the water purification is very important technology and is used all over the world.

All of these applications have one important thing in common: the use of photocatalytic material allows reduction of energy requirements. The only energy used in these applications is solar energy. With use of solar energy as the only source, these methods became environmentally safe, accessible and relatively cheap.

2 THEORY OF PHOTOCATALYSIS

The basic principle of photocatalysis is that when photon falls on surface of photocatalyst the photocatalytic material generates electron-hole pair. Basically, it is the same principle as in any other semiconductor and is commonly used in photovoltaic cells. The generated particles than travel through the volume of this material and when they end on the surface (separately) they can take part in two different reactions. Electrons cause oxidation of substance that is in contact with photocatalytic material and the holes cause reduction of this substance. The important parameter of these materials is band gap. This parameter says how much energy the photon must have to generate electron hole pair.



Figure 1: Principle of photocatalytic water treatment [1]

3 CUPROUS OXIDE

Cuprous oxide is monovalent and most stable oxide of copper metal. Band gap of cuprous oxide is 2.2 eV which corresponds with wavelength 563 nm. That means the wavelength of photon must be 563 nm and lower to generate electron-hole pair. In solar spectrum it is in part with highest fraction of solar energy and goes down to UV part of spectrum with higher energy irradiation. Cuprous oxide has some very useful properties such as environmental acceptability, very low toxicity, is inexpensive and simple to synthesize.

4 PREPARATION

Cuprous oxide nanoparticles were prepared by reducing CuSO₄ using ascorbic acid at room temperature. All used reagents were of analytical grade and were used without any further purification. In this procedure, 40 mL (0.5 mol/L) aqueous solution of NaOH was added into 20 mL aqueous solution of CuSO₄ (1.5 mol/L) with stirring. Then, 50 mL (0.1 mol/L) ascorbic acid aqueous solution was added dropwise into the above solution with vigorous stirring at room temperature. After 30 min reddish precipitate was observed. The particles were separated from the solution by filtration on a sinter glass filter. The product was washed by distilled water and absolute ethanol. The final product was dried in vacuum at 60 °C (more than 6 h) [2].

5 PRACTICAL CONSIDERATIONS AND EXPERIMENTS

For the measurement of photocatalytic effectivity of cuprous oxide used for organic pollutant degradation the methyl orange dye was used. In practice, the spectrophotometer Helios Delta by ThermoFisher Scientific measures visible absorption spectrum of methyl orange solution with photocatalyst cuprous oxide powder added. As a result the degradation ratio was calculated by equation (1).

$$D = \frac{A_0 - A_T}{A_0} \cdot 100\%$$
 (1)

- D ... degradation ratio [%]
- $A_0 \dots absorption of solution in time T = 0 min [-]$
- A_T ... absorption of solution after time T [-]

To see how much photocatalytic material is needed for methyl orange reduction, the different concentrations of cuprous oxide were measured. In Figure 2 the four different absorption spectra are shown. The first absorption spectrum shows methyl orange dye solution degradation without use of any cuprous oxide. The other three graphs show the absorption spectrum of the concentrations 0.5 g/L, 1.5 g/L and 2.0 g/L respectively. The shade of curves goes from lightest to darkest as the individual absorption spectra were measured. The measurement of each spectra was performed every 20 minutes.



Figure 2: Absorption spectra of methyl orange without photocatalyst (1), with cuprous oxide photocatalyst at concentrations of 0.5 g/L (2), 1.5 g/L (3) and 2.0 g/L (4)

Figure 3 shows time dependence of degradation ratio of measured solutions. From this it is obvious that the amount of used photocatalyst is important factor in methyl orange or organic pollutant degradation.



Figure 3: Degradation ratios of measured solutions

6 FUTURE APPLICATIONS

Cuprous oxide as photocatalyst for organic pollutant degradation has a big potential in future applications of water treatment. For this application a photocatalytic system was designed. In Figure 4 this system is illustrated. Basically, the photocatalytic material will be deposited on glass substrate and put into as called photocatalytic chamber. This chamber will be 3D printed with transparent cover so that the solar irradiation can go through and the photocatalytic substrate will be illuminated. The polluted solution of water goes through the system via water pump.



Figure 4: Photocatalytic methyl orange degradation system

7 CONCLUSION

This work contains basic information about photocatalysis and its use for organic pollutant degradation in water treatment applications. In practical and experimental part of this work the measurement of photocatalytic degradation of methyl orange dye was achieved. Results are that the highest ratio of 61.88 % was measured. The degradation ratio depends on concentration of photocatalytic material cuprous oxide. With this it is possible that higher degradation ratio of methyl orange can be observed. At the end of this work the future application is discussed. The photocatalytic methyl orange degradation system has high potential in increasing the effectivity of cuprous oxide as a photocatalyst.

REFERENCES

- [1] *Photocatalysis: Perspective, Mechanism, and Applications*, edited by Saiqa I., Nova Science Publishers, Incorporated, 2019. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/knav/detail.action?docID=5893815.
- [2] Zhang, X., Song, J., Jiao, J., & Mei, X. (2010). Preparation and photocatalytic activity of cuprous oxides. Solid State Sciences, 12(7), 1215–1219. doi:10.1016/j.solidstatesciences.2010.03.009

PERFORMANCE OF POLYVINYLIDENE FLUORIDE-CARBON NANOTUBES COMPOSITE

Denis Misiurev

Master Degree Programme (2. year), FEEC BUT E-mail: xmisiu00@stud.feec.vutbr.cz

Supervised by: Dinara Sobola

E-mail: surname@feec.vutbr.cz

Abstract: The Polyvinylidene Fluoride polymer it is unique materiel, which drown huge amount of attention in nowadays. Even though it has been known since 1969 year and many aspects of the polymer have been studied good enough, there are still much more that need to be studied further, because the Polyvinylidene Fluoride still did not yet reveal his full potential. The strongest side of Polyvinylidene Fluoride is the biggest piezoelectronic respond among all commercially available polymers. Polyvinylidene Fluoride is prepared in various forms: thin films, bulk samples, fibers. PVDF fibers attract the most attention because of high flexibility, lightweight, mechanical stability, chemical inertness.

Keywords: Polymer, carbon nanomaterial, phase composition

1 INTRODUCTION

Polyvinylidene Fluoride (PVDF) is a semicrystalline, non-reactive Thermoplastic that is approximately fifty presents amorphous. The Percentage of crystallinity is based on the Chain ordering defects. PVDF is a Plastic material belongs to fluoropolymer family [1].

Polyvinylidene Fluoride (PVDF) is a semi–crystalline, high purity thermoplastic fluoropolymer. PVDF is readily melt–processible and can be fabricated into parts by injection and compression molding [2].

Thanks to its excellent combination of properties and processability, PVDF has become the largest volume of fluoropolymers after PTFE.

PVDF is available commercially in a wide range of melt flow rates and with various additives to enhance processing or end use properties.

The atomic structure of PVDF is represented by monomer –CH2CF2–. The molecular weight of the monomer is between 16 and 17 kg/mol. The monomer –CH2CF2– has strong electrical dipole moment with regards to electronegativity of fluorine atoms. The monomer forms chains with perpendicular orientation of dipole moments. The PVDF's properties, as with most polymers, has been intensively studied. The main focus of research has been done on two most important aspects, the thirst is polymorphism and second is piezoelectronic respond [2].

The polymer has many electronic applications, especially as jacketing materials for plenum–rated cable used in voice and video devices and alarm systems. The low flame spread and smoke generation of PVDF is a prime asset in these applications.

Emerging applications of PVDF include fuel cell membranes, and components for aircraft.

2 MATERIALS

2.1 POLYVINYLIDENE FLUORIDE

PVDF is semicrystalline material crystallized into five different crystal phases: α , β , γ , δ , ε obtained by different process parameters [1].



Figure 1: Common phases of PVDF [4], SEM image of PVDF

The most stable at room temperature α phase is commonly synthesized among other phases, although the α phase is electrically neutral (non-polar), because of its antiparallel dipoles' alignment [1].

After applying mechanical stretching of alignment, high pressure, changing temperature of crystallization of melt the α phase can be transferred into oriented β phase [2]. The β phase is the most important due to its superior electrical, ferro–electrical and pyro–electrical properties in compare to other phases. Has been found the β phase not only represents huge potential due to the strongest piezoelectric respond among all PVDF's phases, but also the PVDF itself has the strongest piezoelectric response amongst all commercial polymers.

PVDF represents extraordinary mechanicals and chemical parameters such as:

Deformation resistance, absorption resistance, chemically resistant, stability to radiation, high working temperature (from -49° to 302° F), electrical insulator, radiation stability, high Curie point (217.4° F), high purity.

Such combination of properties makes PVDF useful for varieties of application where the polymer is used in:

Aerospace, biosensors, biotechnologies, pharmaceutical, microelectronics, pressure sensors, insulator for batteries.

Well-characterized properties make the PVDF leader of piezoelectric polymers.

2.2 CARBON NANOTUBES

Carbon nanotubes are unique carbon fibers with similar structure of fullerene. As other unique properties have been discovered such as remarkable mechanical and electronic properties unique Roman spectra, thermal conductivity, toughness, interest to nanotubes started to grow and their potential use in wide varieties of applications especially in nano-dimension electronics and medicine [3].



Figure 2: Structure of nanotube [3], produced nanotubes

The structurer of carbon nanotubes is similar to structure of 3D graphite and memorize honey camp. An ideal nanotube represents hexagonal network of carbon atoms rolled up to form seamless

cylinder. These cylinders are represented by layers of fullerene molecules. Base on amount these layers, nanotubes are subdivided on two types: single-walled nanotubes (SWNT) and multi-wall nanotubes (MWNT) [3].

3 ELECTROSPINNING OF PVDF FIBERS

There are many available commercial technologies of producing Nanofibers, however there is one technology which allow produce fine fibers [4].

Electrospinning is unique self–assembly technology offers wide variety of adjustable parameters to produce required fibers. Electrospinning is the thirst technology allowed to produce fibers which formation is given by electrostatic forces rather mechanical. Fibers obtained by using electrospinning are self–organizable [5]. Thanks to rather not completable basic setup, producing huge range of nanofibers and controlling their morphology by changing process parameters the electrospinning gained huge popularity and widely used in nowadays [4].

Parameter	Increasing	Decreasing	Orientation	Morphology	Quality	
Working Distance (WD)	There is strong possibility fibers will not be collect- ed on counter elec- trode.	With decreasing WD the solvent in precursor will not dry out complete- ly, thus thickened drops of precursor will accrue, fur- thermore fibers will not be proper- ly formed.	With decreasing working dis- tance.	It appears, that with decreasing of WD, still wet precursor col- lects even more fibers around itself which leads to creation thickened clus- ters of fibers.	Thickened drops defect that appears at small WD can significantly lower piezoelectronic and tribo- electric effect due to its electroneutrality.	
Voltage	High voltage causes stretching of Taylor cone by higher generated charge which leads to fast- er formation of fibers from smaller amount of precur- sor [6].	Lover voltage shrinks Taylor cone, thus bigger amount of precur- sor can easily degrade on the top of needle and compromise entire process by clog- ging of needle [6].	Does not change	With increasing voltage produc- es fibers become thinner. In case of suc- cessful for- mation [6].	Both thick and thin fibers are used in different ap- plications.	
Number of rotations	By increasing num- bers of rotation counter electrode more fibers are produced.	With lower num- ber of rotations less fibers are produced.	High number of rotations al- lowed to pro- duce oriented (parallel) fiber; chaotically ori- entated fibers are produces by lower number of rotations.	The morphology of parallel and chaotically fi- bers is pretty much same.	In some cases, with high number of rotations, pro- duced fibers have tenden- tion to create areas with nonhomogeneous thick- ness of produced fibers, although fibers produced by lower number of rota- tions have homogeneous thickness.	
Needle diameter	redle inter It has been proved needles with smaller diameter clogging less often rather with bigger one. By using ning Electron Microscopy has been proven, fibers produced by using needles with smaller diameter hav est diameters and vice versa. However, in different cases has been found no correlation of fibers and diameters [7].				bigger one. By using Scan- smaller diameter have tini- elation of fibers and needle	

Table 1: Summarization of processing conditions on produced fibers

Sample	Parameters		D33 (µC/N)	SEM image
29–1,2	Precursor: P WD: Doses: Voltage: Humidity: Rotations: Hot air mode:	VDF 20% 275 20 cm 40–50µl/min 50kV 30->20% 2000 min ⁻¹ On	≈ 20	

Table 2: Characterization table of produced samples

4 THESIS PERSPECTIVES

The main object of the semestral/diploma thesis is based on comparing properties of pure PVDF material with PVDF–CNT composite. The most important aspect of the entire thesis is producing pure PVDF material using technology calls electrospinning. Once material is produced it needs to be evaluated base on piezoelectronic respond. Since PVDF has the strongest respond among all polymers, piezoelectronic respond resp. measurement of d33 parameter is the easiest way to prove functionality of produced material. Parameters of material with the strongest respond will be used to produce composite based on pure PVDF material and Carbon nanotubes. By comparing pure PVDF and PVDF–CNT composite we will be able to prove whether or not Nanotubes has any effect on properties of PVDF.

The table 1 represents parameters which effect the produced material. The effect of singe parameter has been defined experimentally. To determent if there are any changes in Morphology of the produced material the morphology has been studied using Scanning electron Microscopy (SEM). It has been proven number of rotations effect orientation.

The second table contains summary of parameters of the best sample. These parameters will be used to produce PVDF–CNT composite. Unfortunately, the situation around coronavirus makes further progress problematic.

The technology calls Focused Ion Beam (FIB) will provide more detail information regarding cross-section of single fibers. It has been found that certain samples produce by same parameters has sings of hollows inside of fibers which give us a ground to try to produce coaxial fibers using CNT. However, there are still unknowns and speculations about mechanism of their formation which require further studding and running more experiments.



Figure 3: Hollows inside a fiber

5 ACNOWLEDGMENTS

I appreciate for support of this work for Czech foundation Agency (GA 19-17457S).

6 CONCLUSION

Polyvinylidene fluoride one of important polymer with huge scientific interests. PVDF is represented in five crystalline phases: α , β , γ , δ and ϵ . Besides having the strongest piezo- respond of β phase, it also combines pyro-and ferroelectric properties, which is important in many applications. The material has found his place in many areas we all familiar with such as solar panels, batteries, alarm systems, medical masks, filters etc. however, the material has more to offer to all aspects of applied science, manufacture, and medicine.

REFERENCES

- [1] PVDF piezoelectric polymers: characterization and application to thermal energy harvesting [online], available on: https://tel.archives-ouvertes.fr/tel-01241414
- [2] PROCESSING AND CHARACTERIZATION OF PVDF, PVDF-TrFE, AND PVDF-TrFE-PZT COMPOSITES [online], JARED JAMES STROYAN available on: https://www.pdfdrive.com/
- [3] Peter J. F. Harris, Carbon Nanotube Science: Synthesis, Properties and Applications, Cambridge University Press, UK, 2009
- [4] JOACHIM H. Wendorff, SEEMA Agarwal, and Andreas Greiner, Electrospinning: Materials, Processing, and Applications, Singapore, 2012
- [5] *Kazutoshi Fujihara, Zuwei Ma, Wee Eong Teo, Teik-Cheng Lim, Seeram Ramakrishna* An Introduction to Electrospinning and Nanofibers, World Scientific Publishing Co. Pte. Ltd. Singapore, 2005
- [6] *Chitral Jayasanka Angammana*, a Study of the Effects of Solution and Process Parameters on the Electrospinning Process and Nanofiber Morphology, Waterloo, Ontario, Canada, 2011
- [7] Fabrication, Polarization of Electrospun Polyvinylidene Fluoride Electret Fibers and Effect on Capturing Nanoscale Solid Aerosols [online], available on: https://www.mdpi.com/

DETERMINATION OF THE MOST SUITABLE RATIO OF CATHODE MATERIALS FOR THE LITHIUM-SULFUR BATTERY SYSTEM

Petra Benešová

Master Degree Programme (2), FEEC BUT E-mail: xbenes45@vutbr.cz

Supervised by: Kamil Jaššo

E-mail: xjasso00@vutbr.cz

Abstract: This paper presents the topic of lithium-sulfur batteries with a main focus on the influence of different ratio of the cathode materials on the final electrochemical performance of the cell. The theoretical part of the article provides a comprehensive overview of the Li-S technology, main advantages and challenges related to the practical use of Li-S systems, and briefly mentions the typical cathode materials. The objective of the experimental part is to provide a comparison of different ratios of cathode materials in terms of their electrochemical performance, taking into consideration the physicochemical properties of prepared materials.

Keywords: Battery, cell, lithium-sulfur, sulfur, cathode material

1 ÚVOD

Využívání elektrické energie je pro lidstvo v dnešní době tak samozřejmou a neodmyslitelnou součástí každodenního života, že si jej v každodenním shonu téměř neuvědomujeme. S tím souvisí i potřeba tuto energii skladovat, k čemuž slouží právě akumulátory - elektrochemické zdroje elektrické energie. Od dob, kdy Alessandro Volta v roce 1800 sestavil první skutečnou baterii, prošla oblast elektrochemických zdrojů obrovským vývojem a v dnešní době máme k dispozici nezměrné množství baterií o různých chemismech, velikostech a dalších parametrech. Posledním milníkem v této oblasti, který představoval skutečně zásadní pokrok, bylo uvedení prvního lithno-iontového článku firmou Sony v roce 1991. Právě lithno-iontovým bateriím byla z důvodu jejich popularity a širokého komerčního využití v posledních třech desetiletích věnována velká pozornost ze strany vědecké společnosti, a ačkoli se díky nesčetným optimalizacím současné Li-ion baterie blíží svým technologickým limitům, jejich parametry stále nejsou plně dostačující pro energetické požadavky pro příští generaci elektromobilů, HEV a oblasti Urban Air Mobility. Vhodnou alternativou právě pro tyto aplikace by mohly být akumulátory lithium-síra (Li-S). Nízká cena, vysoká dostupnost a velká teoretická specifická kapacita síry činí akumulátory Li-S jedním z nejperspektivnějších kandidátů pro využití ve výše zmíněných aplikacích, nicméně v současné době vyvíjené systémy stále trpí řadou problémů, které brání jejich širšímu využití a kterým je nutné věnovat pozornost.

2 AKUMULÁTORY LITHIUM-SÍRA

Ačkoli je technologie Li-S považována za relativně nový koncept, první zmínka o Li-S bateriích se datuje až do 60. let 20. století. Jako katodový materiál prvních Li-S akumulátorů byla použita čistá síra a anoda byla tvořena kovovým lithiem, pak ale kvůli četným problémům s jejich provozem zájem na dlouhou dobu utichl. V posledních letech se nicméně k systémům Li-S opět vrací pozornost, a to zejména díky jejich vysoké teoretické hustotě energie ~2600 Wh/kg, tedy až téměř 10x vyšší než energetická hustota dnešních komerčně užívaných akumulátorů. Také teoretická měrná kapacita 1672 mAh/g, nízká cena a vysoká dostupnost síry činí tyto akumulátory atraktivním kandidátem pro využití nejen v oblasti elektromobilů, ale např. i nositelné elektroniky [1] [2].

2.1 PROBLÉMY SPOJENÉ S LI-S TECHNOLOGIÍ

Jednou z výzev, se kterými se současně zkoumané Li-S články často potýkají, je ztráta aktivního materiálu spojená se vznikem rozpustných polysulfidů. Vznik těchto polysulfidů během první fáze vybíjecího procesu je na jednu stranu žádoucím jevem, jelikož jsou tímto způsobem postupně odhalovány vnitřní vrstvy síry, na druhou stranu však dochází k nežádoucímu snižování kapacity článku. Po prvotní reakci síry s lithnými ionty totiž dochází k rozpouštění polysulfidů v kapalném elektrolytu a vlivem koncentračního gradientu k jejich difúzi směrem od kladné elektrody, což má za následek jednak postupné nasycování a zvyšování viskozity elektrolytu, ale zejména ztrátu aktivního materiálu kladné elektrody a s tím spojený pokles kapacity [2].

Důsledkem difúze vyšších polysulfidů je také tzv. shuttle-effect. Vyšší polysulfidy rozpuštěné v elektrolytu mohou migrovat až k záporné elektrodě, kde dochází k jejich reakci s kovovým lithiem za vzniku nerozpustných polysulfidů nižších řádů. Ty opět difundují směrem ke kladné elektrodě a jsou oxidovány za tvorby nižších rozpustných forem, které mohou migrovat zpět k anodě a při vybíjení článku reagovat s lithiem, přičemž se tento proces cyklicky opakuje. Tento jev má několik nežádoucích následků, zejména ztrátu aktivního materiálu a tudíž i vratné kapacity nebo intenzivní korozi záporné elektrody [1] [2].

Jelikož procesy probíhající při nabíjení a vybíjení zahrnují fázové přechody pevná látka - kapalina - pevná látka, dochází v Li-S akumulátorech vlivem opakovaného usazování a rozpouštění k významným změnám morfologie povrchu, a také k objemové změně až 79 % vlivem rozdílných hustot elementární síry a jednotlivých polysulfidů. Tyto jevy mohou vést ke ztrátě kontaktu aktivního materiálu s elektrodou. Jedním z možných řešení je využití matric nebo elastických či porézních substrátů, které zlepšují soudržnost kladné elektrody a pokud jsou tvořeny vodivými materiály, zlepšující také elektrický kontakt. Jiným řešením je omezení nabíjecího napětí pouze na hodnotu \sim 2,0-2,1 V, čímž lze zabránit vzniku vyšších polysulfidů a omezit shuttle-effect [3] [4].

Další nevýhodou Li-S akumulátorů je nevodivý charakter elementární síry a polysulfidů, přechod elektronů mezi proudovým kolektorem a aktivním materiálem tedy probíhá pouze v omezené míře a není tak využit plný potenciál vysoké teoretické kapacity síry, limitovány jsou i nabíjecí/vybíjecí proudy. Tento nedostatek mohou do určité míry kompenzovat vodivé příměsi či již zmíněné porézní substráty [3] [4].

2.2 KATODOVÉ MATERIÁLY PRO AKUMULÁTORY LITHIUM-SÍRA

Velké množství v dnešní době vyvíjených a zkoumaných katodových materiálů je na bázi uhlíku. Uhlík se pro tyto účely používá v mnoha různých formách, od porézních struktur až po nanokompozity na bázi grafenu jako např. nanotrubičky, nanovlákna apod. Primární funkcí uhlíku je zvodivění elektrody, jelikož síra sama o sobě je velmi špatný vodič, nicméně v dnešní době zájem směruje k takovým uhlíkovým materiálům, které jsou schopny poskytnout i hostitelskou strukturu, ve které jsou molekuly síry a polysulfidy zachyceny tak, aby byl minimalizován shuttle effect [5].

Jako katodové materiály lze použít i některé sulfidy a oxidy kovů. Sulfidy kovů jednak poskytují zdroj sulfidových aniontů S²⁻, ale také fungují jako kostra umožňující absorbovat molekuly síry a polysulfidy. Stejně tak oxidy kovů, které sice nejsou kvůli malé aktivní ploše ideálními kandidáty, nicméně jsou schopny zabraňovat rozpouštění polysulfidů do elektrolytu. Pro tyto účely bylo zkoumáno velké množství oxidů, např. TiO₂, MnO₂, Al₂O₃ a další. Jedním z přístupů je např. vytvoření duté "kostry", uvnitř níž jsou síra a polysulfidy uloženy [5].

Jednou ze skupin organických materiálů pro katody Li-S akumulátorů jsou vodivé polymery jako např. polypyrrol, polyanilin aj. Tyto materiály jsou zpravidla pružné, tudíž nedochází k poškození struktury vlivem případných objemových změn, ale současně struktura umožňuje difúzi Li⁺ iontů k síře uvnitř. Reakce vodivé matrice s polysulfidy zabraňuje shuttle efektu, a to jak případě použití polymerů jako ochranné povrchové vrstvy katody, tak i v případě, že organický polymer tvoří celé tělo elektrody [5].

3 EXPERIMENTÁLNÍ ČÁST

Cílem experimentu bylo ověřit vliv poměru katodového materiálů na výsledné vlastnosti měřených elektrochemických článků. Jako výchozí byl vybrán poměr 64 % síra, 32 % uhlík (Super P) a 4 % pojivo CMC (karboxymethylcelulóza), v dalších vzorcích je postupně zvyšován obsah síry. Podle teoretických předpokladů by vzorky s vyšším obsahem síry měly disponovat vyšší specifickou kapacitou, nicméně jak již bylo zmíněno výše, nevodivý charakter síry způsobuje zhoršený přenos náboje a měřené články vykazují vyšší vnitřní odpor.

V rámci dosavadního experimentu byly naměřeny tři vzorky katodových materiálů ve složení:

64 % S + 32 % C + 4 % CMC, 76 % S + 20 % C + 4 % CMC, 88 % S + 8 % C + 4 % CMC

Na základě těchto poměrů byly smícháním práškových složek s vodou připraveny elektrodové pasty, které byly míchány v planetovém mlýnu a posléze naneseny na hliníkovou folii, z níž byly po vysušení vyseknuty kruhové elektrody. Měřicí cely byly skládány v rukavicovém boxu s ochrannou Ar atmosférou, jako záporná elektroda bylo použito kovové lithium. Po sestavení cel byla vždy provedena totožná série měření na multikanálovém potenciostatu Biologic VMP3 s cílem prozkoumat a porovnat elektrochemické vlastnosti složených článků.

Součástí měření byl zátěžový test v rámci metody GCPL, tedy cyklického nabíjení a vybíjení článku konstantním proudem ve stanovených potenciálových mezích, potenciálové okno bylo nastaveno v rozmezí 1,8-3,0 V. Obrázek 1 zobrazuje porovnání volumetrických vybíjecích kapacit zkoumaných článků, které byly naměřeny během 50 cyklů GCPL s proměnným vybíjecím proudem. Je zřejmé, že rozdílný obsah síry v jednotlivých pastách má na výsledné kapacity znatelný vliv. Nejvyšší kapacitu vykazuje vzorek obsahující 88 % S, důvodem je nejvyšší obsah aktivního materiálu v celkovém objemu elektrody. Pokud ovšem přepočítáme volumetrickou kapacitu na kapacitu měrnou, tedy na jednotku hmotnosti (viz Obrázek 2), nejlepší vlastnosti vykazuje vzorek s obsahem 64 % S. Tato skutečnost je s nejvyšší pravděpodobností způsobena tím, že se zvyšujícím se obsahem síry v pastě klesá obsah uhlíku, který zajišťuje zvodivění a tedy zlepšuje přenos náboje. To má za následek zvýšení vnitřního odporu článku, což se negativně projevuje na měrné kapacitě.



Obrázek 1: Volumetrické vybíjecí kapacity zkoumaných článků



Obrázek 2: Měrné vybíjecí kapacity zkoumaných článků

Z obrázku 1 také vyplývá, že kapacita se s lineární změnou obsahu síry (krok mezi jednotlivými poměry je vždy 12 % celkové hmotnosti) nemění lineárně, rozdíl mezi poměrem 64 % S a 76 % S je výrazně vyšší než změna mezi 76 % S a 88 % S. Z dosud naměřených dat se tedy zdá, že pravděpodobně existuje jistá maximální hranice dosažitelné kapacity. Tento jev bude možné ověřit po získání více dat, resp. po změření vlastností cel s dalšími poměry katodových materiálů.

4 ZÁVĚR

Za účelem vyhodnocení vlivu poměru jednotlivých složek katodového materiálu na elektrochemické vlastnosti článků Li-S byly navrženy tři různé poměry, na základě nichž byly připraveny kladné elektrody a sestavené cely prošly sérií měření. Naměřené výsledky ukazují, že nejvyšší kapacitu na cm² vykazuje vzorek s 88% obsahem síry, což je způsobeno tím, že v celkovém objemu elektrody obsahuje největší množství aktivního materiálu. V oblasti měrných vybíjecích kapacit ovšem vykazuje nejlepší vlastnosti nejnižší poměr obsahující 64 % S + 32 % C + 4 % CMC. I přes nejvyšší procentní obsah síry 88 % vykazuje vzorek nejnižší měrnou kapacitu, což lze na základě naměřených dat a dostupné literatury odůvodnit právě nízkou vodivostí síry a špatným přenosem náboje.

REFERENCE

- LI, T., BAI, X., GULZAR, U., BAI, Y., et al. A Comprehensive Understanding of Lithium-Sulfur Battery Technology. Advanced Functional Materials [online]. 2019, 29: 190130(1-56). DOI 10.1002/adfm.201901730.
- [2] ZHANG, X., XIE, H., KIM, C., ZAGHIB, K., MAUGER, A., JULIEN, C. Advances in lithium-sulfur batteries. Materials Science and Engineering R [online]. 2017, 121: 1-27. DOI 10.1016/j.mser.2017.09.001.
- [3] FANG, X., PENG, H. A Revolution in Electrodes: Recent Progress in Rechargeable Lithium–Sulfur Batteries. Small [online]. 2015, 11(13): 1488-151. DOI 10.1002/smll.201402354.
- [4] YIN, Y., XIN, S., GUO, Y., WAN, L. Lithium–Sulfur Batteries: Electrochemistry, Materials, and Prospects. Angewandte Chemie International Edition [online]. 2013, 52: 13186-13200. DOI 10.1002/anie.201304762.
- [5] EFTEKHARI, A., KIM, D. Cathode materials for lithium-sulfur batteries: a practical perspective. Journal of Materials Chemistry A [online]. 2017, 5: 17734–17776. DOI: 10.1039/c7ta00799j.

PVDF — AN IDEAL CANDIDATE FOR USE IN NANOGENERATORS

Tatiana Pisarenko

Master's Degree Program (2), FEEC BUT E-mail: xpisar04@stud.feec.vutbr.cz

Supervised by: Dinara Sobola E-mail: sobola@vutbr.cz

Abstract: In this work, the PVDF composite, also known as polyvinylidene fluoride in the form of thin nanofibres, was created. Subsequently, a single-fibre characterization was performed, which proves its piezoelectric properties and describes its structure. Electron microscopy, atomic force microscopy and X-ray photoelectron spectroscopy were chosen as characterization methods. The discussion in this paper deals with the ability of these fibres to use PVDF as a nanogenerator.

Keywords: PVDF, electrospinning, nanofibres, characterization, PFM, STEM, SEM, XPS

1 INTRODUCTION

Polyvinylidene fluoride (PVDF) is a very promising semi-crystalline polymer for various industries, such as medicine, textile industry, electronics, energy harvesting, and many more. Widespread use is in the form of nanofibers. Its unique properties lie in the strong ability to polarise, i.e., the ability to exhibit a piezoelectric effect. Piezoelectric responses are characterized mainly by polar phases: γ , δ , and especially the most decisive β phase [1]. The individual phases differ in the configuration of the homopolymer chain. It very much depends on the parameters of the fibre for the successful formation of the β phase. Thus, if the final product results are to give satisfactory performance, it is desirable to focus directly on the processes already occurring in the single fibre itself [2, 3].

2 MATERIALS, METHODS AND RESULTS

In this experiment, the synthetic 20 % PVDF 275 DMSO/AC fibres were produced by electrospinning method in Contipro 4SPIN device. The properties of the produced nanofibres depend on many parameters set during spinning. For analysis, the samples were created using a continuous rotating collector. These samples differed mainly in the diverse number of spins per minute of the collector drum. The number of 300 rpm was chosen for the first sample and 2000 rpm for the second sample type. Other essential parameters during spinning were temperature 24 °C, humidity 51 %, and 20 cm distance of the emitter from the collector at a voltage of 50 kV. The collector used was a 17 GA needle with a set dose of flow of 30 µl/min. Different collector speeds affect the resulting β phase of the fibres, their organisation, and the average thickness.

The shape of the fibre was described by NTEGRA Prima atomic force microscope (AFM). The scan was performed in the region of $10 \times 10 \,\mu\text{m}$ on a single fibre. Scan velocity was from 8.04 to $13.94 \,\mu\text{m/s}$. It can be seen that not every fibre is necessarily circular in its diametre. It is visible in Figure 1 that both fibres are affected at speeds of 300 rpm and 2000 rpm. The resulting shape can be influenced, for example, by the collector voltage or the type of needle used.



Figure 1: Three and two-dimensional imaging of the fibre at drum speeds of a) 300 rpm and b) 2000 rpm using the AFM microscope.

The piezoelectric response of the fibre were confirmed by piezoelectric force microscopy (PFM) measurements. This was used for the fibre with a more decisive β phase – at the collector cylinder speed of 2000 rpm [4]. The change in polarisation is most visible along the edges of the fibre, described in Figure 2. Here, the colour of these edges changes from light green to deep red in the form of a thin line. The voltage was varied from -5 to 5 V [5].



Figure 2: PFM method used in observation with AFM. A differently polarised fibre is observed mainly at its edges during biasing of a) -5 V, b) 0 V, and c) 5 V. Dipole changes are observed, which are influenced by hysteresis.

Specific parameters of PVDF fibres were observed by scanning electron microscopes (SEM). The first – scanning transmission electron microscope (STEM) – FEI Helios NanoLab 660, monitored the fibre composition using a high-angle annular dark-field (HAADF) imaging detector. The accelerating voltage was set to 30 kV and the current 50 pA, which is a sufficient value for electrons that pass through the polymer fibre.

Figure 3a shows the fibre produced at 300 rpm and Figure 3b at 2000 rpm. The purplish tone of the fibre highlights a different structure that needs to be emphasized. In addition to the variety in different fibre thickness, which is different at first glance, the fibre structure affects not only the strength but also the piezoelectric properties. It is also important to mention the imperfections in the form of balls in the fibre in Figure 3b.



Figure 3: Internal fibre structure at drum speeds a) 300 rpm and b) 2000 rpm observed by STEM and HAADF method.

The second – microscope with a focused ion beam (FIB) was the Tescan Lyra3. Here, the fibres were not observed longitudinally, as in the previous case, but as a cross-section. The accelerating voltage was set to 5 kV for SEM observation. When cutting, the FIB high voltage was 30 kV, and for precise fibre separation, the current was 50 pA. This is a small value for cutting, as the fibre had to be cut very finely to avoid defects. A thin layer of 20 nm thick carbon was also coated before observing due to fibre fixation and preventing charge accumulation.

The cross-section in Figure 4 from a view field of 10 to $0.6 \,\mu m$ confirms the fibre's porous structure, which was also indicated during the observation on STEM. The purplish colour here does not describe significant changes in structure but a thin carbon layer [6]. A similar porous structure occurs in both types of fibres.



Figure 4: Cross-section of one fibre divided into three different magnifications. The yellow border shows the area that was investigated in the following figure. In Figure a) several fibres can be seen, in Figure b) it is already focused on the specific fibre, and its core in Figure c).

Observations of the structure and electrical properties were also supplemented by AXIS Supra Kratos X-ray photoelectron spectroscop (XPS) to determine the fibre's elemental composition and its chemical state. It was focused on the study of the carbon band bonds of the C1s spectrum in Figure 5. The most significant change is almost half the decrease of the C - C/C - H peak [7].



Figure 5: XPS spectra represented the C1s high-resolution energy band's expressed by an envelope, fitted and divided into individual peaks of certain bonds in the material. A more substantial C - C/C - H binding (red peak) is recognised on the fibre produced at a) 300 rpm compared to the fibre produced at b) 2000 rpm.

3 CONCLUSION

This work described the structural, electrical, and chemical characteristics of PVDF. Standard results from publications are based on the entire sheet of nanofibers. The repeatability of these results is often challenging due to the whole specimen's uniformity and many other parameters, which must be constant. The main core of this work was to understand the function of PVDF based on the single fibre and its unique behaviour and properties at the nanoscale, thanks to which it is possible to create a complex structure of an electric energy generator.

ACKNOWLEDGEMENT

Research described in the paper was financially supported by the Grant Agency of Czech Republic under project No. 19-17457S, and by the Ministry of Education, Youth and Sports of the Czech Republic under the project CEITEC 2020 (LQ1601). Part of the work was carried out with the support of CEITEC Nano Research Infrastructure supported by MEYS CR (LM2018110).

REFERENCES

- [1] ISLAM, A.; KHAN, A.; SHAKIR, F.; ISLAM, K. Strengthening of β polymorph in PVDF/FLG and PVDF/GO nanocomposites. *Materials Research Express*, 2020, 7, issue 1. ISSN: 2053-1591.
- [2] PISARENKO, T. Characterization of PVDF nanofibers created by the electrospinning method. Proceedings I of the 26th Conference STUDENT EEICT 2020. Brno: 2020. s. 287-291. ISBN: 978-80-214-5867-3.
- [3] XIN, Y.; ZHU, J.; SUN, H.; XU, Y.; LIU, T.; QIAN, C. A brief review on piezoelectric PVDF nanofibres prepared by electrospinning. *Ferroelectrics*, 2018, 526, issue 1, p. 140–151. ISSN: 0015-0193.

- [4] UEBERSCHLAG, P. PVDF piezoelectric polymer. *Sensor Review*, 2001, 21, issue 2, p. 118–126. ISSN: 0260-2288.
- [5] SENCADAS, V., C. RIBEIRO, I. K. BDIKIN, A. L. KHOLKIN a S. LANCEROS-MENDEZ. Local piezoelectric response of single poly(vinylidene fluoride) electrospun fibers. *Physica status solidi* (a). 2012, 209(12), 2605–2609. ISSN 18626300.
- [6] MOTAMEDI, A.S.; MIRZADEH, H.; HAJIESMAEILBAIGI, F. et al. Effect of electrospinning parameters on morphological properties of PVDF nanofibrous scaffolds. *Progress in Biomaterials*, 2017, 44, issue 3, p. 113–123. ISSN: 2194-0517.
- [7] KASPAR, Pavel, Dinara SOBOLA, Klára ČÁSTKOVÁ, Alexandr KNÁPEK, Daniel BURDA, Farid ORUDZHEV, Rashid DALLAEV, Pavel TOFEL, Tomáš TRČKA, Lubomír GRMELA a Zdeněk HADAŠ. Characterization of Polyvinylidene Fluoride (PVDF) Electrospun Fibers Doped by Carbon Flakes. *Polymers*. 2020, 12(12), 2766.

Magisterské projekty

Silnoproudá elektrotechnika a elektroenergetika

RADIAL BASIS FUNCTION ASSISTED INTERPOLATION FOR ELECTRICAL MACHINE ANALYSIS

Vítězslav Halašta

Master Degree Programme (1), FEEC BUT E-mail: xhalas11@stud.feec.vutbr.cz

Supervised by: Jan Bárta

E-mail: bartaj@feec.vutbr.cz

Abstract: The purpose of this paper is to introduce an interpolation adopting radial basis functions, that was used to analyze permanent magnet synchronous motor for aerospace application. Data used for this interpolation were obtained by performing an electromagnetic calculation using the finite element method.

Keywords: radial basis functions, interpolation, analysis, permanent magnet synchronous motor, aerospace, electromagnetic calculation, finite element method

1 ÚVOD

Využití radiálních bazových funkcí (RBF) je podle [1] zajímavou metodou interpolace, kterou lze využít například pro neuronové sítě, řešení parciálních diferenciálních rovnic nebo rekonstrukci obrazu. V této publikaci budou nejdříve představeny radiální bazové funkce, které budou následně aplikovány v interpolaci případové studie zahrnující analýzu synchronního stroje s permanentními magnety navrženého pro aplikaci v letectví.

Podle [2] je možné RBF interpolant s(x) pro N uzlových bodů hodnot $\{f_k\}_{k=1}^N$ umístěných v $\{x_k\}_{k=1}^N$ definovat následovně

$$s(x) = \sum_{k=1}^{N} c_k \varphi(\|x - x_k\|)$$
(1)

kde výraz $\varphi(||x - x_k||)$ lze zjednodušit jako $\varphi(r)$ a odpovídá příslušné bazové funkci. Jako příklad lze uvést následující thin-plate spline popsaný v [3] vhodný pro interpolaci funkce dvou proměnných.

$$\varphi(r) = r^2 \log\left(r\right) \tag{2}$$

Pro lepší vizualizaci RBF je proces interpolace vyobrazen na Obrázku 1. Radiální bazová funkce typu thin-plate spline byla využita při analýze stroje a následné interpolaci dat, která je popsána v třetí kapitole.



Obrázek 1: Interpolace s využitím radiálních bazových funkcí [2, upraveno].

2 PŘÍPADOVÁ STUDIE

2.1 ANALYZOVANÝ STROJ

Analyzovaný stroj vychází ze stroje popsaného v publikaci [4]. Jedná se o synchronní stroj s permanentními magnety, který byl navržen pro letecký průmysl, konkrétně pro elektrohydraulický systém. Stroj pracuje s rotorem ponořeným v kapalině, má tři fáze, devět drážek, šest pólů a dosahuje momentu 13,75 Nm při 6000 otáčkách za minutu.

Materiály nebyly autorem v publikaci [4] definovány. Jejich volba proběhla s ohledem na podmínky dané normou RTCA-DO160G, která zohledňuje aspekty uvedené v [5] pro aplikaci v letectví. Plechy statoru a rotoru byly realizovány z materiálu M250-35A a byly využity permanentní magnety ze slitiny Samarium-Kobalt, konkrétně magnety Recoma® 18, které zajistily vynikající odolnost vůči vysokým teplotám [6].

2.2 ELEKTROMAGNETICKÝ VÝPOČET

Elektromagnetický výpočet byl proveden pomocí metody konečných prvků v programu Ansys Maxwell 2D. Model včetně sítě pro výpočet metodou konečných prvků je vyobrazen na Obrázku 2 vlevo. Na Obrázku 2 vpravo je vyobrazeno rozložení magnetické indukce ve stroji. Díky symetrii je zřejmé, že bylo možné simulovat pouze třetinu stroje.

Elektromagnetický výpočet byl proveden pro různé hodnoty proudů v d- a q- ose. Cílem bylo popsat chování stroje při změně těchto proudů pomocí 3D grafů. Dále bylo zjišťováno chování stroje při změně otáček a momentu a cílem bylo sestrojit účinnostní mapu pro oblast konstantního momentu.

Výpočet probíhal diskrétně pro různé hodnoty proudů, momentů, otáček, a proto bylo nutné provést interpolaci vypočtených hodnot pro získání většího množství dat pro tvorbu grafických výstupů. K tomu byl využit skript v programovacím jazyku Python, který provedl interpolaci pomocí RBF.





3 VÝSLEDKY INTERPOLACE POMOCÍ RBF

3.1 ANALÝZA POMOCÍ PROUDŮ V D- A Q- OSE

Pro popis chování stroje při změně proudů v d- a q- ose byla vybrána množina těchto proudů, pro které proběhl elektromagnetický výpočet. Tyto body byly následně interpolovány pomocí RBF a došlo k vytvoření 3D grafů vyobrazující vzájemné závislosti mezi momentem, toky, a proudy v d- a q- ose. Vytvořené závislosti jsou vyobrazeny na Obrázku 3. Interpolace odpovídala teoretickému předpokladu, tedy že pro synchronní stroj s permanentními magnety na povrchu byl momentotvorný proud v q- ose. Proudy poté tvořily magnetické toky ve svých příslušných osách. Změna proudu v d- ose dle předpokladu vedla pouze ke změně buzení stroje.



Obrázek 3: 3D grafy vytvořené s využitím RBF interpolace

3.2 ANALÝZA ÚČINNOSTI A TVORBA ÚČINNOSTNÍ MAPY

Na rozdíl od analýzy proudů v d- a q- ose byly výstupem této analýzy 2D grafy s izoliniemi, které zachycují oblasti určité úrovně hodnot dané veličiny. Při analýze účinnosti nebyly data, ze kterých probíhala interpolace srovnány přibližně do roviny, jak tomu bylo v minulém případě, což značně ovlivnilo stabilitu výsledků provedené interpolace. Pokud by byla nevhodně zvolena konkrétní radiální bazová funkce došlo k rozkmitání interpolovaných hodnot daleko mimo množinu hodnot, které jsou předpokládané. Tuto nestabilitu také způsoboval nepoměr mezi měřítky veličin vyobrazených na různých osách, což se projevilo nejvíce u účinnostní mapy, kde se na y-ose nachází hodnoty v rozsahu desítek Nm a na x-ose hodnoty v rozsahu tisíců otáček za minutu.

Jako velmi účinným nástrojem k eliminaci této nestability se ukázal převod veličin do poměrných jednotek vztažených k jmenovitým hodnotám. Tak bylo zajištěno, že obě osy budou v řádech jednotek příslušných veličin. Výsledkem byly 2D izoliniové grafy vyobrazené na Obrázku 4.



Obrázek 4: Závislost účinnosti na proudech v d- a q- ose (vlevo) a účinnostní mapa pro oblast konstantního momentu (vpravo).

Pomocí interpolace RBF bylo znázorněno, jak závisí účinnost stroje na proudech v d- a q- ose (Obrázek 4 vlevo). Zároveň bylo využito výše zmíněného převodu do poměrných jednotek pro sestrojení účinnostní mapy (Obrázek 4 vpravo), kde bylo pomocí RBF funkcí ověřeno, že stroj dosahuje velmi dobré elektromagnetické účinnosti v širokém spektru otáček a dodávaného momentu.

4 ZÁVĚR

Tato publikace ukázala, že radiální bazové funkce mají velký potenciál pro uplatnění v oblasti analýzy elektrických strojů, jelikož jsou schopné interpolovat data, která by se výpočtem metodou konečných prvků získávala velmi pomalu. Další jejich výhodou je, že lze jednoduše přizpůsobit podobu výsledného interpolantu pomocí volby konkrétní radiální bazové funkce tak, aby odpovídal teoretickému předpokladu. V této studii bylo nejlepších výsledků dosaženo pomocí thin-plate spline, ostatní (Gaussova, kubická, lineární aj.) nebyly pro tento konkrétní případ vhodné. Nevýhodou interpolace pomocí radiálních bazových funkcí je jejich nestabilita, která způsobuje, že interpolované hodnoty se nachází mimo rozsah požadovaných nebo fyzikálně možných hodnot. Tomuto problému se dá předejít vhodnou volbou radiální bazové funkce. Pozitivní vliv na stabilitu interpolace má také převod dat do poměrných jednotek, což zajistí, že jsou spolu porovnávány hodnoty blízkých nebo stejných řádů. Radiální bazové funkce lze velmi jednoduše využít při interpolaci pomocí jazyku Python.

PODĚKOVÁNÍ

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-20-6379).

REFERENCE

- SKALA, Vaclav. Fast Interpolation and Approximation of Scattered Multidimensional and Dynamic Data Using Radial Basis Functions [online]. Plzen: WSEAS TRANSACTIONS on MATHEMATICS, 2013 [cit. 2021-03-11]. ISSN 2224-2880. Dostupné z: http://afrodita.zcu.cz/~skala/PUBL/PUBL_2013/2013_RBF-Scattered.pdf
- FLYER, Natasha a Grady B. WRIGHT. A radial basis function method for the shallow water equations on a sphere. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* [online]. 2009, s. 1949-1976 [cit. 2021-03-11]. ISSN 1364-5021. Dostupné z: https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=1011&context=math_facpubs doi:10.1098/rspa.2009.0033
- [3] WEIDENHOLZER, G., S. SILBER, G. JUNGMAYR, G. BRAMERDORFER, H. GRAB-NER a W. AMRHEIN. A flux-based PMSM motor model using RBF interpolation for timestepping simulations. In: *International Electric Machines & Drives Conference* [online]. Chicago, IL, USA: IEEE, 2013, 2013, s. 1418-1423 [cit. 2021-03-11]. ISBN 978-1-4673-4974-1. Dostupné z: doi:10.1109/IEMDC.2013.6556323
- [4] POWELL, David James. *Modelling of high power density electrical machines for aerospace*. University of Sheffield, 2003. Disertační práce. University of Sheffield, department of Electronic and Electrical Engineering.
- [5] RTCA DO-160G for Airborne Equipment | DO-160. DO-160 [online]. Washington, DC: RTCA [cit. 2021-03-25]. Dostupné z: https://do160.org/rtca-do-160g/
- [6] Recoma®: The complete range of SmCo5 and Sm2Co17 alloys. In: Arnold Magnetic Technologies / Global Magnet Manufacturer [online]. Arnold Magnetic Technologies, c2020, c2014 [cit. 2021-03-11]. Dostupné z: https://www.arnoldmagnetics.com/wp-content/uploads/2017/10/Recoma-Combined-160301.pdf

DEVELOPMENT OF THE LOW POWER SYNCHRONOUS MACHINE

Josef Laštovička

Master Degree Programme (1), FEEC BUT E-mail: xlasto02@stud.feec.vutbr.cz

Supervised by: Jan Bárta

E-mail: bartaj@feec.vutbr.cz

Abstract: This paper summarizes the development of a salient pole synchronous machine with a wound rotor. It contains required parameters, analytical calculation using Python programming language, optimalization of defined parameters, finite element method simulations and their results. The last part of the paper is focused on the 3D model of developed machine and the design evaluation. Development was successful and the described work procedure can be repeated at another university.

Keywords: Synchronous machine, Electromagnetic design, Salient pole, Wound rotor

1 ÚVOD

Výuka na většině univerzit se skládá z přednášek, cvičení odborného základu a laboratorních cvičení. I přes velmi dobrou vybavenost univerzitních laboratoří často v technické sféře chybí laboratorní vzorky, na kterých je možné měření provádět. V dnešní době jsou rovněž kladeny vysoké nároky na bezpečnost práce v laboratořích, a proto jsou pracovní stoly často provozovány na bezpečné napětí. To však značně omezuje výběr na trhu elektrických strojů a často je nemožné stroj na toto napětí zakoupit ze sériové výroby a musí být navržen individuálně.

Tento článek řeší vývoj synchronního stroje s vinutým rotorem o nízkém výkonu, který bude provozován na hladině bezpečného napětí. Součástí vývoje byl elektromagnetický návrh, návrh usazení svazku do sériově vyráběné kostry, tvorba výrobní dokumentace, výběr sběracího ústrojí, komunikace s externími dodavateli a redukce pořizovací ceny. Vývoj byl úspěšné ukončen a v současné době se pracuje na výrobě funkčního vzorku stroje.

2 ANALYTICKÝ NÁVRH STROJE

V prvotní fázi návrhu bylo vycházeno z obecných poznatků o synchronních strojích obsažených v publikacích [1] a [2]. Tyto knihy byly prioritně použity při návrhu statoru stroje. Při návrhu rotoru s vyniklými póly bylo čerpáno především z literatury [3] a [4].

2.1 POŽADAVKY NA NÁVRH

Jelikož se jednalo o účelový vývoj stroje pro laboratoř, byly na něj kladeny specifické požadavky. Limitujícím faktorem bylo provozní napětí, které je z důvodu bezpečnosti v laboratoři omezeno na 42 V sdružené hodnoty. Dalším omezujícím faktorem byla osová výška stroje, která bylo zvolena pouze 90 mm, jelikož většina měřících přípravků tuto osovou výšku vyžaduje. Neméně důležitým faktorem byla cena funkčního vzorku, která byla díky použití kostry, ložiskových štítů a plechu statoru ze sériově vyráběného stroje snížena na minimum. O splnění dalších požadavků pojednává postup, kterým se zabývá tento článek. Všechny požadované parametry stroje jsou uvedeny v Tabulce 1.

Parametr	Hodnota	Parametr	Hodnota	
Jmenovité sdružené napětí	42 V	Počet pólů rotoru	4	
Jmenovitý moment	5 Nm	Odhadovaný účiník	0,92	
Počet fází	3	Odhadovaná účinnost	90 %	
Frekvence	50 Hz	Osová výška	90 mm	
Tabulka 1. Požadované parametry stroje				

abulka 1:Požadované parametry stroje.

2.2 PROGRAM V JAZYCE PYTHON

Pro urychlení vývoje stroje a možnost použití i při návrhu dalších strojů byl celý analytický návrhový proces naprogramován v jazyce Python. Program byl pomocí proměnných sestaven tak, aby uživatel zadal požadované parametry a program následně vypočítal první variantu stroje. Následně uživatel posoudí návrh, provede případné změny ve zvolených konstantách a znovu provede výpočet. Je-li s analytickým výsledkem spokojený, může provést zadání výstupních parametrů výpočtu například do programu ANSYS RMxprt, kde je možné si danou geometrii jednoduše vykreslit a dále s ní pracovat. Tento postup shrnuje blokové schéma na Obrázku 1.



Obrázek 1: Blokové schéma návrhu pomocí programu v jazyce Python.

2.3 NÁVRH V PROGRAMU ANSYS RMXPRT

Při návrhu takto malého stroje byl velmi limitující prostor pro vinutí rotoru. Bylo vytvořeno několik variant s rozdílnou šířkou jádra pólu rotoru a z nich byla vybrána nejvhodnější varianta, která tvořila určitý kompromis. Z důvodu zúžení jádra pólu v něm stoupla vypočtená magnetická indukce na 1,2 T a z původně navržených 9 drátů na vodič bylo třeba použít 7.

Dále byla provedena optimalizace radiální délky vzduchové mezery v závislosti na budicím proudu a účinnosti. Z Obrázku 2 je patrné, že musela být zvolena vhodná hodnota s ohledem na oba parametry, tato hodnota činila 0,6 mm, kdy je poměrně malý budicí proud, ale vyšší účinnost. Druhým sledovaným parametrem bylo posunutí zakřivení pólu rotoru v závislosti na harmonickém zkreslení magnetického toku ve vzduchové mezeře. Zde byla zvolena nejvhodnější hodnota 4,2 mm.



Obrázek 2: Grafické závislosti optimalizace vzduchové mezery (vlevo) a posunutí poloměru zakřivení pólu (vpravo).

3 VÝPOČET METODOU KONEČNÝCH PRVKŮ

3.1 CHOD NAPRÁZDNO

První simulace pomocí metody konečných prvků byla zaměřena na chod stroje naprázdno. Jako materiál plechů statoru a rotoru byla zvolena ocel M470-50A. Do programu byla tedy vložena její magnetizační charakteristika z katalogu výrobce. Sledované výsledky byly především průběhy indukovaných napětí a jejich efektivní hodnoty, které korespondovali s předpokládanými hodnotami.

Důležitým výstupem byl také průběh magnetické indukce v železe stroje. Ten byl podle Obrázku 3 rovnoměrný a maximální hodnota v jádru pólu dosahovala akceptovatelných hodnot.

Posledním důležitým výstupem byla charakteristika naprázdno. Ta ukazuje, že hodnoty jmenovitého napětí bylo dosaženo při budicím proudu přibližně 8,5 A, což je příznivé, jelikož vypočtený budicí proud při jmenovitém zatížení byl 10 A. Pokud by hodnota 10 A byla dosažena již při chodu naprázdno, pak by při zatížení proud ještě vzrostl a stroj by se mohl přehřívat.



Obrázek 3: Rozložení magnetické indukce v železe stroje při chodu naprázdno (vlevo) a charakteristika naprázdno (vpravo).

3.2 CHOD PŘI JMENOVITÉM ZATÍŽENÍ

Druhá ze simulací stroje byla zaměřena na chod při jmenovitém zatížení. V tomto stavu byly sledovány především hodnoty magnetické indukce v pólu rotoru. Z grafického rozložení na Obrázku 4 je patrné, že bude docházet k mírnému přesycování v levé části hlavy pólu rotoru, které je způsobené reakcí kotvy. Toto mírné přesycení je akceptovatelné.

Druhou významnou charakteristikou byla momentová charakteristika, která ukazuje, že pracovní bod stroje je nastaven optimálně a dovoluje mírné přetížení stroje bez rizika ztráty synchronismu.

Z průběhu odebíraného proudu ze sítě vychází, že jeho maximální hodnota by byla dosažena při zátěžném úhlu 100°, to by ale bylo již při velkém přetížení s následnou ztrátou synchronismu.



Obrázek 4: Rozložení magnetické indukce v železe stroje při zatížení (vlevo), momentová charakteristika a odebíraný proud ze sítě (vpravo).

4 KONSTRUKČNÍ USPOŘÁDÁNÍ STROJE

V rámci této práce proběhl také vývoj konstrukce stroje. Vyvinuté provedení je vyobrazeno ve formě 3D modelu na Obrázku 5. Tento model ověřuje rozměrové proporce stroje a slouží především k ověření umístění sběracího ústrojí a tvorbě potřebné výrobní dokumentace včetně výpočetního listu.

V rámci modelování bylo navrženo i prodloužení statorové kostry z důvodu umístění sběracího ústrojí. Toto ústrojí bylo vybráno z katalogu externího dodavatele a bude po výrobě stroje namontováno. Pro edukativní účely stroje v laboratořích bylo navrženo i průhledové okénko z plexiskla, které umožní pohled na kartáče stroje.



Obrázek 5: Celkový pohled na model vyvinutého stroje (vlevo) a řez kostrou stroje (vpravo).

5 ZÁVĚR

Tento článek pojednal o návrhu a simulacích synchronního stroje s vyniklými póly a vývoji funkčního vzorku stroje. Součásti návrhu byla i výrobní dokumentace včetně výpočtového listu. Všechny zadané parametry byly dle simulací splněny, a proto byl stroj zadán do výroby. V současné době se tedy pracuje na zhotovení funkčního vzorku, který bude v budoucnu umístěn v laboratoři Fakulty elektrotechniky a komunikačních technologií VUT v Brně.

Předložený návrhový postup a vyvinutá konstrukce je reprodukovatelná a umožňuje tak využití pro návrh a výrobu synchronního stroje pro malé napětí i jinými univerzitními pracovišti.

PODĚKOVÁNÍ

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-20-6379).

REFERENCE

- [1] SEN, Paresh Chandra. *Principles of Electric Machines and Power Electronics*, Third edition. Hoboken: John Wiley, 2014. ISBN 978-1-118-07887-7.
- [2] PYRHONEN, Juha, Tapani JOKINEN a Valeria HRABOVCOVÁ. *Design of rotating electrical machines*, Second edition. Chichester, West Sussex, United Kingdom: Wiley, 2014. ISBN 9781118581575.
- [3] MILLER, Tymothy John Eastham, David A. STATON. *Electric Machine Design using SPEED and Motor-CAD*, Raleigh, 2013.
- [4] KOPYLOV, Igor Petrovič a kol. *Stavba elektrických strojů*, Praha: SNTL, 1988.
SIMULATION OF ELECTROPORATION PROCESS IN STENT OCLUSSION THERAPY

Martin Hemzal

Master Degree Programme (5), FEEC BUT E-mail: xhemza03@stud.feec.vutbr.cz

Supervised by: Veronika Novotná

E-mail: novotnav@vutbr.cz

Abstract: The main focus of this work is to compare the simulations of electroporation in the biliary tract with and without the stent. The aim of this paper is to determine the ablated areas using different combinations of electrodes and different amplitudes of applied voltage. The resuluts of the simulations shows distribution of intnesity of electric field and joule heating. Presented simulations are done using finite element method (FEM) software COMSOL MULTYPHISIC 5.4.

Keywords: Biliary tract, electroporation, FEM, IRE, oclussion, stent

1 ÚVOD

U pacientů s maligní obstrukcí žlučových cest, kteří vyžadují perkutánní drenáž, je nejvhodnější léčbou umístění samoexpandibilního kovového stentu (SEMS). Při aplikaci SEMS dochází dle [1] u 5–40 % pacientů k uzávěru stentu neboli jeho okluzi. Mezi nejčastější příčiny uzávěru stentu patří prorůstaní tumoru stentem (68,8 %), přerůstání okrajů stentu tumorem (11,7 %), hyperplazie, tvorba žlučového kalu a migrace stentu (10,4 %). Při léčbě okluze stentu se dnes využívá mnoho metod, vždy zaleží, čím je okluze způsobena. V některých případech je dostatečné mechanické vyčištění pomocí balónkového katetru. Docházi-li k přerůstání, či prorůstaní stentu je možné aplikovat druhý stent. Jako vhodnější se však jeví použít minimálně invazivních ablačních technik. Mezi ty patří radiofrekvenční ablace, která využívá k destrukci buněk vysokofrekvenční proud v rozsahu 450 MHz až 1000 MHz, nebo mikrovlnná ablace, která využívá dielektrické hystereze k produkci tepla a následnému zničení buněk. Nevýhodou těchto technik je, že dochází k zahřívání tkáně a může dojít k zničení zdravých buněk. Předmětem tohoto článku je zkoumání použití netermální irreverzibilní elektroporace při řešení okluze stentu.

2 ELEKTROPORACE

Elektroporace je jev, při kterém dochází k navyšování membránové propustnosti pro makromolekuly a ionty. Toho je dosaženo pomocí krátkých vysokonapěťových pulzů. Navyšování permeability je spojeno s formací nanopórů v membránách buněk. Za určitých podmínek dochází k rozpadu buněčné membrány a k smrti buňky. Tato metoda se nazývá nevratná neboli ireverzibilní elektroporace (IRE).

Na základě několika experimentů bylo určeno, že elektroporace buňky nastane v okamžik dosažení určité hodnoty membránového napětí. Tato hodnota závisí na vlastnostech použitých pulzů (počet tvar, délka pulzu atd.). Většina autorů uvádí hranici membránového napětí pro elektroporaci mezi 200 mV a 1 V [2]. Aby bylo dosaženo nevratné elektroporace je nutné dosáhnout určité intenzity elektrického pole v daném místě, kdy pro jaterní buňky je tato hodnota 800 V cm⁻¹ [1]. Jelikož buněčná smrt při IRE je způsobena změnou propustnosti buněčné membrány a netermálního procesu, nemá vliv na mimobuněčné části, jako důležité krevní řečiště a nervové spojení, nezpůsobuje zánět ani nežádoucí imunoreakci. Toto umožňuje léčbu nádorů, které by byly normálně považovány za neoperovatelné kvůli blízkosti k těmto strukturám [1].

3 STENT

Stenty jsou trubicovité zdravotnické prostředky z plastu nebo kovu viz Obr.1, které mají zajistit průchodnost blokované trubicovité struktury. V rámci tohoto článku je simulována elektroporace okluze samoexpandibilního stentu (SEMS). Tento typ stentu se vyrábí z různých kovových slitin (nitinol, Elgiloy, platinol). Tyto materiály jsou použity, aby bylo dosaženo adekvátní radiální expanzivní síly, aniž by byla omezena flexibilita a přizpůsobivost žlučovodu. Velikosti SEMS se pohybují v rozmezí 4 až 12 cm v délce a 6 až 10 mm v průměru.



Obrázek 1: Samoexpandiblní stent [3]



Obrázek 2: unikátní balónkový katetr [4]

4 MODEL

V rámci tohoto článku byly vytvořeny dva modely jaterní tkáně o výšce 5 cm a poloměru 4 cm. Jaterní tkání prochází žlučovod a dvě cévy. Do lumenu žlučovodu byl v obou případech vložen unikátní balónkový katetr viz Obr.2. Na jeho povrchu se nachází tři zlaté elektrody svírající mezi sebou úhel 120° s aktivní plochou 10 mm². Vnitřní a vnější plochy byly brány jako elektricky nevodivé. Elektrický potenciál byl vždy přiváděn mezi dvě aktivní elektrody.

První model slouží jako referenční, kdy byl do lumenu žlučovodu vložen jen unikátní balónkový katetr. U druhého modelu viz Obr.3 a Obr.4 byl do lumenu vložen navíc nitinolový stent, kterým prorůstá tkáň. Simulace byly provedeny pomocí metody konečných prvků (FEM) v softwaru COMSOL MULTIPHYSICS 5.4. Parametry pulzů a materiálové konstanty použitých tkání (játra, žlučovod, krev) [5] a nitinolového stentu [6] jsou uvedeny v Tab.1.

Parametr	Hodnota	Jednotka
Výška segmentu jaterní tkáně	5	cm
Elektrická vodivost jaterní tkáně	0,0227	S/m
Relativní permitivita jaterní tkáně	$1,57 \cdot 10^{7}$	-
Relativní permitivita krve	$5,26 \cdot 10^3$	-
Elektrická vodivost krve	0,66	S/m
Elektrická vodivost žlučovodu	1,4	S/m
Relativní permitivita žlučovodu	120	-
Napětí	1500	V
Elektrická vodivost nitinolového stentu	1.10^{6}	S/m
Relativní permitivita nitinolového stentu	1	-

Tabulka 1: Hodnoty materiálových konstant použitých při simulaci



Obrázek 3: Drátový model se stentem



Obrázek 4: Detail žlučovodu s vloženým stentem a balónkovým katetrem

4.1 VÝSLEDKY SIMULACÍ

Na Obr.5, 6 a 7 je zobrazeno rozložení intenzity elektrického pole (kV·cm⁻¹) v obou modelech. V modelu bez stentu Obr.5 se elektrické pole šíří v celém objemu a může dojít k nežádoucímu poškození tkáně. Zatímco u modelu se stentem Obr.5 je šíření elektrického pole omezeno do objemu tkáně mezi katetrem a stentem, jelikož stent zde působí jako Faradayova klec a chrání zdravou tkáň před poškozením. Aby došlo k odstranění nežádoucí tkáně v celém objemu, bylo by třeba provést proceduru IRE několikrát a natáčet diskutovaný balónkový katetr.



Obrázek 5: Rozložení intenzity elektrického pole v modelu bez stentu.

Obrázek 6: Rozložení intenzity elektrického pole v modelu se stentem.

Obrázek 7: Detail rozložení intenzity elektrického pole v modelu se stentem.

Na Obr.8, 9 a 10 je zobrazeno rozložení Jouleových ztrát (kW·cm⁻³). Aby mohlo být určeno, zdali dojde k tepelnému poškození tkáně je třeba provést tepelnou analýzu. Cévy by měly zůstat nepoškozené, protože by měly být chlazeny protékající krví. Z Obr. 6 a 7 je patrné, že Jouleovův efekt je omezen na oblast mezi stentem a balónkovým katetrem, kde je nežádoucí tkáň, což je výhodné, protože by nemělo docházet zahřívání zdravé tkáně. Zároveň je potřebné provést simulace, kdy by se elektrody katetru dotýkaly vodivého stentu a došlo ke zkratu, jelikož by mohlo dojít k tepelnému poškození tkáně.



Obrázek 8: Rozložení Jouleových ztrát v modelu bez stent.

Obrázek 9: Rozložení Jouleových ztrát v modelu se stentem.

)brázek 10: Detail rozložení Jouleových ztrát v modelu se stentem.

5 ZÁVĚR

Metoda IRE se jeví jako slibná terapie při řešení okluze stentu. Díky tomu, že je stent vodivý, tak se elektrické pole uzavírá mezi stentem a balónkovým katetrem, právě v oblasti nežádoucí tkáně. Z toho vyplývá, že okolní zdravá tkáň a citlivé struktury nejsou touto metodou ohroženy. Tato strategie je zejména výhodná v oblastech s komplikovaným přístupem a velkým výskytem cév, nervů a dalších citlivých tkání.

Na základě Obr. 5 lze předpokládat, že hranice pro elektroporaci 800 V·cm⁻¹ je dosažena ve větším objemu než v případě s nitinolovým stentem. To znamená, že může dojít k nežádoucí destrukci zdravé tkáně. V případě simulace bez stentu jsou Jouleovy ztráty největší v nejbližší oblasti elektrod a žlučovodu mezi nimi, ale zasahují i do okolní tkáně. Zatímco v případě simulace se stentem jsou Jouleovy ztráty koncentrované jen mezi elektrodou a stentem.

Další simulace budou vytvořeny na základě požadavků z lékařské praxe. Bude se pracovat s různými konfiguracemi kontaktu elektrod, stentu a tkáně. Stěžejním úkolem bude tranzientní tepelná analýza a určení tepelného namáhání tkáně. Získané výsledky jsou přínosné pro pochopení účinků zkoumané metody a jsou velmi hodnotné pro další výzkum.

REFERENCE

- [1] RUBINSKY, Boris. *Irreversible Electroporation*. 1. Berlin: Springer-Verlag, 2010. ISBN 978-3-642-05419-8.
- [2] RIDTITID, Wiriyaporn a Rungsun RERKNIMITR. Management of an occluded biliary metallic stent. World J Gastrointest Endosc. 2012, 4(5), 157-161. Dostupné z: doi:10.4253/wjge.v4.i5.157
- [3] BONASTENT® biliary stent. In: *Aseptinmed* [online]. [cit. 2021-03-08]. Dostupné z: http://www.aseptinmed.fr/en/stents-metalliques/74-biliary-stent.html
- [4] NOVOTNÁ, Veronika. ANALÝZA ELEKTRICKÝCH A TEPELNÝCH JEVŮ PŘI ELEK-TROPORACI. Brno, 2019. Dizertační práce. Vysoké učení technické. Vedoucí práce Dalibor Červinka.
- [5] Dielectric Properties. *ITIS foundation* [online]. [cit. 2021-01-04]. Dostupné z: https://itis.swiss/virtual-population/tissue-properties/database/dielectric-properties/
- [6] D. Rafiroiu, R. V. Ciupa, A. Iancu, A. Lazar, I. Tiseanu and T. Craciunescu, "Numerical analysis of the electric field and temperature changes around carotid stents," 2011 7TH IN-TERNATIONAL

IMPLEMENTATION OF RES AND ELECTROMOBILITY ON ELECTRICAL PARAMETERS IN MV DISTRIBUTION NET

Stanislav Navratil

Master Degree Programme (2), FEEC BUT E-mail: stanislav.navratil@vut.cz

Supervised by: Michal Ptacek

E-mail: ptacekm@vut.cz

Abstract: This paper is focused on the demonstration of the implementation of decentralized distributed energy sources (mainly renewable energy sources) and electromobility technologies to electrical parameters in Czech medium voltage distribution network. Proposed implementation scenarios are based on EU and national strategic plans for period 2020-2040.

Keywords: renewable sources, electromobility, smart grids, decentralized distributed energy systems

1 ÚVOD

Elektrizační soustava ČR je již od svého vzniku koncipována pro provoz s myšlenkou centrální výroby elektrické energie pomocí klasických energetických zdrojů. V současné době ale napříč tomu dochází k nezanedbatelné implementaci decentrálních zdrojů (DECE), které se často vyznačují nestálým charakterem dodávky. V kombinaci s tím je stále více prosazován trend přechodu na dopravu pomocí elektromobility, což pro distribuční sítě (DS) klade nové technické požadavky. Cílem tohoto příspěvku je upozornit na potenciální problémy vycházející z těchto koncepčních změn a porovnat jejich možné způsoby řešení tak, aby byly zaručeny dostatečné standardy dodávky elektrické energie. Konkrétně se příspěvek zabývá těmito jevy především na úrovni hladiny vysokého napětí (VN).

2 MOŽNÉ KOMPLIKACE V BUDOUCNU A JEJICH NAVRHOVANÁ ŘEŠENÍ

Aby bylo možné kvantifikovat dopady zmíněného vývoje v energetice, je třeba nejprve vytyčit odhad rozvoje implementace těchto prvků v následujících desetiletích. Za tímto účelem byly použity dílčí studie [1], [2] zpracované v rámci Národního akčního plánu pro chytré sítě (NAP SG) [3].

Na základě zmíněné studie [2] je odhadován celkový instalovaný výkon decentrálních zdrojů v roce 2040 přibližně 10,4 GW, což odpovídá nárůstu na téměř 270 % oproti aktuálnímu stavu (2020). Velká část tohoto nárůstů (téměř 46 %) je očekávána v podobě fotovoltaických elektráren (FVE) připojených do sítí nízkého napětí (NN). Nezanedbatelný nárůst se dá dále očekávat u FVE připojovaných do sítí VN, kde je předpokládáno navýšení o téměř 850 MW do roku 2040.

V oblasti rozvoje elektromobility je v roce 2040 odhadován maximální soudobý příkon během denní špičky v ČR způsobený dobíjením na 3,6 GW (varianta vysokého vývoje) [1]. To pro představu odpovídá stavu, kdy by 1/2 všech vozidel v ČR (cca 3 000 000) využívala elektrický pohon.

2.1 KOMPLIKACE S ROZŠÍŘENÍM DECENTRALIZOVANÉ VÝROBY

Jedním z hlavních potenciálních problémů související s nárůstem podílu DECE (především obnovitelných zdrojů) na celkové spotřebě elektřiny, na který aktuální NAP SG [3] poukazuje je snížení rezervy flexibility v elektrizační soustavě (ES). Flexibilita zdrojů v ES je v energetice důležitá pro dodržení základního pravidla, kdy se výroba musí rovnat spotřebě, jinak hrozí změna kvality dodávky energie (odchylka frekvence). Obnovitelné zdroje energie (OZE) jsou ale většinou motivovány provozní podporou pro maximalizaci své výroby, tím se ale snižuje jejich potenciál ke kladné flexibilitě (navýšení výkonu v době odběrových špiček). Tyto negativní vlivy by v budoucnu mohli být alespoň částečně vyváženy pomocí většího uplatnění technologií pro akumulaci elektrické energie. NAP SG se konkrétně zabývá potenciálem využití bateriových uložišť, ty se ale kvůli svého charakteru hodí spíše pro krátkodobou akumulaci (např. Automaticky ovládaný proces obnovení frekvence - aFRP). Pro dlouhodobou akumulaci se momentálně jako vhodnější jeví technologie typu Power to X, její potenciál je ale v současně době teprve zkoumán na pilotních projektech.

Problematika snížení flexibility ES vlivem DECE se projevuje dále v systému automatického frekvenčního odlehčování (SAFO). Ten v současnosti funguje na principu odpojování vývodů (zátěže) na hladinách VN, čímž se zajišťuje rovnovážný stav při nedostatečné výrobě v případě větších systémových poruch. Vlivem rozšíření DECE na hladinách NN a VN však hrozí, že SAFO bude odpojovat se zátěží i velké množství výroby, což může výrazně snížit efektivitu daného systému. Jako řešení tohoto problému se nabízí větší rozšíření dálkově ovládaných spínacích prvků (např. recloser), které by umožnily odpojovat jen vývody na kterých převažuje charakter odběru [3].

Již v současné době jsou přijata opatření, díky nimž jsou mnohé negativní vlivy nárůstu podílu OZE v ES výrazně sníženy. Tato opatření byla zavedena především prostřednictvím Pravidel provozování distribučních soustav (PPDS), která definují závazné požadavky na připojované zdroje. Mezi ně patří například schopnost těchto zdrojů poskytovat statickou a dynamickou podporu sítě. Do statické podpory sítě spadají především principy využívající dodávky nebo spotřeby jalového výkonu daným zdrojem pro udržování napětí (U/Q regulace) ve stanovených mezích, za normálního stavu. Tím lze zabránit ku příkladu možnému nadpětí, které by mohlo vzniknout při změnách toku výkonu v sítích NN a VN. Dynamická podpora sítě je pak schopnost jednotlivých výroben v určité míře překlenout náhlé poklesy napájecího napětí, způsobené především v důsledku symetrických i nesymetrických zkratů. Díky tomu se ve značné míře zabrání hromadnému odpojování dalších zdrojů, což by v extrémních případech mohlo vést až k rozpadu ES.

2.2 KOMPLIKACE S ROZŠÍŘENÍM ELEKTROMOBILITY

Již v toce 2025 jsou dle NAP SG [3] očekávány nezanedbatelné výkonové dopady na DS, především v podobě odběrových špiček a nedostatečné infrastruktury pro dobíjecí stanice. Vznik zmíněných odběrových špiček je způsoben podobností denních harmonogramů majitelů elektromobilů (EV), kdy vzniká požadavek na nabíjení v podobnou denní dobu (např. po příjezdu domů z práce). Řešením je buď cenově motivovat zákazníky k nabíjení mimo exponovaný čas (např. pomocí tarifů), nebo implementovat inteligentní technologie využívající vzájemné komunikace mezi dobíjecí stanicí a elektrizační soustavou k přizpůsobení aktuálního dobíjecího výkonu. Tato opatření již začínají být postupně uplatňována například prostřednictvím požadavku PPDS, kde je specifikováno, že dobíjecí stanice s nabíjecím výkonem překračujícím 22 kW musí obsahovat rozhraní pro sledování nebo řízení celkového odběru.

Elektromobilita dle NAP SG poskytuje na druhou stranu i potenciální možnost rozvoje trhu s flexibilitou prostřednictvím využití akumulačních kapacit elektromobilů. V této oblasti je však mnoho problému jako nadbytečné opotřebení baterií, nebo technická náročnost na dobíjecí stanice. Proto by bylo pravděpodobně třeba motivovat (např. finančně) majitelé EV k poskytnutí bateriových kapacit k těmto účelům.

3 ANALÝZA VLIVU IMPLEMENTACE OZE A NABÍJECÍCH STANIC V DS VN

3.1 POPIS DS A NÁVRH SCÉNÁŘŮ

Praktickou částí tohoto příspěvku je simulace vývoje OZE a elektromobility na vybrané části sítě VN nacházející se v Jihomoravském kraji (viz schéma na Obrázku 1). Jedná se o smíšenou síť, kde je přibližně 8,5 km kabelových a 18,4 km venkovních úseků. Soustava napájí sedm obcí jejíž počty obyvatel se pohybují v rozmezí 400 až 2500. Zmíněné studie [1], [2] v dané oblasti předpo-kládají v důsledku dobrých atmosférických podmínek a dojezdových vzdáleností nadprůměrný nárůst implementace FVE a elektromobility v následujících dvaceti letech.

Simulace vlivu OZE a elektromobility byla provedena pomocí výpočtu ustáleného chodu v SW Bizon Projektant. Zatížení prvků sítě bylo určeno z naměřených výkonových toků na vývodu z transformační stanice VVN/VN, které byly dále rozpočítány dle instalovaných výkonů jednotlivých distribučních transformátorů (TR) 22/0,4 kV. Pro potřeby analýzy došlo v souladu s požadavkem provozovatele distribuční soustavy k anonymizaci dat o DS. V rámci prvotní analýzy byly předmětem hodnocení dva scénáře implementace, a to:

Scénář 1: V rámci výpočtu vlivu OZE byl uvažován stav, kdy je předpokládán nejvyšší dodávaný výkon prostřednictvím FVE v sítích NN a VN (červenec, 12:00 hod.). Do dané soustavy byla následně připojena výroba $P_{FVE,(+)}$ (viz Tabulka 1) vycházející z odhadu studie [2] pro rok 2040. V tomto scénáři není uvažovaná implementace nabíjecích stanic (NS).

Scénář 2: V rámci výpočtu vlivu elektromobility byl uvažován stav, kdy je předpokládána nejvyšší odběrová špička (leden, 19:00 hod.). Do dané soustavy byly následně připojeny dodatečné odběry nab. stanic $P_{NS,(-)}$ (viz Tabulka 1) vycházející z odhadu studie [1] pro rok 2040, kdy byla uvažována varianta vysokého rozvoje elektromobility. V tomto scénáři není zahrnuta implementace OZE.

Scénář	Hladina		NN						VN
1	Rok	Obec 1	Obec 2	Obec 3	Obec 4	Obec 5	Obec 6	Obec 7	-
$P_{\rm FVE}$ (kW)	2020	+274	+202	+261	+45	+145	+144	+99	+3285
(11.11.)	2040	+1401	+1033	+1335	+231	+740	+735	+504	+3915
Scénář 2	Rok	Obec 1	Obec 2	Obec 3	Obec 4	Obec 5	Obec 6	Obec 7	Celkem
$P_{\rm NS}$	2020	-22	-16	-21	-4	-12	-12	-8	-95
(kW)	2040	-1970	-1452	-1876	-325	-1040	-1033	-708	-8405

 Tabulka 1:
 Změna výkonových poměrů v dané síti vlivem elektromobility a OZE

3.2 VÝSLEDKY PRO SCÉNÁŘ 1

Dle výsledků ustáleného chodu vznikne v soustavě výkonový přetok do hladiny 110 kV o velikosti téměř 6,5 MW. Pro porovnání se součet všech odběru v této sítí ve stejnou dobu blíží k hodnotě 2,4 MW. Během výpočtu však nebyla uvažována žádná forma akumulace nebo omezování výkonu jednotlivých FVE. Tento přetok by také mohlo částečně zmírnit započítání odběrů elektromobility. Z pohledu zatížení se v síti na prvcích VN a VVN nevyskytnul žádný nedostatek (maximální zatížení na TR 22/0,4 kV činilo 42 %).

Napěťové poměry v uzlech VN byly dodrženy v celé síti, jelikož byly splněny meze $U_n \pm 10$ %. V uzlu, kde byla nevíce koncentrovaná výroba FVE došlo k navýšení napětí na 108 % U_n . Do výpočtu však nebyla zahrnuta U/Q regulace, díky které by se napěťové poměry mohly značně zlepšit.

Na hladině NN se může projevit potenciální problém související se změnou toku výkonu mezi dobou značné výroby FVE (slunečný den) a zanedbatelné výroby FVE (noc, zataženo), kdy dochází ke skokové změně napětí ve vybraných částech sítě. Jedním z řešení je použití TR 22/04 kV umožňující automatické přepínání odboček transformátoru pod zatížením (OLTC).

3.3 VÝSLEDKY PRO SCÉNÁŘ 2

Dle SW analýzy došlo k vzrůstu odebíraného výkonu na vývodu z transformační stanice VVN/VN z 4,9 MW (2020) na 14 MW (2040). To pro představu odpovídá situaci, kdy je v dané oblasti uvažováno přibližně 3000 vozidel na elektropohon. Během tohoto stavu však došlo k přetížení kmenové linky, které v nejvíce namáhaném úseku odpovídalo hodnotě 101 % (viz Obrázek 1). Bylo by tedy nutné buď zmíněný úsek posílit, nebo přesunout značnou část nabíjecího výkonu na jinou denní dobu. Napětí ve všech uzlech na hladině VN se pohybovalo v daných mezích $U_n \pm 10$ %.



Obrázek 1: Odhad maximálního zatížení dané sítě při implementaci elektromobility

4 ZÁVĚR

Provedená analýza se v první fázi věnovala oddělené implementaci OZE a nabíjecích stanic pro EV a výhradně byla zaměřená na hodnocení výkonové bilance a napěťových poměrů. Výsledky poukazují na důležitost opatření jako jsou rekonstrukce linek, U/Q regulace, akumulace, omezování výroby OZE nebo řízení nabíjecích výkonů. V rámci následujících analýz budou hodnoceny další navržené scénáře, které budou zaměřeny především na společnou implementaci OZE a nabíjecích stanic, pakliže tyto technologie instalované společně mohou mít například potenciál ovlivnit případné negativní vlivy spojené právě s jejich oddělenou implementací.

REFERENCE

- [1] Dílčí studie pro pracovní tým A25: Predikce vývoje elektromobility v ČR. 1. Praha: Euroenergy, 2018, 146 s. Dostupné také z: <u>https://www.mpo.cz/assets/cz/energetika/strategicke-a-koncepcni-dokumenty/narodni-akcni-plan-pro-chytre-site/2019/10/Studie-NAP-SG-A25_Elektromobilita.pdf</u>
- [2] Výpočty dopadu rozvoje decentrálních výroben doprovozu distribuþní a přenosové soustavy.
 1. Brno: EGÚ Brno, 2017, 28 s. Dostupné také z: https://www.mpo.cz/assets/cz/energetika/strategicke-a-koncepcni-dokumenty/narodni-akcniplan-pro-chytre-site/2020/5/A9-A13-Vypocty-dopadu-rozvoje-decentralnich-vyroben.pdf
- [3] ČR. *Národní akční plán pro chytré sítě*. In: . Praha: Ministerstvo průmyslo a obchodu, 2015. Dostupné také z: <u>https://www.mpo.cz/assets/cz/energetika/elektroenergetika/2016/11/Narodni-akcni-plan-pro-chytre-site.pdf</u>

ENERGY FLOW ANALYSIS IN LV GRID WITH HIGH LEVEL PENETRATION OF PV INSTALATIONS

Robin Kolarik

Master Degree Programme (2), FEEC BUT E-mail: xkolar69@stud.feec.vutbr.cz

Supervised by: Martin Paar

E-mail: paar@feec.vutbr.cz

Abstract: The article is focused on the analysis of daily energy flow in LV grid with PV installations, electric vehicles and the character of the building consumption. PV systems are situated on the roofs of the houses and use maximum roofs technical potential. Electric vehicles are divided to three categories and by each house are situated one electric car. Load profile of each houses are based on the standardized load profile, whitch is allocated according to distribution tariff. The aim of this article is to determine daily power flow analysis in municipality LV grid.

Keywords: distribution network, electric vehicle, load diagram, PV system

1 ÚVOD

Článek pojednává o časové simulaci průběhu toku výkonu v části obce, vybraná část pokrývá 41 rodinných domů. Jedná se o standardní obecní zástavbu řadových rodinných domů. Střecha objektů je pokryta ve své maximální využitelné ploše fotovoltaickými (FV) panely a u každého objektu je uvažováno dobíjení rodinného elektromobilu (EV).

2 ASPEKTY SIMULACE

Jednotlivé objekty jsou charakterizovány průběhem spotřeby, průběhem výroby ze střešní instalace fotovoltaické elektrárny (FVE) a dobíjením elektromobilu. Časové okno simulace bylo zvoleno 24 hodin, které je rozděleno na 5minutové intervaly. Vzhledem k hodinovým průměrným výkonům z dostupných dat, byla provedena interpolace průběhů v programu MATLAB metodou "pchip", která zachovává původní hodnoty průběhu a dopočítává zbylé hodnoty pomocí kubické interpolace.

2.1 TYPOVÉ DIAGRAMY DODÁVKY ELEKTRICKÉ ENERGIE

Ve stávajícím způsobu měření odběru elektrické energie tvoří velkou část odběratelé, u kterých není využíváno průběhové měření. Na základě dlouhodobého sledování průběhů u velkého počtu těchto zákazníků byly sestaveny modelové průběhy spotřeby těchto odběratelů. Typové diagramy dodávky (TDD) tak vyjadřují odhadovanou bilanci elektrické energie v čase, a jejich úkolem je modelovat předpokládaný průběh.

Společnost OTE a.s. na svých stránkách poskytuje osm typů ročních TDD s hodinovým intervalem průměrných výkonů [1]. Jednotlivé typy TDD zohledňují různorodé chování spotřebitelů elektrické energie. Například spotřebitel, který využívá elektrickou energii na vytápění i ohřev teplé vody (TV) bude mít jiný charakter odběru než spotřebitel, který využívá pro vytápění a ohřev TV jinou formu než elektrickou. Chování spotřebitelů též zohledňuje provozovatel distribuční soustavy, který na základě předpokládaného využití elektrické energie a velikosti spotřeby v objektu stanoví spotřebiteli distribuční sazbu. Jednotlivé distribuční sazby jsou zařazeny do příslušné třídy TDD.

Přehled vybraných příkladů tříd TDD [1]

•	TDD1	Podnikatel	odběr bez tepelného využití elektrické energie	C01d, C02d
•	TDD4	Domácnost	odběr bez tepelného využití elektrické energie	D01d, D02d
•	TDD5	Domácnost	odběr s akumulačním spotřebičem	D25d, D26d
•	TDD7	Domácnost	odběr s přímotopným vytápěním/ tepelným čerpadlem	D45d, D55d
•	IDD/	Domacnost	odber s primotopnym vytapenim/ tepelnym cerpadlem	D45d

Z důvodu přehlednosti nejsou v přehledu distribučních sazeb uvedeny všechny distribuční sazby. Pro vybrané kategorie TDD jsou na obrázcích níže (**Obrázek 1** a **Obrázek 2**) vyneseny interpolované křivky denního zatížení pro den 31.5.2021 a roční spotřebou objektu 1 MWh



Obrázek 1: průběh spotřeby TDD1 a TDD4



Obrázek 2: průběh spotřeby TDD5 a TDD7

DODÁVANÝ VÝKON Z FVE

Pro zjištění předpokládaného průběhu výroby elektrické energie ze střešních FVE je využit výpočetní program PVGIS[2]. Poskytuje průběh předpokládané roční výroby FVE v hodinových intervalech. Vstupními hodnotami do programu je místo instalace, sklon, natočení a instalovaný výkon FV panelů, případně ztráty systému. Vzhledem k rozdílnému natočení střech objektů ve vybrané síti, byly zjištěny průběhy předpokládané výroby z FV panelů pro 8 různých směrů natočení panelů od 0° do 360° po kroku 45°. Jednotlivé průběhy byly zjištěny pro instalovaný výkon FV panelů $P_{inst} = 1$ kWp. Tyto průběhy jsou brány jako výchozí a následně pro konkrétní objekt bude stanovena předpokládaná výroba FVE v závislosti na maximálním možném instalovaném výkonu střešní instalace objektu. Průběh dodávky výkonu z FVE pro vybrané směry natočení, jsou vyneseny na obrázcích níže. (**Obrázek 3, Obrázek 4**)



Obrázek 3: výroba FVE pro sever a jih, 1 kWp



Obrázek 4: výroba FVE po východ a západ, 1 kWp

ELEKTROMOBIL

V obci bude uvažováno s využitím tří typů EV. Kategorie A počítá s malým EV s průměrným denním nájezdem 30 km/den a spotřebou 131 Wh/km, zastoupení 20 %. Kategorie B odpovídá rodinnému EV s denním nájezdem 45 km/den a spotřebou 150 Wh/km, zastoupení 65 %. Poslední kategorie C odpovídá firemnímu použití EV s denním nájezdem 90 km/den a spotřebou 190 Wh/km, zastoupení 15 %. Dobíjení EV je uvažováno tak, aby vykompenzovalo denní nájezd ujetých kilometrů a křivka dobíjení EV zohledňuje skutečnost různého času připojování EV do DS. Nejvíce připojení EV do sítě je uvažováno mezi 15 až 18 hodinou.[3] [4] V současné verzi výpočtu není uvažováno řízené dobíjení EV.

3 UKÁZKOVÝ OBJEKT

Jedná se o typ rodinného domu, který je nejvíce zastoupen v simulované oblasti. Objekt má sedlovou střechu orientovanou na východ – západ. Celková plocha střechy je rovna 100 m². Roční spotřeba elektrické energie objektu činí 6410 kWh se sazbou D25d (TDD5). Průběh spotřeby objektu pro daný den je zobrazen červenou křivkou (**Obrázek 5**). Je získán vynásobením jednotlivých průměrných odběrů v čase přepočteného TDD5 a celkovou roční spotřebou.

$$P_{(t)obj} = P_{(t)TDD5} \cdot P_{spot \check{r}eba} \tag{1}$$

Průběh dodávaného výkonu z FVE umístěné na objektu je rozdělen na dva průběhy, jeden průběh zahrnuje dodávku z východní strany instalace a druhý průběh ze západní strany instalace. Předpokládáme symetrickou sedlovou střechu, stejný podíl střechy je orientován na východ i západ, dále je zahrnut koeficient využití střechy $K_{\nu} = 0.5$. [1][5][6].

Uvažujeme panel o výkonu 300 Wp a plochou 1,8 m². Instalační plocha FV panelu je vypočtena z rozměrů panelu uvedené výrobcem a zahrnutí nutné instalační mezery mezi panely. Na jednu stranu střechy je instalován výkon $P_{inst} = 3,9$ kWp. Průběh dodávky výkonu z FVE je znázorněn modrou křivkou (**Obrázek 5**) a celková výkonová bilance ukázkového objektu je zobrazena na obrázku 6.



Obrázek 5: průběh výroby FVE a spotřeby objektu, průběh nabíjení EV



4 SIMULACE OBLASTI

Na obrázku níže (**Obrázek 7**) je zobrazena výsledná výkonová bilance modelované sítě Celkem se v dané lokalitě nachází 41 fotovoltaických střešních instalací. Na severní stranu střech je instalovaný výkon všech FV panelů roven 24,9 kWp, na jižní stranu 85,8 kWp, na západní stranu 171,6 kWp a na východní stranu je orientován výkon 203,1 kWp.



Obrázek 7: výkonová bilance modelované sítě 31.5.2021, pondělí. (P <0 dodávka, P >0 odběr)

Dominantní vliv na průběh budou mít instalace orientované na východ a západ. Tato skutečnost je důsledkem dvou velmi výrazných špičkových dodávek do sítě v 10 hodin a následně ve 14 hodin. Tyto časy výkonových špiček odpovídají také průběhům předpokládané dodávky z FV panelů na obrázku v kapitole 0. Od 15 té hodiny se zvyšuje spotřeba vlivem připojování EV do sítě. V simulovaný den 31.5.2021 je z FV instalací pokryta téměř veškerá spotřeba EV.

5 ZÁVĚR

Z výsledných průběhů vyplývá potenciál soběstačnosti zvoleného objektu v daném dni, pokud by byla správně zvolená a navržená akumulace, která není u daného objektu v současném návrhu uvažována. Modelová síť je napájena z distribučního transformátoru VN/NN (22/04 kV) o výkonu 400 kVA. Dle výsledků simulace by v daném dni by tak nedošlo k nežádoucímu přetížení napájecího transformátoru. Vedení je uvažováno v celé své délce jako AYKY 3x185+70 a vzhledem k topologii sítě nedošlo k jeho přetížení nadměrným průchodem proudu. Napětí v jednotlivých vybraných uzlech sítě zůstalo po celou dobu simulace v mezích stanovených v PPDS. [8]

Jako pokračování tohoto projektu je snaha o automatizaci procesu. Schopnost načíst již existující síť, přiřadit TDD jednotlivým objektům podle distribuční sazby, přiřadit střešní FV instalaci a následně vyhodnotit předpokládanou výkonovou bilanci v kterýkoli den v roce pro vybranou oblast.

REFERENCE

- [1] STATISTIKA: Normalizovaný TDD. *OTE* [online]. [cit. 2021-03-10]. Dostupné z: https://www.ote-cr.cz/cs/statistika/typove-diagramy-dodavek-elektriny/normalizovane-tdd?date=2021-03-10
- [2] PHOTOVOLTAIC GEOGRAPHICAL INFORMATION SYSTEM. *EU SCIENCE HUB* [online]. [cit. 2021-03-10]. Dostupné z: https://re.jrc.ec.europa.eu/pvg_tools/en/#PVP
- [3] FILIP, Robin. *Bateriové systémy pro distribuční sítě*. Brno, 2019. Dostupné také z: https://www.vutbr.cz/studenti/zav-prace/detail/119142. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav elektroenergetiky. Vedoucí práce Martin Paar.
- [4] FIGENBAUM, Erik a Marika KOLBENSTVEDT. Learning from Norwegian Battery Electric and Plug-in Hybrid Vehicle users: Results from a survey of vehicle owners [online]. Oslo: Institute of Transport Economics (TØI), 2016 [cit. 2021-03-09]. ISBN 978-82-480-1718-9. https://www.toi.no/publications/learning-from-norwegian-battery-electric-and-plug-in-hybrid-vehicle-users-results-from-a-survey-of-vehicle-owners-article33869-29.html
- [5] Jakubes J., Járka V.: Studie "Potenciál solární energetiky v České republice". ENACO. Praha 2015, 58 s.
- [6] Čambala P., Hrubý M., Muselík O., Špaček T., Procházka J.: Oponentní posudek k vybraným tématům z návrhu Národního Klimaticko-Energetického Plánu (NKEP) pro oblast FVE. EGÚ Brno a. s. 2018, 36 s.
- [7] ŠTEFEK, Martin. *Potencionální produkce elektrické energie ze střešních fotovoltaických elektráren v obci do 3000 obyvatel.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2019. 74 s. Vedoucí bakalářské práce Ing. Martin Paar, Ph.D.
- [8] Pravidla provozování distribučních soustav: Příloha č.4 Pravidla pro paralelní provoz výroben a akumulačních zařízení se sítí provozovatele distribuční soustavy. 2021. Dostupné také z: https://www.egd.cz/predpisy-smlouvy-pro-elektrinu

Magisterské projekty

Zpracování signálů, obrazu a dat

COMPARISON AND EVALUATION SYSTEM FOR BEAT TRACKING ALGORITHMS

Karolína Staňková

Master Degree Programme (2), FEEC BUT E-mail: xstank03@stud.feec.vutbr.cz

Supervised by: Matěj Ištvánek E-mail: xistva02@stud.feec.vutbr.cz

Abstract: This work deals with systems for detecting rhythmic structures of music recordings. The field of retrieving information from music (MIR) is developing rapidly and has more accurate results than ever before. It allows us to examine the harmonic and tonal properties of music, rhythm, tempo, etc. Various algorithms are used in the detection of rhythmic structures. However, today, most new methods use neural networks. The thesis aims to summarize the current research results of systems for detecting music times and tempo in MIR, to describe methods of calculating and evaluating the parameters of music recordings, and to implement a program that allows comparison of available detection systems. The result of the work is a script in the Python language, which uses five different systems to detect the rhythmic structure of test recordings. It then checks the outputs of the algorithms according to the given reference and compares the given systems with each other using several evaluation quantities.

Keywords: Beat Tracking, dataset, evaluation methods, Music Information Retrieval, neural networks

1 ÚVOD

Detekce rytmických struktur hudebních nahrávek je jednou z podoblastí oboru *Music Information Retrieval* (MIR), tedy získávání informací z hudby. Možnost přesně detekovat tempo a rytmus jakékoli skladby a na základě těchto kritérií pak nahrávky porovnávat mezi sebou nabízí mnohá využití, například nové způsoby výzkumu historických nahrávek, či přesnější výsledky doporučovacích algoritmů, které nabízí některé komerční streamovací služby.

Beat tracking, tedy detekci dob v zadané skladbě, umožňuje několik různých dostupných systémů, jež vznikaly jak na akademické půdě, tak v komerční sféře. Mnohé z nich je možné využít volně, tedy jako open-source. Nicméně jejich praktické použití je limitováno, neboť systémy fungují odlišně a neexistuje zatím žádný způsob, jak jich na zvolený dataset volat více najednou. Některé se sice dají využít ve formě plug-in modulů pro existující programy pro editaci hudby, avšak porovnat (alespoň poslechem) tyto systémy mezi sebou, či dokonce využít objektivního srovnání pomocí různých hodnoticích veličin, je dnes zatím velmi komplikované.

Diplomová práce, jíž tento článek představuje, si klade za cíl zmíněná omezení překonat a nabízí program v jazyce Python, který umožňuje na libovolném datasetu porovnat pět různých systémů pro detekci rytmické struktury skladeb, a tyto systémy mezi sebou porovnává pomocí několika hodnoticích veličin. Struktura tohoto programu je uvedena na blokovém schématu 1.

2 STATE-OF-THE-ART SYSTÉMY

Většina systémů vychází z myšlenky, že pozice dob ve skladbě závisí na dvou skutečnostech: zaprvé, hledané časové okamžiky se obvykle shodují s hudebními dobami (které se v nahrávce často dají



Obrázek 1: Blokové schéma programu.

snadno rozpoznat jako místa s velkou energií – například údery do bicích), a zadruhé, tyto okamžiky mají mezi sebou stabilní časový interval (mají určitou fázi). V čem se pak beat tracking systémy liší, jsou způsoby výpočtů těchto parametrů.

Jedním ze systémů, s nimiž zmíněný program pracuje, je knihovna pro Python s názvem *madmom* [1]. Jedná se o komplexní systém pro získávání informací z hudby, přičemž detekce rytmického systému je jednou z jeho součástí. Pro určení pozic dob využívá obousměrnou neuronovou síť LSTM pro analýzu signálu po rámcích. Vstupy této sítě jsou melovské spektrogramy (vypočítané z Fourierovy transformace s různou délkou okna) a jejich relativní rozdíly. Výstupem neuronové sítě je pak aktivační funkce, ze které je pomocí autokorelace určeno tempo a fáze nahrávky. Pro výpočet pozic dob využívá systém madmom dva algoritmy – první z nich spoléhá na to, že tempo se během nahrávky nemění (v systému je nazýván *Beat Detector*) a vypočítává pozice dob ze všech vzorků aktivační funkce, druhý pak počítá s fluktuacemi tempa a pozice dob určuje pouze z nejbližších šesti sekund skladby (*Beat Tracker*). Ukázka detekovaných dob systémem madmom (algoritmus *Beat Tracker*) v grafu časové závislosti a ve spektrogramu je na obrázku 2.



Obrázek 2: Graf závislosti úrovně nahrávky písně *Rain* na čase a spektrogram nahrávky se zvýrazněnými časy dob detekovanými systémem madmom.

2.1 DALŠÍ POUŽITÉ SYSTÉMY

Program dále využívá systému *librosa* [6], který pro výpočet pozic dob využívá obálku intenzity začátků událostí (*onset strength envelope*). Tuto funkci systém koreluje s jejími různě zpožď ovanými verzemi, aby zjistil fázi, a na jejím základě určil časy dob. Oproti systému madmom je librosa méně komplexní a nedokáže si poradit s výraznějšími záměrnými fluktuacemi tempa v rámci skladby.

Třetím využívaným systémem je knihovna pro C++ a Python *essentia* [2], jež postupuje podobně jako *librosa.* K určení dob využívá nejprve detekci začátků hudebních událostí, k čemuž zpracovává informace z více různých detektorů, založených například na melovských spektrogramech či komplexních spektrálních diferencích, a z nich určuje časy dob pomocí skrytých markovových modelů.

Výstupem jsou ty časy dob, na nichž mezi detektory panuje největší shoda.

Čtvrtým použitým systémem je knihovna *aubio*. K detekci dob využívá podobně jako librosa obálku intenzity začátků hudebních událostí a její autokorelaci, jíž ovšem navíc filtruje bankou hřebenových filtrů a váhuje pomocí Rayleygho a Gaussova váhovacího okna.

2.2 HODNOCENÍ SYSTÉMŮ

Vyhodnocení výsledků systémů je možné pomocí subjektivní či objektivní evaluace. V prvním případě je potřeba zkušený posluchač a poměrně velké množství času, přičemž kritéria i přísnost hodnocení se mohou různit. Proto je pro dosažení reprodukovatelnějších výsledků lepší zvolit objektivní hodnocení, pomocí hodnoticích veličin. Aby bylo možné vypočítat jejich hodnoty, musíme nejprve získat tzv. *ground truth*, tedy základní pravdu. Tu lze získat několika způsoby, v rámci diplomové práce byly nejprve časy dob ve skladbách určeny pomocí plug-in modulu v editačním programu Sonic Visualiser, a následně byly ručně zpřesněny. Takto získané referenční časy dob nazýváme anotacemi (nebo tzv. labels). Jejich porovnáváním s detekovanými časy získáme výsledky hodnoticích veličin jako je například F-score, P-score, Precision, Recall [4], apod. Grafický výstup programu je znázorněn na obrázku 3.



Obrázek 3: Ukázka srovnání systémů pomocí metody F-score (průměrná hodnota za všechny testované nahrávky)

3 VÝSLEDKY PRÁCE

Výsledným produktem této práce je program v jazyce Python s grafickým rozhraním (viz obrázek 4), který umožňuje na libovolné databázi skladeb vyzkoušet a porovnat všech pět zmíněných systémů pro detekci časů dob (zde uvedené knihovny byly sice jen čtyři, madmom, librosa, essentia a aubio, nicméně systém madmom je v programu zařazen dvakrát – jednou jako *Beat Detector*, podruhé jako *Beat Tracker* – pro ověření schopnosti systémů adaptovat se na případné změny tempa v rámci skladby). Uživatel zvolí, které z nabízených systémů chce na svou databázi využít, a případně zda chce nějak upravit jejich parametry (například velikost časového okna, výchozí tempo, apod. – tyto parametry se mírně liší u každého systému) a stiskem tlačítka spustí analýzu nahrávek. Výsledné časy dob se ukládají jako textové soubory do výstupní složky, aby bylo možné je dále využít. Po provedení analýzy si může uživatel zobrazit libovolnou ze skladeb v grafu (v časové závislosti nebo jako spektrogram) se zvýrazněnými časy skladeb tak, jak je určil daný systém, a zobrazit si též vyhodnocení výsledků. Po kliknutí na příslušné tlačítko vypočítá program všechny hodnoticí veličiny a každou hodnotu zobrazuje ve sloupcovém grafu, který umožňuje přehledné srovnání všech využitých systémů (jak naznačuje obrázek 3).

SP_Stankova - S System pro detekci rytmicke struktury skladeb						
Vyberte slozku s datasetem k analyze						
Vyberte slozku s anotacemi						
Librosa	Parametry	Spustit analyzu				
🗆 Madmom Beat Tracker	Parametry					
Madmom Beat Detector	Parametry	Vykreslit graf				
🗆 Essentia						
☐ Aubio	Parametry	Vyhodnotit				
System pripraven. Pro spusteni analyzy prosim vyberte slozku s nahravkami. Pro spusteni vyhodnoceni prosim vyberte slozku s nahravkami a slozku s anotacemi.						
Vybrana slozka s nahravkami: /home/karolinka/PycharmProjects/diplmka/audio						
Vybrana slozka s anotacemi: /home/karolinka/PycharmProjects/diplmka/annotations						

Obrázek 4: Ukázka grafického rozhraní vytvořeného programu.

4 ZÁVĚR

Cílem této práce byla sumarizace, srovnání a vyhodnocení dostupných systémů pro detekci dob. Za tímto účelem vznikl program v jazyce Python, jenž umožňuje analyzovat nahrávky několika různými systémy pomocí jednoduchého grafického rozhraní. V rámci další práce budou tyto systémy porovnány mezi sebou pomocí testovacího datasetu a budou vyhodnoceny pomocí evaluačních veličin. Dále bude připraven a ručně anotován druhý dataset sestávající ze skladeb s fluktuujícím či nepravidelným tempem a méně přehlednou rytmikou, na němž budou systémy také testovány.

REFERENCE

- [1] BÖCK Sebastian, SCHEDL Markus, *Enhanced Beat Tracking with Context-Aware Neural Networks*, Proceedings of the 14th International Conference on Digital Audio Effects (DAFx), 2011.
- [2] BOGDANOV, Dmitry, N WACK, Emilia GÓMEZ, et al. ESSENTIA: an Audio Analysis Library for Music Information Retrieval. Proceedings - 14th International Society for Music Information Retrieval Conference [online]. 2013(11) [cit. 2020-11-04]. Dostupné z: http://hdl.handle. net/10230/32252
- [3] BROSSIER, Paul M. Automatic Annotation of Musical Audio for Interactive Applications. London, 2006. PhD dissertation. Centre for Digital Music, Queen Mary University of London. Vedoucí práce Dr. Mark Plumbley and Prof. Mark Sandler. Dostupné z: https://aubio.org/phd/thesis/ brossier06thesis.pdf
- [4] DAVIS, Mathew, Norberto DEGARA QUINTELA a Mark PLUMBLEY. *Evaluation Methods for Musical Audio Beat Tracking Algorithms* [online]. 2009(10) [cit. 2020-10-30]. Dostupné z: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.152.6936&rep=rep1&type=pdf
- [5] DEGARA QUINTELA, Norberto, Enrique ARGONES RÚA, Antonio PENA, Soledad TORRES-GUIJARRO, Matthew DAVIES a Mark PLUMBLEY. *Reliability-Informed Beat Tracking of Musical Signals*. IEEE Transactions on Audio, Speech and Language Processing [online]. 01/2012, 2012(20), 290-301 [cit. 2020-11-04]. Dostupné z: https://ieeexplore.ieee.org/document/ 5934584
- [6] ELLIS, Daniel PW. *Beat tracking by dynamic programming*, Journal of New Music Research 36.1 (2007): 51-60. Dostupné z: http://labrosa.ee.columbia.edu/projects/beattrack/

BEAT TRACKING SYSTEM BASED ON A NEURAL NETWORK

Tomáš Suchánek

Master Degree Programme (2), FEEC BUT E-mail: xsucha13@vutbr.cz

Supervised by: Matěj Ištvánek E-mail: xistva02@stud.feec.vutbr.cz

Abstract: This thesis deals with systems for tempo and beat detection in music recordings, whose functionality is based on neural networks. The basic structure of such systems is briefly described and the emphasis is then placed on a comparison of recurrent and temporal convolutional networks, which have proven to be the most suitable for this task. The main outcome of this work is then proposal and comparison of modified temporal convolutional network with other state-of-the-art networks in a beat tracking system. The results suggest that simplification in existing architectures could benefit from faster training times, while it maintains or slightly improves the accuracy of a detection system.

Keywords: Beat tracking, machine learning, neural network, signal processing

1 ÚVOD

Detekce tempa a dob je jednou z podoblastí vědního oboru *Music Information Retrieval* (MIR), který si klade za cíl extrakci dále uplatnitelných informací z hudebních nahrávek. Ačkoli lze tuto extrakci s velkou spolehlivostí provádět manuálně, je v kontextu posledních let a množství dat vytvářen tlak na co nejrychlejší a nejpřesnější automatizaci těchto postupů. S navyšující se dostupností výpočetní techniky se pak dostávají do popředí neuronové sítě, s jejichž možnostmi lze počítačovým systémům mj. vtisknout právě schopnost identifikace rytmu a hudebních dob. Získané informace se pak uplatňují při synchronizaci moderních technologií na specifické události v hudbě, při doporučování obsahu u streamovacích služeb, nebo v neposlední řadě také při porovnávání různých interpretací totožných skladeb. Tato práce si stanovuje za cíl výzkum dosavadních přístupů a zaměřuje se na efektivitu různých neuronových sítí v těchto systémech. Vlastní systém s různými architekturami sítí je také výsledkem a předmětem závěrečné diskuze této práce.

2 STRUKTURA DETEKČNÍCH SYSTÉMŮ

Proces detekce dob ve své podstatě vychází z myšlenky, že skladba ve frekvenčním spektru obsahuje opakující se vzorec zvukových událostí, na jejichž základě lze tempo a doby odvodit. Cílem systému je tedy nalezení těchto vzorců a odvození nejpravděpodobnější sekvence hudebních dob, které takové vzorce vysvětlují. Zvukový signál je pak za tímto účelem zpracováván zpravidla ve třech bodech.

Nejprve je provedena jeho transformace do časově-frekvenční reprezentace, nejčastěji mel spektrogramu, který v porovnání s běžným spektrogramem blíže simuluje nelinearitu a maskování lidského slyšení. Za účelem zisku robustnější informace o vývoji spektra v čase jsou pak mel spektrogramy s různými délkami segmentů kombinovány do jediné časově-frekvenční reprezentace, která slouží jako vstup do další části systému.

V té již figurují neuronové sítě, jejichž úkolem je na základě předem daných časových anotací dob každé skladby vyvodit vztah mezi těmito anotacemi a událostmi ve zvukovém spektru, přičemž jejich následným výstupem je vyjádření pravděpodobnosti, s jakou se v každém časovém segmentu vstupní

reprezentace vyskytuje doba. Ačkoli je zřejmé, že z výsledných dat lze pomocí algoritmů na hledání lokálních maxim sestavit žádaný výstup, může takový proces vzhledem k určité nejistotě neuronové sítě u různých skladeb vykazovat horší přesnost v důsledku falešných detekcí.

Tento problém aktuálně nejúspěšněji eliminuje využití pravděpodobnostních modelů [1], s jejichž vhodným přizpůsobením lze snadněji identifikovat i změny v tempu nebo metrické struktuře skladby. Tato práce, podobně jako mnohé další, využívá algoritmus založený na skrytých Markovových modelech. Ten v předem nastavených podmínkách, jako je například maximální tempové zrychlení mezi sousedními dobami, vyvozuje nejpravděpodobnější sekvenci dob, která těmto podmínkám a vstupním datům daných neuronovou sítí vyhovuje. Obrázek 1 ukazuje příklad výstupu sítě a odvození konečných pozic dob pravděpodobnostním modelem.



Obrázek 1: Výstup neuronové sítě a pravděpodobnostního modelu.

2.1 VYUŽÍVANÉ NEURONOVÉ SÍTĚ

Požadavky na neuronové sítě pro efektivní detekci lze shrnout dvěma body. Prvním z nich je možnost zpracování datové posloupnosti libovolné délky, druhým je možnost vytváření interních souvislostí mezi časově proměnnými událostmi. Pro tyto účely se již v jiných pracích [2, 3] prokázaly jako nejefektivnější obousměrné rekurentní sítě s buňkami LSTM a temporální konvoluční sítě TCN. První jmenované si prostřednictvím zpětných vazeb vytvářejí paměť, která je pro detekci dob zásadní, přičemž struktura LSTM buněk tuto paměť dále navyšuje. TCN sítě pak paměť nemají, nicméně jejich architektura disponuje konvolučními filtry, které se s každou další vrstvou sítě rozšiřují. Hodnota výstupní vrstvy pro určitý časový segment vstupní reprezentace pak v důsledku závisí i na okolních vstupních datech, jejichž pomyslné časové rozpětí je definováno právě hloubkou sítě a počáteční šířkou konvolučního filtru. Princip této sítě je blíže vyjádřen na obrázku 2. Tyto sítě se také zásadně odlišují procesem učení – zatímco u rekurentních sítí je chyba na výstupu zpětně šířena přes všechny časové kroky vstupní reprezentace, u sítě konvolučního typu je chyba šířena pouze přes pevně daný počet vrstev a jejich parametrů, z čehož plyne markantní zrychlení učícího se procesu především u delších nahrávek.

2.2 NÁVRH SYSTÉMU

V rámci práce byl vytvořen systém, ve kterém byly pro další srovnání testovány čtyři architektury neuronových sítí. Algoritmus jako takový sdílí pro každou síť stejnou fázi předzpracování vstupních dat a také konečné odvození sekvence dob, které vychází z prací [1, 3] a je dostupné v knihovně



Obrázek 2: Struktura TCN sítě. Černé čáry vyznačují závislosti prvků mezi sousedními vrstvami. Na příkladu je použit konvoluční filtr o šířce k = 3 s dilatačními faktory d = 1, 2, 4, čímž se rozumí roztažení filtru v každé z dalších vrstev.

madmom pro jazyk Python. Vstupní zvuková data jsou převedena do tří mel spektrogramů s délkami segmentů N = 1024,2048,4096, fixním posunem analyzačního okna h = 441 a M = 20 mel frekvenčními pásmy na jeden segment. Každý mel spektrogram je dále zvláště filtrovaný mediánovým filtrem s posuvným oknem o délce $N_{med} = N/100$, přičemž výsledná reprezentace se dále odečítá od té původní. Jednotlivé spektrogramy jsou poté spojeny do jediné reprezentace, která tímto postupem získává rozměr M = 120 mel frekvenčních pásem s časovým rozlišením 100 fps při $f_{vz} = 44, 1$ kHz. Mezi testované sítě pak patří dvě replikované architektury z prací [2, 3] (dále CNNTCN a BLSTM) a dvě vlastní modifikované architektury (dále BGRU a STCN), jejichž strukturu popisuje obrázek 3. Rozdíl mezi BGRU a BLSTM tkví pouze v typu buněk, CNNTCN od STCN pak odlišuje kromě jiného nastavení vnitřních parametrů také předcházející 2D konvoluční blok, který slouží k dalšímu předzpracování vstupních dat.



Obrázek 3: Architektura sítí BGRU (vlevo) a STCN (vpravo). Počet vzorů jedné vstupní dávky označuje *d*, jejich délku pak značí *s*. Výstupní data jsou dále vstupem pravděpodobnostního modelu.

3 VÝSLEDKY

Všechny sítě byly trénovány na datasetu čítajícím 220 skladeb o délce 30 s pomocí 5složkové křížové validace, přičemž celý proces byl třikrát opakován, získané výsledky byly zprůměrovány a jsou vypsány v tabulce 1. Pro vytvoření základního náhledu na výkonnost systému byla použita nejběžnější metrika F-score, přičemž hodnota F-score = 1 vyjadřuje absolutní shodu detekce s příslušnou anotací skladby. Z výsledků vyplývá, že navržená síť STCN dosahuje srovnatelného či mírně lepšího skóre než další testované sítě a to při kratší době zpracování jedné iterace vzorových dat při učení, což je výhodné při učení sítě na velkých datových sadách. Z jiných prací se stejným zaměřením však vyplývá, že různé sítě mohou vykazovat různé skóre na např. žánrově odlišných datasetech a pro potvrzení nebo naopak vyvrácení dosažených výsledků je zapotřebí testování zopakovat na větším množství dat. Je také vhodné podotknout, že časy učení byly prozatím dosaženy na jednotce CPU a v případě využití grafické karty budou časy nadále menší a mohou se do jisté míry poměrově lišit.

 Tabulka 1:
 Srovnání přesnosti detekčního systému s různými neuronovými sítěmi.

Síť	F-score	Doba učení
BLSTM	0,664	10,91 s/it
BGRU	0,653	10,86 s/it
CNNTCN	0,681	4,13 s/it
STCN	0,685	1,36 s/it

4 ZÁVĚR

Práce shrnuje zásadní poznatky o systémech na detekci dob založených na využití neuronové sítě a zprostředkovává vhled do přesnosti systému s různými síť ovými architekturami. Praktickým výstupem práce je funkční algoritmus, pomocí kterého pak byly jednotlivé sítě testovány a vyhodnoceny, přičemž jedna z navržených sítí vykazuje konkurenceschopnou přesnost v porovnání se zavedenými state-of-the-art architekturami při rychlejším čase jejího učení. V další fázi práce bude zkoumán vliv úprav předzpracování dat na přesnost detekce a sítě budou podrobeny rozsáhlejšímu testování na robustnější datové sadě.

REFERENCE

- [1] KREBS, Florian, Sebastian BOCK a Gerhard WIDMER. Rhythmic pattern modeling for beat and downbeat tracking in musical audio [online]. 2013 [cit. 2020-11-11]. Dostupné z: http: //phenicx.upf.edu/system/files/publications/Krebs_ISMIR_2013.pdf
- [2] DAVIES, Matthew a Sebastian BOCK. Temporal convolutional networks for musical audio beat tracking. European Signal Processing Conference (EUSIPCO) [online]. 2019 [cit. 2020-12-05]. Dostupné z: http://telecom.inesctec.pt/~mdavies/pdfs/ DaviesBoeck19-eusipco.pdf
- [3] BOCK, Sebastian, Florian KREBS a Gerhard WIDMER. Joint beat and downbeat tracking with recurrent neural networks. 17th International Society for Music Information Retrieval Conference [online]. 2016 [cit. 2021-03-11]. Dostupné z: http://www.cp.jku.at/research/ papers/Boeck_etal_ISMIR_2016.pdf

PROGRAM FOR ASSISTANCE IN LEARNING ENGLISH PRONUNCIATION

Jan Malucha

Master Degree Programme (1), FEEC BUT E-mail: xmaluc00@vutbr.cz

Supervised by: Milan Sigmund

E-mail: sigmund@feec.vutbr.cz

Abstract: This study presents a compact MATLAB tool to assist students in learning the correct pronunciation of English language, with use of three known methods for estimation of speech signal voicing. A short introduction into speech voicing analysis is given and the developed program is briefly described. The practical use of the program was verified by test with non-native speaker.

Keywords: speech signal, pronunciation, voicing

1 ÚVOD

Vlivem pandemické situace se pojem "distanční výuka" v současnosti stává velmi často skloňovaným tématem. S rozvojem komunikačních technologií se toto téma začíná objevovat i v diskuzích o budoucnosti vzdělávání a vyvstávají otázky, zda vzdělávání na dálku zůstane nadále pevnou součástí běžného studia. Obzvláště ve sféře středoškolského vzdělávání toto světové trendy skutečně naznačují, a má tudíž smysl zaměřit pozornost na vývoj komplexních nástrojů k podpoře distanční formy výuky. Zejména v oblasti výuky cizích jazyků lze s výhodou aplikovat poznatky teorie zpracování řečových signálů, jedné z disciplín obecného zpracování a analýzy signálů, široce využívané v oblastech od rádiových komunikací až po zdravotnické aplikace. Výuka cizích jazyků klade velký důraz na mluvenou a poslechovou formu výuky, a je zde tedy možné zefektivnit zdokonalování se ve správné výslovnosti např. anglického jazyka za podpory komplexních nástrojů analyzující řečový signál získaný skrze mikrofon.

2 ZNĚLOST A NEZNĚLOST ŘEČI

Samotný řečový signál je časovou reprezentací zachycující mluvenou řeč. Z fyzikálního hlediska jde o reprezentaci chvění přenosového média, zpravidla vzduchu, vyvolané budičem, jímž je hlasové ústrojí člověka. Jeho podoba se více či méně může lišit pro různé jazyky. V případě většiny indoevropských jazyků však můžeme dané řečové signály lingvisticky považovat za časové průběhy reprezentující slova (tedy spojení hlásek) oddělená úseky ticha. Fonologicky jednotlivé hlásky dále dělíme na znělé a neznělé, a to podle podstaty jejich tvorby [1] – znělé hlásky jsou tvořeny kvaziperiodickým kmitáním hlasivek a mohou tudíž mít základní frekvenci, resp. melodii (např. hláska /a/), oproti tomu hlásky neznělé nejsou tvořeny hlasivkami a typicky mají charakter šumu (např. hláska /š/). Úseky ticha jsou kromě pauz mezi slovy či větami typické pro rozhraní mezi určitými hláskami (např. /dk/ ve slově ředkev), což je způsobeno přechodem mezi různými stavy hlasového ústrojí.

Budeme-li se zabývat anglickým jazykem, je třeba říci, že na rozdíl od českého jazyka je angličtina až velmi citlivá na správnou výslovnost hlásek v rámci daného slova. Při špatné výslovnosti či jejich záměně totiž může velmi často dojít k naprosté změně jeho významu (např. rozdíl mezi slovy *ship [lod*] a *sheep [ovce]*). Zároveň je třeba věnovat zvýšenou pozornost hláskám, které se v českém jazyce vůbec nevyskytují [2], a jež při řeči pro správné vyznění zpravidla vyžadují jistý cvik. Znělost/neznělost je tedy jedním z hlavních atributů správné výslovnosti.

3 CÍLE PRÁCE

Předmětem tohoto příspěvku je vytvoření programu k analýze řečových signálů, jehož praktickou funkcí má být vyhodnocení a evaluace parametru znělosti a neznělosti v průběhu vstupního řečového signálu a srovnání jeho vlastností s parametry ideálního protějšku. Výstupem má být grafické znázornění výsledků, přizpůsobených k interpretaci ve vztahu k výuce cizího jazyka.

4 VYTVOŘENÝ PROGRAM

Metodami vhodnými a často používanými k určování znělosti a neznělosti jsou: STE (short time energy), ZCR (zero crossing rate) a HNR (harmonic-to-noise ratio). Všechny pracují na základě určování míry šumu v daném segmentu řečového signálu - metoda STE využívá výpočet krátkodobé energie signálu [1], ZCR určuje četnost změny polarity signálu [3], resp. průchodů rovnovážnou polohou, a HNR analyzuje poměr periodické složky signálu ke složce šumové [4]. Výpočetní algoritmy jsou rozšířeny o schopnost určení úseků ticha, což omezí vliv pauz v různých variantách daného signálu, které by jinak byly určeny jako neznělé a mohly by značně zkreslovat výsledky. Program byl vytvořen v prostředí MATLAB.

Typickou vlastností neznělé hlásky je její šumový charakter, což způsobuje vyšší položení na frekvenční ose oproti znělým hláskám periodického charakteru, zároveň vykazuje nižší energii a náhodnost průběhu, zatímco znělá hláska vykazuje velmi málo šumu a dominuje periodická složka. Úsek ticha se svou podobou nejvíce blíží nulové hodnotě. Analýza probíhá po velmi krátkých úsecích signálu, aby byla zajištěna stacionarita šumové složky. Program tedy nejprve rozčlení vstupní signál na krátkodobé úseky o délce 10 ms a ty jednotlivě zpracuje vybranou výpočetní metodou, čímž je porovná se šumem a následným prahováním je označí za znělé/neznělé či ticho. Souborem výstupních dat analýzy signálu je četnost výskytu úseků znělosti, neznělosti a ticha. Tato četnost je porovnána s četností vzorového signálu pomocí Euklidovské vzdálenosti a výsledkem celé procedury pak je hodnota podobnosti mezi analyzovaným a vzorovým signálem.





5 PRAKTICKÉ VYUŽITÍ A TESTOVÁNÍ PROGRAMU

Praktický význam programu spočívá v podpoře distanční výuky cizích jazyků, a to i bez účasti vyučujícího. Příkladem může být kontrola správné výslovnosti anglické věty. Anglický jazyk je velmi zvučný a citlivý na správnou srozumitelnost, do čehož negativně vstupuje neschopnost správné artikulace znělých a neznělých pasáží jednotlivých slov. To se pro mluvčího uvyklého na slovanské jazyky může projevovat např. zkracováním znělých samohlásek, přílišným zvýrazňováním některých neznělých souhlásek či zadrháváním při přechodech mezi jednotlivými hláskami.

Pro analýzu těchto vlivů byla vybrána následující anglická věta: "*Wall cavities need to be inspected by an expert to ensure walls are secure and any damaged wall-cavity insulation will also need to be removed.*" Její záznam, namluvený rodilým mluvčím, má délku trvání cca 9 sekund. Obr. 2 zobrazuje krátkou část časového průběhu této věty společně s grafickým znázorněním pasáží vyhodno-cených jako znělé/tiché/neznělé – kladná hodnota potvrzuje znělost, nulová hodnota značí ticho a záporná neznělost. Analýza byla provedena za použití výpočetní metody STE, která se ukázala být optimální pro daný účel díky efektivní schopnosti rozlišit krátkodobé úseky neznělosti a ticha. Metoda ZCR se v tomto ohledu projevila jako méně spolehlivá vlivem relativně silného šumu při pořizování nahrávek, metoda HNR se pro detekci ticha ukázala být takřka nevhodná.



Obrázek 2: Řečový signál rodilého mluvčího

Výše uvedená nahrávka byla přehrávána českému studentovi, jenž je profesionálním učitelem angličtiny hodnocen jako průměrný až lehce nadprůměrný, za účelem zdokonalování jeho výslovnosti. Student následně nahrál 34 pokusů o správné vyslovení dané věty. Na Obr. 3 je trojrozměrný graf s osami interpretujícími procentuální četnost znělosti, neznělosti a ticha. Červený bod značí poměr četnosti pro vzorový signál namluvený rodilým mluvčím, modré body značí signály jednotlivých 34 pokusů českého studenta a jsou číslovány podle pořadí pokusu.



Obrázek 3: Postupné zlepšování výslovnosti anglické věty

Jak lze vidět, v prvních pokusech kvalita výslovnosti studenta výrazně kolísá, přiblížení se správné výslovnosti se děje spíše náhodně. Se zvyšujícím se počtem opakování, a postupném zdokonalování, se poloha pokusů soustřeďuje kolem bodu nedaleko polohy vzorového signálu a výslovnost se stabilizuje, jak ukazuje Obr. 4. To, že se poslední pokusy nesoustřeďují přímo u vzorové věty, je dáno stálou několikaprocentní odlišností přízvuků. K dosažení dokonalého britského přízvuku by bylo zapotřebí např. strávit určitý čas v zemi typické tímto přízvukem či dlouhodobější nácvik.



Obrázek 4: Zlepšování výslovnosti s opakováním (procentuální podobnost s rodilým mluvčím)

Program je možné rozšířit o funkci slovního hodnocení výslovnosti na základě shody pokusu se vzorem; procentuální hranice a dané hodnocení určuje učitel. Zpětná vazba pro studenta pak bude srovnání histogramu vzoru a pokusu, procentuální podobnost a slovní hodnocení, viz Obr. 5.



Obrázek 5: Vybrané pokusy správné výslovnosti

6 ZÁVĚR

Oproti mnoha jiným oblastem studia klade studium jazyků velký důraz na mluvenou a poslechovou stránku výuky. Z tohoto důvodu má jeho distanční forma studia za podpory výpočetní techniky velký potenciál. Předložený program analyzuje řečové signály pomocí již existujících matematických metod, přidanou hodnotou je zde zejména uživateli srozumitelné grafické znázornění přibližování se správné výslovnosti z hlediska znělosti/neznělosti. Nástroj tak nabízí možnost atraktivního zefektivnění výuky cizích jazyků. Dalším krokem ve vývoji programu by pro zpřesnění analýzy mohlo být rozšíření o schopnost identifikace jednotlivých hlásek a vyhodnocení podobnosti s hláskami vzorového signálu.

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory projektu interního grantu VUT v Brně FEKT-S-20-6361.

REFERENCE

- [1] J. Psutka, L. Müller, J. Matoušek, and V. Radová, Mluvíme s počítačem česky. Academia, Praha, 2006.
- [2] M. Reed and J. Levis, The Handbook of English Pronunciation. John Wiley & Sons, Hoboken, 2019.
- [3] R. G. Bachu, S. Kopparthi, B. Adapa, and B. D. Barkana, Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy. In Proc. Advanced Techniques in Computing Sciences. Springer, Dordrecht, 2010, pp. 279-282.
- [4] J. P. Teixeira, C. Oliveira, and C. Lopes, Vocal acoustic analysis-jitter, shimmer and HNR parameters, Procedia Technology, 2013, vol. 9, pp. 1112-1122.

Doktorské projekty

Biomedicínské inženýrství a bioinformatika

OPERON-EXPRESSER: THE INNOVATED GENE EXPRES-SION-BASED ALGORITHM FOR OPERON STRUCTURES INFERENCE

Jana Schwarzerová

Doctoral Degree Programme (1), FEEC BUT E-mail: xschwa16@stud.feec.vutbr.cz

Supervised by: Karel Sedlář

E-mail: sedlar@feec.vutbr.cz

Abstract: Current biotechnological research of bacterial genomes has huge potential due to the use of next-generation sequencing (NGS) platforms. NGS era opens paths for analysis of data for sufficient description of microorganisms with ecology and biotechnology potential in the future. Although some tools for inference of specific structures in bacterial genomes exist, their pipelines and methodologies are often based only on searching bacterial genome databases and comparison with model microorganisms. This paper deals with the design of a new algorithm for operon structures inference in bacterial genomes. The algorithm combines searching in bacterial databases and gene expression information processing. The algorithm was implemented in R language and tested on *Clostridium beijerinckii* NRRL B-598. The bacterium is a typical performer in the field of biofuels production due to its ability to produce butanol. Thanks to that this bacterial organism can be of great potential from an ecological and biotechnological point of view. The paper also provides a comparison of operon structures derived by Operon-mapper and its extending by Operon-expresser.

Keywords: Bacterial genome, Operon prediction, Operon-mapper, Gene expression information, Biotechnology, Transcriptome, RNA-Seq, *Clostridium beijerinckii* NRRL B-598

1 INTRODUCTION

Operon [1] represents group of genes that are transcribed together as a single mRNA which is in a bacterial genome. Currently, it is a half of century has passed since the discovery of operons but the importance of operons in bacterial gene networks and the relationship between their organization and gene expression remain poorly understood [1]. This situation needs a better methodology for understanding based on in silico tools for exact inference of operon structures which will combine two approaches. The first approach represents tools based on searching bioinformatics databases for important genome structures of already described bacteria and comparing them to the unknown bacterial genome. The second approach relies on gene expression information and correlation [2] between gene expression information in one operon.

RNA-Seq [3] is one of the most used sequencing methods which originated from the NGS era. Although, sequencing methods are constantly developed to overcome older techniques e.g. microarrays. All these methods give important details of gene expression information. In this study, a new methodology for inference of operon structures is designed, described, and immediately tested to *Clostridium beijerinckii* NRRL B-598. This bacterium belongs to the group of non-model organisms, which can have huge potential for an application in solving environmental challenges. Specifically, *C. beijerinckii* NRRL B-598 is a butanol producer, which is in demand, because there is a focus on sustainable microbial production of biobased fuels.

2 METHODS

Operon is the cluster of genes that have the same promotor and the genes are transcribed and regulated as a single large mRNA including multiple structural genes, see Figure 1. Transcription unit (TU) [4] is a concept that was defined to make the understanding of operon function easier. TU is obtained from the genes that transcribe and regulate simultaneously. The identification of TUs is a challenge for resolving the understanding of transcriptional regulation.



Figure 1: Operon structure includes promotor (P), operator (O) and a group of genes which are transcribed and regulated as a single large mRNA.

2.1 **OPERON-MAPPER:** THE WEB SERVER FOR OPERON STRUCTURES INFERENCE

Operon-mapper [5] is an open-source tool in a web server that predicts the operons of any bacterial or archaeal genome sequence. This tool predicts operon using knowledge about an intergenic distance of neighboring genes as well as the functional relationships of their protein-coding products. Operon-mapper is based on an artificial neural network (ANN). This algorithm was tested on a set of experimentally defined operons in *Escherichia coli* and *Bacillus subtilis* and reached accuracies of 94,6 % and 93,3 % [5]. ANN has inputs that are the intergenic distance of contiguous genes and score which thinks of the functional relationships between the protein products.

2.2 OPERON EXPRESSER: THE INNOVATED GENE EXPRESSION-BASED ALGORITHM FOR OPERON STRUCTURES INFERENCE

The Operon-expresser based on finding the gene expression value of one gene and the gene expression value from upstream or downstream gene in sequence. After that the correlation coefficient between these genes is calculated. The correlation coefficient is obtained from the Pearson correlation formula [6]:

$$r = \frac{\sum(x - m_x) (y - m_y)}{\sqrt{\sum(x - m_x)^2 (y - m_y)^2}},$$
(1)

where m_x and m_y are the means of x and y variables. The x variable represents the gene expression information of the selected gene and the y variable represents the gene expression information of the next gene which is upstream or downstream gene in sequence. However, to seed up the algorithm and use prior knowledge the Operon-expresser algorithm can use precedent information of predicted operons and the algorithm uses operon. It means, it counts with group of genes – operon instead of one gene.

If the correlation coefficient is more than 75 % and the distance between these operons is less than a defined threshold in bp, we assume that it is one operon and these operons are concatenated. The threshold was defined as a mean distance between the predicted operons. This process repeats until all correlation coefficients are more than 75 % or the distance between operons is higher than 5 000 bp. The algorithm was created in R script using R/Bioconductor Biostrings [7] and R/ Hmisc [8].

This algorithm was used in the pipeline for the extension of Operon-mapper output. It is important to combine the information about transcription and translation with genomics data. In Figure 2, the first branch is connected with Operon-mapper [5]. The second branch rely on gene expression information, which is obtained from RNA-Seq data. The raw RNA-Seq data are prepared in a count table format that describes gene express information.



Figure 2: Pipeline of the new methodology for inference operon structures

In the beginning, the bacterial genome sequence is taken as input data for the Operon-mapper [5]. Thanks to Operon mapper, the list of predicted operon structures is obtained in *.csv* format. No information on gene expression is obtained in the prediction, yet. The subsequent step is filtering pseudogenes and transcription ID genes to locus tags. The filtering and transcription marked by locus tag is implemented in R script using R/Bioconductor genomeIntervals [8], R/Bioconductor Biostrings [7], and R/stringr [8]. After the transcription step gene expression information is added in the format of count table. Next step represents division operons using computing correlation gene expression value. The final step includes Operon-expresser algorithm, which was developed in R language. The algorithm uses prior knowledge about obtained operons and counts with group of genes. It means operon instead of one gene. The final output represents predicted operon structures which are obtained knowledge from bacterial genomes databases based on machine learning prediction, Operon mapper [5], and also gene expression information. The schema of the whole pipeline is visualized in Figure 3.



Figure 3: Step by step pipeline of new methodology for inference operon structures

2.3 MATERIALS

The developed algorithm for inference of operon structures was applied on the dataset from *C. beijerinckii* NRRL B-598. The dataset includes gene expression information from seven different replicates which were sequencing in six time-points that cover all metabolic stages and genome format *.fna* [11]. The replicates include six transcriptomes. Two replicates represent standard cultivation and they are mentioned in the study by Sedlar et al. [12], two replicates represent also standard cultivation which are mentioned in the study by Patakova et al. [13] and two replicates represent the response to butanol-shock and are described in the study by Sedlar et. al. [14]. Samples were sequenced using Illumina NextSeq500. The output from NextSeq was raw data in *.fastq* format. These raw data are pre-processed using a reproducible analytical pipeline for using raw RNA-Seq data from non-model organisms [15] and count tables represented gene express information were obtained.

3 RESULTS AND DISCUSSION

In the beginning Operon mapper [5] was applied to genome from *C. beijerinckii* NRRL B-598 number CP011966.3 [11]. The list of the first prediction operon structures was obtained. The first prediction of operon structures includes 3351 operons. It means 3219 operons without pseudogenes. In the next step, the Operon expresser toolbox of R scripts was used. The first script performs filtering pseudogenes and transcription ID genes to locus tags using R/Bioconductor genomeIntervals [8], R/Bioconductor Biostrings [7] and R/tidyverse stringr [16]. The marker of locus tags is matching with marker in created count tables from study by Schwarzerova [15] represents gene expression information. Then grow seed algorithm was applied and so the combination of searching databases based on machine learning algorithms and gene expression information was united. The algorithm was implemented in R using R/Bioconductor Biostrings [7] and R/ Hmisc [8]. The threshold of correlation for division into two operons was set less than 75 %. The neighbor length of two locus tag was checked as less than 5 000 bp. These thresholds were set heuristically. After application of Operon-expresser algorithm, the total number of obtained operon structures is 3778.

on structures of Operon-expresser				
	Parameters	Operon-manner	Operon-expresser	Difference

Table 1: Comparison of basic static quantities from operon structures of Operon-mapper and oper-

Parameters	Operon-mapper	Operon-expresser	Difference
Total number of operons	3219	3778	559
Maximum number of genes	38	38	0
Minimum number of genes	1	1	0
Mean number of genes	1.992	1.659	0.333
Standard deviation number of genes	2.467	2.197	0.270

Table 1 shows a basic static analysis of operon structures which has been obtained from Operonmapper and its extension of Operon-expresser. The maximum and the minimum number of genes in one operon is same in both approaches. The dataset of operon obtained from Operon-mapper has maximum number of genes in eighty-second operon in genome sequence, but in dataset of operon from Operon-expresser has maximum number of genes in seventy-nine operon in genome sequence. Although the genes that are included into by both operons are same. The change is possible to observe in the mean and standard deviation number of genes in one operon. The overall matching of genes in each of operons between datasets from Operon-mapper and datasets from Operonexpresser is 52 %. This low percentual number is caused by property of Operon-expresser algorithm thanks to that this algorithm divides operons in the absence of linear dependence in added gene expression information. It reflects the specification of the boundaries between the individual operons relied on gene express information from RNA-Seq.

4 CONCLUSIONS

Tools for operons structures inference are based only on searching databases and comparing already known operons and genomes with unknown operons and genomes. Despite the fact that these algorithms are using advanced methods of machine learning and deep learning, none of these algorithms does not take into account gene expression information. One of the latest approaches predicts operon using knowledges about intergenic distance of neighboring genes as well as the functional relationships of their protein-coding products. This approach uses Operon-mapper. However, thanks to NGS era which brought huge analysis data for sufficient description in microorganisms, the gene expression information is affordable. In this paper the innovated gene expression-based algorithm for operon structures inference called Operon-expresser was described. The algorithm was implemented in R and used as an extending to the Operon-mapper. This is the first time ever that a methodology which considers gene expression information using *in-silico* tools has been used for operon structure inference. The test was provided on dataset from bacterial strain *C. bei*- *jerinckii NRRL B-598.* The basic statistical analysis shows the maximum and the minimum number of individual locus tags in one operon does not change significantly, it deduces that the Operon-expresser allows the specification of boundaries between the individual operons relied on gene expression information from RNA-Seq.

ACKNOWLEDGEMENT

Computional resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures

REFERENCES

- [1] LIM, Han N., Yeong LEE a Razika HUSSEIN. Fundamental relationship between operon organization and gene expression (2011)
- [2] BURGI, Hans-Beat a Jack D. DUNITZ. Structure correlation. 1. ISBN 3-527-29042-7.
- [3] Wang, Z. et. Al. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57–63 (2009).
- [4] Scitable by nature EDUCATION: transcription unit-[online]. 2005-[Accessed 2021-01- 30] Available from: https://www.nature.com/scitable/definition/transcription-unit-260/
- [5] B, Taboada, Estrada K, Ciria R a Merino E. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. 2018
- [6] *STHDA Statistical tools for high-throughput data analysis: Correlation Test Between Two Variables in R.* Available from: <u>http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r</u>
- [7] Pagès H, Aboyoun P, Gentleman R, DebRoy S (2020). *Biostrings: Efficient manipulation of biological strings*. R package version 2.58.0, <u>https://bioconductor.org/packages/Biostrings</u>.
- [8] HARRELL, Frank E. *Package 'Hmisc'* 2020 Available from: <u>https://CRAN.R-project.org/package=Hmisc</u>
- [9] Gagneur J, Toedling J, Bourgon R, Delhomme N (2020). *genomeIntervals: Operations on genomic intervals.* R package version 1.46.0.
- [10] Wickham, H. (2019b). stringr: Simple, consistent wrappers for common string operations. Retrieved from <u>https://CRAN.R-project.org/package=stringr</u>
- [11] KOLEK, Jan et. al. Draft Genome Sequence of Clostridium pasteurianum NRRL B-598, a Potential Butanol or Hydrogen Producer [online].
- [12] SEDLAR, Karel et. al. Transcription profiling of butanol producer Clostridium beijerinckii NRRL B-598 using RNA-Seq. 30 May 2018.
- [13] PATAKOVA, Petra et. al. Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in Clostridium beijerinckii NRRL B-598 at the transcriptomic level. 2019
- [14] SEDLAR, Karel et al. A transcriptional response of Clostridium beijerinckii NRRL B-598 to a butanol shock. 2019.
- [15] SCHWARZEROVÁ, J. Reproducible analytical pipeline for using raw RNA-Seq data from non-model organisms. ISBN: 978-80-214-5867-3.
- [16] WICKHAM, Hadley. Stringr: Simple, Consistent Wrappers for Common String Operations [online]. 2019

MONITORING THE PARAMETERS OF CELLS MIGRATING IN PSEUDO-3D EXTRACELLULAR MATRIX

Inna Zumberg

Doctoral Degree Programme (1), FEEC BUT E-mail: xzumbe00@stud.feec.vutbr.cz

Supervised by: Larisa Chmelikova, Vratislav Cmiel

E-mail: chmelikoval@vut.cz, cmiel@vut.cz

Abstract: Cell migration plays an essential role in a number of biological processes, such as embryogenesis, the immune response, wound healing and inflammation. Most research regarding cell migration is based on experiments using two-dimensional (2D) cell cultures, and the detailed molecular and biophysical mechanisms of these processes are already well known. However, much less is known about cell behaviour in the three-dimensional (3D) environment of living tissues. Pseudo-three-dimensional (pseudo-3D) cell cultures bridge the gap between 2D and 3D geometries and combine their advantages. This review presents the quantitative evaluation of human mesenchymal stem cells (hMSCs) migration parameters in pseudo-3D artificial extracellular matrix (ECM) of various compositions.

Keywords: hMSCs, Pseudo-3D, Extracellular matrix, Collagen, Confocal microscopy.

1 INTRODUCTION

Various methods are currently used to study cell migration. In vitro model systems based on cell culturing are most often used. The experiments are most often performed on 2D cell cultures, where the cells are in a form of a monolayer on different surfaces, most often on a plastic or a glass [1, 2]. Such experiments include, for example, a scratch assay, where a scratch is made into a cell monolayer and the cell migration towards the centre of the scratch is monitored. The advantages of 2D cell cultures include simple and economical implementation; shorter time required for cell culture formation in the range of minutes to hours; and high speed of proliferation. However, these 2D systems are not surrounded by their natural microenvironment, and their responses to various chemicals or other stimuli may differ significantly from those found in a living organism. Cells lack components from the ECM, including proteins (especially collagen (CG)), glycosaminoglycans and glycoproteins (laminin, fibronectin (FN)).

However, differences between cell behaviour in culture and *in vivo* have gradually led to the transition to 3D models that better represent the microenvironment of living tissues [3]. Here, the shape of cells, the interactions between them and the environment, better mimic the natural environment. Such models contain not only proliferating cells, but also resting cells and apoptotic, hypoxic or necrotic cells. These models have the ability to better reflect normal differentiation, cell behaviour, and intercellular interactions. Cells cultured in a 3D environment show adhesion to the intercellular matrix with their entire surface. This environment not only physically supports cells, but also provides information that can affect cell differentiation in some way.

These important features make 3D cell cultures more physiologically relevant and predictive than 2D cultures. However, these 3D systems have their limitations, especially in microscopic imaging and data evaluation. Imaging may become difficult depending on the material transparency and scaffold size. The creation of pseudo-3D cell cultures allows to combine the ease of 2D systems while overcoming some limitations of 3D systems. Therefore, a pseudo-3D culture was defined as a culture where cells are embedded in the ECM but in contact with coverslip. [4]

A network of ECM molecules are basic building blocks of such 3D and pseudo-3D models, which is often in the form of a gel. The construction material must have a specific structure and spatial arrangement of molecules to ensure cell growth and migration. Sufficient porosity should be provided for uniform cell proliferation, migration and growth, both in the scaffold space and over time. An important parameter of the material is its biocompatibility. This is compatibility with living tissue, with the material not being toxic, harmful, or physiologically reactive. The cells must be able to adhere to the material, migrate to the surface and grow and proliferate in the scaffold. Such material must also be non-immunogenic, so as not to cause any adverse reactions (inflammation, cytotoxicity, mutagenicity, etc.) [5]. The materials used to form these support matrices are generally of two types. In the first case, these are biological polymers such as fibrin, collagen, hydrogel, chitosan. In the second case, synthetic polymers such as polylactic acid (PLA) or similar materials such as polyglycolic acid (PGA) and polycaprolactone (PCL) can be used.

Collagen, together with elastin and hyaluronic acid are the most important components of the ECM that are predominantly produced by fibroblasts. The collagen gel is bioactive and has cell adhesion promoting properties [6]. Another principal part of the ECM is fibronectin. It is a the most abundant molecule of the extracellular matrix and it belongs to the most studied ECM components. FN increases cell adhesion and proliferation; affects cell migration pathways *in vivo* and in culture; and specifically stimulates cell proliferation. FN also plays a role in cell morphology, cytoskeletal organisation, haemostasis, and wound healing. The study [7] showed that the addition of FN to cells cultured in a collagen gel led to an increase in the number of cells and their motility.

The goal of our experiment was to focus on the study of cell migration in pseudo-3D collagen gel, simulating an ECM environment in order to determine the influence of the gel environment on the parameters of cell migration. In addition, this experiment was extended to study the effect of FN on cell migration. The hMSC migration was monitored using a confocal microscope. The hMSC cell line that is considered the gold standard for clinical research was chosen for this experiment.

2 MATERIALS AND METHODS

2.1 DESCRIPTION OF THE EXPERIMENT

The ECM was prepared from rat tail collagen type I (Ibidi GmbH) and FN (Sigma-Aldrich s.r.o.). In this paper hMSCs isolated from a human adipose tissue obtained by abdominal liposuction were used. Cells were cultured in a low glucose Dulbecco's Modified Eagle Medium (DMEM) containing 5% foetal bovine serum (FBS) and 1% of penicillin-streptomycin (100 U·ml⁻¹:100 μ g·ml⁻¹) at 37°C and 5% CO₂. All chemicals were purchased from Sigma-Aldrich (St. Louis, Missouri, USA).

In our experiment, cells were seeded in one well of a 24-well plate (Sigma-Aldrich). When cell confluency reached 80-90%, then cells were stained with CellTrackerTM Green CMFDA Dye (CMFDA; InvitrogenTM) in a final concentration of 1 μ M in serum-free medium for 15 min. Cells were gently washed by phosphate-buffered saline (PBS) before and after staining. Then, 750 μ l of rat tail collagen type I (ibidi GmbH) in a final concentration of 1 mg·ml⁻¹ were prepared in a micro-tube and placed on ice. Then 200 μ l of prepared collagen gel was placed into the second microtube and mixed with 80 μ l of FN in a concentration of 10 μ g·ml⁻¹. In the third microtube 200 μ l of collagen gel were mixed with 80 μ l of FN in a concentration of 50 μ g·ml⁻¹.

To study the hMSC migration in ECM, 8-well chambered cover glasses (#1.5H, Cellvis) were chosen in order to observe all of the studied ECM types simultaneously. One of the wells was filled with 500 μ l of growth medium (control). The next three wells were filled with 250 μ l of each prepared gel (CG and CG/FN). Then for gel polymerization, the cover glass was placed into a humidified incubator (37°C, 5% CO₂) for one hour. At this moment, the previously stained cells were passaged using accutase (PAA Laboratories GmbH). When the gel polymerization was complete, then 125 μ l of growth medium was added to the surface of the gels. Then cells were seeded in the well with growth medium and the three wells with the gels at a density of $4 \cdot 10^3$ cells·cm⁻². After that, the cover glass was placed into an incubator for four hours. During this time cells were penetrated into the gels and adhered on the bottom of the cover glass. After this, the cover glass was placed into the confocal microscope incubator chamber (The Stage Top Chamber, OKOLAB). Five hours after cell seeding, the image acquisition was started. Thus, the atmosphere of 5% CO₂ and 37°C temperature was maintained during the entire length of the experiment.

2.2 DATA ACQUISITION

The data were acquired using a Leica TCS SP8X confocal microscope equipped with White Light Laser (WLL). The samples in the micro-chamber were observed using a lens with 10X magnification. Cells were labelled with CellTrackerTM Green CMFDA fluorescent dye, which can be retained in living cells through several generations (72 hours). Excitation wavelength was set to 490 nm and an emission range of 500–540 nm, corresponding to the dye spectral properties. To follow the hMSC migration, time-lapse data were acquired; the stacks of 67 images were obtained every 15 min over a period of 16 hours and 30 min. The volume included in an image stack is of the size 1.16×1.16 mm with a spatial resolution of 512×512 pixels for each image.

2.3 DATA ANALYSIS

The aim of processing the image sequences obtained from the experiments is to determine the necessary migration parameters, such as the total distance travelled, the net distance and the speed of cell motility. Based on these parameters, a comparison of the four tested groups and evaluation of the results of the experiments was performed. An algorithm in the MATLAB software environment (version R2020a) was designed for the semiautomatic analysis of confocal image stacks. The base of the algorithm (Figure 1) is the cycle of the opening microscopic images and marking the positions of selected cells, until all images from the entire set are processed. The given cell density and time interval of image acquisition were sufficient to ensure that manual cell marking was objective and that there were no errors in determining the position of a selected cell throughout the image sequence. Cells were cultured between a cover glass substratum and soft gel layer, and the experiments did not indicate cell migration into the gel layer, so, the cells were always in a single focal plane. The obtained groups of coordinates are further used for numerical calculations and representations of the obtained traces (directional roses, Figure 3).



Figure 1: Scheme of the proposed algorithm.

In the first step, the cells of interest were manually marked in each first image of the sequence. Then the cells were tracked using processing of all images in the sequence. Due to the different image sizes, the data were normalized by converting pixels to micrometres (1 px = 2.275μ m). The recalculated routes were used as input data for plotting in the directional rose. The total distance (also as Accumulated distance) travelled is then also calculated from the obtained tracks:

$$d_{total} = \sum_{i=1}^{n-1} \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2}$$
(1)

where x_i and y_i are the coordinates of the cell in the i-th image, x_{i+1} and y_{i+1} denote the coordinates of the same cell in the (i+1)-th image, n denotes the total number of images in the sequence. The net distance (from the start to the end position, also as Euclidean distance) is calculated as

$$d_{net} = \sum_{i=1}^{n-1} \sqrt{(x_{end} - x_{ini})^2 + (y_{end} - y_{ini})^2}$$
(2)

Knowledge of the time interval of image obtaining allows to calculate the velocity of individual cells and also the velocity of all monitored cells in a given experiment.

3 RESULTS AND DISCUSSION

The presented algorithm was used to evaluate the migration of hMSCs in a pseudo-3D environment. Three independent experiments were performed with cells in collagen gel with a collagen concentration of 1 mg·ml⁻¹. The principle of each experiment consisted of long-term scanning in four wells, which formed four experimental groups. After applying the proposed algorithm to microscopic image stacks, the migration parameters were obtained (Figure 2).



Figure 2: Migration parameters obtained during the experiments.

The numerical results were further used for analysis using statistical tests. The aim was to verify whether the average value of these parameters was approximately the same in all groups. A one-way ANOVA (MATLAB, anova1) was used for evaluation. A test on the identity (homogeneity) of the variances (F-test in MATLAB, vartestn) and Shapiro-Wilk test were used at first to verify the assumption of a normal distribution and agreement of the variance of the data in the groups. The results show that the difference between the individual groups is insignificant (p-values > 0.05).

The control group included cells that were cultured only in the presence of medium. Cells in the other groups were cultured in a collagen gel, that was a higher density environment. The result showed that there was a significant difference between the groups in all experiments – cells cultured in collagen gel move at a higher velocity (Figure 2).
The aim of further statistical tests was to determine whether the addition of FN affects the direction of cell migration and the speed of cell motility (Figure 3). The same statistical tests were used as before. The control group contained cells that were cultured in collagen gel, and other groups included collagen with FN at 10% concentration and 50% concentration. The effect of FN on cell migration speed was significant in the first and third experiment, but there was no difference among the groups in the second experiment.



Figure 3: Cell traces (directional roses) performed in experiments.

In [7] it was also suggested that FN added to the collagen gel affects cell migration pathways *in vivo* and in culture. To verify this, an evaluation of the net distance travelled by the cell during the experiment, which reflects the directionality of the migration, can be performed. But the effect of FN on the direction of hMCS motility was not statistically confirmed in any experiment.

4 CONCLUSION

Collagen is the predominant extracellular matrix glycoprotein in most animals. Reconstituted collagen gels are commonly used in many standard in vitro 3D assays. This paper describes an experiment where collagen was used to simulate the pseudo-3D environment of an extracellular matrix. It was found that cells cultivated under a collagen gel layer migrate faster than in a medium, but cells did not migrate into the collagen. The addition of fibronectin to the gel also has a positive effect on the speed of cell migration. However, the addition of fibronectin did not affect cell migration pathways.

REFERENCES

- [1] Doyle, A.D., Petrie, R.J., Kutys, M.L., Yamada, K.M.: Dimensions in cell migration. Current Opinion in Cell Biology, 25 (5), s. 642–649, 2013.
- [2] Ashby, W.J., Zijlstra, A.: Established and novel methods of interrogating two-dimensional cell migration. Integr Biol (Camb.), 4 (11), s. 1338–1350, 2012.
- [3] Even-Ram, S., Yamada, K.M.: Cell migration in 3D matrix. Current Opinion in Cell Biology, 17, s. 524–532, 2005.
- [4] Li, Y., Kilian, K.A.: Bridging the Gap: From 2D Cell Culture to 3D Microengineered Extracellular Matrices. Adv Healthc Mater, 4 (18), s. 2780–2796, 2015.
- [5] O'Brien, F.J: Biomaterials & scaffolds for tissue engineering. Materials Today, 14 (3), s. 88–95, 2011.
- [6] Wolf, K., Alexander, S., Schacht, V., Coussens, L.M., von Andrian, U.H., van Rheenen, J., Deryugina, E., Friedl, P.: Collagen-based cell migration models in vitro and in vivo. Seminars in Cell & Developmental Biology, 20 (8), s. 931–941, 2009.
- [7] Sevilla, C.A., Dalecki, D., Hocking, D.C.: Extracellular matrix fibronectin stimulates the self-assembly of microtissues on native collagen gels. Tissue Engineering: Part A, 16 (12), s. 3805–3819, 2010.

MULTIPLE INSTANCE LEARNING FRAMEWORK USED FOR ECG PREMATURE CONTRACTION LOCALIZATION

Petra Novotna

Doctoral Degree Programme (4th), FEEC BUT E-mail: novotnap@vutbr.cz

Supervised by: Marina Ronzhina E-mail: ronzhina@vut.cz

Abstract: We propose the model combining convolutional neural network with multiple instance learning in order to localize the premature atrial contraction and premature ventricular contraction. The model is based on ResNet architecture modified for 1D signal processing. Model was trained on China Physiological Signal Challenge 2018 database extended by manually labeled ground truth positions of premature complexes. The presented method did not reach satisfying results in PAC localization (with dice = 0.127 for avg-pooling implementation). On the other hand, results of localization of PVCs were comparable with other published studies (with dice = 0.952 for avg-pooling implementation).

Keywords: EEICT, ECG, PAC, PVC, CNN, MIL, arrhytmia, localization

1 INTRODUCTION

Premature atrial contractions (PAC) are one of the common diagnosed arrhythmias characterized by premature heartbeat originated in atria. When occurred isolated, they are not life-threatening. When combined with more underlying medical conditions, PAC could result in early death. Detection of PAC is often combined with early diagnostics of atrial fibrillation (AF) [1]. PACs, from the point of view AF detection, are recorded and quantified by 24 h Holter recordings, and may thus serve as a surrogate marker for paroxysmal AF [2].

Primary, PAC were detected by decision rules mostly based on rhythm features [3]. Since PAC is morphologically very similar to normal QRS complex, morphological features are not usually used for detection (contrary to detection of premature ventricular contractions, PVC). On the other hand, in most of the previously published approaches of PAC detection, temporal features based on RR intervals were used [4]. Such approach requires one-to-one QRS complex detection, which could end up being very time consuming.

Deep-learning approaches are current state-of-the-art methods for PAC classification and detection. Here, we propose method based on deep neural network. Our model uses only global labeling from train dataset and does not require QRS complex detection for identification of local abnormalities in ECG. We trained the model on data demonstrated mentioned arrhythmias (PAC and PVC). We, thus, tested the model capability to localize pathologies with characteristic temporal and morphological manifestation, respectively. Presented results are the attempt to extend the previous work [5]. Briefly, the algorithm for PAC localization contains several steps. First, ECG signal is classified into selected arrhythmia types. Second, the feature signal is derived using so called multiple instance learning (MIL) tool. Finally, PAC are localized via peak detection in the feature signal. Following chapters describe the proposed method in more detail.

2 DATASET

Recordings analysed in this paper come from publically available database from China Physiological Signal Challenge 2018 (CPSC2018) [6]. The database consists of recordings assigned to 9 different arrhythmias. To create the model, we used only three labels (global annotations): Normal, PVC and PAC. Altogether, 2145 signals were used (653, 574 and 918 for PVC, PAC and Normal, respectively). The database was split into the training and testing datasets in the 8:2 ratio. All the provided signals included 12 lead recordings. Proposed model is also suited for the use of reduced number of recorded leads. In the database, only global annotations (e.g. label "PVC" means that there are some PVCs in the whole ECG) are available. To evaluate the results of PAC/PVC localization, the local true PAC/PVC positions are needed. We, therefore, analysed the whole dataset manually and added the positions of pathological complexes in the database.

3 MODEL ARCHITECTURE

3.1 CLASSIFICATION NETWORK

Convolutional neural network (CNN) was chosen for presented algorithm. The model is based on residual neural network (ResNet) [7]. ResNet CNN is mostly used for image processing, so we modified it for signal processing. All parts, which are originally designed to be applied on 2D data (images) were replaced by operators suitable for 1D data (signals). Classification network follow up is presented on Figure 1. The input ECG enters the model to CNN network, which consists of repeating block of convolutional layer, batch normalization and ReLu activation function. Max pooling or average pooling are applied after. All convolutional blocks are followed by sigmoid activation function and global max pooling layer, out of which final label for the signal is gained.



Figure 1: Architecture of proposed model. CNN Model: $Conv(3, i \times 12)$ - convolution layer with filter size 3 and $i \times 12$ filters; MaxPool(2,2) - max pooling with filter size 2 and stride 2.

3.2 MIL FRAMEWORK

MIL framework [8] process the output of previous applied convolutional layers. The effect of convolution is in the multiple sub-sampling of the input ECG. The output signal can be referred as MIL feature signal. The feature signal contains the information of pathology (PAC/PVC) position likelihood: the peaks in the feature signal correspond with the part(s) of the input ECG contributing the most significantly to the final classification. MIL feature signal is of variable length (bag of instances) and is projected to a single output label by global max pooling. Global pooling layer was followed by standard weighted cross-entropy loss (WCE), which is defined as:

$$WCE_{MIL} = w_{pos}t\log(\max_{i}s_i) + w_{neg}(1-t)\log(1-\max_{i}s_i),$$
(1)

where s_i is MIL feature signal (which is the output of the sigmoid activation function), t is a binary signal label, w_{pos} and w_{neg} are weights for positive and negative classes, respectively. The weight itself is in inverse proportion to the frequency of class label. The feature signal, representing the arrhythmia position likelihood, is an input for further processing.

3.3 MODEL IMPLEMENTATION DETAILS

Since the dataset used for training the model is not initially balanced, data augmentation was needed. Various augmentation techniques were applied on ECGs. Particularly, each lead was randomly multiplied by 0.3 and signal was stretched by 20 %. Circular shifting of selected signal parts was performed, too. Batch size, which is constant, does not reach the size of every analysed signal, so zero-padding was applied. More implementation specifications of the model are: optimisation with Adam optimiser ($\beta_1 = 0.9$; $\beta_2 = 0.999$) [9]); decoupled weight decay regularization ($\lambda = 10^{-5}$) [10]; initial learning rate 0.001 (with every 50 epochs multiplied by 0.1). Weighted cross-entropy loss was applied in the training phase. Batch size was initially set to 32 and weights were set via Xavier initialisation [11].

3.4 ARRHYTHMIA LOCALIZATION

Premature complexes localization is performed by the peak detector applied on MIL feature signal. The detector was set up with three specific parameters: threshold, minimal distance and peak prominence [12]. The parameters' optimal values were set to: minimum-maximum of all feature signals (likelihood maps) for each arrhythmia, respectively; 0 - min-max range as for peak prominence; 0 - 2 s for minimal distance between peaks. These values were obtained by Python implementation of Bayesian optimization [13].

4 RESULTS AND DISCUSSION

Results reached by presented model are summarized in Table 1. Main goal of this study was to extend our previous model in order to reach the results comparable with another recent reports. From Table 1 (see gray rows), this was achieved for PVC localization performed by the model with max pooling layer (e.g. compare with [14]). The results for PAC localization (black rows in the Table 1), were not satisfying. The reason for such difference in localization of PVC and PAC may lay in their characteristics, mostly the morphological point of view. On Figure 2, we can see that QRS complexes labeled as PAC (both red and grey) are not as different in it's shape, duration and power, compared to the normal QRS complexes. Features distinguishing PAC from normal QRS are mainly temporal. From Table 1, it is obvious that presented MIL framework is not as sensitive to temporal pathology manifestations as to morphological ones. As a result, PACs cannot be accurately localized via obtained MIL feature signal.

In future work, we well focus on the improvement of PAC localization. We suppose that the main

temporal feature contributing to the PAC recognition is the ECG part corresponding to the compensation pause. We therefore believe that appropriate post-processing of MIL feature signal could result in improved detection efficacy. Further enhancement of the method may consist in extension of the model for most common arrhythmias (such as atrioventricular blocks, atrial fibrillation, atrial flutter, etc.).

Method	Arrhytmia	Prec.	Recall	Dice	ACC
MIL - max-pool	PAC	0.032	0.187	0.055	0.506
MIL - avg-pool	PAC	0.097	0.181	0.127	0.789
MIL - max-pool	PVC	0.922	0.805	0.810	0.975
MIL - avg-pool	PVC	0.877	0.752	0.952	0.967

Table 1: Results of PAC localization compared to PVC localization, coming from the same model.



Figure 2: MIL feature signal and PAC beats detection. PAC position likelihood is represented by upper blue signal, where the amplitude of the signal represents the contribution to the final classification. One PAC was detected, one was not.

5 CONCLUSION

In conclusion, proposed method was previously proven to work well on localization of PVC, which has significantly different morphology as compared to normal QRS complex. In case of PAC, the results were not satisfactory. One of the most benefits of presented model is almost whole absence of pre- and post-processing of the output. Adding the post-processing part could result in better outcome of the model for PAC localization, which will be the subject of further research.

REFERENCES

[1] CHONG, Boon-Hor, Vincent PONG, Kwok-Fai LAM, et al. Frequent premature atrial complexes predict new occurrence of atrial fibrillation and adverse cardiovascular events. *Clinical Research:*

Atrial Fibrillation. 2012, 14(-), 942–947.

- [2] TODO, Kenichi, Tomonori IWATA, Ryosuke DOIJIRI, et al. Frequent Premature Atrial Contractions in Cryptogenic Stroke Predict Atrial Fibrillation Detection with Insertable Cardiac Monitoring. *Cerebrovascular Diseases*. 2020, **49**(2), 144-150. ISSN 1015-9770.
- [3] VISINESCU, M., BASHOUR, C. A., WAZNI, O. and B. GOPAKUMARAN. Automatic Detection of Conducted Premature Atrial Contractions to Predict Atrial Fibrillation in Patients after Cardiac Surgery. *Computers in Cardiology*. 2004. **31**(4), 429-432. ISSN 0276â'6547.
- [4] HAN, Dong, Syed Khairul BASHAR, Fahimeh MOHAGHEGHIAN, Eric DING, Cody WHIT-COMB, David D. MCMANUS a Ki H. CHON. Premature Atrial and Ventricular Contraction Detection Using Photoplethysmographic Data from a Smartwatch. *Sensors*. 2020, 20(19), -. ISSN 1424-8220.
- [5] NOVOTNA, Petra, Tomas VICAR, Jakub HEJC, Marina RONZHINA and Jana KOLAROVA. Deep-Learning Premature Contraction Localization in 12-lead ECG From Whole Signal Annotations. In: *Computing in Cardiology 2020*. 47. Rimini, Italy: IEEE, 2021, 2020-12-30, s. -. ISSN 2325-887X.
- [6] LIU, Feifei, Chengyu LIU, Lina ZHAO, et al. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. *Journal of Medical Imaging and Health Informatics*. 2018, 8(7), 1368-1373. ISSN 2156-7018.
- [7] HE, Kaiming, Xiangyu ZHANG, Shaoqing REN and Jian SUN. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). -: IEEE, 2016, 2016, s. 770-778. ISBN 978-1-4673-8851-1. ISSN 1063-6919.
- [8] CARBONNEAU, Marc-André, Veronika CHEPLYGINA, Eric GRANGER and Ghyslain GAGNON. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*. 2018, 77, 329-353. ISSN 00313203.
- [9] KINGMA, Diederik P. and Jimmy BA. Adam: A method for stochastic optimization. arXiv preprint arXiv14126980 2014.
- [10] LOSHCHILOV, Ilya and Frank HUTTER. Decoupled weight decay regularization. arXiv preprint arXiv171105101 2017.
- [11] GLOROT Xavier and Yoshua BENGIO. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings 9:249-256, 2010.
- [12] AHN, Jungryul, Myoung-Hwan CHOI, Kwangsoo KIM, Solomon S. SENOK, Dong-il Dan CHO, Kyo-in KOO and Yongsook GOO. The advantage of topographic prominence-adopted filter for the detection of short-latency spikes of retinal ganglion cells. *The Korean Journal of Physiology Pharmacology*. 2017, **21**(5), 555-563. ISSN 1226-4512.
- [13] NOGUEIRA, Fernando. Bayesian Optimization: Open source constrained global optimization tool for Python. 2014. URL: https://github.com/fmfn/BayesianOptimization.
- [14] ZHAO, Zhongyao, Xingyao WANG, Zhipeng CAI, Jianqing LI a Chengyu LIU. PVC Recognition for Wearable ECGs Using Modified Frequency Slice Wavelet Transform and Convolutional Neural Network. In: LI, Jianqing a Chengyu LIU. *Computing in Cardiology 2019*. Singapore: IEEE, 2020, s. -. ISBN 978-1-7281-6936-1. ISSN 2325-887X.

Doktorské projekty

Komunikační technologie a informační bezpečnost I.

COMPROMITATION OF NETWORK INFRASTRUCTURE THROUGH THE DEFAULT SETTINGS OF WINDOWS OS

Daniel Paučo

Doctoral Degree Programme (1st), FEEC BUT E-mail: xpauco00@stud.feec.vutbr.cz

> Supervised by: Lukáš Malina E-mail: malina@feec.vutbr.cz

Abstract: The number of cyber-attacks grows every single day. In fact, in 2020, the Microsoft reported about 64% increment of the number of reported vulnerabilities in the last 5 years. In this article, we present and demonstrate some recent cyber-attacks and security threats that can compromise whole domain network in many current enterprises. These attacks use mainly default Windows OS services, settings and protocols.

1 INTRODUCTION

Computer networks are nowadays expanded almost everywhere. From households or shops to factories, there are computers that are connected to the Internet from the majority or they are at least interconnected in the Local Area Network (LAN). As these systems are connected to the internet, the risk of compromitation is increasing. If they are not connected to the internet but still communicating in a LAN, there is still risk of compromitation for example from a perspective of a dissatisfied employee who has access to the network.

The main objective of this article is to describe and analyze default settings of Windows OS protocols, analyze the attacks aimed at these protocols, and describe the countermeasures to mitigate the critical information leakage leading to system compromitation. These attacks were chosen because they are considered as "low-hanging fruits". It means, that they are easy to realize and could potentionaly bring sensitive information. The attacks performed during this research were aimed at Windows Server 2008, 2012 and 2016.

This article describes a network infrastructure based on Windows OS and its protocols. It comes to protocols that could lead to information leakage when set up improperly. These information could be used for example to authenticate to the system. Scenarios and attacks described in this article simmulate the attacker located in the internal network of the given infastructure. The main objective of the attacker is to compromise Domain Controller (DC) that runs as a control point of the whole domain. The role of the domain controller is to store authentication data for the whole domain, control the access to the domain services, etc.

2 ATTACK SCENARIO PREPARATION

To demonstrate described attacks there was designed and implemented cyber exercise with a scenario of penetration test of certain company. This exercise was implemented in virtualization platform VMWare. There is a network infrastructure consisting of Demilitarized Zone (DMZ) containing Linux servers and internal network (SRV) consisting of Windows Server 2012 and Windows Server 2016.

The Figure 1 represents the scheme of the laboratory network. The simulation of penetration tester starts out of the company network by a reconnaissance phase where the pentester explores Domain



Figure 1: Scheme of the network infrastructure

Name System (DNS) server and web server situated in DMZ. This web server contains preinstalled advanced web shell that has to be found and used to get reverse shell to the pentesters machine (10.0.76.66). This compromised machine could be used as a pivot to the internal network of the company what is the next step of the exercise. This intrusion is possible due to improper system configurations and its protocols. Files server has Remote Desktop (RDP) service enabled that is not secured well and can be used through the proxychains that forwards connection to internal network. This service allows connections from the DMZ so as the pentester connects to the system, he/she can abuse prepared vulnerability in Utilman.exe. All systems are interconnected in the domain BREW.EX. The main goal of the pentester is to compromise DC and dump the credentials of each domain user. The pentester gets the access to the DC by using the Pass the Hash attack in which he uses credentials of domain administrator who was logged in the files server to configure some stuff. The last step of the exercise is to create a Golden ticket that serves to sign kerberos tickets and grant the pentester access to the all domain services.

3 WINDOWS OS DEFAULT SETTINGS

The most of current ICT systems connected to the network uses name resolution to be capable of communicate between each other and to work properly as well. To give an example, if a client wants to communicate with a name address of "example.com", it must be done a name resolution in the network to set up a proper routing of data flow. This communication by default looks like following. The client who wants to communicate through the network firstly tries to resolve the name to the IP address using its cache memory. If the cache memory doesn't contain required record, the client sends the query to the DNS server. If neither DNS server does not resolve the name, the client sends the query across the whole network segment. To achieve this, it uses Link-Local Multicast Name Resolution (LLMNR), NetBIOS-NS and mDNS protocols. In a case that no system knows the answer, the resolutions fails. However if there is an attacker listening in the network, he/she accepts the query, responds to the client and deceive it that he/she knows the address as can be seen in Figure 2. The motivation behind this behaviour is gaining authentication data from the client who sent the query.



Figure 2: Client is sending the query to the network

3.1 INTERCEPTION OF AUTHENTICATION DATA

As it can be seen from Figure 2 if there is the attacker in the network, he can spoof an authoritative source resolving names to the IP addresses and force the clients to communicate with his/her system. The attacks aimed at these services are called LLMNR/NBT-NS Poisoning and Server Message Block (SMB) Relay and they are classified as Man-in-the-Middle (MitM) attacks.

The attack scenario focused on LLMNR/NBT-NS/mDNS is following. The attacker spoofs the authoritative source serving to resolve names in the network and answers to sent queries pretending he knows the identity of the destination system. By this manner he poisons mentioned services and the clients will communicate with him. If the destination service requires authentication, the client sends its username and NT Lan Manager (NTLM) hash to the address provided by attacker. The first option what the attacker could do with the gained hash is to crack it offline. The second option is to forward the credentials to the destination system and create a session with it allowing him to run a code. This option can be done within the poisoned communication or independently from it. See [1][2] for more details.

Figure 3 shows the process of poisoning. The client sent query for the name SERVICE and the attacker responded with poisoned answer. The client accepted spoofed answer and tried to authenticate to his bogus service. The attacker obtained his IP address, username and NTLMv2 hash that he can use for offline cracking.

The attack scenario focused on Web Proxy Auto-Discovery (WPAD) protocol works in similar way. Each modern web browser has an option to set up proxy. As the article is focused on Windows domain, we used up-to-date Microsoft Edge web browser which is a default web browser of Windows 10 OS. After entering the web address the browser always checks the proxy settings. The default configuration is set to set up proxy automatically so with the each query for the new address there is a query sent to the network that is looking for proxy configuration file. If there is not DNS record for

<pre>[+] Listening for events [*] [NBT-NS] Poisoned answer sent to 172.16.0.3 for name SERVICE (service: File Server)</pre>					
[*] [LLMNR] Poisoned answer sent to 172.16.0.3 for name service					
[*] [MDNS] Poisoned answer sent to 172.16.0.3 for name service local					
[*] [LLMNR] Poisoned answer sent to 172.16.0.3 for name service					
[SMB] NTLMv2-SSP Client : 172.16.0.3					
[SMB] NTLMv2-SSP Username : WINDEV2008EVAL\User					
[SMB] NTLMv2-SSP Hash : User::WINDEV2008EVAL:4dfe7bebe0d8f40f:E6DA585E5D	82C6				
EE8B7649B84D0B7EDA:010100000000000000653150DE09D201D52EC3E8E21E06B1000000000	2000				
80053004D004200330001001E00570049004E002D00500052004800340039003200520051004	1004				
60056000400140053004D00420033002E006C006F00630061006C0003003400570049004E002	D005				
00052004800340039003200520051004100460056002E0053004D00420033002E006C006F006	3006				
1006C000500140053004D00420033002E006C006F00630061006C0007000800C0653150DE09D	2010				
60004000200000080030003000000000000000000	E2CB				
F8DC5011E8554E8AE1ACA94E979F630DC073F530A001000000000000000000000000000000000	0000				
900180063006900660073002F0073006500720076006900630065000000000000000000					

Figure 3: Interception of NTLM hash using Responder

the address "wpad.domain.ex", the query is sent through the LLMNR/NBT-NS/mDNS protocols as can be seen in the Figure 4.

1 0.000000	192.168.0.235	213.46.172.38	DNS	69 Standard query 0x36be A wpad.home
2 0.000578	192.168.0.235	224.0.0.251	MDNS	70 Standard query 0x0000 A wpad.local, "QM" question
3 0.001037	fe80::192f:fc99:d0b	ff02::fb	MDNS	90 Standard query 0x0000 A wpad.local, "QM" question
4 0.001397	192.168.0.235	224.0.0.251	MDNS	70 Standard query 0x0000 A wpad.local, "QM" question
5 0.001817	fe80::192f:fc99:d0b	. ff02::fb	MDNS	90 Standard query 0x0000 A wpad.local, "QM" question
6 0.002557				98 <ignored></ignored>
7 0.003776				98 <ignored></ignored>
8 0.013364	213.46.172.38	192.168.0.235	DNS	144 Standard query response 0x36be No such name A wpad.home SOA a.root-servers.net
9 0.013925	192.168.0.235	192.168.0.255	NBNS	92 Name query NB WPAD<00>
10 0.014212	192.168.0.235	224.0.0.251	MDNS	70 Standard query 0x0000 A wpad.local, "QM" question
11 0.014736	fe80::192f:fc99:d0b	ff02::fb	MDNS	90 Standard query 0x0000 A wpad.local, "QM" question
12 0.015528	fe80::192f:fc99:d0b	ff02::1:3	LLMNR	84 Standard query 0x4ec7 A wpad
13 0.015706	192.168.0.235	224.0.0.252	LLMNR	64 Standard query 0x4ec7 A wpad
14 0.015955	192.168.0.235	192.168.0.255	NBNS	92 Name query NB WPAD<00>

Figure 4: Interception of requests for WPAD in Wireshark

These queries are abused by the attacker who runs the HTTP server and answers the queries. The web victim's browser tries to connect to the attacker's HTTP server to download the *wpad.dat* file however this action requires the authentication what means that client sends its username and NTLM hash.

Figure 4 shows the common procedure what is done in the background when the user enters the address to the address bar. The procedure corresponds with the procedure described in Section 3.

3.2 PASS THE HASH ATTACK

This attack is focused on bypassing the authentication mechanism and most often is used against Windows OS. The attack depends on a Single Sign-On (SSO) functionality of authentication mechanism used by authentication protocols as NTLM or Kerberos. With SSO, users can enter their passwords once to be able to use resources they have been given rights to, without prompting them for their passwords again. For more details see [3].

In authentication process, the client's password is converted to the hash and after this action it is discarded. In the following steps it is used only the obtained NTLM hash. This is the essence of the Pass the Hash (PtH) attack. The attacker does not need to have knowledge of the user's password. Everything he needs to authenticate to the server is the knowledge of user password's hash and the username.

To perform PtH attack we chose Metasploit and it's module named psexec to perform the attack with. Psexec is the utility from the Sysinternals package that serves to administer systems remotely [4].

4 MITIGATION

The best way to mitigate these type of attack in which the sensitive data are leaked is to disable use of protocols like LLMNR, NBT-NS and mDNS. These protocols are used only in a case that the DNS server is not available or the DNS record does not exist but despite of this are these protocols enabled by default in Windows OS. All mentioned protocols are nowadays used only behalf of legacy systems in the network. So if there are no systems that run Windows 2000 and older, these protocols are not necessary [5]. The next recommendation linked to this threat is to set up the DNS record for the WPAD protocol.

The second step could be to completely disable NTLM authentication in the network and deploy protocols like Kerberos. Another option is to enforce SMB signing at all devices in the network. SMB signing serves to authenticate the source and the authenticity of each SMB packet. So if the authentication fails, the packet is dropped and this would mitigate the scenario of MitM attack focused on SMB service [2].

There are more recommendations how to mitigate this type of attacks as for example the network segmentation or the use of the model with the lowest privileges. See the explicit description of this model at [6].

5 CONCLUSION

This article described and analyzed the methods of Windows domain compromitation through the default Windows OS settings. The main objective of the article was to analyze the protocols of Windows OS that are enabled by default and could lead to the compromitation of the whole domain. Our motivation behind writing up this article was an annual increase of cyber attacks and an effort to teach users and administrators how to mitigate this simple attacks with possible critical impact.

REFERENCES

- [1] SALVATI, M. (2017, June 2). *Practical guide to NTLM Relaying in 2017 (A.K.A getting a foothold in under 5 minutes)* [online]. [cit. 2020-11-25]. Available at: https://byt3bl33d3r.github.io/practical-guide-to-ntlm-relaying-in-2017-aka-getting-a-foothold-in-under-5-minutes.html
- [2] KUEHN, E. (2018, April 11). *Ever Run a Relay? Why SMB Relays Should Be On Your Mind* [online]. [cit. 2020-11-25]. Available at: https://blog.secureideas.com/2018/04/ever-run-a-relay-why-smb-relays-should-be-on-your-mind.html
- [3] EWAIDA, B. (2010, January 21). *Pass-the-Hash attacks: Tools and Mitigation* [online]. [cit. 2020-12-12]. Available at: https://www.sans.org/reading-room/whitepapers/testing/pass-the-hash-attacks-tools-mitigation-33283
- [4] MICROSOFT: *Windows Name resolution*. Microsoft 29.06.2016 [online]. [cit. 2020-12-12]. Available at: https://docs.microsoft.com/en-us/sysinternals/downloads/psexec
- [5] MICROSOFT: Windows Name resolution. Microsoft 29.06.2016 [online]. [cit. 2020-12-12]. Available at: https://docs.microsoft.com/en-us/previous-versions/windows/it-pro/windows-server-2003/cc728457(v=ws.10)?redirectedfrom=MSDN
- [6] BEYOND TRUST: *Least Privilege*. BeyondTrust [online]. [cit. 2020-12-16]. Available at: https://www.beyondtrust.com/resources/glossary/least-privilege

ANALYSIS AND DETECTION OF SLOWCOMM AND SLOW NEXT DOS ATTACKS

Marek Sikora

Doctoral Degree Programme (4.), FEEC BUT E-mail: marek.sikora@vutbr.cz

Supervised by: Vaclav Zeman E-mail: zeman@feec.vutbr.cz

Abstract: This paper describes the proposal of Slowcomm and Slow Next attack models and their implementation. The attack generator is used to discover the vulnerabilities of several Internet services in an experimental network. Based on the described characteristics of attacks, these attacks' detection methods are investigated and implemented as an intrusion detection system (IDS).

Keywords: Slow DoS attacks, Slowcomm, Slow Next, attack generator, IDS, intrusion detection system, detection signatures.

1 INTRODUCTION

This paper is focused on the slow DoS (Denial of Service) attacks that are independent of the application layer protocol. DoS attacks are trying to prevent legitimate users from accessing a specific service. DoS attacks have gradually evolved from primitive lower-layer flood attacks to sophisticated application-layer attacks. Nowadays, attackers most often attack user-end applications using unsecured vulnerabilities. Due to lower-layer protocols' valid use, effective application attack detection is one of the leading research challenges [1].

1.1 RELATED WORKS

Almost no scientific works deal more deeply with the Slowcomm and Slow Next attacks, so this paper article is based primarily on [2] and [3], which present a novel of these attacks. However, several related works deal with the categorization of slow DoS attacks [1], general detection techniques [4], or summarize methods of prevention and mitigation [5].

1.2 CONTRIBUTION

Due to the absence of a freely available Slowcomm and Slow Next attack generator, this paper presents its implementation of an attack generator. Following current research, this article proposes and implements specific detection signatures for these attacks. Improving network systems security is the primary goal of this paper.

2 SLOW DOS ATTACKS

Slow DoS attacks are characterized by low traffic and similarity to the traffic of legitimate users with a slow Internet connection [1]. Therefore, they can be performed from only one computer, mobile phone, smart or IoT device. The slow DoS attacks strategy consists of establishing the maximum number of connections to the server and keeping them open as long as possible. In this way, the server's input queues will be flooded, which leads to the dropping requests of legitimate users.

2.1 SLOWCOMM

The Slowcomm attack mechanism is universal and can be used to exploit various application protocols. The attacker establishes the maximum number of concurrent connections to the target server and sends an incomplete valid request. Instead of sending another piece of data, the attacker waits for the maximum time allowed for the connection not to be closed and then sends the next incomplete portion of the request. In this way, the attacker prolongs the transmission of the entire request for as long as possible. The average data transfer rate is usually less than 1 KB/s [2]. Based on the server configuration, the server can close the connection if it does not reach the allowed values, such as the processing time or the data rate. However, the attacker monitors the state of the connection, tries to adapt to the server's response, and re-establish the closed connections to re-occupy the victim's free resources. Thus, a legitimate user does not have the opportunity to establish communication with the server. This attack is practically very similar to the Slowloris attack. However, compared to it, Slowcomm is more sophisticated - it needs less bandwidth; it can attack more application protocols and respond to connection closures [2]. The course of the attack against the Hypertext Transfer Protocol (HTTP) server is shown in Figure 1.

2.2 SLOW NEXT

Similar to Slowcomm, this attack can be applied to various application protocols. It aims to establish a maximum connection with the target server and keep them occupied as long as possible. The attacker sends valid requests to the server, and the server then legitimately responds. The attack exploits the *next* parameter, which is the time between the end of the server response and the beginning of the next



Figure 1: Slowcoi

Figure 2: Slow Next attack model.

3 ATTACK GENERATOR

The generator is implemented as a Python 3.0 script, and it uses the Sockets library to connect to the target system. In this paper, the generator is used to attack HTTP, File Transfer Protocol (FTP), and Secure Shell (SSH) protocols.

Mandatory parameters for launching an attack are attack selection, target IP address, target port, and the number of connections the generator should establish. It is also possible to specify the data (payload) that the requests will contain, the frequency of requests, and the number of threads, including the initialization speed (only for Slow Next).

4 ATTACK DETECTION

Based on the obtained attack samples, it was possible to propose Slowcomm and Slow Next attacks detection signatures and implement an intrusion detection system (IDS). This detector was created in Python using the Scapy and LibPcap libraries.

Detection is based on the interception of communications and finding suspicious connections that match the attack's signature. Suspicious connections are stored in a database for subsequent evaluation of the situation. Each signature has a set of defined network communication thresholds for attack detection. Setting these thresholds is key to detection accuracy. Accuracy also depends on the volume of data collected. Because both attacks simulate a legitimate user with a slow connection, they are not detectable at the first occurrence. Detection thresholds are set based on values a normal user cannot reach, but an attack does. Both attacks are detected when more than 20 connections from one IP address are established. To detect Slowcomm attacks, network communication must contain 20 or more unfinished requests at the same time and a time limit of 10 seconds for processing all requests of each connection. To detect Slow Next attacks, network communication must contain more than two seconds between requests within one connection, the spacing between requests is constant with a tolerance of 0.1 seconds, requests size is constant with a tolerance of 10 characters. If any of the conditions are met, network traffic corresponds to the behavior of a Slowcomm or Slow Next attack.

5 TESTING

The test environment consists of four virtual machines - an attacker, a legitimate client, and two servers. In all tested scenarios, the attacker is trying to cause a denial of service to the user.

The targets of the attack were chosen: vsftpd (FTP server), OpenSSH (SSH server), Microsoft IIS, Apache, Nginx, and lighttpd (web servers). All servers were left in the default configuration. The Apache server had active modules reqtimeout and mpm_prefork.

A summary of the results of vulnerability testing is shown in Table 1. More detailed information on the tests is described in chapters 5.1 and 5.2. Detailed test results are also shown in Figure 3 and 4. Web server availability is expressed by a green line that takes only two values: maximum – the server is available; minimum – the server is unavailable. The attacker's established Transmission Control Protocol (TCP) connections are expressed by the orange line.

Server	Slowcomm	Slow Next
Apache 2.4.29	1	✓
Nginx 1.14.0	1	×
MS IIS 10.0	×	×
lighttpd 1.4.45	1	1
vsftpd 3.0.3	1	1
OpenSSH	1	×

Table 1: Discovered vulnerabilities.

5.1 SLOWCOMM ATTACK

The Slowcomm attack on the Apache server was configured to open 300 concurrent connections and send another piece of data every 7 seconds. The web server handles approximately 225 connections, so it is vulnerable to this attack. Even if the reqtimeout module terminates the connection after 20 to 40 seconds, the attack generator immediately re-establishes the connection. The course of this attack scenario is shown in Figure 3.

When testing the vulnerabilities of other web servers, a vulnerability was also found for Nginx and lighttpd servers. The only Microsoft IIS that uses asynchronously processed thread groups was completely immune to the attack [6]. The Nginx server, which has a different event-driven architecture than Apache, was able to handle many more connections with minimal memory requirements [7], but after establishing 800 connections, it began terminating all connections. This paradoxically led to so much server load that it made the website unavailable.

The FTP attack scenario contained the initial data payload USER <name>, representing an incomplete FTP message. The following pieces of data always contained one random character. When the attacker established 2000 concurrent connections, the legitimate user could no longer establish a session with the server.

The attack on SSH revealed independence of request content. Data only needs to be incomplete or invalid. Compared to other tested services, SSH requires more computing capacity to establish communication. Establishing 12 connections was sufficient to perform a DoS attack.

5.2 SLOW NEXT ATTACK

The Slow Next attack on the Apache server was configured to open 700 connections and send portions of data every 4 seconds. After establishing approximately 680 connections, the DoS attack was successful. The web server was flooded, and a legitimate user could not connect to the server. The course of this attack is shown in Figure 4. The difference against the Slowcomm attack could be observed in the data rate. Slow Next generates an average data flow of 20 kB/s, thanks to the transmission of valid requests and responses.

When deploying the attack on other Web servers, only the lightpd vulnerability was discovered. Microsoft IIS and Nginx have not been affected.

The Slow Next attack on FTP was similar to the Slowcomm attack. The server became unavailable when the attacker established around 2000 connections. The Slow Next attack on SSH failed. SSH requires only valid communication. When an SSH message is repeated, the server terminates the communication. Slow Next could not be used for the username entry phase either, as the user has only three attempts and a time limit of 60 seconds. Also, the client must first be prompted by the server to enter his login.



Figure 3: Slowcomm attack on Apache server.

Figure 4: Slow Next attack on Apache server.

5.3 **DETECTION**

In the case of non-distributed attack scenarios, the detector detected an attack immediately when 20 established connections with invalid content from one source IP address are exceeded. Detection

times differed depending on attack type and settings. The attacks against HTTP were detected approximately 0.2 seconds after launch. In the case of a distributed DoS simulation, each attacker's computer established a single connection. Attacking connections initially appears to be legitimate users, but their behavior over time meets the criteria for classifying an attack. In the HTTP scenario, each Slowcomm attack connection was detected 10 seconds after launch when the allowed time to complete the request was exceeded. Slow Next attack was detected approximately 60 seconds after launch, once enough client requests have been accumulated and their distribution has been analyzed.

6 CONCLUSION

The Slowcomm and Slow Next attack generator proved the vulnerability of the tested HTTP, FTP and SSH servers. Based on the obtained attack samples, IDS was implemented and tested in a laboratory network. In all test scenarios, the attack detection was successful. However, the speed of a distributed attack detection is highly dependent on the appropriate setting of the allowable parameters of network traffic. A detection delay may vary. Future work may focus on improving the accuracy of detection and analysis of the occurrence of false positives in the network with real traffic.

7 ACKNOWLEDGMENT

The research was supported by the Technology Agency of the Czech Republic under grant no. FW01010474.

REFERENCES

- Cambiaso, E., Papaleo, G., Chiola, G., Aiello, M.: Slow DoS attacks: definition and categorisation, International Journal of Trust Management in Computing and Communications, 2013, DOI: 10.1504/IJTMCC.2013.056440
- [2] Cambiaso, E., Papaleo, G., Aiello, M.: Slowcomm: Design, development and performance evaluation of a new slow DoS attack, Journal of Information Security and Applications, 2017, DOI: 10.1016/j.jisa.2017.05.005
- [3] Cambiaso, E., Papaleo, G., Chiola, G., Aiello, M.: Designing and Modeling the Slow Next DoS Attack, International Joint Conference. Cham: Springer International Publishing, 2015, pages 249–259, Advances in Intelligent Systems and Computing, DOI: 10.1007/978-3-319-19713-5_22
- [4] Aiello, M., Cambiaso, E., Scaglione, S., Papaleo, G.,: A similarity based approach for application DoS attacks detection, 2013 IEEE Symposium on Computers and Communications (ISCC), Split, Croatia, 2013, DOI: 10.1109/ISCC.2013.6754984
- [5] Mahjabin, T., Xiao, Y., Sun, G., Jiang, W.: A survey of distributed denial-of-service attack, prevention, and mitigation techniques, International Journal of Distributed Sensor Networks, 2017, DOI: 10.1177/1550147717741463
- [6] Tanwar, S.: Hands-On Parallel Programming with C# 8 and .NET Core 3 [online], Birmingham: Packt Publishing, 2019, p. 251, [cited 2020-12-09], ISBN 978-1-78913-241-0, URL: <https: //bit.ly/3du0Zkj>
- [7] Garrett, O.: Inside NGINX: How We Designed for Performance & Scale, NGINX [online], Seattle: F5 Networks, 2015 [cited 2020-12-09], URL: <https://www.nginx.com/blog/ inside-nginx-how-we-designed-for-performance-scale/>

VERIFICATION OF TRANSMISSION PARAMETERS OF MULTICORE OPTICAL FIBER

Michal Látal

Doctoral Degree Programme (2), FEEC BUT E-mail: xlatal08@stud.feec.vutbr.cz

> Supervised by: Petr Münster E-mail: munster@feec.vutbr.cz

Abstract: With the ever-increasing demands on the transmission parameters of telecommunication optical systems, such as transmission capacity, new challenges arise for scientists to meet these requirements. One of the possible solutions to increase the transmission capacity of telecommunication networks is the multicore optical fiber, which is the current trend in modern space division multiplex. First and second chapters of the article deal with the modern concept of space division multiplex. A selected seven-core optical fiber was tested for insertion loss, attenuation, length for individual cores, and fiber continuity using the direct and backscattering method. The achieved results of the performed measurements are summarized and commented in the third chapter of the article.

Keywords: attenuation, crosstalk, fiber, loss, multicore, space division multiplex

1 INTRODUCTION

In telecommunication systems, a number of technological discoveries and advances have taken place over the last few decades, which have enabled an increase in the information capacity of transmission networks. In the early 1980s, the low-loss silica optical fiber was developed. One of the milestones in the development of the increase in information capacity in transmission networks was the development of a fiber amplifier, a method of amplifying an optical signal using the Erbium-doped fiber amplifier (EDFA). Another significant advance was the method of increasing the number of transmission channels per fiber, which was aided by the method called Wavelength Division Multiplex (WDM). This method enabled an even wider use of already laid and installed optical fiber networks. The use of the mentioned technologies in combination with newly laid optical fiber networks made it possible to fulfill the idea of bringing optical fiber to the end user in the house. This idea was called fiber to the home (FTTH) [1]. For current optical fiber networks and systems, existing optical fibers represent the limits of their capabilities in terms of the amount of data transferred. It is therefore still necessary to look for new ways to handle the growing amount of transmitted data.

The maximum transmission capacity of optical fiber systems and networks can be expressed using the Shannon-Harltey theorem. Using the Shannon-Hartley theorem, the maximum rate at which data can be sent in a single transmission channel can be expressed based on the bandwidth of the transmission channel and the signal to noise ratio. Maximum transmission capacity of optical fiber networks has many different limitations which are related to physical limitations of optical fibers to bind only a certain amount of energy able to transmitt in the fiber before non linear effects start to occur [2]. Due to this, transmission capacity higher than 100 Tbit/s cannot be achieved with standard single-mode optical fibers (SMF) [3]. With ever-increasing demands on transmission capacity in the field of information transmission such as the Internet of Things (IoT), Industry 4.0, cloud storage, transmissions in 5G mobile networks and also demands for high-quality HD, 4K, 8K streaming services and online video conferencing, transmission capacity can become insufficient for future applications of optical transmission networks.

One of the possible solutions to increase the transmission capacity of optical fiber systems is the use of optical cables with a larger number of integrated fibers. However, problems can arise in densely populated urban areas, where the already existing optical infrastructure is on the edge of its spatial capacity, where it is not possible to increase the number of optical cable routes. Building new routes for optical cables is very demanding, laborious and, above all, expensive. So current trends should lead to the minimization of installation costs and to the miniaturization and integration of modern technologies.



Figure 1: Evolution of transmission capacity in optical fibers [3]

2 SPACE DIVISION MULTIPLEX

One solution to the problem of maximum transmission capacity could be a solution using the telecommunication system called Space Division Multiplex (SDM). Space division multiplex uses individual spatial channels for transmission, using multiple cores and/or modes in one common optical fiber. The main advantage of Space division multiplex technology is that its use will not increase the number of installed optical cables and at the same time reduce the cost of installing new optical routes. These benefits make this technology probably the most efficient way to deal with the ever-increasing demands on the transmission capacity of telecommunication networks.

Space division multiplex can be implemented in several ways. One way to implement a space division multiplex is the Mode Division Multiplex (MDM) using optical fibers with several modes as individual transmission channels. Another possibility is the single-mode multicore optical fiber.

Using combinations of these possibilities, researchers in previous years achieved, for example, a transmission capacity of 2.15 Pbit/s using a 22-core single-mode optical fiber with a length of 31 km and 64QAM modulation in each core [4] and 2.05 Pbit/s ($360WDM \times 114SDM \times 50$ Gbit/s) using a Super-Nyquist-WDM DP-QPDSK signal over a 9.8 km 6-mode 19-core optical fiber [5]. The researchers in article [6] using a 6-mode 19-core optical fiber, this time with a length of 11.3 km, were able to achieve a transmission capacity of up to 10.16 Pbit/s.

In 2020, Google began laying an underwater fiber optic cable connecting the shores of the United States (Virginia Beach) and France (Saint-Hilaire-de-Riez). The submarine optical cable with a length of about 6500 km called "Dunant" uses the advantages of the space division multiplex technology. Dunant uses 12 pairs of fibers along with a number of technical innovations aimed at maximizing the bandwidth, thanks to which a transmission capacity of up to 250 Tbit/s can be achieved. At the beginning of February this year, the Dunant submarine optical cable was put into operation, making it the fastest connection between two continents [7].

3 **MEASUREMENTS AND RESULTS**

At the Department of Telecommunications, Brno University of Technology, we had a multicore optical fiber IPT-CORE developed by the InPhoTech company available to verify the properties. The multicore fiber was provided for testing by CESNET association. The IPT-CORE multicore optical fiber contains seven individual separate cores, where each fiber core meets the requirements of transmission parameters in accordance with the ITU-T Recommendation G.652, so you can take full advantage of this recommendation for efficient multiplexing using WDM. An important element in the IPT-CORE system is the passive fan-in/fan-out component, which allows at both ends of the fiber to send and receive information from/to each fiber core separately, thus effectively ensuring the functionality of seven cores in one fiber. As shown in Figure 2, all seven individual cores were connected to an optical switchboard at ports 1-7 and 18-24.



Figure 2: Wiring diagram of individual fiber cores

Verification of the multicore optical fiber was performed by first verifying the transmission properties, specifically measuring losses and crosstalk of individual cores by a direct method and then by verification of fiber continuity, fiber (core) lengths, attenuation and loss using the optical reflectometry OTDR (Optical Time Domain Reflectometry).

The first measurement by the direct method was performed using the source of optical radiation type FLS-600 by EXFO and the optical power meters type PM-420, PM-800 by OPTOKON and FPM-600 by EXFO. Three measurements were performed, always for the same source of optical radiation with a wavelength of 1550 nm and different optical power meters. The performed measurements were then compared with the values given in the datasheet of the fiber manufacturer. The measured values of insertion loss of individual cores and catalog values of the fiber are recorded in Table 1.

Fiber core	1–18	2–19	3–20	4–21	5-22	6-23	7–24
PM420	6.05	_	6.35	8.87	6.39	5.56	6.58
PM-800	6.09	—	6.16	8.95	5.85	5.66	6.49
FPM-600	6.98	—	5.91	8.58	5.77	5.30	6.13
Datasheet ¹	5.70	5.20	6.10	9.30	5.70	8.50	8.10

¹INPHOTECH. 2020. Multicore Fiber with Fan In/Out datasheet. SN: 7CF2F201112.

For measurements by the direct method, the difference between the values of individual measurements did not change significantly. For individual measurements the difference between the values (maximum and minimum loss values) was at most 0.93 dB. The maximum and minimum values of the differences between the measured values and the values given in the datasheet by the manufacturer differed in the range from -3.20 to 1.28 dB. The differences between the measured and datasheet values of the insertion loss do not indicate the quality of the measurement or the error rate of the measurement. It can be assumed that for the needs of the manufacturer or the datasheet, other, more sophisticated methods were used to measure the insertion loss of a multicore optical fiber and more accurate measuring instruments were applied.

In the second part of the measurement, the continuity of the fiber was verified using the OTDR method. The length of the fiber was verified, specifically the lengths of individual cores, loss and attenuation. The measurements were performed using the OTDR module SOT-A80 by Atomowave. Measurements were performed for individual cores 1–7 in both directions, with set parameters: wavelength 1550 nm, range 1 km, pulse width 25 ns and measurement time 30 s. The length of the launch cable was 500 m. The measured values of the individual parameters are compared in Table 2.

Dir. $A \rightarrow B$	Loss	Attenuation	Length	Dir. $B \rightarrow A$	Loss	Attenuation	Length
Fiber core	[dB]	[dB/km]	[m]	Fiber core	[dB]	[dB/km]	[m]
1	0.051	0.163	312	1	0.060	0.192	312
2	0.050	0.160	312	2	-	-	-
3	0.047	0.151	312	3	0.063	0.201	312
4	0.041	0.131	312	4	0.156	0.494	316
5	0.046	0.148	311	5	0.364	1.170	311
6	0.058	0.183	315	6	0.054	0.174	311
7	0.052	0.166	312	7	0.050	0.159	312
Min	0.041	0.131	311	Min	0.050	0.159	311
Max	0.058	0.183	315	Max	0.364	1.170	316
Median	0.050	0.160	312	Median	0.062	0.197	312
Average	0.049	0.157	312	Average	0.125	0.398	312

Table 2: Val	lues from OTDR	measurements f	for all fiber	r cores ii	n both directions

The graphical dependence of power on distance is shown in Figure 3, specifically for the core number 7. The graph in Figure 3 shows approximately 500 m of launch fiber (event 1–2) after which a measured multicore fiber with a length of approximately 300 m (event 2–3) was connected.



Figure 3: Graph of the dependence of backscattered optical power on distance

The resulting OTDR measurements for the loss, attenuation and length transmission parameters for the $A \rightarrow B$ direction were as expected. From the measured maximum and minimum values of the transmission parameters, it can be seen that the values did not change significantly during the whole measurement and did not fluctuate to extremes, so the average value was very close to the median value. The average values for the $A \rightarrow B$ direction were 0.049 dB for the loss, 0.157 dB/km for the attenuation and 312 m for the length.

On the contrary, it was for the $B \rightarrow A$ direction where the measurement results did not meet expectations. For the second fiber core (measured from port 19), it was not possible to measure the transmission parameters at all. The cause could be an insufficiently connected connector from the inside of the optical switchboard. Due to the warranty of the whole product it was not possible to correct this fact. Furthermore, the fiber cores 4 and 5 (measured from ports 21 and 22) showed a greater rate of signal loss and thus increased attenuation for a given length. The fifth fiber core, for which the maximum values of the loss and the attenuation reached 0.364 dB and 1.170 dB/km respectively, would not even meet the minimum requirements for transmission quality of certain applications.

4 CONCLUSION

Multi-core fibers represent an innovative solution designed for space division multiplex. They can transmit multiple times (depending on the number of cores) more information than conventional single-core optical fibers, while maintaining the same geometric and transmission parameters. Within the article, individual transmission parameters of the selected multicore optical fiber were tested. The possible occurrence of crosstalk between individual cores, which did not occur during the measurement, was also tested. The chosen design of multi-core optical fiber allows its easy and fast deployment in industrial and telecommunication applications that can be implemented in existing networks.

ACKNOWLEDGEMENT

The multicore fiber from the CESNET association was used for the measurements.

REFERENCES

- AL-AMRI, M. D., M. M. EL-GOMATI and S. M. ZUBAIRY. 2016. Optical Communication: Its History and Recent Progress: Optics in our time [online]. New York, NY: Springer Berlin Heidelberg. Available at: https://link.springer.com/chapter/10.1007/978-3-319-31903-2_8
- [2] MITRA, Partha P. and Jason B. STARK. 2001. Nonlinear limits to the information capacity of optical fibre communications. Nature. 411(6841), p. 1027-1030.
- [3] RICHARDSON, D. J., J. M. FINI and L. E. NELSON. 2013. Space-division multiplexing in optical fibres. Nature Photonics. 7(5), p. 354-362.
- [4] PUTTNAM, B. J., R. S. LUIS, W. KLAUS, et al. 2015. 2.15 Pb/s transmission using a 22 core homogeneous single-mode multi-core fiber and wideband optical comb. In: 2015 European Conference on Optical Communication (ECOC). Spain: IEEE, 201, p. 1-3.
- [5] SOMA, D., K. IGARASHI, Y. WAKAYAMA, et al. 2015. 2.05 Peta-bit/s super-nyquist-WDM SDM transmission using 9.8-km 6-mode 19-core fiber in full C band. In: 2015 European Conference on Optical Communication (ECOC). Spain: IEEE, 201, p. 1-3.
- [6] SOMA, Daiki, Yuta WAKAYAMA, Shohei BEPPU, et al. 2018. 10.16-Peta-bit/s Dense SDM/WDM Transmission over 6-Mode 19-Core Fiber across the C L Band. Journal of Lightwave Technology., p. 1-1.
- [7] LARDINOIS, Frederic. 2021. Google's new subsea cable between the US and Europe is now online. TechCrunch [online]. Available at: https://techcrunch.com/2021/02/03/googles-new-subseacable-between-the-u-s-and-europe-is-now-online/

GPON ATTACKS AND ERRORS CLASSIFICATION

Adrián Tomašov

Doctoral Degree Programme (1st), FEEC BUT E-mail: xtomas32@vutbr.cz

Supervised by: Tomáš Horváth E-mail: horvath@feec.vutbr.cz

Abstract: This paper focuses on various types of attacks and errors in an activation process of Gigabitcapable passive optical networks. The process sends messages via Physical Layer Operation Administration and Maintenance header field inside the transmitted frame. An exemplar network communication is captured by a special hardware-accelerated network interface card capable of processing optical signals from passive optical networks. The captured data is filtered of irrelevant parts and messages and correctly formatted into a suitable shape for a neural network. The filtered data is divided into small sequences called time windows and analyzed using a recurrent neural network-based on Gated recurrent unit cells. A new neural network model is designed to classify sequences into several categories: additional message, missing message, error inside (noisy) message, and message order error. All of these categories represent a certain type of attack or error. The proposed model can distinguish message sequences into these categories with high accuracy resulting in revealing a possible attacker or drift from protocol recommendation.

Keywords: Activation Process, GPON, GRU, Recurrent Neural Network, PLOAM

1 INTRODUCTION

Passive optical networks are currently the most promising solution providing broadband internet connectivity, video streaming, etc. The distributional network consists only of passive components (e.g., fiber cables, splitters) and has a tree-like topology, where Optical Line Termination (OLT) is the root and leaves are devices called Optical Network Unit (ONU). The messages are broadcasted into the network and received by all members of communication. Thus, it is necessary to keep confidentiality of transmitted data and synchronization due to multiple access.

Gigabit-capable Passive Optical Networks (GPON) protocol is defined by a set of recommendations by Inter-national Telecommunication Union (ITU) [1]. A recommendation is not as strong as standard, so device vendors can modify this protocol to match their requirements without proper description. This allows attackers to send malicious messages of various formats through the specific parts of the frame without deciding whether the message is a part of the protocol or a malicious message sent by an attacker. With a specific targeted attack, the attacker would prevent clients from communicating through the network, change the priority of data flow and proceed with different denial of service or abuse them for his benefit.

One of the weak points is the activation process. This process is defined through a finite state machine, and a transition between states is triggered by specific messages located in a Physical Layer Operation Administration and Maintenance (PLOAM) field in a frame header. This process's main purpose is to synchronize newly activated ONUs and prepare them for transmission. It also handles error states (loss of signal, loss of framing), resulting in restarting this process and temporary outage.

This paper focuses on the development of a classification model based on Machine Learning (ML) algorithms. The model can learn and recognize messages and communication patterns from exemplar

communication, which represents trusty communication. The model analyzes message sequences in GPON and classifies every malicious or strange behavior different from the exemplar one. This type of analysis is specific and requires a particular type of neural network capable of time series evaluation.

Neural networks are not a novelty in ML; however, they were limited by computational resources until recently. Today technology supports hardware acceleration allowing the development of complex machine learning models applied in various fields.

Time series analysis is a specific ML problem requiring a modified model for a certain data format. Models based on recurrent neural networks are often used for such analysis. These models consist of recurrent cells capable of sharing internal state/states to themselves through sequence processing. Xueqi Zhang et al. (2020) [2] evaluate and compare traditional predictive methods against a recurrent neural network based on Long Short-Term Memory (LSTM) cells. These models predict an amount of ambient town noise. The best accuracy has the model with LSTM layers.

Xin Wei et al. (2021) [3] compare neural networks based on LSTM, Gated Recurrent Unit (GRU), Simple Recurrent Neural Network (SimpleRNN) and based on several dense layers. These models predict pore-water pressure depending on the amount of rainfall water in a certain time segment. LSTM and GRU layer-based models have clearly the best accuracy in such prediction. At the same time, authors refer to GRU cell effectively, which requires about 40% fewer resources compared to LSTM based model.

2 ANALYZED DATA

The most important part of each ML project used for data analysis is the learning dataset. Even a perfect neural network design would have poor performance, accuracy, and generality with insufficient dataset quality. Therefore, it is necessary to keep enough attention to the data itself, which are used as examples for the ML model. The first step is to analyze the dataset using mathematical and statistical tools to reveal any possible hidden error or bias. This section describes data capturing and data pre-processing algorithms and the final data shape suitable for the neural network. In the end, there are function descriptions used for generating attacks and errors applied to captured communication.

2.1 DATA CAPTURE

This work is based on data captured from a real GPON network. The frames are captured by a custom Network Interface Card (NIC) containing Field Programmable Gate Array (FPGA) capable of converting signals from the optical domain into a suitable format for machine processing. The converted data from the optical network are sent into a data server, which parses the frames. The extracted information is being correctly formatted and stored in a database. This work is based on the database data, which are exported in JavaScript Object Notation (JSON) format.

The analysis is focused on the ONU activation process. Therefore, during the data capturing, ONUs are randomly disconnected from a power supply, or attenuation of optical connection is randomly changed. These events randomly execute the whole activation process several times by each ONU, which provides a wide range of the activation process usage and suitable data for the neural network.

2.2 DATA PRE-PROCESSING AND FORMAT

The captured dataset is filtered of irrelevant messages and message fields before preparing as input for the neural network. Considering the work is focused on attacks and errors in GPON activation process, the most relevant data are located in PLOAM field in the frame header. Other parts and message headers are deleted because they do not have any influence on the activation process. Consequently,

the rest of the data¹ is being inspected further. Messages with ID 11 are 99.98% of all captured messages and deleted from the dataset. Otherwise, they would cause enormous distortion, and trained models would not learn any useful patterns.

The output of the filtration process is a sequence of 13 bytes messages. This sequence is too long to be used directly as an input of the neural network model. Therefore, the sequence is divided into smaller segments called time windows. The number of messages in segments is 30, which is long enough to contain the whole activation process for several ONU units and short enough to be used as input for the learning process. After division, there is a list of time windows with shape 13×30 bytes. In the end, the dataset is normalized into float numbers by dividing by 255.

2.3 ERROR/ATTACK GENERATION

Several functions are used to generate data into the learning dataset containing searched errors. Each function's input is a cloned instance of captured messages, which is being modified by each function directly. The modified sequences are generated time windows with a specific width suitable for the neural network input. The generated error datasets are:

- **Deleted message error** dataset is generated by deleting important messages of the activation process. Important messages have these IDs: 4, 8, 10, and 18. This function iterates through each important message and deletes all occurrences of that message in a separate instance of captured message sequences. The output of this function is almost four times longer sequence compared to the input sequence.
- Additional message error generation process is also focused on messages of the activation process. This function iterates through message: 4, 8, 10, and 18, and for each occurrence of the current message, this function doubles the message. The output sequence is more than four times longer than the input sequence.
- **Order error** is generated using a function, which creates time windows according to the given format from the input sequence. The next step inverts messages in time windows, which clearly generates corrupted message sequences.
- **Invalid message error** containing invalid messages is generated by adding Gaussian noise into the captured communication with a mean equal to 0 and standard deviation equal to 0.2. The generated message sequence is different from the captured communication and contains various errors simulating a malicious device or possible attack.

3 NEURAL NETWORK MODEL

Inspiration for the model is natural language processing (text), speech recognition (voice), and various other time series classification models. Most of them use neural networks based on convolutional or recurrent layers. Based on similar researches and experiments during this study, the proposed neural network model is based on GRU recurrent layer, which has enough tools for long and complex sequences [4]. The GRU cell design is simpler compared to LSTM cell, so it requires fewer hardware resources, but still can solve the exploding/vanishing gradient problem gently.

3.1 DESIGN

Considering the data characteristic and format, there are two GRU recurrent layers with 64 cells at the beginning of the neural network followed by two dense layers. GRU layers can analyze time

¹The rest of the data contains PLOAM messages only.

series, which is exactly the problem this model investigates. The remaining dense layers conclude the outcomes of recurrent layers and classify the sample into the correct category. The number of neurons in the last dense layer is equal to the number of classified categories.

3.2 LEARNING

An exemplary dataset learns the model with generated sequences described in Section 2.3 and uses supervised learning. There is no early stop function used during learning, but the model's current state is saved after each learning epoch. The batch size parameter is set to 128. This process is executed exactly 200 epochs to reveal the tendency of loss and accuracy metrics when the network is being over-fitted. The optimizer used for learning is *Adam*. The loss function used for an error evaluation is sparse categorical cross-entropy, which can convert expected labels from numbers to one-hot encoding and evaluate the network error.

The learning dataset is obviously imbalanced, which is visible in data methods described in Section 2.3. Thus, class weights are computed to eliminate an error caused by the imbalanced dataset. These weights are used only during learning, and they scale the loss value for each neuron according to category differently. The weights are calculated using compute_class_weight from sklearn.utils.class_weight module. The whole dataset is divided into three disjunctive sets in a ratio of 70:15:15 to fairly evaluate the quality and other parameters of the final model.

3.3 **RESULTS EVALUATION**

The proposed neural network can recognize attacks with high accuracy. Considering the time window size, it can recognize possible attacks or errors through the PLOAM field in the ongoing communications. The network quality is evaluated using confusion matrices showing to which category each sample is classified. The confusion matrix evaluated on the testing dataset is visible in Figure 1. The figure proves that the recurrent neural network's proposed architecture accurately classifies message sequences into all categories. However, there are some false predictions of additional and missing dataset samples, possibly because these datasets may contain sequences of massages without modified parts. Thus, these sequences are the same as sequences in the exemplary datasets and should be deleted from datasets representing attacks and errors.



Figure 1: The confusion matrix of test set evaluation.

4 CONCLUSION

This work investigates the neural network design used for the analysis of PLOAM messages. Based on similar works focusing on time series analysis, the proposed model is based on layers with GRU cells. These cells are more effective than LSTM, but with almost the same accuracy of classification.

The exemplar communication is captured in the real GPON network in the laboratory environment. This data contains many useless parts, fields, and message types, and the data, if filtered to consists of related information only. The remaining data generate other datasets containing various types of errors or attacks on the activation process. All datasets are formatted into the time windows suitable for neural network learning.

The recurrent neural network's proposed design shows very high accuracy in analyzing the activation process, which is visible in the confusion matrix of the test dataset evaluation and the accuracy evolution of the validation dataset during learning.

This work is one of the first in the field of machine learning analysis of GPON protocol and reveals possibilities of developing an Intrusion Prevention System (IPS) and Intrusion Detection System (IDS) based on machine learning models. Furthermore, most importantly, this reveals the possibility of real-time inspection of the communication inside a passive optical network. This gives many options for enhancing this work in several topics:

- Anomaly detection models based on unsupervised learning, especially auto-encoders.
- Evaluation possibility of a convolutional network for messages sequence analysis.
- Enhance data capture process to gain ONT Management Control Interface messages.

ACKNOWLEDGEMENT

The research described in this paper was financed by a grant from the Ministry of the Interior of the Czech Republic, Program of Security Research, VI20192022135, PID VI3VS/746 for "Deep hardware detection of network traffic of next generation passive optical network in critical infrastructures".

REFERENCES

- [1] ITU-T. Gigabit-capable passive optical networks (g-pon): Transmission convergence layer specification. [online], 01 2014.
- [2] Xueqi Zhang, Meng Zhao, and Rencai Dong. Time-series prediction of environmental noise for urban iot based on long short-term memory recurrent neural network. *Applied Sciences*, 10(3):1144, 2020.
- [3] Xin Wei, Lulu Zhang, Hao-Qing Yang, Limin Zhang, and Yang-Ping Yao. Machine learning for pore-water pressure time-series prediction: Application of recurrent neural networks. *Geoscience Frontiers*, 12(1):453 467, 2021.
- [4] R. Dey and F. M. Salem. Gate-variants of gated recurrent unit (gru) neural networks. In 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), pages 1597– 1600, 2017.

SHARED DETECTION OF CYBERATTACKS ON VOIP EXCHANGES.

Jiri Jezek

Doctoral Degree Programme (1), FEEC BUT E-mail: xjezek13@vutbr.cz

> Supervised by: Pavel Silhavy E-mail: silhavy@feec.vutbr.cz

Abstract: In the last decade, security in Voice over IP raises on importance. This article focuses on shared defence systems in VoIP, summarizes the existing work related to this topic. In standard systems with basic security, it is common to monitor attacks and prevent them with certain countermeasures. A problem occurs when multiple PBXs are connected to one network because the individual PBXs do not know that other network entities are attacked. It can lead to the attack spreading to other devices connected to one network and significantly slowing the whole network. This paper proposes a solution focused on sharing information about an attacker with other network entities by using method MESSAGE from SIP protocol.

1 INTRODUCTION

Voice over IP (VoIP) is a technology designed to transmit multimedia (mostly audio) data over Internet Protocol (IP) networks. VoIP systems gradually replace traditional telecommunication systems. The main reason is that VoIP offers low cost and high quality of service for multimedia communications. Thanks to the development of 5G networks, VoIP is expected to become the dominant technology for communication based on voice [1].

For implementing VoIP, we can use plenty of protocols. Today mostly used is SIP (Session Initial Protocol) for signalling [9] in cooperation with SDP (description of relation) [10] and RTP (transmission of voice data) [11].

VoIP is a growing industry, and it makes it an attractive target for attackers. In 2017, a 29.2 billion US dollar loss was reported related to VoIP attacks, according to the Communications Fraud Control Association (CFCA). One of the leading causes of telecom fraud is Private Branch Exchange (PBX) hacking and toll fraud [8],[1]. Securing PBX is demanding for many organizations. Due to their usage attacks in popular services like web, email, ssh, traditional intrusion detection systems are not primarily meant to be used on PBX, so they are not as sufficient as they may be. Even with the Session Border Controller or edge gateway, PBX is still susceptible to a brute-force attack on SIP accounts with weak passwords. To secure the PBX and SBC correctly, human expertise is required to calibre rules and analyze the logs for abnormalities.

There can be a problem with detection attacks. PBX must detect the upcoming attack and apply defence mechanisms, and this will take some time. For higher efficiency of defence mechanisms, there can be some form of shared defence, where one PBX can contact others in the network and provide them with information about the attack and how to defence against it [2].

2 RELATED WORKS

Researchers have been investigating the options of shared defence mechanisms for PBX. In the experiment with VoIP based PBX honeypot, researchers confirmed that attacks on PBX are growing. This study has found out that almost 19 million SIP messages occurred on the experimental honeypot

in only ten days. This experiment also reacted to a similar experiment from 2009 to 2012, where between 3 years was realized 47.5 million SIP messages. We can see a considerable increase in attacks in this experiment. [3].

In another study, the authors proposed a solution with the database, which were connected nodes, which should be protected. For preventing attacks, was used system fail2ban. The proposed solution was for the general use of any systems, not specially for VoIP systems [4].

The next study is focused on the decentralized type of database – blockchain. The concrete solution runs on Ethereum blockchain. In their study, they also use fail2ban for preventing attacks. This work focuses especially on VoIP systems where information about attackers are written by smart contracts to Ethereum blockchain, to where are connected other PBX by a special interface [2].

3 PROPOSED SOLUTION

This paper proposes a solution that focuses on giving information about potential attacks using SIP protocol. Firstly, we need Intrusion Prevention System (IPS), which can identify and react to incoming attacks by applying defined rules. Based on previous researches on this topic, it seems that the best IPS in our case is Fail2ban. Fail2ban scans log files and ban IP addresses reporting too many failed login attempts. This happens by updating firewall rules to disable attempts for connection from those IP addresses for a configurable time [5].

The proposed solution is focused on sending information about an attacker using SIP protocol. If we look at SIP methods, concrete on requests, we can find some methods that can be used to provide information about an attacker, which was detected by IPS. To provide information about an attacker seems to be the best method MESSAGE, which can be easily detected and readable. Method MESSAGE is normally used for instant messaging. On image 1 you can see the flow of SIP MESSAGE request. Using MESSAGE method have also advantage that the recipient only accepts the message



Figure 1: SIP message flow

and responds by 200 OK (when the message was sent correctly) [6]. In this case, we do not need to devise a feedback system because it is provided natively by SIP protocol (200 OK response). With 200 OK response, is sender informed that every PBX on the network got his message with information about the attacker.

Now we know in what method we will provide data with to others PBX in the network. We need to define the format of information about attackers sent by the method MESSAGE. The main information is the attacker IP address, which is the key for subsequent blocking of attackers by the firewall. The next field can be the type of detected attack (for example, Toll Fraud). Important information can be the rate of attack in attempts per second. For correct receive, it is needed to distinguish if it is a message used by a user for instant messaging or an alert message signalising that PBX is under attack. For this, we will need the format message body starting with some special characters, which

Table 1: Main fields in SIP MESSAGE				
Field	Description			
ipattacker	Attacker IP address.			
attack	Type of attack (i.e Toll Fraud,)			
rate	Reported attempts per second			

are recognisable for receiving party. Special characters to recognise can be practically anything important is recognizability for both participants of the session. The final body of request MESSAGE then can look like this:

<sipdeff>ipattacker=192.168.0.1;attack=tollFraud;rate=10000;<sipdeff>

SIP protocol is a text-based protocol. So practically anybody can read and modify them (i.e. Man in the middle attack). This might be a big problem because we send an important message leading to blocking attackers IP addresses. This vulnerability of SIP can be abused for blocking legit users IP addresses instead of attacker IP address. For securing data exists protocol SIPS (Secured SIP). In this protocol are SIP messages encrypted by using TLS (Transport Layer Security) protocol. Securing SIP messages is necessary for the correct function of the proposed solution.

In table 1, you can see the proposed data, which will be sent by PBX, which is under attack. In the end, it is needed to detect the message with the IP address of the attacker. This depends on the used PBX system. We will need several scripts. First to generate and send a message to other PBXs in our network. To generate SIP MESSAGE, we need to use the SIP message generator, for example, open-source tool SIPp [7], where we can define SIP message format. The MESSAGE method's body must be defined with the correct format to be recognizable for other PBXs. On PBXs must also be defined script, which after receiving information about the attacker, will apply firewall rules.

In figure 2 you can see the scheme of proposed solution. In the first step, the attacker wants to attack PBX2. After some time, IPS reads from logs that there can be malicious activity and add a firewall rule to block the attacker IP address. Then PBX2, with the help of their scripts, generate SIP MESSAGE which will be encrypted by TLS protocol, with information about the attacker and send it to other PBX in the network (PBX1 and PBX2). Other PBXs, by using the scripts, add new firewall rules to attacker IP address. At this moment, the attacker cannot access any PBX in the network by his IP address.

4 DISCUSSION

The work proposes the theoretical solution to the growing problem of VoIP attacking. The proposed solution is focused on the network with a larger amount of PBXs. The disadvantage of the solution with one centralized database [4] is that attacker can directly attack that database and disable communication with PBXs (by DoS attacks), or even worse, gain access to the database and by adding some false data harm legitimate users. This is solved by sending messages with information about detected attacks to other PBXs directly by SIP, where is also a back control mechanism by confirmation received message by SIP protocol.

As mentioned above, the disadvantage also removes the second solution proposed in SIPChain [2]. There is a problem solved by the decentralized database, concrete by using Ethereum blockchain. The disadvantage of using SIPChain is that it must use another interface to communicate with the blockchain. This could lead to a potential security issue. There can also be a problem if some attacker starts providing false information about attackers to blockchain used as a database, and legit users can be blocked, or it can cause a large overhead.

The proposed solution's disadvantages can be compatibility between PBXs and processing information about attackers obtained from the message method. There is also the need to secure SIP messages,



Figure 2: Scheme of proposed solution

as was mentioned in the previous chapter. This will be probably a little slower than with not secured SIP messages, but it is necessary.

4.1 FUTURE WORK

Future work will focus on developing a protocol working above SIP, based on the proposed solution in this paper. The whole solution will also be practically tested and evaluated with the chosen software PBXs. Depending on the results will be the solution subsequently modified or added more security features.

We can write this into several points:

- Implement IPS to work with our chosen PBX system.
- Define the full format of the data in MESSAGE body.
- Write scripts to control receiving and sending SIP MESSAGES.
- Practically tests the whole solution.
- Make corrections in response to the tests.

5 CONCLUSION

This work focuses on the problem of shared security in the network of PBX. In the second chapter are briefly described already existing works on this topic. The first mentioned work is about the increase in the number of PBX attacks in the last years. The other two works focus on different solutions related to this topic. The next chapter proposes a theoretical solution, which focuses on

sending information about the attacker using the SIP protocol. The main idea is that when some PBX is under attack, with the help of IPS blocks the attacker IP address and generate and send SIP MESSAGE to others PBXs in the network. In SIP MESSAGE method are encapsulated information about the attacker in the defined format. When PBXs accept SIP MESSAGE, they apply firewall rules according to the obtained information. This theoretical solution will lay the foundation for future work, which will be focused on the implementation of the proposed solution. The effort is to create a shared defence protocol working above the SIP protocol by using method MESSAGE for carrying these data. This paper also discussed the advantages and disadvantages of the proposed solution to compare with existing methods, which have been described in the second chapter.

ACKNOWLEDGEMENT

The research was supported by the project FEKT-S-20-6312 "Research on electronic communications and information and systems and their use to secure critical infrastructures".

REFERENCES

- [1] Nazih, W.; Elkilani, W.S.; Dhahri, H.; Abdelkader, T. Survey of Countering DoS/DDoS Attacks on SIP Based VoIP Networks. Electronics 2020, 9, 1827. https://doi.org/10.3390/electronics9111827
- [2] A. Febro, H. Xiao and J. Spring, "SIPchain: SIP Defense Cluster With Blockchain," 2019 Principles, Systems and Applications of IP Telecommunications (IPTComm), Chicago, IL, USA, 2019, pp. 1-8, doi: 10.1109/IPTCOMM.2019.8920874.
- [3] McInnes, N., G. Wills and E. Zaluska. 'Analysis of threats on a VoIP based PBX honeypot.' (2019).
- [4] M. Ford et al., "A process to transfer Fail2ban data to an adaptive enterprise intrusion detection and prevention system," SoutheastCon 2016, Norfolk, VA, USA, 2016, pp. 1-4, doi: 10.1109/SECON.2016.7506771.
- [5] Fail2ban. Fail2ban [online]. 2016 [cit. 2021-03-07]. Available from: http://www.fail2ban.org/wiki/index.php/Main_Page
- [6] SIP method message. VoIP-info [online]. [cit. 2021-03-07]. Available from: https://www.voip-info.org/sip-method-message/
- [7] GAYRAUD, Richard a Olivier JACQUES. SIPp. SIPp [online]. 2014, 04/20/2014 [cit. 2021-03-07]. Dostupné z: http://sipp.sourceforge.net/
- [8] J. Osenbaugh, "Telecom fraud on the rise: What enterprises need to know." https://www.nojitter.com/security/telecom-fraud-rise-what-enterprises-need-know, February 2019. Web. [cit. 2021-03-07].
- [9] SIP: Session Initiation Protocol [online]. Network Working Group, 2002 [cit. 2021-03-07]. Dostupné z: https://tools.ietf.org/html/rfc3261
- [10] SDP: Session Description Protocol [online]. Network Working Group, 2006 [cit. 2021-03-07]. Dostupné z: https://tools.ietf.org/html/rfc2327
- [11] RTP: a Transport Protocol for Real-Time Applications [online]. Network Working Group, 2003 [cit. 2021-03-07]. Dostupné z: https://tools.ietf.org/html/rfc3550

OPTICAL FIBER INFRASTRUCTURE PROTECTION DEMONSTRATION SYSTEM USING DISTRIBUTED OPTICAL SENSING

Petr Dejdar

Doctoral Degree Programme (3), FEEC BUT E-mail: xdejda00@stud.feec.vutbr.cz

> Supervised by: Petr Münster E-mail: munster@feec.vutbr.cz

Abstract: The paper deals with optical fiber distributed sensing based on Mach-Zehnder interferometry for infrastructure protection. The system is used in the landscape model, where digging can be simulated. The developed system measures vibrations close to the optical fiber and evaluates the exceeding of the limit value of the signal deviation, on the basis of which the alarm is triggered. This demonstration model is used to demonstrate the possibility of protecting the optical fiber infrastructure from mechanical damage, such as optical cable cuts caused by digging activity. The paper describes both the design of the demonstration system and the GUI for setting and evaluating alarms.

Keywords: data acquisition, GUI, interferometry, optical fiber, photodetector, sensor

1 INTRODUCTION

With the growing need for increased transmission speeds in data networks, optical fibers are being used more frequently. With the growing number of optical cable installations, new possibilities offer the use of this medium for non-data applications, such as the transmission of ultrastable quantities or the sensing of acoustic/mechanical vibrations and temperature in the immediate vicinity of optical cables. Optical fiber sensors are now mainly used with new installations, but in special cases, these can also be used on existing infrastructures, in some cases also on fibers with active data transmission. Reduce of the costs for the installation of new optical cables thus decrease overall costs. Optical fiber sensors, such as, high temperature, humidity, strong electromagnetic interference, etc [1]. In addition, due to their size, accuracy, and wide range of measurement options, such as temperature, pressure, refractive index, or vibration [1][2], these sensors are increasingly used in non-standard conditions, such as aircraft and space shuttle sensors [3], or for measuring the properties of high voltage transformers [4].

There are many types of optical sensors. One important group are distributed optical sensors. These sensors can measure quantities along the entire length of the fiber and basically can be divided into two basic categories. The first category uses backscattering. The Rayleigh, Brillouin, and Raman scatterings are the most suitable. Each type of scattering is used for a different application. Circulators are used for signal analyzes which is reflected back from the measured optical fiber [5]. The second category uses interferometers. The most commonly used interferometer is the Mach-Zehnder interferometer. However, there are many others, such as Michelson, Sagnac, or Fabry-Perot interferometer [6], which is suitable for point measurements. Distributed sensors are more suitable for more sophisticated measuring systems, for example for measuring acoustic vibrations or for measuring the magnetic field.

2 OPTICAL FIBER SENSING SYSTEM

Vibration sensors are a key part of interferometric sensors. The Mach-Zehnder interferometer is shown in the Figure 1. Due to its simplicity, this interferometer is used in affordable sensor systems. A simple Mach-Zehnder interferometer can be used together with operational amplifiers and other components for vibration detection, when the limit is exceeded, the alarm is triggered [7]. Which is essentially a primitive but effective security feature. A Mach-Zehnder interferometer is also suitable for low-cost vibration detection of infrastructures such as buildings or bridges [8].



Figure 1: Scheme of Mach-Zehnder interferometer.

2.1 LASER

The basis of the interferometer that was used a was low-noise, 1550 nm laser Koheron LD100, shown in the Figure 2. The current is set by a precision trimmer from 0 to 55 mA and operates at a voltage from 3.3 V to 5 V. The LD100 is equipped with a modulation input connected to alternating current for current modulation from 1 kHz to 200 MHz. The LD100 typically has a spectral bandwidth of 3 MHz. The signal from the laser is not modulated in any way in the interferometer connection and light into the fiber is coupled with a constant wavelength [9].

2.2 PHOTODETECTOR

A Koheron PD100 photodetector is used to convert the light signal into an electrical, shown in the Figure 2. PD100 photodetectors are amplified photodetectors with a bandwidth of about 100 MHz. The photodetector contains an InGaAs photodiode with a wavelength range of 900–1700 nm and its power supply from 6 to 15 V. Small signal bandwidth is 1.6 kHz - 105 MHz at 3 dB and input current noise density is $9 \text{ pA}/\sqrt{\text{Hz}}$ at 10 MHz [10]. A special single-mode optical fiber with a 900 vm buffer, which is thinner than conventional fibers, was used to construct the interferometer. The cladding of the fiber has diameter of $900 \mu \text{m}$. These fibers were used as measuring and reference fibers. In this case, the measuring fiber is passed over the upper part of the model, and the reference fiber is placed in the lower part of the model so that these fibers do not interact with each other.

2.3 DATA ACQUISITION UNIT

Analog Discovery 2 (AD2) was used to evaluate data from a Mach-Zehnder interferometer. The selected connection corresponds to the basic diagram of this interferometer, which is shown in Figure 1. LabVIEW for GUI programming was used, where a library supplied by the software manufacturer is provided.

AD2 is a multifunctional USB device combining a two-channel oscilloscope, a functional analog waveform generator, a 16-channel logic analyzer, a pattern generator, including a programmable power supply. This USB-powered device allows you to create and test analog and digital circuits in any environment [11].



Figure 2: Connection of Mach-Zehnder interferometer.

3 MODEL FOR DEMONSTRATION

A landscape model for optical fiber guidance was made, as can be seen in Figure 3. The model consists of a wooden vessel into which an interferometer has been implemented. The interferometer is surrounded by a damping material so that it is not affected by any external influence and is stored in the model. The upper part of the model presents a landscape where one half is a meadow/field and the other is sand, representing the soil. The sensing arm of the interferometer is guided under this sandbox. The two halves are shielded from each other by different materials, under the "meadow" there is damping cotton wool, and therefore this part serves as a resting zone. When the model of digger travels through the rest zone, the sensor system does not react, and when the digger travels from the rest zone to the other half of the sand, the sensor system reports that there is increased movement in the area, it detects this movement due to increased vibrations. The final feature is that when initiating digger work in the fiber area, the sensor system evaluates it as a dangerous movement and sends a warning to the graphical user interface (GUI) in Figure 4.



Figure 3: Model of the landscape for demonstration of sensing vibrations.

4 GUI FOR THE MODEL

A program in LabVIEW was created for this model, which is used for processing measured data, including analysis of data. This program ensures that the entire sensor system works in real-time. Within this program, the following parts were solved. Communication and settings of Analog Discovery 2. Sensing and control of the signal in real-time. Detection of the signal threshold level and subsequent analysis. Acquiring and saving measured data to a file and converting the signal to audio signal.

The front panel is divided into several parts. Dominant is the display of measured data, where it is possible to see how vibrations cause signal deviation. The threshold for motion detection is used to detect the digger approaching the sand, and the Threshold for vibration detection is used to detect possible fiber disturbance. These thresholds can be set in case the digger model would replace another machine or a digger with a different weight. To the left of the graph are the scale and color settings of the graph. There are also a STOP button to end the program and a button to save the signal to a file and an audio file. At the bottom of the GUI is a top view of the model, where the colors of the sand and fiber change (in Figure 4 yellow sand, and red fiber) depending on whether a motion is detected in the fiber area. There is also a hazard statement with date and time. Based on the intensity of the vibrations, the Optical fiber in danger window also pops up, which can be canceled using the Continue or Call operator button if there is an immediate threat to the fiber. The front panel also contains FEEC and Optolab logos, because the model is primarily intended for the presentation of research, as can be seen in Figure 4.



Figure 4: Front panel of graphical user interface.

5 CONCLUSION

Interferometric distributed fiber optic sensing systems have been described in the paper. A Mach-Zender interferometer was chosen to demonstrate the use of the interferometer as a vibration detector. The next part of the paper describes the individual components of the sensor system and the board for data acquisition and preprocessing. The landscape model demonstrating the excavation work is described below. The last part describes a graphical user interface for evaluating the exceeding of the threshold value and thus triggering an alarm. The GUI is also able to export data or possibly convert data to a way audio file. The model was tested and will be used to demonstrate the use of optical
distributed sensors at promotional events organized by BUT. This model was included in the video for Open Days BUT FEEC. This model has reliably demonstrated the possibility of securing optical fibers using an optical fiber interferometer and is the basis for the application of interferometers in practice as security devices to increase the security of optical fibers or use to secure objects and perimeter.

ACKNOWLEDGEMENT

Research described in this paper was financed by the grant of the Ministry of the Interior of the Czech Republic, Program of Security Research, VI20192022146 (Distributed fiber optic sensing system for use in perimeter and line structures protection).

REFERENCES

- [1] X. Pan, Y. Dong, J. Zheng, J. Wen, F. Pang, Z. Chen, Y. Shang, and T. Wang, "Enhanced FBG Temperature Sensitivity in PbS-Doped Silica Optical Fiber" in *Journal of Lightwave Technol*ogy, vol. 37, no. 18, pp. 4902-4907, Sep. 2019.
- [2] Y. Liu, Y. Wang, D. Yang, J. Wu, T. Zhang, D. Yu, J. Zhenan, and H. Fu, "Hollow-Core Fiber-Based All-Fiber FPI Sensor for Simultaneous Measurement of Air Pressure and Temperature" in *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11236-11241, Dec. 2019.
- [3] W. L. Richards, A. R. Parker, W. L. Ko, A. Piazza, and P. Chan, *Application of Fiber Optic Instrumentation*. Springfield, Virginia: National Technical Information Service, 2012.
- [4] M. Lenner, A. Frank, L. Yang, T. M. Roininen, and K. Bohnert, "Long-Term Reliability of Fiber-Optic Current Sensors" in *IEEE Sensors Journal*, vol. 20, no. 2, pp. 823-832, Jan. 2020.
- [5] D. A. Krohn, T. W. MacDougall, and A. Mendez, "Fiber Optic Sensors: Fundamentals and Applications", 4th edition. Bellingham, Washington: SPIE Press, 2015.
- [6] L. Wang and N. Fang, "Applications of Fiber-Optic Interferometry Technology in Sensor Fields", in *Optical Interferometry*, Feb. 2017.
- [7] J. Guo and Y. Chen, "An Optical Fiber Warning System Based on Vibration Sensing" in 2018 14th International Conference on Computational Intelligence and Security (CIS), 2018, pp. 93-95.
- [8] N. K. Chen and K. F. Hsiao, "Micro fiber Mach-Zenhder interferometer vibration sensor" in 2017 16th International Conference on Optical Communications and Networks (IC-OCN), 2017, pp. 1-2.
- [9] Koheron "1550 nm low-noise DFB laser"[Online]. Available: https://assets.koheron.com/ datasheets/koheron_ld100-laser.pdf
- [10] Koheron "Low noise photodetector"[Online]. Available: https://assets.koheron.com/datasheets/ koheron_pd100-photodetection.pdf
- [11] Digilent "Analog Discovery 2 Reference Manual"[Online]. 2015. Available: https://reference. digilentinc.com/_media/reference/instrumentation/analog-discovery-2/ad2_rm.pdf.

REVIEW COMPARISON AND ANALYSIS OF POX AND RYU CONTROLLER SDN

Shujairi Murtadha

Doctoral Degree Program (1), FEEC BUT xshuja00@vutbr.cz

Supervised by Vladislav Škorpil

E-mail: skorpil@feec.vutbr.cz

Abstract: SDN it is a technology that changed the concept of Conventional Networks by making all control functions in one central place and centralized decision-making. The Control Plane is an essential and important part of the SDN architecture, as well as working on Controllers, knowing their types, and comparing them in terms of performance and efficiency. this paper deals with the performance comparison of two Controllers POX and RYU.

Keywords:: SDN, POX, RYU

1 INTRODUCTION

Software-Defined Network (SDN) is a new networking paradigm whose architecture shifted from its old distribution model to a centralized design. The latter structure is characterized by the separation of control planes and data. The Control Plane comprises the Controller, while the data encompasses forwarding elements, Routers, and Switches [22]. SDNs have recorded a massive deployment and growth in varying forms of networks. Their utilization is in cellular and wide area networks, internet of things and wireless networks, and datacenter networks [26]. Compared to the old system model, SDN has a particular uniqueness entailing multilayered structure for safeguarding efficient traffic flow and forwarding within its working [18].

The multilayered structure is replacing the traditional complex and inflexible networks with innovative Controllers, fixing different limitations, such as mobility, scalability, security, debugging, and manual configuration. Some of the SDN Controllers include Floodlight, POX, NOX, Beacon, Open Daylight, ONOS, and RYU, and they act as brains of SDN management [11]. This discussion critically analyzes and compares the RYU and POX controllers in which the work distinguishes the two controllers' functionality to highlight their differences and similarities.

2 POX ANALYSIS

POX is an SDN Controller that originated from NOX, an indigenous OpenFlow Controller. Its repository has various branches [14]. According to [4], POX is Apache-licensed and written in Python. Individuals can utilize the controller to understand SDN concepts, for it is easy. The Controller is used for faster prototyping and development of modern network applications. Besides, the design often comes pre-installed by the Mininet tactic machine. A user can use the POX Controller to turn dumb OpenFlow gadgets into a firewall, load balancer, switch, or hub device [8]. The regulator permits easy means of running SDN or OpenFlow experiments. One of its unique attributes is that the system can be passed within varying parameters according to experimental or real topologies, allowing its user to experiment with a Mininet emulator, testbeds, or actual hardware [8]. The advantage of using the POX Controller is its requirement for minimal memory space for its performance compared to other archetypes. The structure has considerable throughput outcomes than other SDN Controllers. POX provides productivity, effective documentation, language support, open-source, and

platform [19]. This model supports visualization tools, graphic user interfaces, and simple architecture [23]. Command-line interface Syntax, together with varying components and options, is required during POX operations [7]. The POX Controller can provide better results than NOX If properly operated.

3 RYU ANALYSIS

The RYU Development Team (2020) considers the RYU Controller a component-based application elaborated on the networking framework. RYU offers freeware aspects that are adequately defined, simplifying designers' ability to develop modern control applications and network management (RYU Development Team, 2020. The RYU Controller has a simple structure due to its Mininet command (RYU project team, n.d.). The design has a unique way of operation. It functions under Apache 2 licenses and an open source, written based on Python, and deployed and supported by NTT cloud information centers. RYU's primary source code is located on GitHub, and the Open RYU society is its supporter and provider [6]. One might also be interested in its composition. RYU applications comprise a variety of components crucial for SDN applications. Under the RYU Controller, an individual can adjust existing apparatuses and develop new elements. The reason is RYU Controller entails a software environment to facilitate building a network distribution within a Mininet system, denoting its advantages when executing operations requiring SDN architecture [9]. According to [16], one of RYU's strengths is its supports for Nicira extensions and numerous southbound procedures for controlling devices like Configuration Protocol, OpenFlow Management, and NETCONF. Still, the controllers also comprise Router Northbound Applications and a Simple Switch [17]. The Northbound software regulates networks via centralized controllers, permitting other systems to be more flexible and dynamic while adjusting to the structure's conditions.

4 COMPARISON

Pox, a modern python-based NOX version, has great-level SDN APIC supporting virtualization and includes a query-able topology graph. Unlike RYU, POX is not entirely written using Python. Some of the RYU components include messaging, reusable libraries, application management, event management, and OpenFlow support (SDN Tutorial, n.d.). These components have only been proved to be more existent within the RYU control and not the POX control. Regrading POX, its features are different Python applications that can be invoked if the controller starts from the command line [10]. Additionally, POX components implement software functionality. The Python-based controller is vastly utilized within the research community because of how easily it can be programmed. The structure provides a platform for rapid development and prototyping software that manages network gadgets within an OpenFlow network ("Mininet and Pox"). The Controller can be remotely connected with various applications like intrusion detection, load balancer, firewall, and routing, and Mininet [3]. The application can be installed directly from its repository, GitHub, and it is built within Mininet. Contrary, the RYU Controller was developed from a Japanese word that meant flow. This regulator, like POX, is Python-based and an OpenFlow with an efficient API via which developers can manage, control, and program applications. Still, OF-conf, OpenFlow, and Netconf, which are RYU protocols, can be utilized to configure network devices [25]. The RYU has its code accessible within its Apache 2.0 license besides well-organized documentation for facilitating diverse SDN contexts [15]. The provided attributes denote exceptional strengths and limitations of SDN scenarios in their different functionalities.

In most cases, RYU is utilized to gather statistical data from switches. Hence, it can be configured as a switch, firewall router, and traffic monitor. Both controllers have the forwarding, control, and application layers. [1] state that the three layers exchange information through the southbound and northbound application programs. Southbound is often between the forwarding and applications layer. At the same time, the northbound is between the control and applications layer. Both regulator's data plan, architecture, control logic, and control plan are implemented using a centralized approach managed by an intelligent design [5]. POX currently exchanges data with Open Flow v1 Switches,

with outstanding support from Open vSwitch. Conversely, RYU uses its components that include messaging supporting architecture developed using other languages [6]. Another exceptionality is that controllers help SDN separate data plane from Control Plane [2]. These attributes indicate the structure facilitating control logic within a central Controller. Regardless of the difference between the two Controllers, they play a crucial role in SDN operations. Figure1 shows the properties of both controllers.

Features	ΡΟΧ	RYU	
GUI	Yes	Yes	
OpenFlow Version	V1.0	V1.0, v1.2, v1.3 v1.4, v1.5, Niciria extensions	
Language Support	Python	Python	
REST API	Yes	Yes	
Platform Support	Linux, Windows, Mac	Linux	
License Provider	Apache	Apache	
Learning Curve	Easy	Moderate	
Distributed	No	Yes	

Table 1: Properties of POX and RYU

5 CONCLUSION

Network operators locate and operate centralized consoles, so their performance is important to network management. In this paper, we study, analyze, and compare two controls (POX, RYU) and their characteristics.

In our future work, we aim to create, complex simulations, and test the efficiency of multiple controllers, using tools that create the load on the controller also using a DDoS attack to find out the efficiency of each of the controllers.

REFERENCES

- Abdullah, M. Z., Al-awad, N. A., & Hussein, F. W. (2018). Performance comparison and evaluation of different software defined network controllers. International Journal of Computing & Network Technology, 06(02), 36-41. doi:10.12785/ijcnt/060201
- [2] Ali, J., Lee, G., Roh, B., Ryu, D. K., & Park, G. (2020). Software-defined networking approaches for link failure recovery: A survey. Sustainability, 12(10). doi:10.3390/su12104255
- [3] Ali, J., Lee, S., & Roh, B. (2018). Performance analysis of POX and Ryu with different SDN Topologies. Proceedings of the 2018 International Conference on Information Science and System - ICISS '18. Retrieved from https://booksc.xyz/book/72816877/06d6d0
- [4] Altangerel, G., Chuluuntsetseg, T., & Yamkhin, D. (2019). Performance analysis of IPv6 transition technologies and transition methods. The 13th International Forum on Strategic Technology (IFOST 2018). Retrieved from https://www.researchgate.net/publication/332466286_Performance_analysis_of_SDN_controllers_POX_Floodlight_and_Opendaylight
- [5] Asadollahi, S., Goswami, B., & Gonsai, A. M. (2017). Software-defined network, controller comparison. International Journal of Innovative Research in Computer and Communication Engineering, 5(2), 211-217. Retrieved from https://www.researchgate.net/publication/320346909_Software_Defined_Network_Controller_Comparison

- [6] Asadollahi, S., Goswami, B., & Sameer, M. (2019). Ryu controller's scalability experiment on software defined networks. 2018 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC). doi:10.1109/ICCTAC.2018.8370397 Retrieved from
- Bholebawa, I. Z., & Dalal, U. D. (2017). Performance analysis of SDN/OpenFlow controllers: POX versus floodlight. Wireless Personal Communications, 98(2). doi:10.1007/s11277-017-4939-z
- [8] Chauhan, N., & Sood, m. (2019). Performance analysis of POX, open v switch, and opendaylight SDN controllers on the cloud. International Journal of Innovative Technology and Exploring Engineering, 8(8S3), 332-339. Retrieved from https://www.ijitee.org/wp-content/uploads/papers/v8i8s3/H10900688S319.pdf
- [9] Eissa, H. A., Bozed, K. A., & Younis, H. (2019). Software Defined Networking. In 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA) (pp. 620-625). IEEE. doi:10.1109/STA.2019.8717234
- [10] Elmedani, M. K. (2017). Performance evaluation of software-defined networking compares to traditional networks (Master's degree. Sudan University of Science and Technology). Retrieved from http://repository.sustech.edu/bitstream/handle/123456789/18764/Performance%20Evaluation%20of%20Software....pdf?sequence=1&isAllowed=y
- [11] Islam, T., Islam, N., & Refat, A. (2020). Node to node performance evaluation through the RYU SDN controller. Wireless Personal Communications. Retrieved from https://www.researchgate.net/publication/338703576_Node_to_Node_Performance_Evaluation_through_RYU_SDN_Controller
- [12] Kaur, S., Singh, J., & Ghumman, N. S. (n.d.). Network programmability using the POX controller. Department of Computer Science and Engineering, 134-138. Retrieved from http://www.baburd.com.np/material/NG/Paper2-SDN-POX-Controller.pdf
- [13] Mininet and pox. (n.d.). Retrieved from https://pld.cs.luc.edu/courses/netmgmt/sum17/notes/mininet_and_pox.html
- [14] Muragaa, W. H., Seman, K., & Marhusin, M. F. (2017). A POX controller module to prepare a list of flow header information extracted from SDN traffic. International Journal of Computer and Systems Engineering, 11(12). Retrieved from https://zenodo.org/record/1132991/files/10008318.pdf
- [15] Nazir, F., Humayun, Q., Ahmad, R. B., & Elias, S. J. (2017). Software-Defined Network Testbed Using ZodiacFX a Hardware Switch for OpenFlow. EAI Endorsed Transactions on Scalable Information Systems, 4(14), 1-6. doi:10.4108/eai.25-9-2017.153150
- [16] O'Briain, D. (2019). RYU SDN. Testbed manual version 1.6. Institute of Technology Carlow. Retrieved from http://www.obriain.com/training/sdn/RYU_Soft_Testbed_v1.6_odt.pdf
- [17] Pham, D., & But, J. (2016). The Ryu action node v1.01. CAIA Technical report 161006A. Retrieved from http://caia.swin.edu.au/reports/161006A/CAIA-TR-161006A.pdf
- [18] Quincozes, S. E., Soares, A. A., Oliveira, W., Cordeiro, E., Lima, R. A., Muchaluat-Saade, D., & Albuquerque, C. (2019). Survey and comparison of SDN controllers for teleportation and control power systems. Retrieved from http://dl.ifip.org/db/conf/lanoms/lanoms2019/196234_1.pdf
- [19] Rowshanrad, S., Abdi, V., & Keshtgari, M. (2016). Performance evaluation of SDN controllers: Floodlight and OpenDaylight. IIUM Engineering Journal, 17(2), 47-57. doi:10.31436/iiumej.v17i2.615
- [20] RYU Development Team. (2020). RYU Documentation. Release 4.34. Retrieved from https://readthedocs.org/projects/ryu/downloads/pdf/latest/
- [21] RYU project team. (n.d.). RYU SDN framework. Using OpenFlow 1.3. Retrieved from https://osrg.github.io/ryu-book/en/Ryubook.pdf
- [22] Salman, O., Elhajj, I. H., Kayssi, A., & Chehab, A. (2016). SDN controllers: A comparative study. 2016 18th Mediterranean Electrotechnical Conference (MELECON). Retrieved from https://www.researchgate.net/publication/304457462_SDN_controllers_A_comparative_study

- [23] Saritakumar, N., Srinivasan, A. V., Albert, E., & Rani, S. S. (2019). Performance evaluation of the pox controller for software-defined networks. International Journal of Innovative Technology and Exploring Engineering, 8(9S2), 679-684. Retrieved from https://www.ijitee.org/wp-content/uploads/papers/v8i9S2/I11390789S219.pdf
- [24] SDN Tutorial. (n.d.). SDN controllers. Retrieved from https://cs.wmich.edu/~alfuqaha/spring16/cs6560/modern/lec4-SDN.pdf
- [25] Tseng, Y. (2018). Securing network applications in software-defined networks (Diss., Paris Descartes University).
- [26] Zhu, L., Karim, M., Sharif, K., Li, F., Du, X., & Guizani, M. (2019). SDN controllers: Benchmarking & performance evaluation. Retrieved from https://arxiv.org/pdf/1902.04491.pdf

Doktorské projekty

Komunikační technologie a informační bezpečnost II.

POSSIBLE REPLACEMENT OF BULK REMOTE SENSING AT THE POINT OF DELIVERY

Lukáš Benešl

Doctoral Degree Programme (1.), FEEC BUT E-mail: xbenes44@stud.vutbr.cz

> Supervised by: Petr Mlýnek E-mail: mlynek@vutbr.cz

Abstract: This article deals with a possible replacement of HDO (Bulk Remote Reading), from the distributor's point of view and the customer's point of view. A measurement, which simulates communication was performed from the smart meter to the end device (photo-voltaic panels, charging station for an electric car). The simulated measurement was divided into five scenarios, which are discussed at the end.

Keywords: Powerline communication, HDO replacement, single carrier PLC, Directive 2012/27/EU, V2G, charging station, photo-voltaic panels, SmartBox, smart meter

1 INTRODUCTION

Today, there is an increasing awareness of Directive 2012/27/EU on energy efficiency, which obliges the Member States of the European Union to allow remote smart meter reading at the beginning of 2027 for all meters. This legislation aims to increase energy efficiency from 20% to 32.5% by the end of 2030 [1]. Therefore, distribution companies operating in the Czech Republic must be prepared for the replacement of a large number of electricity meters, which cannot read remotely yet. However, it is not just that Smart Meters should be able to read values remotely only. In this regard, the Smart Grid Task Force set up by the European Commission has identified 10 specific functionalities in particular that can be enabled by Smart Meters:

- Provide the readings directly to the consumers or to a third party.
- Update readings frequently enough to use energy savings schemes.
- Allow remote reading by the operator.
- Provide 2-way communication for maintenance and control.
- Allow frequent enough readings to be used for network planning.
- Support advanced tariff schemes.
- Allow remote ON/OFF control of power supply and/or flow or power limitation.
- Provide secure data communication.
- Allow fraud detection and prevention.
- Provide import/export and reactive metering.

The decision to install intelligent meters across the Member States follows the long-term cost-benefit analysis presented in the third energy package [2].

1.1 PRODUCTION AND CONSUMPTION MANAGEMENT

Further development of HDO technology is not possible. Also, HDO cannot control other new devices, for example chargers for electric cars or photo-voltaic panels. The distributor of electricity must be able to control the production or consumption of electricity behind a smart meter. The supply points will be changed to so-called hybrid points soon. It means that these points will supply energy to the distribution network sometimes and at another time the energy will be consumed. The distributor must be able to control the energy flow in all directions. The customer has photo-voltaic panels at home. These panels produce electricity if it is not currently used, it starts to accumulate in the battery storage. The most appropriate solution to be sold to the distributor/aggregator is when the batteries are fully charged. Also electric cars could supply V2G (Vehicle-to-Grid) energy if necessary. Alternatively, the vehicle will be charged only by the optimal speed from photo-voltaic panels or home battery storage [3].

In the future decade, the distributor should be able to manage the charging station of the electric cars. In case of an unstable network, the proceedings should allow rechargeable power adjustment in five steps at least to prevent sudden large consumption. Also, the distributor will need to limit or regulate energy supplies from photo-voltaic [4] in the energy overflow [5].

Control of these hybrid points would be allowed only if the relay output counts increase in the smart meter or use direct control from the smart meter. In another case, an additional device can be used. It is frequently called a Smart Box or Control Box.

1.2 WIRELESS OR CABLE COMMUNICATION FOR SMART METERING

The communication of the smart meter towards the customer is possible either as wireless or wired. Each of the options has its specifications, which are advantageous but they also have limitations.

Wireless communication has an advantage that there is no need to build a physical communication medium and it is possible to move the end device if is it necessary. The disadvantage is that there may not always be a good signal among the communicating elements. Encryption must be used, so it can be more demanding on computing power. Wireless technology can be easily disrupted.

Wired communication is more reliable, so it can achieve high transmission speeds. Also communication cannot be disrupted. The end device can be easily powered by a transmission medium. The disadvantage is the need to build a communication medium and the end device must be firmly in one place [6].

If PLC (Powerline Communication) technology will be used, it is no longer necessary to build new transmission medium and power distribution can be used for communication. The maximum distance for powerline communication can be up to 1500 meters [7].

2 METHODOLOGY

PLC was selected for the household measurement with a single carrier and narrowband with multi-carrier.

Narrowband PLC - Technology operates in the 3–500 kHz frequency band, which includes the Chinese band 3–500 kHz and the Japanese ARIB band 10–450 kHz, the European CENELEC band 3–148.5 kHz, the US FCC band 9–500 kHz [8, 9]. According to the data bit rate, this technology can be further divided into:

- Low Data Rate (LDR): These are technologies with a single carrier and a data rate of several kbps. Typical examples of LDR NB-PLC are LonWorks standards, IEC 61334, X10, Home-Plug C&C, and SITRED.
- High Data Rate (HDR): These are multi-carrier technologies with data rates from tens of kbps to 500 kbps. Typical examples are technologies based on ITU-T standards by G.hn, IEEE P1901.2, PRIME and G3-PLC.

In our case, the ModemTec MT49R modem was used for a single carrier, which can communicate in the range from 60 kHz to 145 kHz. Modem uses D-BPSK (Differential Binary Phase-Shift-Keying) modulation. The baud rate on the PHY (physical layer) is 10 kbps, which never changes and the application layer is 5.33 kbps, which also is never changed. The modem only connects to one phase.

2.1 TOPOLOGY OF MEASUREMENT

Measurement topology was compiled about household AC distribution. The default measuring point was in the case of the first measurement. Then the second modem was moved to other sockets that are marked in Figure 1.



Figure 1: Measured topology of the house, including the garage.

The measurement contains five scenarios that define the different distances of 0, 10, 30, 60, and 100 meters. For the first to the fourth scenario, the cable is under the plaster, specifically the cable CYKY-J $3x2.5 mm^2$. The fifth scenario with the longest cabling leads through two electricity meters that have a joint to the same transmission tower. The air cable is with the aluminum core AES $4x16 mm^2$, the next part of the cable is in the ground, the type of cable is unknown.

The measurement was always at the same time for five hours. The start was always at 3 PM and the end was at 8 PM. During this time, all home electronics (microwave, dishwasher, television) could be switched on and off. Values of SNR (Signal-to-Noise ratio) were recorded every two seconds.

3 RESULTS

Measurement with one carrier achieved very high values. SNR of about 50 dB was achieved for the first three scenarios. For the fourth scenario, values dropped nearly to half than the previous scenarios, the values were slightly over 20 dB. The last scenario is based on communication through two electricity meters and a distance of about 100 meters, there were outages during communication. Table 1 summarizes all measured values.

3.1 FIRST AND SECOND SCENARIO

The first scenario is based on communication in the same socket. This value is therefore the reference value. In a real topology, this is the highest possible SNR value, which should have the lowest noise.

SNR [dB]	1. scenario	2. scenario	3. scenario	4. scenario	5. scenario
Local-Remote	54.4	51.8	48.3	23.8	17.3
Remote-Local	56.0	49.8	47.7	21.7	15.8

Table 1: SNR for single carrier PLC.

The distance between the modems in the second scenario is 10 meters. The decrease in values is not significant here.

3.2 THIRD SCENARIO

This scenario show communication up to a distance of 30 meters. From the measurements, it can be concluded that the socket circuits are connected probably on the same phase. The decrease of SNR value between the first and third scenarios is only 7 dB. A television and a router were connected between this route. These devices may slightly increase the noise spectrum background.

3.3 FOURTH SCENARIO

The fourth scenario declares a measured distance of 60 meters. There is a drastic reduction in the SNR value by more than half. It is, therefore, possible that the fourth scenario communicates at different phases.

3.4 FIFTH SCENARIO

The fifth scenario was tasked with the functionality of communication up to the garage, where could be an electric car with a charger (wallbox) potentially. By the fact that the garage is connected through the second house supply point. On this route, there is an air cable and two electricity meters. The socket in the garage is probably not on the same phase. During communication, there were outages, but modems could partly communicate together.

4 DISCUSSION

It is therefore obvious that modems have great resistance to interference and they can communicate through two electricity meters. The single-phase PLC modem allows communication over one kilometer.



Figure 2: Chart of Single Carrier PLC, SNR dependence on the distance.

Figure 2 shows the dependence of the SNR value on the communication distance. Likely the communication will not need to be routed through the smart meter to another consumption point. However, this measurement is proof that the technology can adapt to relatively diverse cabling with different deposition methods.

5 CONCLUSION

This measurement verified that it will be possible to use PLC technology for direct control of consumption and production behind the supply point. The fifth scenario simulated the distance of a possible charging station for an electric car or a photo-voltaic panels. Communication was possible even through two electricity meters. This specific case will not be very common in the distribution network, so there is no need to worry about communication between the smart meters and the end device.

In further development, it is necessary to define the requirements for the customer itself to ensure sufficient connectivity and quality communication. It is necessary to determine if it is more appropriate to implement control into a smart meter or replace the HDO system with smarter equipment. It will be necessary to allow the device to collect data for analytical methods and prediction for dynamic tariffs and the possibility of trading energy surplus.

REFERENCES

- Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC. In: 25.10.2012. Available at: https://bit.ly/3uWy5ES
- [2] PRETTICO, G., et al. Distribution System Operators Observatory 2018. Publications Office of the European Union, 2019.
- [3] Van Mierlo, J.; Berecibar, M.; El Baghdadi, M.; De Cauwer, C.; Messagie, M.; Coosemans, T.; Jacobs, V.A.; Hegazy, O. Beyond the State of the Art of Electric Vehicles: A Fact-Based Paper of the Current and Prospective Electric Vehicle Technologies. World Electr. Veh. J. 2021, 12, 20. https://doi.org/10.3390/wevj12010020.
- [4] H. Hatta, M. Asari and H. Kobayashi. Study of energy management for decreasing reverse power flow from photovoltaic power systems. 2009 IEEE PES/IAS Conference on Sustainable Alternative Energy (SAE), Valencia, Spain, 2009, pp. 1-5, doi: 10.1109/SAE.2009.5534838.
- [5] Iioka, Daisuke; Fujii, Takahiro; Tanaka, Toshio; Harimoto, Tsuyoshi; Motoyama, Junpei. 2020. Voltage Reduction in Medium Voltage Distribution Systems Using Constant Power Factor Control of PV PCS. Energies 13, no. 20: 5430. https://doi.org/10.3390/en13205430
- [6] KABALCI, Yasin. A survey on smart metering and smart grid communication. Renewable and Sustainable Energy Reviews. 2016, 57, 302-318. ISSN 1364-0321. Available at: https://doi.org/10.1016/j.rser.2015.12.114
- [7] T. Sangsuwan, S. Thepphaeng and C. Pirak, Experimental performance analysis of powerline communication technologies in AMI systems, The 20th Asia-Pacific Conference on Communication (APCC2014), Pattaya, Thailand, 2014, pp. 382-386, doi: 10.1109/APCC.2014.7092841.
- [8] M. A. O. Kharraz, C. Lavenu, P. Jensen, D. Picard and M. Serhir. Characterization of the input impedance of household appliances in the FCC frequency band. 2017 IEEE International Symposium on Power Line Communications and its Applications (ISPLC), Madrid, Spain, 2017, pp. 1-6, doi: 10.1109/ISPLC.2017.7897103.
- [9] D. Cooper and T. Jeans. Narrowband, low data rate communications on the low-voltage mains in the CENELEC frequencies. II. Multiplicative signal fading and efficient modulation schemes. In IEEE Transactions on Power Delivery, vol. 17, no. 3, pp. 724-729, July 2002, doi: 10.1109/TP-WRD.2002.1022795.

SMART ENERGY MONITORING SYSTEM BASED ON POWER LINE COMMUNICATION TECHNOLOGY

Martin Rusz

Doctoral Degree Programme (1), FEEC BUT E-mail: xruszm00@vutbr.cz

> Supervised by: Petr Mlynek E-mail: mlynek@feec.vutbr.cz

Abstract: The aim of this paper is to describe the design of a smart monitoring system based on Power Line Communication technology. As part of the description of technologies and the entire design, technologies commonly used today are described in more detail. The main features of the whole system are mainly reliability, scalability, security. For the purposes of the design, several measurements of different communication technologies were performed. The results of these measurements are important for evaluating the design of the entire system.

Keywords: Monitoring System, Gateway, Power Line Communication, Smart Socket, ZigBee

1 INTRODUCTION

With the growing number of electrical appliances as well as the increasing electricity consumption of households, it is necessary to apply more intelligent systems for monitoring and managing energy consumption. Thanks to these systems, it is possible to increase overall efficiency, reduce electricity consumption and also reduce CO_2 emissions. The European Union has also committed itself to these steps in the Energy Efficient Plan 20-20-20 [1].

Nowadays, concepts such as Smart Buildings, Smart Cities or Industry 4.0 are becoming more and more widespread. These concepts are characterized primarily by the ability to control electrical equipment such as household appliances, heating or air conditioning. Another key feature is the signaling of non-standard events, response to non-standard events and, in advanced systems, also the prediction of possible events [2, 3]. In new buildings this system can be included in the design itself and designed to provide all the necessary functionalities. However, for existing buildings, in most cases it is necessary to install intelligent control systems additionally [4].

The intelligent measurement system, which will be described in this paper, should enable the implementation of all necessary protocols occurring in the Smart Buildings environment and also ensure a sufficient level of security. All the requirements described above are met by the considered Power Line Communication (PLC) technology. Advantage of this technology is the use of the existing electrical network as a transmission medium, which will simplifies the application of the entire system. Throughput, efficiency, packet loss, interference immunity and communication range are the key parameters that need to be determined in the deployment of this technology [5].

2 COMPARISON OF COMMUNICATION TECHNOLOGIES

This section briefly describes the tested Zigbee and PLC technologies. The basic parameters of individual communication technologies will be presented in this part on which the currently available solutions of smart sockets, or entire intelligent control systems, are based. For comparison, the results of measurements will also be presented here, thanks to which it will be possible to evaluate the advantages and disadvantages of individual communication technologies.

2.1 ZIGBEE

ZigBee is a wireless communication technology provided by the IEEE 802.15.4 standard. It uses the license bands 868 MHz, 902–928 MHz and 2.4 GHz for communication. Depending on the frequency band used, transmission speeds of 20, 40, 250 kbps are then defined. Its main advantages include reliability, simplicity, easy implementation, low energy consumption and affordability. This technology is suitable for building large-scale wireless networks thanks to the use of Ad-Hoc routing, which allows communication between the controlling node network and the terminal device without a direct connection between them. Modem network topologies can be star, tree or mesh. For secure communication is used 128 bit AES encryption [6].

For testing in laboratory conditions was used a simple topology, which can be seen in the figure 1. The topology consists of a configuration PC with Radwin software. The control modem was configured via a USB to RS232 converter. For measurement were selected CDX800 modems using the 869 MHz transmission band with a maximum speed of 24 kbps.



Figure 1: Topology for laboratory ZigBee measurement.

2.2 POWER LINE COMMUNICATION

PLC technology uses an electrical distribution network as a transmission communication channel. Communication can take place at the level of low voltage 230 V, 400 V (LV), high voltage 22 kV (HV) or very high voltage 110–220 kV (VHV). The principle of the technology is based on the transmission of a useful signal, which is modulated on the carrier frequency of the electrical network. For this technology, the useful signal has a much higher frequency and lower amplitude than the mains signal. Depending on the bandwidth and use, the technology is divided into two basic types [7]:

- Narrowband PLC (NB-PLC): bandwidth 3–148.5 kHz according to the valid European standard CENELEC, data rate from 10 kbps to 500 kbps, communication range up to several kilometers
- Broadband PLC (BB-PLC): bandwidth 1.8–250 MHz, data rate from several Mbps up to Gbps, communication range up to 300 meters

For the design of communication devices of the smart energy monitoring system was chosen the broadband PLC standard HomePlug AV (HP-AV) on which the used integrated circuit AR7420 is built. The topology, which can be seen in the figure 2, was designed for measurements in the laboratory. The topology consists of several parts:

- Separating transformer serves to suppress unwanted noise occurring on the power line
- BPL modem development kit based on integrated circuit AR7420
- Network analyzer EXFO FTB-1 PRO for measuring transmission parameters
- Noise generator white noise generator PROMAX PROPOWER-1 (1-50 MHz)



Figure 2: Topology for laboratory PLC measurement.

2.3 COMPARISON AND DISCUSSION

The average measured values of both devices were selected from the performed measurements. Because these are two different technologies, it is difficult to compare them. However, it can be seen from the table 1 that the CDX800 ZigBee modem achieved very low data rates and the total message transmission time was much longer compared to PLC technology. Zigbee systems are now very commercially widespread and focused primarily on signaling, control and measurement in the Smart-Buildings concept environment. Broadband PLCs are a possible alternative to these wireless systems, mainly due to the sufficient range of communication. transmission speeds and delays. Thanks to the sufficient data transmission capacity, they can also be used, for example, for camera systems, connection of terminal equipment to the Internet, building small local networks and other similar application. There is currently no similar commercial solution based on broadband PLC technology.

Device:	Distance [m]:	Data rate:	Round Trip Time [ms]:	Efficiency [%]:
CDX800	75	19.8 kbps	197.67	98.5
AR7420	100	58.2 Mbps	2.289	100

Table 1: Comparison of ZigBee and Broadband PLC devices.

3 DESIGN OF A MONITORING SYSTEM

This section describes each part of smart energy monitoring system. Specifically, the key features of individual devices and their purpose in the design. Figure 3 shows simplified block diagram of the whole system. The whole system consists of several blocks, which represent hierarchically divided elements of the system.

The blue blocks in block diagram represent the smart sockets that make up the lowest layer of the system. Their task is to measure electrical parameters and basic computing operations that do not require powerful computing units. Another important task is the mediation of broadband communication for end devices, such as user PCs. A terminal device that requires broadband communication is connected to the smart socket via an Ethernet interface with an RJ-45 connector. The design of the smart socket is described in more detail in section 4.

The purple part of the diagram shows the communication interface between the smart sockets and the master monitoring center. This part is represented by multi-purpose gates. Up to 16 smart sockets



Figure 3: Block diagram of monitoring system.

can be connected to each gateway via BB-PLC. The main task of the gateway is to forward the measured data via wireless or wired interfaces to the superior monitoring system. The gateway can be configured locally and the required data can also be read from it locally from the connected sockets. Multipurpose gateway is described in more detail in section 5.

The top layer of the block diagram is the monitoring center. Accumulates and further processes received data from individual gateways. The parameters and functions of this section depend on the application. The most important task of this part is the implementation of a database system that will be store the measured data and allow the monitoring system to evaluate and respond to network events, power consumption, fault prediction and other functions based on this data.

4 PROTOTYPE OF SMART SOCKET

As part of the development of the entire system, a functional prototype of a smart socket based on BB-PLC technology was designed. The prototype was designed on the basis of the AR7420 integrated circuit, which has a sufficient transmission speed (up to 100 Mbps at the application layer and a communication range of up to 300 m). A communication module from the WisPLC Pro development kit was used for the development. The prototype will be extended by a measuring part, which is needed to measure electrical parameters (voltage, current) and non-electrical quantities (temperature, humidity). The voltage measurement will be realized by a suitably designed voltage divider, the current measurement will be realized by an integrated circuit, which works on the basis of the Hall effect. A switching relay will be implemented for switching appliances, which will provide control of the connected appliances.

5 MULTIPURPOSE GATEWAY

The main task of the multi-purpose gateway is manage the network of connected smart sockets. From the point of view of the system, it is an intermediary between the individual sockets and the superior monitoring center. The gateway will contain a broadband PLC chipset, which will ensure communica-

tion with the installed sockets. It will be possible to use a several technologies to transfer information to the superior part of the system. The LTE interface will be used for wireless communication, which will be used mainly in applications with a large distance between the individual elements of the system. For local applications, it will be possible to use the Ethernet interface via the RJ-45 connector.

The multi-purpose gateway will also include an RS485 serial interface for connecting external devices, such as an electricity meter, batteries, PV panels, wallbox chargers. Their task is to measure electrical parameters, switching on/off of appliances and other similar tasks. This interface will also be used for local gateway configuration, firmware update or debugging. For the possibility of remote switching of several independent electrical devices, several controlled relays, so called realy box, will be implemented.

6 CONCLUSION

This paper describes the design of an intelligent measuring system based on broadband PLC technology. This solution responds to the current situation of the need to control and regulate production and consumption behind the point of consumption. Solution presented in the article using PLC is ideal for places where a wireless solution is not possible and installation of new cables is complicated. Based on the measurements and prof of the concept of the prototype, high transmission speeds necessary for the deployment of security and advanced functionalities (surveillance, camera systems, etc.) are achieved. The next step in this work is to put the whole system into test operation, to design a methodology for testing this system and optimization of individual parts.

REFERENCES

- [1] Report on revision of the energy efficiency action plan. https://www.europarl.europa. eu/doceo/document/A-7-2010-0331_EN.html?redirect.
- [2] J. Al Dakheel, C. Del Pero, N. Aste, and F. Leonforte. Smart buildings features and key performance indicators: A review. *Sustainable Cities and Society*, 61:102328, 2020, Available from: https://doi.org/10.1016/j.scs.2020.102328.
- [3] F. M. Bhutta. Application of smart energy technologies in building sector future prospects. In 2017 International Conference on Energy Conservation and Efficiency (ICECE), pages 7–10, 2017, Available from: https://doi.org/10.1109/ECE.2017.8248820.
- [4] A. Altayeva, B. Omarov, and Y. I. Cho. Multi-objective optimization for smart building energy and comfort management as a case study of smart city platform. In 2017 IEEE 19th International Conference on High Performance Computing and Communications, pages 627–628, 2017, Available from: https://doi.org/10.1109/HPCC-SmartCity-DSS.2017.86.
- [5] P. Mlynek, M. Rusz, L. Benesl, J. Slacik, and P. Musil. Possibilities of broadband power line communications for smart home and smart building applications. *Sensors (Switzerland)*, 21(1):1– 17, 2021, Available from: https://doi.org/10.3390/s21010240.
- [6] S. F. Shende, R. P. Deshmukh, and P. D. Dorge. Performance improvement in zigbee cluster tree network. In 2017 International Conference on Communication and Signal Processing (ICCSP), pages 0308–0312, 2017, Available from: https://doi.org/10.1109/ICCSP.2017.8286367.
- [7] P. Mlynek, J. Misurec, Z. Kolka, J. Slacik, and R. Fujdiak. Narrowband power line communication for smart metering and street lighting control. *IFAC-PapersOnLine*, 48(4):215–219, 2015.
 13th IFAC and IEEE Conference on Programmable Devices and Embedded Systems, Available from: https://doi.org/10.1016/j.ifacol.2015.07.035.

INDUSTRIAL NETWORK SECURITY MODULE

Karel Kuchař, Eva Holasová

Doctoral Degree Programme (I), FEEC BUT E-mail: {xkucha24, xholas08}@vutbr.cz

Supervised by: Radek Fujdiak E-mail: fujdiak@vutbr.cz

Abstract: This article is focused on a fast and efficient evaluation method of communication of the Modbus/TCP protocol. Modbus/TCP does not implement authentication or communication encryption. Therefore, a Modbus Security module was created, which allows sniffing specific network traffic and parsing particular information from the packets. This information is stored in the database using PostgreSQL on each master and slave station. It evaluates whether there is an attack on the network by comparing information in individual databases. There is an additional authentication of individual stations using the created SSH connection between databases. Everything is visualised using the Grafana tool.

Keywords: Modbus/TCP, Sniffer, Scapy, Python, PostreSQL, Grafana, Security, SSH

1 INTRODUCTION

This article is focusing on a fast and efficient method of providing additional authentication in an industrial network. Other papers are focusing on using artificial intelligence or machine learning to distinguish ongoing attacks or anomalies, e.g., Radoglou-Grammatikis et al. [1] are in their paper focusing on DoS (Denial of Service), MitM (Man in the Middle), and other attacks via machine learning methods placed on the slave station. In [2], Marian Nicolesc et al. introduced a Modbus logger and analyzer capable to work with a database. In [3], Nyasore et al. compare the impact of deep packet inspection. This inspection might be crucial in attack detection. In comparison to the mentioned papers, our proposed method can distinguish incoming traffic from a legitimated node. Moreover, the slave station can detect network cyber-attacks such as DoS, MitM, Injection attacks, and more with minimal impact or requirements on the used devices.

2 INDUSTRIAL PROTOCOL MODBUS/TCP

The Modbus protocol [4] is one of the oldest industrial protocols. For demonstration, we select the most widespread industrial protocol - the Modbus/TCP [5–7], which operates on the application layer of the ISO/OSI model. The communication in this protocol is master-slave and challenge response. The Modbus/TCP protocol uses function codes. Function codes are used for specifying what will be transmitted between parts of the system. Each function code transmits kinds of values, binary values, numbers, or files.

Modbus/TCP was created for systems in very secure environments. This protocol is not in the base version implementing authentication of transmission. Moreover, there is not supported encryption of transmitted information. This means that every communication via this protocol must be inspected. Due to the lack of security and interconnection of IT and OT, industrial networks using Modbus/TCP protocol (not only this industrial protocol) are vulnerable to many cyber attacks. Every part of the Modbus network might be the potential target of cyber attacks.

Based on *Data breach investigation report* [8], one of the most common cyber attacks targeting an industrial network are DoS Attacks (ca. 55%), error caution (ca. 10%) and other types of attacks, such as phishing (ca. 15%) or ransomware (ca. 5%).

3 MODBUS/TCP SECURITY MODULE

To provide additional security to the industrial network, the Modbus Security module was created. This module handles traffic corresponding to Modbus/TCP protocol. The sniffing module using Scapy¹. was created. This module allows sniffing specific network traffic and parses particular information from the packet. The experimental testing was focused on the most common function codes and their parameters that are transferred via Modbus/TCP packets. Table 1 shows the selected parameters specific for individual function codes. Besides, every packet has information about transaction ID, protocol ID, length, and unit ID. These parameters combined with parameters from the Ethernet layer, IP, and TCP layer are used to log traffic in relation to Modbus protocol.

FC	Register type	Type	Specific for Eurotion Code
IC	Register type	Турс	Specific for Function Code
1	Read Coil	request	startAddr, quantity
2	Read Discrete Input	request	startAddr, quantity
3	Read Holding Registers	request	startAddr, quantity
5	Write Single Coil	request	outputAddr, outputValue
6	Write Single Holding Register	request	registerAddr, registerValue
15	Write Multiple Coils	request	startAddr, quantityOutput, byteCount, outputsValue
1	Read Coil	response	byteCount, coilstatus
2	Read Discrete Input	response	byteCount, inputStatus
3	Read Holding Registers	response	byteCount, registerVal
5	Write Single Coil	response	outputAddr, outputValue
6	Write Single Holding Register	response	registerAddr, registerValue
15	Write Multiple Coils	response	startAddr, quantityOutput

Table 1: Overview of selected Modbus/TCP parameters

To perform the experimental testing, an experimental network setup was created. Figure 1 shows an experimental network, where the master station and slave station are virtualized using VMware software. As the operating system, Ubuntu 20.04 was chosen. Within these stations, a Security module was placed. Modbus/TCP communication was performed via Pymodbus library². The Modbus communication is transmitted via Ethernet (MitM is already established), SQL communication is transmitted via a secured channel. Firstly, the Security module sniffs individual parameters from communication from MAC, IP, TCP, and Modbus/TCP layers are stored in the local database (located in Security module 1). Due to many of records, PostgreSQL database³ was used. Timestamp (with an accuracy of ms) and ID are added to the database too. Then, Modbus/TCP communication is transferred via an Ethernet communication medium to the slave side. In the slave station (Security module 2), the sniffing steps are replicated. All accessible information is stored in a local database. However, it is also possible to use a Python script to compare the individual parameters stored in the database. In case of a mismatch in the records, a security alert is generated.

In the case of experimental testing, Security modules are located on virtual machines. However, in a real network, it is possible to place these devices separately. If the attacker tries to read the value on the slave station, where the start address is set to 10, see Listing 1, this attempt is written into the database. The record contains information about ID number, time (black marked), and information

¹Scapy library available on https://scapy.net

²Pymodbus library available on https//pypi.org/project/pymodbus/

³PostgreSQL available on https://www.postgresql.org



Figure 1: Network setup, principle of Security module

taken from Modbus layer (red marked), TCP layer (blue marked), IP layer (brown marked), and MAC layer (orange marked). Based on the captured incoming Modbus/TCP traffic, the *sequence number* is used as the searching element. A Python script running on the slave side will contact the database located in Security module 1 via a secured channel created using SSH (using preshared keys). The sent query finds the last record with the corresponding sequence number. If there is no match, a security alert is generated and all query is shown, see Listing 1. This type of alert is generated every time when the attacker performs attacks such as DoS, injection attack, MAC spoofing. In case of these attacks, Modbus/TCP request/response is not recorded in Security module 1's database.

Listing 1: Generated alert, no match found

If the attacker acts as MitM and tries to remove or modify Modbus/TCP packets, a security alert is generated, see Listing 2. In this case, a removal attack was performed. There are three records where function code 3 was used in Security module 1's database (red marked), but only two records in Security module 2's database (blue marked). When there is a match in the sequence number and there is no violation in the parameters, then the traffic is labeled as legitimate traffic.

Listing 2: Generated alert, different parameters

```
1 *** Security Violation ***
2 Problem detected in: Function Code.
3 Value of master is: [(3,), (3,), (3,)] value of slave is: [(3,), (3,)]
```

The critical part of this approach is the security of the slave station. In the case of control of the master station by an attacker, this method is not effective. The crucial is the place where the Security module is placed. The closer the module is placed to each device, the more effective it provides protection.

The Security module performed in this paper might work as a hidden element in the network (passive sniffing) or as an active part. The active part means that firstly it verifies the transferred packet's content, and then, the packet is forwarded to the slave station. Suppose the Security module is working in "passive" form and a security violation is detected. In that case, it is possible to "repair" this state using the generated new request to the slave station and set values to the previous settings. The delay caused by writing to the database and comparing values is ca. 6 ms. This approach's advantage is that

the time synchronisation of Security module 1 and 2 is not needed to do traffic inspection. And the only relevant variant of the packet is the same as the last record in Security module 1. Any deviation is treated as a security risk.

Based on the performed testing, it was necessary to adjust the record's reading speed from the database located on Security module 1. If there is no delay, false security alerts are generated. After adjusting this delay to 1 s, more than 142,000 Modbus/TCP packets were transferred and only 14 security alerts were generated. And 11 of these packets were of the *response* type, with further delay it is possible to eliminate these cases. Although these 9 packets are counted, false positives make up ca. 0.01 % of all packets. A security alert is generated for each packet with modified parameters or for a packet sent from a device other than Security module 1.

4 GRAFANA

To visualise individual records, open-source analytical and monitoring tool Grafana⁴ was used. This software can cooperate with the PostgreSQL database. Therefore, it allows binding databases from Security modules 1 and 2. It allows to visualise the content of databases, set thresholds, etc. To work with Modbus communication, when logging the Modbus/TCP packet parameters, ID and timestamp are added. To be able to precise visualisation, the time must be used to the nearest millisecond. In the case of the accuracy of seconds, it is impossible to show all records taken from databases. This time is taken from the Security modules' operating system when Modbus/TCP packet is detected.

Due to transferring *registerval* parameter sniffed from Modbus/TCP communication as *text*. The values had to be retyped to visualise these values, see Listing 3. This query removes the square brackets created when writing to the field and casts the value. Figure 2 shows the visualised function codes of detected and recorded Modbus/TCP packets taken from the database located in the Security module 2 in the timeline.



Figure 2: Grafana graph showing function codes in the timeline

⁴Grafana available on https://grafana.com/

5 CONSLUSION

This article was focused on methods providing additional security to industrial networks presented on Modbus/TCP protocol via a Security module cooperating with the sniffing module. Sniffed data containing information taken from MAC, IP, TCP, and Modbus layers is stored in local databases within individual Security modules. Besides, it implements Python script that creates a secured tunnel to Security module 1 (master side). Data is obtained through this tunnel using a database query. This request filters the records according to the sequence number of the packet captured by Security module 1 and returns the last one. Based on the comparison of these two records, the security alert might be generated in the case of a mismatch or missing record. Filtering is performed based on a sequence number due to (relatively) uniqueness and no time synchronisation is required. The proposed method allows fast attack detection without the need to work with higher layers that may cause additional delay. The described method in this article seems to be effective because the level of false positives is negligible. Using this method, the slave station can detect a high amount of potential attacks targeting the slave station. The Grafana tool was used to visualise individual records from databases.

ACKNOWLEDGEMENT

The described research is part of the grant project registered under no. FV40366 and funded by the Ministry of Industry and Trade of the Czech Republic.

REFERENCES

- P. Radoglou-Grammatikis, I. Siniosoglou, T. Liatifis, A. Kourouniadis, K. Rompolos, and P. Sarigiannidis, "Implementation and detection of modbus cyberattacks," in 2020 9th International Conference on Modern Circuits and Systems Technologies (MOCAST), 2020, pp. 1–4.
- [2] D. M. Nicolescu, R. Tătăroiu, D. C. Trancă, and G. C. Pătru, "Logger and analyser for modbusbased industrial networks," in 2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet), 2020, pp. 1–4.
- [3] O. N. Nyasore, P. Zavarsky, B. Swar, R. Naiyeju, and S. Dabra, "Deep packet inspection in industrial automation control system to mitigate attacks exploiting modbus/tcp vulnerabilities," in 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), 2020, pp. 241–245.
- [4] "Modbus application protocol specificationv1.1a," 2004. [Online]. Available: https://www.modbus.org/docs/Modbus_Application_Protocol_V1_1a.pdf
- [5] R. Amoah, S. Camtepe, and E. Foo, "Securing dnp3 broadcast communications in scada systems," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1474–1485, 2016.
- [6] C. Cremers, M. Dehnel-Wild, and K. Milner, "Secure authentication in the grid: A formal analysis of DNP3 SAv5," *Journal of Computer Security*, vol. 27, no. 2, pp. 203–232, Mar. 2019. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JCS-181139
- [7] E. Bou-Harb, N. Ghani, A. Erradi, and K. Shaban, "Passive inference of attacks on cps communication protocols," *Journal of Information Security and Applications*, vol. 43, pp. 110– 122, 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2214212616302113
- [8] verizon.com, "2020 data breach investigations report," pp. 119–120. [Online]. Available: https://enterprise.verizon.com/resources/reports/2020-data-breach-investigations-report.pdf

IMPACT OF ACTIVE SCANNING ON THE INDUSTRIAL CONTROL NETWORKS

Ondřej Pospíšil

Doctoral Degree Programme (2), FEEC BUT E-mail: xpospi89@stud.feec.vutbr.cz

Supervised by: Radek Fujdiak E-mail: fujdiak@feec.vutbr.cz

Abstract: This article deals with the impact of active scanning on industrial networks. The impact on industrial networks is commented from the perspective of the penetration tester methodology. This topic is important because active scan tools are affordable and easy to use, and their intrusive impact on industrial devices can be critical. The article's main goal was to evaluate the impact on the industrial network from the penetration tester point of view using the most popular tools for active network scanning. In order to demonstrate and evaluate the results, an industrial testbed based on real industrial hardware was built for the article. The article also demonstrated how to use the information obtained by scanning for a Denial of Service attack.

Keywords: ICS, scanning, Nmap, Zmap, PLC, DoS, HMI

1 INTRODUCTION

Network scanning allows visualizing the network configuration and communication infrastructure. Network reconnaissance, whether an active or passive scan, is a significant part of cybersecurity. By scanning, it is possible to get an overview of the network and identify possible security holes. It is one of the fundamental pillars of cybersecurity, and it is the essential skill of a penetration tester. From the point of view of the penetration tester, the methodology, according to Certified Ethical Hacker (CEH) [1], contains seven steps:

- 1. Check for live systems: The step of determining which hosts are active on the network.
- 2. Check for open ports: The detected IP addresses are used to search for open ports.
- 3. Scan beyond Intrusion Detection System (IDS): Avoid detection from IDS.
- 4. **Perform banner grabbing:** Gathering information about running services on hosts.
- 5. Scan for vulnerabilities: Use the information from the scan to examine vulnerabilities.
- 6. Draw network diagrams: Reconstruct the network architecture.
- 7. Prepare proxies: Due to anonymity.

However, there is a problem with scanning industrial networks. Unlike conventional IT networks, industrial networks face security risks in these scanning procedures. This is due to the use of communication-sensitive industrial equipment and their configuration, where the emphasis is on a real-time operation. Network scanning can cause a complete failure of a Programmable Logical Controller (PLC) or increase the time delay, leading to unexpected situations in the process. An example is a robotic arm process. In this process, a slight delay means that the whole process can get into an unexpected or even dangerous situation. The solution may be to use security measures such as IDS, firewalls, and anti-malware as in IT networks. However, due to their nature and specific operating systems, this is not easy for industrial devices.

The key aim of this article is the impact of the behavior of active scanners in industrial networks. This behavior is described from the point of view of a penetration tester. The article demonstrates the possibilities in scanning industrial networks and what impact his reconnaissance may have. It also demonstrates the possibility of using information obtained from scanning to perform a Denial of Service (DoS) attack. The following list details the main contributions of the article:

- Impact of active scanning on industrial networks.
- Impact assessment based on behavior according to the penetration tester methodology.
- Demonstration of a DoS attack based on the information obtained from the reconnaissance.

The article is structured as follows: The introduction outlines the issues and summarizes related works status. The following chapter describes the difference between passive and active scanning and the active scanning tools used for work. The next chapter deals with the description of the testbed environment. In conclusion, the results obtained by the performed experiment were evaluated.

1.1 RELATED WORKS:

The number of articles dealing with the possible impacts of scanning tools on industrial networks is minimal. The methodology for searching and selecting articles dealing with this issue was as follows: During the 9th of March 2021, a final set of keywords (*"Industrial control system*" AND (Reconnaissance or scan*)*) was used to query Scopus and Web of Science (WoS). This query identified 81 articles for Scopus and 52 for WoS. The results duplicates from databases were removed. After that, the systematic literature review identified a total of 15 articles. Only three articles [2, 3, 4] dealt with the issue of scanning in more depth, and only one [3] of them dealt with the impact of scanning on the industrial networks.

2 NETWORK SCANNING

According to Bou-Harb et al. [5] network scanning can be divided into three different categories: (1) Nature of cyber scanning, (2) Cyber Scanning Strategies, (3) Cyber Scanning Approaches. This article focused on the first category, on the division of scanning into passive and active.

- **Passive scanning:** Identifies network services by monitoring the traffic generated by servers and clients as they pass through the observation point. Specialized hardware, software, port mirroring, or hardware taps can be used. Alternatively, specialized software can be used. An example of the most commonly used passive scanning software is Wireshark.
- Active scanning: It identifies devices in the network by sending probe packets and then monitors the individual hosts' responses. An example of a probe packet might be a typical TCP handshaking procedure for establishing a connection. With the active scanning, information about the operating system can be obtained thanks to fingerprinting and possible processes and applications running on the hosts.

Thus, active scanning can provide information about open ports and whether these ports are protected. The main disadvantage of active scanning is that it is very intrusive because it requires the device's responses. This creates a problem when scanning industrial networks where the devices are very sensitive. This scanning can be revealed thanks to IDS, which are not a standard part of industrial networks. In contrast, passive scanning's main advantage is that it is not intrusive and cannot be detected by a third party. Within passive scanning, it is also possible to find out more information about what is happening on the network, especially since it can be turned on for a very long time without any detection. The main disadvantage of passive scanning is that it detects only active services. The article further deals only with active scanning.

2.1 TOOLS FOR ACTIVE SCANNING

Two prevalent open-source tools for active network scanning (Nmap and Zmap) were chosen for the article. The selection emphasized open-source applications as well as their documentation and popularity. To demonstrate how easy and affordable these tools are to use.

- **Nmap:** It is one of the most widely used active scanners in the world of penetration testers and attackers. Supports several types of scanning such as TCP-SYN Scan, Service Detection Scan, HTTP Banner Grab.
- **Zmap:** is a fast single packet network scanner designed for Internet-wide network surveys. Supports TCP-SYN Scan, ICMP Ping Sweep, NTP Scan.

3 TESTBED

To examine the impact of active scans on the industrial network, a testbed was created on real industrial hardware. The diagram of the testbed and communication architecture see in the Figure 1. Due to the fact that this problem concerns industrial equipment and elements in industrial networks in general, and also in order not to demonstrate vulnerabilities to a particular manufacturer, the description of the equipment and the scheme are described in general.



Figure 1: Scheme of industrial control system network tesbed.

The engineering workstation station is used for communication with devices, analysis, and supervision. It is also used to update processes and programs running on the devices. The PLC is a control unit for the entire network and provides the logic of its work. If the PLC communication is affected, the whole network is disrupted either by increasing the delay, significantly impacting the final process or a complete outage. There is also a Human Machine Interface (HMI) in the network with which the operator can monitor and make changes in the network's operation so that he does not have to access the engineering workstation, which is often elsewhere. The demonstration of the process output was servomotor with an encoder chosen. For this servomotor to work, a servo driver directly designed for this engine is connected to it. Servodrive is a frequency converter that enables communication with the PLC. The individual parts are connected using an industrial switch. The whole testbed is powered by a DC source with 24 V output and 5 A. The Raspberry Pi was used as an scanning device.

4 IMPACT OF ACTIVE SCANNING ON THE INDUSTRIAL NETWORK

The network's scanning will be described from the penetration tester point of view, and the impact on this network will be commented. The Nmap and Zmap tools have been selected for active scanning. First, the individual tools were tested to see how significant intrusive impact they would have.

4.1 ZMAP

During the initial test, it was found that the Zmap tool has a very intrusive impact on the entire network. The command was used to test the scan: *sudo Zmap -p 102 -N 100*. As can be seen

in Figure 2, this command disabled the PLC communication, and therefore, it is not possible to scan an industrial network at all with this scanning tool.

		Reply from 19 Reply from 19 Reply from 19 Request timed	2.168.0.1: by 2.168.0.1: by 2.168.0.1: by 2.168.0.1: by	tes=32 time=5ms T tes=32 time=2ms T tes=32 time=2ms T	TL=30 TL=30 TL=30
🍟 Onlin	🔓 Opera	Device/module	Connection establis	Message	Details
2 5	a ²	PLC_1	Direct	Not reachable	Establish new online connection
a ²⁷		GSD device_1	PLC_1	Not reachable	Establish new online connection

Figure 2: PLC communication failure.

4.2 NMAP

Scanning with Nmap had no as much intrusive effect. Therefore, a possible procedure of the penetration tester (described in the introduction) in the industrial network and what this procedure can have an impact is demonstrated here.

- 1. Check for live systems: The command *Nmap -sP 192.168.0.0/24* was used to search for live devices, and all devices were detected. Most attempts went without a harder impact on the network, but in one case out of many, the PLC communication failed, and therefore, the other devices stopped working.
- 2. Check for open ports: The command *sudo Nmap -sS -p- 192.168.0.0/24* was used to search for open ports. In most cases, it was not possible to capture the servo driver at .05. One of many attempts was to disrupt communication. Mostly without impact. To get more information, the -v (verbose) flag was used, so the *sudo Nmap -v -sS -p- -T5 192.168.0.0/24* command was used. At speed T5, the network usually remains stable. However, if the T4 speed is set or the targeted scan is directly on the device, more frequent outages occur. If there is no outage, there is a significant delay. The error messages are the same as for the Zmap scan.
- 3. Scan beyond IDS: This step was not necessary.
- 4. **Perform banner grabbing:** For gathering more information, the operating system detection command (*sudo Nmap -O*) was tested. The individual IP addresses were gradually tested. It was not possible to obtain operating system information from industrial devices. There was not a single network outage during this testing.
- 5. Scan for vulnerabilities: From the information obtained, it was possible to identify the devices in the network partially. It was possible to identify open ports on devices and some processes. It was also found that there is an open port 102 on which the iso-tsap process was running, which is vulnerable to a DoS attack.
- 6. **Draw network diagrams:** From the obtained information, it was possible to create a reconstruction of the network architecture.
- 7. **Prepare proxies:** This step was not necessary.

4.3 DOS DEMONSTRATION

From the information obtained, it was possible to test the DoS attack. The attack was aimed at the HMI device to demonstrate how the device can be loaded with queries, and it is then unable to respond. Due to this fact, the operator is then not able to control the network using the HMI. A program called hping3 was chosen to demonstrate the DoS attack.

A packet flood was sent to the HMI 192.168.0.3 using the command: *sudo hping3 -V -c 200000 -d* $150 - S - p \ 120 - flood \ 192.168.0.3$. As a result, the device was taken out of service for the time of sending packets, and it was not possible to control the running of the servo motor with it. After the attack ended, the state returned to its original state, and it was possible to control the servo motor again using the HMI. Evidence of failure can be seen in Figure 3.

	R R R R R	eply from 192 equest timed o equest timed o equest timed o eply from 192 equest timed o	.168.0.3: byte but. but. but. .168.0.3: byte but.	es=32 time=4ms TT es=32 time<1ms TT	L=64 L=64
🍟 Onlin	🔓 Opera	Device/module	Connection establis	Message	Details
د		HMI_RT_1	Direct	Not reachable	Establish new online connection

Figure 3: HMI communication failure.

5 CONCLUSION

It can be seen from the results in the article that it is not appropriate to use active network scanners for scanning industrial networks. The first described tool Zmap was very intrusive for the network communication. The impact on the device was with each attempt to scan. The second described tool was Nmap. This tool did not have as significant an intrusive impact as Zmap. For this reason, the procedure of a penetration tester methodology was demonstrated on this tool. In step 1) Check for live systems; a communication failure was detected in only one of 40 attempts. In step 2) Check for open ports; the impact was more significant, especially when using the -v (verbose) flag. The article also was demonstrated how to gather information. With this informations, a DoS attack was performed on the HMI. As a result, the operator was not allowed to control the device using this terminal. It is not advisable to use active scanning tools to improve security (gathering information about open ports and information about devices) in a real industrial network. From the network administrator's point of view, for the reasons mentioned above, it is better to switch to long-term passive scanning from which more information can be found.

ACKNOWLEDGMENT

The described research is part of the grant project registered under no. FV40366 and funded by the Ministry of Industry and Trade of the Czech Republic.

REFERENCES

- [1] WALKER, Matt. 2019. *CEH Certified Ethical Hacker All-in-One Exam Guide*. 4th ed. New York: McGraw-Hill Education.
- [2] TAMURA, Kensuke and Kanta MATSUURA. 2019. Improvement of Anomaly Detection Performance Using Packet Flow Regularity in Industrial Control Networks. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences: Special Section on Cryptography and Information Security.* 2019 (E102.A), 65–73.
- [3] COFFEY, Kyle, Richard SMITH, Leandros MAGLARAS and Helge JANICKE. 2018. Vulnerability Analysis of Network Scanning on SCADA Systems. *Security and Communication Networks*. 2018, 1–21.
- [4] ANTROBUS, Rob, Benjamin GREEN, Sylvain FREY and Awais RASHID. 2019. The Forgotten I in IIoT: a vulnerability scanner for industrial internet of things. *Living in the Internet of Things* (*IoT 2019*). Institution of Engineering and Technology, , 1-8.
- [5] BOU-HARB, Elias, Mourad DEBBABI and Chadi ASSI. 2014. Cyber Scanning: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials* [online]. 16(3), 1496-1519.

SECURITY OF WEB APPLICATIONS IN PHP

Tomas Slunsky

Doctoral Degree Programme (2), FEEC BUT E-mail: xsluns01@stud.feec.vutbr.cz

> Supervised by: Jiri Prinosil E-mail: prinosil@feec.vutbr.cz

Abstract: This article deals with the security of web applications, focussing on vulnerabilities in web applications written in PHP language. This work reveals existing security issues, demonstrates the impact of them and propose solution with more approaches. The solution focuses mainly on the level of network filtering with Intrusion Detection System (IDS) or Intrusion Prevention Systems (IPS). There are more issue solution approaches and it will therefore be possible to propose the best one and describe it more.

Keywords: web, web application, vulnerabilities, security issues, HTTP, PHP, OWASP, exploit, IDS, IPS, server-side programming language

1 INTRODUCTION

Nowadays the importance of the Web and web applications are growing continuously. The main reason for that is the important fact - informations from the Web are easily accessible from anywhere in the world and it is practically the most used software for this reason at all. Other reasons for the growth were the simplicity how web applications can be deployed today. Other important fact is the entire communication process provided by the HTTP (Hypertext Transfer Protocol) protocol and others related to it. Due to these circumstances many e-commerce companies appeared and build entire business on this principle. Changes are also happening in the case of innovation, where companies are integrating web technologies, from industry to banking, etc.

With the increasing demands on web presentations, it was necessary to introduce new technologies, thanks to which it was possible to better interact with the user and offer more advanced services and functions. Since the beginning of the Internet, web presentations have moved forward enormously. But the requirements have increased too. With the introduction of new technologies, there are also associated security issues and vulnerabilities. Security emphasis is needed much more than ever.

Web application security has become an increasingly important topic. From this reason the Open Web Application Security Project (OWASP) was established to improve security. OWASP maintains the top 10 web application vulnerabilities.

The goal of this article is to describe the vulnerability of web applications in PHP and describe advanced techniques for dealing with these vulnerabilities. Intrusion Detection System (IDS) / Intrusion Prevention Systems (IPS) technology, its possibilities and a proposal for improvement will be discussed in more detail.

1.1 VULNERABILITY

A website vulnerability in this article meaning is a misconfiguration or weakness in a website or web application code that allows an attacker to gain some level of control of the website, and possibly the whole server in the worst case.

1.2 PHP

PHP (Hypertext Preprocessor) is a language used to create dynamic web applications. There are many web frameworks written in PHP for web development. A web framework provides a structure and starting point for creating web application. There are many things solved yet. They offer ready-made tools for authentication, security and have built-in libraries and tools. The time required for development is less. Examples of PHP frameworks are Nette, Laravel, CodeIgniter and others.

2 ANALYSIS OF CURRENT WEB VULNERABILITIES

Nowadays PHP is powering more than 75% of the top ten million websites and is the most widely used server side programming language in Web applications [3]. On PHP is running huge projects as Wikipedia or Facebook for instance. The reason why PHP is a so popular general-purpose scripting language for web is because of his suitability for web development and is free, fast, flexible and also pragmatic. Server-side programming language usage [6] is shown in Fig 1.



Obrázek 1: Usage statistics of server-side programming languages for websites

2.1 OWASP TOP 10

According to OWASP [5] the most fundamental vulnerabilities in recent years are attacks [4] such as:

A1 - Injection An injection is a very widespread attack and can occur when untrusted data is sent to an interpreter as some part of a query or command. [4]. The attack can affect various systems such as SQL (Structured Query Language), NoSQL (non-SQL), LDAP (Lightweight Directory Access Protocol) and can cause an interpreter to execute unwanted commands or access data without proper permission.

Example of vulnerability is shown bellow.

```
mysql_query("SELECT * FROM user WHERE id_user = " . $_GET['id']);
```

At this moment, attack can be done with GET request to endpoint.

http://example.com/ourscript.php?id=25~ORDER~BY~2

- A2 Broken Authentication Web application features related to user authentication may be implemented incorrectly. This can lead to an attacker being able to steal a user's identity. By exploiting this vulnerability, attacker is able to perform all of user's operations. The user's identity can be stolen temporarily or even permanently.
- A3 Sensitive Data Exposure This is one of the most critical security threats. It occurs in case a web application does not sufficiently protect sensitive information from users who aren't allowed to access to data. These are implementation errors that attackers use to gain access to sensitive data stored by applications.
- A4 XML External Entities (XXE) This vulnerability can allow an attacker to view files on the application server filesystem, remote code execution or internal port scanning via XML External Entities. Vulnerability is caused by poorly configured XML (Extensible Markup Language) processors. The problem is little known yet a many of today's security tests do not take it into account. At the same time, the impact of its misuse can be very serious.
- A5 Broken Access Control Authenticated users are allowed to do everything without any restrictions because of they are often not properly enforced and attackers can exploit these flaws to access unauthorized functionality and/or data.
- **A6 Security Misconfiguration** It's the most commonly seen issue caused by insecure default, ad hoc or incomplete configurations. Another important thing is also to fix the error using the patches and keep the system up to date.
- A7 Cross-Site Scripting (XSS) A security vulnerability occurs when a web application inserts untrusted data into a client browser's output without proper validation or escaping. For instance, it can be done by saving untrusted data into web application database via some of the input forms. By exploiting this weakness, the attacker is able to steal the user's session, modify it page content, redirect users and use the user's browser to run malicious code. XSS attacks are a serious danger to web applications and can have major consequences in the case of a successful attack. XSS attack is demonstrated in Fig 2.



Obrázek 2: Demonstration of XSS attack. [7]

- **A8 Insecure Deserialization** User-controllable untrusted data is deserialized by a website. This situation potentially enables an attacker to modify serialized objects in order to pass malicious data into the code of application.
- **A9 Using Components with Known Vulnerabilities** Most of today's frameworks are using different third party libraries. If one of these libraries contains a vulnerability and that one is exploited, it can lead to an application crash, etc.
- A10 Insufficient Logging & Monitoring The consequence of this vulnerability is there is no immediate response for bug and potential attackers can exploit this vulnerability for long time until it is typically detected by a third party [2].

3 NEW PHP NETTE VULNERABILITY

3.1 NETTE CVE-2020–15227

A few months ago, at the end of 2020, there was a big vulnerability discovered in the Nette framework. Nette is a full featured component based framework developed by David Grudl and czech PHP community. Nette is powering many websites in the Czech Republic and abroad and runs large websites, e-shops and projects.

The issue that is occurred in the framework was the variant of injection. In certain circumstances it was possible remote code execution via this fatal security issue [1], in other words, there was a code injection belonging to the A1 OWASP category.

The bug affected all versions of nette from version 2.0.19 to version 3.0.6 practically and the attack was carried out using a specially compiled URL (Uniform Resource Locator), which was able to execute PHP code then.

It is not known yet how long this bug could have been exploited by the attackers and whether any damage has occurred so far.

4 SOLUTION WITH IPS/IDS DETECTION

The IPS system can be divided according detection method. Signatures detection method uses its own database of marks, which are strings specific to a given type of attack. An IPS system that uses stateful tag detection monitors network traffic for compliance with these types of attack-specific tags. Once a match is found, the IPS system takes the appropriate action.

Based on the previously described attacks and due to the possibilities of how attacks are performed, it is advisable to choose an IPS system for the protection of the web application, which will use signature detection. For better attacks filtering, the signature database is updated using machine learning. Machine learning deals with algorithms that allow to change internal state. Thanks to machine learning, it is possible to adapt to changes in the surrounding network environment. It can better respond to new attacks and thus increase the security of web applications. The basic task types of machine learning can be summarized in the following three points:

- cluster analysis of data with similar properties
- classification of input data into classes
- estimation of the numerical value of the output according to the input.

5 CONCLUSION

This article outlines current vulnerabilities in web frameworks. Furthermore a specific case of vulnerability in Nette framework was described. Article also discusses the possibilities of defending these vulnerabilities using IPS/IDS and suggests effective ways to mitigate the effects of the security issues.

REFERENCE

- CVE-2020–15227: Potential Remote Code Execution Vulnerability. In: Nette Comfortable and Safe Web Development in PHP [online]. nette.org: nette.org, 2020, 2020 [cit. 2021-02-14]. Available from: https://blog.nette.org/en/cve-2020-15227-potential-remote-code-executionvulnerability
- [2] Insecure deserialization. In: Web Application Security, Testing, & Scanning PortSwigger [online]. portswigger.net: PortSwigger, 2020, 2020 [cit. 2021-02-14]. Available from: https://portswigger.net/web-security/deserialization
- [3] M. Backes, K. Rieck, M. Skoruppa, B. Stock and F. Yamaguchi, "Efficient and Flexible Discovery of PHP Application Vulnerabilities,"2017 IEEE European Symposium on Security and Privacy (EuroS&P), Paris, France, 2017, pp. 334-349, doi: 10.1109/EuroSP.2017.14.
- [4] OWASP Top Ten. In: OWASP Foundation | Open Source Foundation for Application Security [online]. owasp.org: owasp.org, 2020 [cit. 2021-02-14]. Available from: https://owasp.org/www-project-top-ten/
- [5] OWASP. Open Source Foundation for Application Security [online]. Owasp.org: OWASP Foundation, 2021 [cit. 2021-03-24]. Available from: https://owasp.org/
- of [6] Usage statistics server-side programming websites. In: languages for W3Techs [online]. online: W3Techs, 2021 [cit. 2021-02-13]. Available from: https://w3techs.com/technologies/overview/programming_language
- [7] XSS. In: Defender's Notes [online]. notes.defendergb.org: Defender's Notes, 2020, 2020 [cit. 2021-02-14]. Available from: https://notes.defendergb.org/web-sec/vuln/xss

IMPLEMENTATION OF FSK MODULATED SIGNAL RECEIVER USING SOFTWARE DEFINED RADIO

Jan Pospisil

Doctoral Degree Programme (2), FEEC BUT E-mail: xpospi90@vutbr.cz

Supervised by: Radek Fujdiak E-mail: fujdiak@feec.vutbr.cz

Abstract: This paper discusses recent trends in a wireless communication system using Software Defined Radio (SDR). SDR, in this case, the BladeRF 2.0, in combination with a software layer written in the Python language, is taking care of signal demodulation. BladeRF 2.0 is a device that provides a powerful waveform development platform expected by industry professionals. The BladeRF 2.0 is used to design a Frequency Shift Keying (FSK) receiver (in the form of binary data). As a transmitter, we created a custom-designed end-device using CC1101 RF (Radio Frequency) module. The scenario in the current experiment is that the SDR acts as a receiver for the transmitting end-node.

Keywords: BladeRF, SDR, Software Defined Radio, signal demodulation

1 INTRODUCTION

The RF (Radio Frequency) signal is used for wireless communication between electronic devices. It's first necessary to apply a suitable modulation technique to convert a baseband signal into an RF signal to transmit it through an RF environment. Modulation is the process of changing one or more properties of a periodic waveform, called a carrier signal, with a separate signal called a modulation signal, which usually contains information to be transmitted. Within the signal transmission, its failures occur; this is due to, e.g., interference, attenuation, and reflections. Frequency Shift Keying (FSK) is a frequency modulation system in which digital information is transmitted through a discrete change in carrier frequency. Thus, the value of logical zero is one frequency, and the value of logical one is another frequency.

In contrast to relative work articles where the authors of [1] focus on the implementation of Quadrature Phase Shift Keying (QPSK) modulation using Field-Programmable Gate Array (FPGA) and verify the results using simulation. In another paper, the authors apply QPSK modulation [2], capable of partial re-configuration and theoretical verification. The authors of the Frequency modulation (FM) modulator [3] focus on the FPGA implementation.

In the field of Software Defined Radio (SDR), a relatively large number of articles are available, including the ones mentioned above, which deal with the implementation of signal processing in various variants. The topic of SDR technology is relevant, especially in the last 20 years. However, most scientific publications focus on theoretical and simulation results and thus lack real-life implementation. In contrast to the above, our paper focuses on implementing a demodulator verified in real-life conditions. Due to the relatively high price of SDR hardware, the real-life deployment of SDR technology is still slow. SDR devices are mostly used for scientific research. As part of the future vision, SDR equipment should replace complex single-purpose systems. The main advantage is saving funds when implementing a new communication technology into the already deployed infrastructure that uses SDR. Where, in contrast to conventional single-purpose systems, the upgrade can be performed mainly in the software.

2 SOFTWARE DEFINED RADIO (SDR) AND THE END-DEVICE

The SDR device receives a signal from the ambient RF environment at configured parameters in the form of the carrier frequency (f), bandwidth (BW), sampling rate (SR), and gain (G). After setting these parameters, the device starts capturing the digitized analog signal. In this case, it is necessary to perform the implementation of signal processing by the software.

 Table 1: BladeRF 2.0 micro A9 basic parameters [4].

Frequency range Bandwidth	Sampling rate
1 9 8	Sumpling fute
BladeRF 2.0 47 MHz – 6 GHz 56 MHz	61.44 Msps

The BladeRF 2.0 micro xA9 device (see Figure 1) has, in addition to the parameters mentioned in Table 1, also a USB 3.0 port (Cypress FX3) and allows 2x2 Multiple-Input Multiple-Output (MIMO) communication. This device's advantage is the presence of an Altera Cyclone FPGA circuit that can be used for direct signal processing [4].



Figure 1: Function block diagram of the BladeRF micro xA9 device [4].

The BladeRF contains the RF 2×2 transceiver AD9361. The Rx signal path from AD9361 passes downconverted signals to the baseband receiver section. The baseband Rx signal path is composed of two programmable analog low-pass filters, a 12-bit ADC (Analog-to-Digital Converter), and four stages of decimating filters (see Figure 2) [5].



Figure 2: Functional block diagram of one RX channel of AD9361 [5].

The output from the BladeRF is the data flow of the digitized captured radio signal samples. The data flow consists of a sequence of individual samples in the form of two 16-bit values that represent the component I and Q. The output data rate depends directly on the sampling frequency. When the maximum sampling frequency is set (61.44 MHz), the device then generates ≈ 2.0 Gbps. As the data flow increases, so do the demands for the computer's performance for processing this data. As part

of an advanced implementation, the computational load can be reduced by the FPGA engagement for the necessary signal processing that might be performed directly in the BladeRF. The computer then receives the already processed data. The FPGA signal processing implementation is considering to be done in the future. For the real-life transmitting end-device and BladeRF SDR device, see Figure 3.



Figure 3: End-device periodically sends data using FSK modulation on the left and the receiving SDR BladeRF on the right.

The end-device consists of an STM32L432KC microcontroller (MCU) powered via micro USB port, and CC1101 RF integrated circuit wired via Serial Peripheral Interface (SPI). Table 2 summarizes the end-device configuration. It should be noted that the device does not use any data encryption. Otherwise, the received data would not be readable without knowing the decryption key. In a real-life scenario, the test end-device might be replaced by any regular end-device that uses the FSK modulation after modifying the receiver's signal parameters, such as carrier frequency, bandwidth, deviation, and data rate. The testing transmission distance was for the best results (low error rate) in the length of centimeters.

 Table 2: The end-device configuration.

Radio	Carrier frequency	Bandwidth	Data rate	Frequency deviation	TX power
CC1101	868 MHz	203 kHz	4800 bps	47.607 kHz	12.5 dBm

3 PYTHON SOFTWARE FSK DEMODULATION

We used the Universal Radio Hacker (URH) tool to receive the samples from BladeRF. URH stores the samples in the file with name extension: complex32s. As the extension indicates, the file contains complex samples composed of the two signed 16-bit integers for the I and Q values. The Python script further handles the processing part. For all the signal processing parts, see Figure 4.



Figure 4: Basic blocks of signal processing for FSK demodulation based on envelope detection.

We smoothed the received signal (see Figure 5) by the moving average filter with a window length of 10 samples. Due to the receiver's high sampling rate (2.5 Msps) and transmitter's low data rate (4.8 kbps), a negligible signal degradation occurred by filtration.



Figure 5: Raw FSK signal received samples – the first 4 000 samples.

In the second step, we applied the first-order discrete differentiation function. We achieved the envelope detection using a Hilbert transformation. Finally, we filtered the signal with a low-pass 100 tap FIR filter with a cutoff frequency of 1 x bitrate to obtain the optimal results (see Figure 6).



Figure 6: Hilbert transformation and low-pass filtering of the signal (red) compared to the smoothed FSK signal (blue) – the first 25 ms.

The last step in the signal processing part is the slicing and bit recognition. First of all, we took the mean of the entire signal as a decision threshold. Every time that signal crosses the mean value, the slicer creates an edge (see Figure 7 – the red line). This operation results in square signal output. At this moment, we cannot directly convert the square signal into bits due to the sequences of the same values. These need to be split first by the symbol length. The script goes through the edges and, given the symbol lengths (based on adaptive symbol length measurement), breaks the same-value sequences and recognizes each bit.



Figure 7: Slicing (red) envelope signal (blue) based on the threshold mean value (orange) and the edge detection – the first 25 ms.

The following Figure 8 shows the whole received signal with the edge detection. The bit recognition result can be seen more in detail in Table 3. As can be observed at the beginning of the frame, a preamble (4 B) is received, then a header containing synchronization data and payload length (5 B). Then a user payload (14 B) is received. Finally, a footer in the form of a Cyclic Redundancy Check (CRC, 2 B) is added.


Figure 8: Slicing filtered envelope signal (red) based on the threshold mean value (orange) and the edge detection - complete view.

Index	0	1	2	3	4	5	6	7	8	9	10	11	12
Hex	AA	AA	AA	AA	D3	91	D3	91	0E	48	65	6C	6C
Ascii	а	а	а	а	Ó	Â	Ó	Â	Â	Н	e	1	1
Index	13	14	15	16	17	18	19	20	21	22	23	24	
Hex	6F	2C	20	57	6F	72	6C	64	21	0A	5C	97	
Ascii	0	,		W	0	r	1	d	!	<lf></lf>	١	Â	

Table 3: Received data (byte index, hexadecimal and ASCII value).

4 CONCLUSION

Using a software-defined radio, we successfully put a system for receiving the FSK modulated signal into operation. The advantage of this design is the possibility of a feasible way to modify the signal processing part (i.e., modulator/demodulator). Compared to all the hardware used in a traditional communication system, we created this in the software. We successfully transmitted the data from the real-life end-device to the BladeRF SDR and decoded it. Therefore, it is clear that other signal modulation types, even more complex, are possible to be implemented. Our future focus is on developing the signal processing directly on the FPGA within the SDR to reduce the computer's load.

ACKNOWLEDGEMENT

The described research is part of the grant project registered under no. TJ02000332 and funded by the Technology Agency of the Czech Republic.

REFERENCES

- QPSK Modulator and Demodulator Using FPGA for SDR MUKESH, Mandadkar; ABHISHEK, Lokhande; BHAMBARE, R. R. QPSK modulator and demodulator using FPGA for SDR. *International. Journal of Engineering Research and Applications*, 2014, 4.4: 394-397.
- [2] KUMAR, KA Arun. FPGA implementation of PSK modems using partial re-configuration for SDR and CR applications. In: 2012 Annual IEEE India Conference (INDICON). IEEE, 2012. p. 205-209.
- [3] HATAI, Indranil; CHAKRABARTI, Indrajit. A new high-performance digital FM modulator and demodulator for software-defined radio and its FPGA implementation. *International Journal of Reconfigurable Computing*, 2011, 2011.
- [4] *Nuand LLC*, BladeRF 2.0 [online]. [Accessed 8 March 2021]. Available from: https://www.nuand.com/bladerf-2-0-micro/
- [5] AD9361 Functional Block Diagram [online]. [Accessed 8 March 2021]. Available from: https://www.analog.com/en/products/ad9361.html

START BUILDING YOUR OWN CYBER RANGE – CYBER ARENA

Tomas Stodulka

Doctoral Degree Programme (1), FEEC BUT E-mail: xstodu07@vutbr.cz

Supervised by: Zdenek Martinasek E-mail: martinasek@feec.vutbr.cz

Abstract: The research deals with the importance of education against cyber crimes in today's world. One of the solutions is through so-called cyber ranges. First two chapters briefly explain the current problematic and present some representatives of cyber ranges. The work continues by stress testing of the selected cloud computing platform that is the most suitable for cyber range purposes. Based on the results, servers/devices are purchased, which form the basis of the future Cyber Arena in BUT.

Keywords: Cyber Arena, cyber range, cyber security, OpenStack, stress testing.

1 INTRODUCTION

We live in a world where technology and vast amounts of data play a significant role in everyday life. Recently, they are also developing very rapidly and it is expected that their growth will continue at the same or even higher pace. However, there is also a higher probability of security risk. This fact gives the attackers a substantial advantage over the defenders. Another critical point is that today's attackers may not have the complex knowledge to perform cyber attacks. There are many automated tools available on the Internet that are not complicated to use. Or conversely, use the created exploit for the vulnerability (e.g. in an outdated version of the system). That is why we are increasingly encountering cyber attacks, because even a beginner can become an attacker today.

The company usually has a security policy customized to its own needs. However, even the most secure system can be compromised. Attacks can come at any time and their nature can change immediately. Highly skilled attackers are constantly evolving, testing, innovating, and developing new tools. No one is immune from government organizations to the smallest family business. Therefore, it is necessary to educate public or professional experts to be able to plan for potential attacks and prepare for their consequences and generally to know how to limit such security risks. These purposes are greatly fulfilled by cyber range – platform, responsible for the complete preparation of various simulated scenarios, where users are exposed to diverse challenges [1].

As these platforms are usually paid for a considerable amount of money, there is ongoing work on the own cyber range in BUT named **Cyber Arena**. This platform will be used in the education of a couple of university courses to improve students' awareness of the cyber crimes mentioned earlier.

2 CYBER RANGES IN CZECH REPUBLIC AND WORLD

Cyber ranges may be owned by governments, industries, and academic institutions. Because they manage virtual resources, their usage is not limited to the organization's local network, but could be available also by Internet across the world, so they can be used by private as well as public organizations along with students, researchers, trainers and education providers. There are tens of platforms available today that can be classified on the basis of infrastructure association as public (CyberBit, KYPO), private (IBM, KYPO), or federated (NATO, CRATE) [1]. Below is described the largest cyber range platform in Czech Republic and another one important from around the world.

KYPO Cyber Range Platform (KYPO CRP) is developed by Masaryk University since 2013, and is entirely based on state-of-the-art approaches such as containers, infrastructures as a code, microservices, and open-source software, including cloud computing technology – OpenStack. KYPO CRP focuses on two use cases: (1) creates a pool of virtual environments using user-prepared definitions; (2) provides GUI for training, each run of training has an assigned virtual environment created in case 1. At the end of the year 2020, KYPO CRP becomes open-source [2].

CyberBit was founded in 2015. Trainees operate in a hyper-realistic environment which includes a configurable corporate network, commercial security tools, emulated traffic, and simulated attacks with various levels of difficulty. CyberBit provides the creation of sessions for training the latest online threats, such as ransomware, web server shutdown, and denial of service attacks. Within the sessions, this platform also provides: instructors, a wide range of precreated scenarios, and all without the need of implementation on own servers. The main disadvantage of this platform is its price, and as a long-term solution it is therefore not entirely suitable [3].

Cyber range must be capable of a wide range of functions. The most important is the virtualization of secure environments for users, including communication infrastructure (firewalls, routers, networks, VPNs – Virtual Private Networks, etc.), servers with remote access (VMs – Virtual Machines, or containers), and segregating these structures into separate projects – all in an automated way. Based on [4], the cloud computing platform OpenStack is chosen as the core of the Cyber Arena.

3 PERFORMANCE TESTING AND RESULTS

Within the measurement, scenarios are created to determine how much the given scenario depends on HW resources such as: disk, CPU, RAM, or network utilization. Based on these results, servers for the cyber arena will be purchased. As a test environment, the entire structure of the Cyber Arena is deployed on 1 VM with significantly weaker HW parameters, so that the overload condition can be easily reached during the measurement. Parameters of mentioned VM are: 8 CPUs, 32 GB of RAM, and 200 GB of disk. Test scenarios are divided into three groups that are described below.

The **first group** of scenarios focuses on the complexity of subcommands (small scenarios) of Open-Stack, to effectively separate which processes have the greatest impact on the HW load. The individual scenarios and their consumption of HW resources are described in table 1. The effect of scenarios on CPU load cannot be clearly determined – the results are greatly influenced by OpenStack actions performed at different intervals. The last three instances were created using the cirros image with parameters: 512 MB of RAM, 1 CPU, 1 GB of disk, and connection to a private virtual network.

The **second group** focused on the creation of a complex scenario with stress test implementation. The scenario consists of 2 VMs in separate networks – client and HTTP/FTP server. The client then constantly downloads a large file and accesses to the web page. During the measurement, CPU was overloaded up to 2.8 times in 7 concurrently running scenarios. Despite the considerable overload, the system was able to handle the load and there were no errors, collisions, or packet losses on the individual virtual interfaces. Instead, the total generated traffic was decreased to 42.1 % of the previous speed (from 16.48 Mbit/s to 7.05 Mbit/s).

From previous measurements, it is still not possible to unambiguously determine what effect the individual load scenarios have on HW resources. The previous scenario is therefore divided into four practical scenarios, which represent the **third group** of measurements: (1) S21 – create a sandbox of an attacker (kali-linux-top10 image) with connection to 1 virtual network and 1 router; (2) S22 – same as S21, plus with a created vulnerability web server (webgoat-8.0 image); (3) S23 – same as S22, but each sandbox is attached to its own virtual network; (4) S24 – same as S23, but the router now becomes a firewall with some rules implemented. As the stress testing, scripts simulating the basic behavior of a student in the role of an attacker are prepared. First of all, the student scans

Scenario	RAM	Disk	CPU
S <i>x</i>	[MiB]	[MiB]	[%]
S1: create virtual network (N)	9,1	≈ 0	can't be determined
S2: create cirtual router (R)	pprox 0	≈ 0	can't be determined
S3: attach N to R	8,1	0,1	can't be determined
S4: create firewall (FW) rule	0,6	≈ 0	can't be determined
S5: create FW policy (FWP)	≈ 0	≈ 0	can't be determined
S6: add FW rule to FWP	≈ 0	2,3	can't be determined
S7: create FW	14,7	0,1	can't be determined
S8: add FWP to FW	≈ 0	≈ 0	can't be determined
S9: apply FW to R	3,9	0,1	can't be determined
S10: create new project	≈ 0	30,3	can't be determined
S11: create user in project	0,4	≈ 0	can't be determined
S12: create VM	239,8	56,7	≈ 1,63
S13: create container	91,8	0,5	≈ 0,19
S14: create sandbox container	191	1,7	$\approx 0,40$

Table 1: Consumption of HW resources by OpenStack commands

adjacent networks looking for a vulnerable server with open ports (using nmap). The moment he finds it, he registers, logs in and at random short intervals, he loads pages containing training for various web vulnerabilities. Web browsing is handled using 2 tools: (1) curl (gets only partial content of pages); (2) phantomjs (supports JavaScript code). Part of the results for these steps is shown in graphs 1.





Figure 1: HW resources usage within nmap stress testing

All graphs and measured data are part of the research report due to the large scale. Based on the measurements, it can be concluded that the creation of a virtual infrastructure has the highest impact on it. OpenStack archives logs that are too large when the log files reach a size (approximately 400–520 MiB), this activity was recorded many times in the measurements (negative values of disk usage). Conversely, the activities of instances are the most dependent on CPU computing power. If the system is overloaded, the process operation is slowed down. As a result, the flow of data across virtual networks or the generation of data flow is reduced.

From the 3rd group of scenarios also follows: (1) The first deployment of the scenario has a much higher RAM usage than the following (consumed initially by OpenStack services); (2) The communication through a router in S23 scenario doesn't have a significant effect on HW resources compared to S22 (slight increased RAM is recorded only for nmap loads); (3) Communication via firewall (S24) also has almost no effect on HW resources compared to the router (S23); (4) The higher load in the S21 scenario for the nmap load is due to the higher scanning frequency, as there is no web server in the network and 1 load cycle then takes 10 times less compared to S22–S24. It also means that scanning the network itself is more difficult than scanning the server ports; (5) Significant use of the S23 disk was caused probably by unknown causes. In addition, mentioned disk usage was always only during the stress testing. After stopping the load, the size of the disk returned to a similar state as before the measurement; (6) The higher data flow in scenarios S23 and S24 is due to the twice number of interfaces between communications.

4 BUILDING THE CYBER ARENA

Based on section 3, several of HW components were purchased. For the implementation of the Cyber Arena, a workplace containing: 3 HP servers, a Smart UPS unit, and Mikrotik optical router with support for SFP+ and SFP28 technologies, was created. The workplace diagram is shown in figure 2 a), and the current physical involvement is shown in figure 2 b).







(b) physical involvement

Figure 2: Cyber Arena cloud

Linux operating system with CentOS 8.2 distribution was installed on all servers. Hardware parameters after OS installation are filled in table 2. The HW capabilities of the Cyber Arena are equal to approximately 1/34 of CPU performance, 1/17 of RAM, and 1/50 of disk (including hardware RAID1 and RAID5 redundancy) compared to the previous experimental environment. Cloud Core router has implemented own scripts for automated backup and firmware upgrades also with a sophisticated firewall to prevent students or viruses to escape from Cyber Arena. Moreover, it serves for VPN connections for students and administrators. After they connect to the private network, they will be able to access all applications and tools available in Cyber Arena.

Hostname CPUs		RAM Disk (SSD+HDD)		Ethernet	Fiber
	[-]	[GB]	[GB]	ports [-]	ports [-]
arena-ctrl01	6	32 +16 swap	0+2700	2	2
arena-comp01	256	256 +4 swap	960+2700	4	2
arena-comp02	256	256 +4 swap	960+2700	4	2

 Table 2: Servers HW parameters

5 CONCLUSION

Cyber ranges represent an innovative solution designed for preparing security environments to effectively train users against today's cyber crimes. They can emulate thousands of scenarios for both beginners and professionals and in the academic sphere, the cyber range offers an integral opportunity to try out theoretical information in practice. Within the article, three sets of scenarios were created including stress testing that simulate the basic behave of students. Based on the measured results, devices and servers representing the basis for the Cyber Arena were purchased. At present, they are fully operational to provide the creation of security environments. Furthermore, Cyber Arena has implemented tools to enable: central logging using Kibana and database Elasticsearch; and monitoring HW and OpenStack resources using Grafana visualization and Prometheus data collection.

There are ongoing works on a web-based training application that will guide students through prepared cyber games. There is also a need to simplify complex scenarios creation – the only possible solution is offered by the Heat OpenStack orchestration service. Thus, another work in progress is to automate the creation of Heat templates. In the future, the mentioned works will be completed and the first cyber games will be implemented.

ACKNOWLEDGEMENT

The presented research was financed by the Ministry of Interior of the Czech Republic under grant no. VI20192022132.

REFERENCES

- [1] PRIYADARSHINI, I. 2018 Features and Architecture of The Modern Cyber Range: A Qualitative Analysis and Survey. Master thesis. Faculty of the University of Delaware.
- [2] *KYPO Cyber Range Platform*. Masaryk University [online] c2020 [cite 14/3/2021]. Available at: https://docs.crp.kypo.muni.cz/.
- [3] Cyber Range for Higher Education Build or Buy?: White Paper CYBERBIT [online]. c2019 [cit. 11.10.2018]. Available at: https://www.cyberbit.com/resource/cyber-range-for-higher-education-build-or-buy/>.
- [4] STODULKA, T.; MARTINASEK, Z.; FUJDIAK, R. Cloudove platformy a jejich srovnani: Kubernetes, OpenStack a OpenShift. Elektrorevue – Internet magazine (http://www.elektrorevue.cz), 2020, vol. 22, n. 5, p. 145-155. ISSN: 1213-1539.

SOFTWARE DEFINED NETWORK FOR SMART HOME

Shujairi Murtadha

Doctoral Degree Programme (1), FEEC BUT xshuja00@vutbr.cz

Supervised by: Vladislav Škorpil

E-mail: skorpil@feec.vutbr.cz

Abstract: As the world embraces Smart Home Technologies, there is a need for an efficient management platform for these services. Accordingly, this article proposes the Software-Defined Smart Home (SDSH) platform, which offers openness, virtualization, and centralization. The platform can integrate heterogeneous devices and offer flexibility between user demands and family scenes. This article offers a detailed explanation of SDSH and using it as a Firewall, explores potential applications, and identifies challenges and opportunities associated with its adoption.

Keywords: SDN, Smart Home, SDSH, IoT, Raspberry Pi

1 INTRODUCTION

Over the recent past, the world has witnessed the proliferation of Smart Home solutions. Smart home technologies are defined as digitally enhanced devices that provide high-quality services [1]. Moreover, Smart homes consider a promised approach for energy efficiency, climate change, and others [2]. the Office of Gas and Electricity Markets in the United Kingdom (UK) demonstrated that smart homes provided a solution for the decarbonization of electricity and other related programs [3]. the prevalence of smart homes' technology reached about 7.5% of households and the expected revenues about of 44.2 billion in 2018 [3]. Moreover, the expectations that the smart homes ached about 30% (84 million) of households by 2022; with France, Germany, and the United Kingdom leading the European market [4].

Indeed, the growth of the Smart Home market has continued to grow ever since. The field of academics has been examining the gap between the need and adoption of Smart Home applications. One of the key findings identified is that users are experiencing challenges managing the ever-growing list of devices in their homes [5]. The sheer diversity of smart home devices means that issues of interoperability and poor integration are common. As a result, Dixon et al. emphasized the need for the creation of dynamic smart home solutions [5]. As smart devices and the complexity of applications increase, the capability to meet the needs of common users for smart homes remains problematic.

2 SOFTWARE-DEFINED NETWORK (SDN)

The hierarchical network architecture that has been successful for open systems, is not able to deal with the complexities associated with closed systems. At the same time, global internet traffic has been increasing exponentially [6]. Today, the typical user demands greater bandwidth and additional services. Accordingly, there is a need for a scalable, high-performance network architecture to support the growing demand and enable flexibility. In 2018, McKeown presented the concept of Software-Defined Networking (SDN) that comprises data and control planes [7]. While the Data Plane comprises the hardware abstraction and physical Infrastructure Layers, the Control Plane consists of network applications and operating systems [7]. OpenFlow, which is a standardized communication protocol, decouples these two planes [8]. Some of the advantages of the SDN architecture include openness, the ability to deploy the network operating systems and applications on servers that adopt

X86 architecture and control data forwarding using OpenFlow, and the capacity to use OpenFlow to decouple the Control and Data Planes and virtualize the network [7]. As such, the concept of SDN has found applications in routers due to the need to develop flexible, open, and modularized devices [9]. The idea has also found application in cloud computing and data center applications [10,11]. Researchers have also explored the possibility of applying the concept of SDN to the Internet of Things [12]. The application of SDN could help in developing an IoT-oriented system structure design to address challenges such as service quality and network heterogeneity [13,14]. However, SDN does not adequately address two major issues: rapid developments in the Smart Home field and the complexity of application scenes. Accordingly, the Software-Defined Smart Home (SDSH) is proposed.

3 SOFTWARE-DEFINED SMART HOME (SDSH)

The Software-Defined Smart Home (SDSH) adopts an approach that comprises three levels: Application Layer, Controller, and Smart hardware layers. Whereas the three layers in SDN technology are similarly offset by the 3 layers in SDSH technology proposed with some additions.

The proposed device in the Control layer is an SDSH Controller that consists of an Arduino Microcontroller or a Raspberry Pi. One of the benefits of this controller is that if the SDN Controller falls, the Microcontroller can manage the Home Network, store its information, and add a level of security by verifying the identity of the User also it can work as a firewall separating the home network from the SDN Controller.

Layer types	SDN	SDSH	Security types
Application Layer	Application, Management	Application, Management	Service security
Control	SDN controller	SDN Controller	System security
Layer (Control plane)	Open vSwitch, Ryu NOX/POX	SDSH Microcontroller Raspberry Pi 4, Arduino	Subsystem Security
Infrastructure Laver		Smart Devices	Device security
(Data plane)		Wi-Fi, Bluetooth, 433MHz	Communication security

Figure 1: Architecture SDN vs SDSH

4 SDSH MICROCONTROLLER AS A FIERWALL

Due to the inability of the sensors to contain the software such as the firewall as an example, or the ability to programing, it represents one of the vulnerabilities that could cause penetration through it and entering the network to the main SDN Controller and tampering fully networks In this section the suggestion is the device (SDSH Microcontroller) to function as a Firewall, not just for home network management as the main part of the SDSH Microcontroller is the Raspberry Pi.

In other words, Raspberry Pi is a small computer (Microcontroller) the size of which does not exceed the size of credit cards with a low cost that can be connected to a computer screen or TV screen, and a mouse and keyboard can be connected to it to facilitate control, and to perform many tasks such as home automation: it can be controlled By turning on and off the lights, temperatures, and humidity in the home using smartphone applications, and storing all the data related to that easily through some sensors, Raspberry Pi and some coding. The important thing is that it is a programmable control device that could contain the software as well as open-source. Figure 2 Shows the SDSH Firewall in the Home Network. Figure 3 represents the Flowchart for SDSH Firewall Implementation



Figure 2: SDSH Firewall for the Smart Home Network



Figure 3: Flowchart for SDSH Firewall Implementation

5 CHALLENGES AND OPPORTUNITIES

Besides the technological gaps, the adoption of SDSH is likely to encounter several problems and present some opportunities. The main challenge is the risk of privacy infringement. Indeed, the Internet of Things (IoT) paradigm has been criticized because of poor family privacy protection. Devices such as sensors and cameras collect vast quantities of personal information, which are fed into machine learning algorithms to support automation [15]. Although measures such as the removal of personal identifiers such as names and addresses are often employed, hackers can reconstruct this information [16]. The SDSH model also proposes an interconnected approach which means that one intrusion could have far-reaching consequences within a network in terms of security and safety. Accordingly, strong security mechanisms must be adopted together with this architecture. The security should not just be limited to technical aspects but must also consider human-related system vulnerabilities [18].

A final opportunity is that SDSH can be integrated with the current Smart Home platform. A key aspect of SDSH is its openness, which allows interconnections with other platforms [17]. It also

includes virtualization technology that minimizes the heterogeneity and complexity associated with connecting different protocols or systems [19]. The SDSH platform can also offer different services for all external services or other platforms hence improving user experience.

6 FUTURE WORK

The future work will be conducting a practical study of the implementation of SDSH Microcontroller integrated into smart home sensors, the study will be considering several performance parameters that will be discussed and analyzed. to design an efficient system that capable of self-recovery in the case of the SDN-failure, artificial intelligence algorithms will be integrated into the system for error detection and correction and to prevent and predict a cyber-attack.

7 CONCLUSION

The next phase of improvements in smart home technologies should focus on addressing challenges such as differences in user requirements, diverse hardware, and the need for monitoring. Accordingly, this paper presents SDSH, which is based on the SDN's strategies of optimization, centralization, and virtualization. SDSH is based on a controller that integrates with external services and connects with different smart devices. It also allows APIs to connect to third-party services. Some of the technologies central to the operation of SDSH include demands-acquiring technology, system recourse virtualization, network formation, and AI-based decision-making. In addition to discovering the requirements of users, SDSH offers a smart platform for controlling devices. Although SDSH is a promising architecture, key challenges such as the possibility of cyber-attacks and privacy infringement must be addressed.

REFERENCES

- [1] Strengers Y, Nicholls L. Convenience and energy consumption in the smart home of the future: industry visions from Australia and beyond. Energy Res. Soc. Sci. 2017; 32:86–93.
- [2] ovacool, Benjamin K., and Dylan D. Furszyfer Del Rio. "Smart home technologies in Europe: a critical review of concepts, benefits, risks and policies." Renewable and sustainable energy reviews120 (2020): 109663.
- [3] Ofgem. Upgrading our energy system: smart systems and flexibility plan: progress update. 2018.
- [4] Sforza M. Twenty-two million smart homes in Europe: from science-fiction to reality. cityfied. Jan-2019 [Online]. Available: http://www.buildup.eu/en/new.
- [5] McKeown, N.: Software-Defined Networking, INFOCOM keynote speech, vol. 17, no. 2, 2009, pp. 30–32.
- [6] Masayoshi, K., et al.: Maturing of OpenFlow and software-defined networking through deployments, Computer Networks vol. 61, 2014, pp. 151-175.
- [7] Xu, K. et al.: Toward a Practical Reconfi gurable Router: A Software Component Development Approach, IEEE Network, vol. 28, no. 5, 2014, pp. 74–80.
- [8] Yang, H. et al.: Cso: Cross Stratum Optimization for Optical as a Service, IEEE Commun. Mag., vol. 53, no. 8, Aug. 2015, pp. 130–39.
- [9] Banikazemi, M. et al.: Meridian: an SDN Platform for Cloud Network Services, IEEE Commun. Mag., vol. 51, no. 2, Feb. 2013, pp. 120–27.

- [10] Valdivieso, Á., L. et al.: SDN: Evolution and Opportunities in the Development IOT Applications, Int'l. J. Distrib. Sensor Networks, vol. 2014, 2014.
- [11] Qin, Z.: A Software Defi ned Networking Architecture for the Internet-of-Things, 2014 IEEE NOMS, 2014, pp. 1–9.
- [12] Jararweh, Y. et al.: Sdiot: A Software Defi ned Based Internet of Things Framework, J. Ambient Intelligence and Humanized Computing, vol. 6, no. 4, 2015, pp. 453–61.
- [13] Lee, H.: Home IoT resistance: Extended privacy and vulnerability perspective, Telematics and Informatics, vol. 49, 2020, pp. 101377.
- [14] H. J. et al.: February. De-identification and privacy issues on bigdata transformation, in 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), 2020, pp. 514-519.
- [15] Alam et al.: IoT virtualization: A survey of software definition & function virtualization techniques for internet of things, arXiv preprint arXiv:1902.10910, 2019.
- [16] Abomhara, M. and Køien, G. M.: Cyber security and the internet of things: vulnerabilities, threats, intruders and attacks, Journal of Cyber Security and Mobility, 2015, pp. 65-88.
- [17] Zhang, D. et al.: Heteroedge: taming the heterogeneity of edge computing system in social sensing, in Proceedings of the International Conference on Internet of Things Design and Implementation, 2019, (pp. 37-48).
- [18] Dixon, C. et al.: An Operating System for the Home, Proc. 9th USENIX Conf. Networked Systems Design and Implementation, 2012, pp. 25–25.
- [19] C. V. N. Index, "Forecast and Methodology, 2013–2018, 2013."

Doktorské projekty

Kybernetika a automatizace I.

FINITE CONTROL SET BASED ON THE VOLTAGE VECTOR REPRESENTATION OF SWITCHING STATE

Michal Kozubík

Doctoral Degree Programme (2), FEEC BUT E-mail: xkozub05@stud.feec.vutbr.cz

Supervised by: Pavel Václavek E-mail: vaclavek@feec.vutbr.cz

Abstract: This paper introduces novel approach in the implementing of finite control set model predictive control for the permanent magnet synchronous motor. This approach is based on the pregenerating of voltage vectors based on the specific switching state of voltage source inverter. The limitation of the pre-generating is based on the resolution of encoder used to measure mechanical angle of rotor. Proposed algorithm is described and tested in PIL simulation using Jetson Nano, for the control algorithm execution, and Simscape model of motor.

Keywords: finite control set, model predictive control, voltage vectors, optimization, general-purpose computing

1 INTRODUCTION

Finite control set (FCS) model predictive control (MPC) of the permanent magnet synchronous motor (PMSM) is based on the calculation of the optimal switching state of voltage source inverter (VSI) [3]. The computation of all possible combinations across the whole prediction horizon has huge computational demands. One of the aims of the research in the field is the acceleration of the computation. One of approaches is the reduction of the control set [2]. The approach described in this paper is based on the [1]. Further acceleration is done by using pre-generated vectors.

The paper is organized as follows. First chapter analyzes the problem of evaluating model with imbued switching states. Following by the second chapter, which introduces the approach of pregenerated voltage vectors. Third chapter describes the usage of introduced idea in algorithmic way. The last chapter discusses the simulation results of the applied algorithm.

2 ANALYSIS

The prediction of the future states of the controlled system is essential part of predictive control. The evaluation of the model is necessary for the correct prediction. During control of fast systems, such as PMSM, it is necessary to perform the evaluation as fast as possible.

The difference equations describing current in dq-frame - $i_d(k+1)$, $i_q(k+1)$ - are

$$i_{d}(k+1) = \left(1 - T_{s}\frac{R_{s}}{L_{d}}\right)i_{d}(k) + T_{s}P_{p}\frac{L_{q}}{L_{d}}\omega_{m}(k)i_{q}(k) + \frac{T_{s}}{L_{d}}\left(\sqrt{\frac{2}{3}}\cos\vartheta_{e}(k)\right)u_{A}(s(k)) + \frac{T_{s}}{L_{d}}\left(-\frac{1}{\sqrt{6}}\cos\vartheta_{e}(k) + \frac{1}{2}\sin\vartheta_{e}(k)\right)u_{B}(s(k)) + \frac{T_{s}}{L_{d}}\left(-\frac{1}{\sqrt{6}}\cos\vartheta_{e}(k) - \frac{1}{2}\sin\vartheta_{e}(k)\right)u_{C}(s(k))$$

$$(1)$$

$$i_q(k+1) = \left(1 - T_s \frac{R_s}{L_q}\right) i_q(k) - T_s P_p \frac{1}{L_q} \left(L_d i_d(k) - \Psi_{PM}\right) \omega_m(k) + \frac{T_s}{L_q} \left(-\sqrt{\frac{2}{3}} \cos \vartheta_e(k)\right) u_A(s(k)) + \frac{T_s}{L_d} \left(\frac{1}{\sqrt{6}} \sin \vartheta_e(k) + \frac{1}{2} \cos \vartheta_e(k)\right) u_B(s(k)) + \frac{T_s}{L_d} \left(\frac{1}{\sqrt{6}} \sin \vartheta_e(k) - \frac{1}{2} \sin \vartheta_e(k)\right) u_C(s(k))$$

$$(2)$$

where

i_d, i_q are stator current components in dq frame,	L_d, L_q are rotor inductance components,
ω_m is rotor mechanical angular speed,	P_p is number of pole pairs,
ϑ_e is rotor electrical angle,	Ψ_{PM} us permanent magnet flux,
u_A, u_B, u_C are phase voltages,	T_s is sampling period,
R_s is stator winding resistance,	s(k) is switching state of VSI.

The necessity of evaluating of functions sin and cos combined with computation of required multiplications in every step of prediction increases the computational demands of the optimization algorithm. This leads to the prolonging of computational time and makes the solving of optimization problem harder in real-time.

Therefore, requirements of the real-time computation lead to need for simplifications, such as vector representation described in following section.

3 VECTOR REPRESENTATION

The parts of equations (1) and (2) with phase voltages are representation of voltages $u_d(k)$ and $u_q(k)$ given by the information about switching state s(k) and electrical angle $\vartheta_e(k)$.

By the definition of the 2-level VSI, there exist two possible states on every phase - ON (1) or OFF (0). For three phases this makes total of eight combinations. Every combination can be transformed to the vector of voltage in dq-frame using Park and Clarke transformation

$$\begin{bmatrix} u_d(k) \\ u_q(k) \end{bmatrix} = U_{DC} \begin{bmatrix} \cos(\vartheta_e) & \sin(\vartheta_e) & 0 \\ -\sin(\vartheta_e) & \cos(\vartheta_e) & 0 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} s_A(k) \\ s_B(k) \\ s_C(k) \end{bmatrix},$$
(3)

where U_{DC} is supply voltage and s_A, s_B and s_C are switching states of respective phases.

The transformation shows the combinations $s(k) = [0 \ 0 \ 0]^T$ and $s(k) = [1 \ 1 \ 1]^T$ result in same voltage vector $u = [0 \ 0]^T$. This makes total of seven possible vectors of voltage. Mentioned vector remains zero for every possible electrical angle ϑ_e . Only six vectors remain dependent on the position of rotor.

These vectors are shown in the figure 1. The vectors are named according to the switching state, e.g. (100) - phase A on, phase B - off and phase C - off. The subscript denotes the rotor angle used for the vector generation.



Figure 1: Change of the voltage vector based on the electrical angle ϑ_e

The figure shows the change of the generated vectors for the specific angle ϑ_e . After the angle passes the multiple of 60 degrees, the switching combination generating given vector shifts.

4 ALGORITHM

Drawing on the information mentioned before, the finite control set algorithm can be developed. First of all, the proper finite control set must be defined. Control value of developed algorithm is switching states s(k). Model used in optimization problem uses mapping

$$s(k) \to \begin{bmatrix} u_d \left(\mathfrak{d}_e(k) \right) \\ u_q \left(\mathfrak{d}_e(k) \right) \end{bmatrix}.$$
(4)

Now, it is necessary to generate the vectors of voltage. Drawing on the information in the figure 1, the values must be pre-calculated for only one switching state.

The pre-generating of vectors comes with one complication. The real count of possible voltage vectors is infinite. To solve this issue, the parameters of practical realization are utilized. Every encoder used for the measurement of the rotor angle has finite resolution. In the developed algorithm, the number of pre-generated vectors is connected with the defined resolution of encoder.

For the resolution *r*, the r + 1 voltage vectors are pre-generated. Additional vector is zero vector. If the zero vector is chosen, the switching state $s(k) = [0 \ 0 \ 0]^T$ is selected.

The one of essential parts of the control algorithm is prediction. During this phase, all possible combinations of switching states are evaluated. Then the optimal combination is selected and switching state is applied to VSI.

For proper function of algorithm, it is necessary to define the way of finding out the currently used vectors. For the initial position $\vartheta_e = 0$, the position p of n-th used vector in pre-generated vectors is

$$p = \left\lfloor \frac{n \cdot r}{6} \right\rfloor. \tag{5}$$

During movement of rotor, shift l occurs. Its value is based on the difference between actual angular position and the multiple of 60 degrees d

$$l = \frac{d}{60}r.$$
 (6)

Thus, the position of used *n*-th vector is

$$p = \left\lfloor \frac{n \cdot r}{6} + \frac{d}{60}r \right\rfloor.$$
(7)

The optimization is performed with properly selected voltage vectors. It is necessary to transform the result back to switching state.

The result is the number *n* of given vector for the $\vartheta_e = 0$. Due to the rotation, it is necessary to perform shifting given by the multiplicity of actual position and 60 degrees *m*. The order is based on the order of vectors in the figure 1. The exception to this is the zero vector.

Principle of the transformation for the case of n = 1 and m = 2 is shown in the following figure.



Figure 2: Transformation of vector to switching state

5 RESULTS

Proposed algorithm was tested in the PIL simulation. Simscape model of PMSM and VSI was used for the simulation of their behavior. For the algorithm execution, Jetson Nano was used. The communication was established via UDP protocol. Table 1 shows the parameters of simulated PMSM.

Parameter	Value	
R_s	0.822 Ω	
L_d	0.016 H	
L_q	0.024 H	
Ψ_{PM}	0.097×10^{-3}	Wb
Pp	5	
J	0.870×10^{-3}	kg m ²
ω_r	$150 \text{rad} \text{s}^{-1}$	
T_R	7.275 Nm	
I_R	10 A	
U_{DC}	200 V	

Table 1: 1	Parameters	of PMSM
------------	------------	---------

The simulation run with the encoder with the resolution r = 4096. The experiment tested, whether the algorithm is able to track the reference angular speed. The derivative of the initial ramp was purposely chosen higher than the rated torque of the motor to test, whether is the algorithm able to keep the current within its limit. The Figure 3 a) shows the results of the reference tracking experiment. The results show algorithm was able to ensure the reference tracking. The 3 b) shows the value of current was kept within the limit of $I_R = 10$ A.

The results are dependent on the chosen period. The sampling period was selected based on the length of execution of algorithm. First part is the time required by the solution of optimization problem. The measurement has show the value around 10 μ s, which is 33 % improvement, when compared to the approach in [1]. The time required by the communication and data manipulation did not change. Thus, the sampling period was set to 100 μ s.



Figure 3: Results of the tracking reference experiment

6 CONCLUSION

This paper introduced the novel approach in the implementation of finite control set model predictive control. This approach is based on the pre-calculation of possible voltage vectors generated by the switching state of the voltage source inverter. The number of pre-calculated vectors is connected with the resolution of used encoder. The strategy of selecting specific vectors a backward transformation to the switching state is described.

The algorithm based on the introduce method was tested in PIL simulation. The experiment has confirmed the ability of algorithm to ensure the reference tracking and to keep the stator current within its limits.

ACKNOWLEDGMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- KOZUBIK, M., AND VACLAVEK, P. Speed control of pmsm with finite control set model predictive control using general-purpose computing on gpu. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society* (2020), IEEE, pp. 379–383.
- [2] PREINDL, M., AND BOLOGNANI, S. Model predictive direct torque control with finite control set for pmsm drive systems, part 2: Field weakening operation. *IEEE Transactions on Industrial Informatics* 9, 2 (2013), 648–657.
- [3] RODRIGUEZ, J., KAZMIERKOWSKI, M. P., ESPINOZA, J. R., ZANCHETTA, P., ABU-RUB, H., YOUNG, H. A., AND ROJAS, C. A. State of the art of finite control set model predictive control in power electronics. *IEEE Transactions on Industrial Informatics* 9, 2 (2013), 1003–1016.

A UNITY BASED INDUSTRIAL PROCESS SIMULATION CONTROLLED VIA VIRTUAL PLC

David Michalík

Doctoral Degree Programme (2nd year), FEEC BUT E-mail: david.michalik@vut.cz

> Supervised by: Petr Fiedler E-mail: fiedlerp@vutbr.cz

Abstract: This paper describes the development of an industrial process simulation application that is used for educational purposes. It is based on the Unity game engine and Game4Automation plugin, which offers a base framework and assets to establish the industrial process in a virtual environment. The application is then connected via Modbus TCP to a project created in the CODESYS programming environment, which is also responsible for the simulation of a control system - PLC.

Keywords: unity, game engine, industrial process simulation, virtual plc, education methods

1 INTRODUCTION

Game engines have increased their palette of functions since their creation in the 1990s. They offer a framework for game development, cinematic industry, education and even science. Nowadays, the two most widely used game engines, Unity and Unreal Engine 4, have found their way into the development of scientific applications, such as driving simulators. They offer a practically unlimited development potential of delivering a custom experience to the user. With the pandemic situation in the world as it is, online learning and software simulations have become one of the most used methods for education in universities. In this paper, a method to handle the situation of restricted access to school laboratories is proposed. In the Materials and Methods section, the Unity game engine is briefly introduced with its capabilities, together with its **Game4Automation** plugin by in2Sight GmbH that enables the simulation of industrial process and control. For the control IDE, the **CODESYS** is utilized, for it offers a free alternative of programming IDEs for industrial applications and enables the use of a simulated controller. These two vital components are financially available for students and therefore come with no license policy problems. In the results section, the overview of the developed application is presented.

2 MATERIALS AND METHODS

There are several key components that are used to complement the whole project. The Unity game engine, Game4Automation plugin and the CODESYS development environment.

2.1 UNITY GAME ENGINE

Unity is a game engine by Unity Technologies developed in 2005 and has since become one of the most widely used frameworks for developing 2D or 3D applications, be it by small indie developers or large companies. The applications in Unity3D are based on C# or JavaScript. The Unity3D editor offers a graphical user interface and enables the developers to easily access specific components, objects or assets and program/modify them according to the project's needs. This game engine comes with a proprietary licensing policy. As long as the projects developed are not used for commercial

purposes, Unity Technologies does not claim any revenue produced by the application. Also, the framework is available for educational and personal purposes without any fees [1].

One of the key advantages of the Unity game engine is the community of developers and available assets/plugins. They are easily accessible through the Asset Store and a developer can include these assets or plugins in a project. The pricing of these assets is individual. The main object of a developed application is the Scene. An opened scene can be observed in Figure 1. In this scene, GameObjects can be inserted and are handled through the hierarchy system that specifies parent-child relations to the objects present. GameObjects can be physical objects represented by a 3D model (these can be imported from many widely used 3D modelling programs, such as Autodesk, 3dsMax or Blender), or virtual objects that handle the behavior of some components or even objects (e.g. sky simulation, animations). Each of these GameObjects has several attributes and components, that can be modified or added to change the behavior of the said objects in the scene.

The editor comes with several advanced components and tools, such as the PhysX physics engine, which is a vital component for this paper, for it enables the simulation of collision events and therefore gives a user a simulated representation of real-world object behavior.



Figure 1: Unity3D Game Engine.

2.2 GAME4AUTOMATION PLUGIN

Game4Automation is a Unity3D plugin (compatible with Unity 2019.4) made by in2Sight GmbH and offers a solution for the simulation of industrial processes. It is a framework for developing a more simplified digital twin of a real project. It includes assets widely used in the industry, such as conveyor belts, manipulators, robots and handles their behavior via the control of several components. The main advantage of this plugin lies in the possibility to connect the simulation in Unity to a controller via communication standard, TCP-IP, Modbus or OPC UA for example.

The features of the plugin include [2]:

- CAD Link CAD Interface to Import CAD data based on STEP and 3MF format
- Scripts for Interfaces, Signals, Drives, Drive behavior models, Sensors, Picking, Sources, Sink
- TCP-IP Interfaces to Siemens Hardware like Sinumerik, Simatic S7-300, S7-400, S7-1200, S7-1500 and Simotion

- Modbus Master Interface
- OPC UA client interface
- Interface to RoboDK, a virtual industrial robot controller
- Game4Automation base component for easy Scene navigation with the mouse and saving of camera positions

And many more. Full documentation can be found on the developer's website [2]. The plugin comes with multiple licensing options, Starter and Professional, which differ in available features and are paid for, and the Education that is not available for commercial use.

2.3 CODESYS - VIRTUAL PLC SIM

For controlling the simulation, the CODESYS Development System IEC 61131-3 is used. It serves as a software platform for tasks in industrial automation technology. In our case, the main advantage is the option to simulate a control device. Therefore, students can operate the simulation at their homes without the requirement of having a physical device present. The CODESYS platform offers multiple programming options determined by IEC 61131-3, like FBD, LD, IL, ST or SFC. It also supports multiple industrial communication protocols, such as OPC-UA, PROFINET, EtherNet/IP or Modbus TCP, which is used in our case [3].

3 RESULTS

Application overview

An application based on Unity was developed to simulate an industrial process. As seen in Figure 2, there are two main parts to this application. The CODESYS project, which simulates a PLC with a user's control programme, and the Unity application, which serves as a physical simulation of the manufacturing process. The application gives a user visual information about the environment and also specific information about several sensors or devices through virtual variables, which are connected to the PLC sim through Modbus TCP.



Figure 2: Application diagram

The main goal of this simulation is to replicate an industrial process from the point of view of a system designer and a programmer. It serves an educational purpose to allow students to access physical

devices, such as PLCs, I/O devices or robotic arms. In this way, a proper control requirements of said process can be simulated on any desktop computer with the support of Unity and CODESYS.



Figure 3: Game4Automation Simulation Scene

Simulated industrial process

As seen in Figure 3, the factory scene consists of several components:

- 1 Small conveyor belt
- 2 Manipulator
- 3 Big conveyor belt
- 4 Robot manipulator
- 5 Box storage
- 6 Storage handler
- 7 Storage units

These components put together a process in which the product is stacked, handled and stored properly. It is up to the programmer to design a system to control all of these components appropriately. Unity's physical engine renders the physical reality of the world, so if the user does not handle these factors, the process shall fail. At the small conveyor belt, the Source object is present. This object is responsible for generating an item of the selected product. By default, the product is represented by the can model. User can change the model according to their needs. The important component of the conveyor belt is the Drive script, which handles the movement of the conveyor. A user can control the direction, target position, target speed and the initial start/stop commands through exposed variables. At the end of the small conveyor belt is a Sensory object. The range and cone of this sensor can be manipulated and it returns a boolean value based on the presence of the product.

The robot picks empty boxes from the box storage and puts them on the big conveyor belt. It has multiple positions set via the animation state machine that handles the conditions of state transitions. Each state represents a specific position of the robot's axes. A gripper is attached to the end of the robot. The gripper has a sensor attached with a Grip component that handles picking up the item detected via the sensor. In the next step, the user has to program the behavior of the Manipulator to stack the items inside the box. The Manipulator operates in two dimensions and has a grip attached to it (which works like the robot gripper). Similar to the conveyor belt, the Manipulator is operated via the Drive script. When the items are stacked properly, the big conveyor belt moves the box to the storage handler. This handler has a custom script that controls its movements in three dimensions, enabling a user to program it to move the box with stacked items to a previously specified storage unit position. For variety purposes in students' projects, this position is generated based on their IDs. If the control system operating all of the aforementioned components is set up properly, it should follow the process as explained.

CODESYS control

A list of variables that operate components implemented in the scene is presented to the students. Game4Automation plugin enables the creation of various data types, including bool, float or integer. These can be set to be either input or output variables. Through Modbus TCP, it is possible to map these variables in Unity to variables created in CODESYS project and control the objects via this protocol.

4 CONCLUSION

An application simulating an industrial process has been developed. It serves educational purposes for students to simulate the programmable environment of handling a PLC control and gives them feedback through the application developed in Unity. For that, the Game4Automation plugin is used, which offers a set of assets and functions to make it possible. In the future, the application will be extended with more means of controlling the objects and more scenes that represent different types of manufacturing processes.

5 ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- [1] P. Megha, L. Nachammai, and T. M. S. Ganesan, "3D game development using unity game engine", *International Journal of Scientific & Engineering Research*, vol. 9, no. 3, pp. 1353–1356, 2018.
- [2] *Game4automation documentation*, 2019. [Online]. Available: https://game4automation.com/documentation/current/index.html (visited on 10/03/2021).
- [3] Codesys online help. [Online]. Available: https://help.codesys.com/ (visited on 10/03/2021).

DERIVATION AND PRACTICAL COMPARISON OF RECURSIVE LS AND TLS SYSTEM IDENTIFICATION METHODS

Dominik Friml

Doctoral Degree Programme (1), FEEC BUT E-mail: Dominik.Friml@vutbr.cz

Supervised by: Pavel Václavek, Jakub Dokoupil E-mail: vaclavek@feec.vutbr.cz, dokoupilj@feec.vutbr.cz

Abstract: The least squares (LS) type of methods are the most widely used methods in system identification despite their obvious imperfection. Such methods use a regressor, that is supposed not to contain any error, notwithstanding that it is constructed from measured data. This can be solved by using the total least squares (TLS) type of methods. Derivation of both batch and recursive methods of LS and TLS for identification and their practical comparison is presented in this paper.

Keywords: error in variables, recursive least squares, recursive total least squares, akaike information criteria, bayes information criteria, identification, Rayleigh quotient, inverse iteration

1 INTRODUCTION

The problem with error in variables is shown to be inherent for all system identification processes [?]. The least squares type of methods are thus not generally applicable and should be used with caution [?]. On the other hand total least squares type of identification methods does not suffer from this drawback. This article will describe batch and recursive system identification methods using LS and TLS. Derived algorithms are implemented in hardware in the loop system and the identified models are compared based on Akaike and Bayes information criteria [?,?].

2 IDENTIFIED SYSTEM

Let us consider a linear system described by

$$\overline{\mathbf{y}} = \overline{\mathbf{h}}^T \mathbf{\theta},\tag{1}$$

where $\theta \in \mathbb{R}^n$ is the column vector of unknown and identified parameters, $\overline{\mathbf{h}} \in \mathbb{R}^n$ is regression vector and $\overline{y} \in \mathbb{R}$ is system output scalar value.

Since this linear system can be used to describe linear differential equation of $\frac{n}{2}$ -th order, it is suitable as model of real linear, time invariant, SISO system.

$$\overline{\mathbf{y}}(t) = \overline{\mathbf{h}}(t)^T \mathbf{\theta} = \overline{u}(t-1)b_1 + \overline{u}(t-2)b_2 + \dots + \overline{u}(t-n)b_n + \mathbf{y}(t-1)a_1 + \overline{\mathbf{y}}(t-2)a_2 + \dots + \overline{\mathbf{y}}(t-n)a_n,$$
(2)

where $u(t - \frac{n}{2})$ is *t*-th input value delayed by $\frac{n}{2}$ sample periods. In such case, following substitutions are assumed

$$\overline{\mathbf{h}}(t) = \left[\overline{u}(t-1), \overline{u}(t-2), \dots, \overline{u}(t-\frac{n}{2}), \overline{y}(t-1), \overline{y}(t-2), \dots, \overline{y}(t-\frac{n}{2})\right]^{T}$$

$$\boldsymbol{\theta} = \left[b_{1}, b_{2}, \dots, b_{n}, a_{1}, a_{2}, \dots, a_{n}\right]^{T}$$
(3)

The target of identification is to find such parameters θ , that the difference between model output and measured data is in some sense minimal.

3 LEAST SQUARES IDENTIFICATION

Let us assume, that measured output of system $\tilde{y}(t)$ consists of true output value $\bar{y}(t)$ and error $e_y(t)$. Let us further assume, that the error is random and has character of normally distributed white noise with zero mean and unknown variance σ_y . Measured data for least squares (LS) are thus expected to be generated by following model

$$\widetilde{y}(t) = \overline{y}(t) + e_y(t) = \overline{\mathbf{h}}(t)^T \mathbf{\theta} + e_y(t) \quad ; \ e_y \sim \mathcal{N}(0, \mathbf{\sigma}_y) \tag{4}$$

Batch of measured data can be described by

$$\widetilde{\mathbf{y}} = \overline{\mathbf{H}}\boldsymbol{\theta} + \boldsymbol{e}_{\mathbf{y}}(t),\tag{5}$$

where $[\tilde{y}(t), \tilde{y}(t+1), \ldots, \tilde{y}(t+k)]^T = \tilde{\mathbf{y}} \in \mathbb{R}^k$ and $[\bar{\mathbf{h}}(t), \bar{\mathbf{h}}(t+1), \ldots, \bar{\mathbf{h}}(t+k)]^T = \bar{\mathbf{H}} \in \mathbb{R}^{k \times n}$ is vector of measured output data and matrix consisting of known regressors, respectivel. *k* is number of measurements. This equation in general defines overdetermined system of equations. Least squares identification tries to find such $\hat{\theta}$, so that the square of the difference between measured outputs $\tilde{\mathbf{y}}$ and $\bar{\mathbf{H}}\hat{\theta}$ is minimal.

$$\min_{\hat{\theta}} J(\hat{\theta}) = \min_{\hat{\theta}} \| \widetilde{\mathbf{y}} - \overline{\mathbf{H}} \hat{\theta} \|_2$$
(6)

Minima of optimization function $J(\hat{\theta})$ can be shown to be in,

$$\hat{\boldsymbol{\theta}} = \left(\overline{\mathbf{H}}^T \overline{\mathbf{H}}\right)^{-1} \overline{\mathbf{H}}^T \widetilde{\mathbf{y}}.$$
(7)

4 TOTAL LEAST SQUARES IDENTIFICATION

Total least squares identification can be explained as an extension of the least squares identification, where the noise, or more generally error, is expected not only on the output variable but also on the regressor.

This consideration is more fitting to our case, because regressor $\overline{\mathbf{h}}$ in equation (??) consist of the delayed measured data $\tilde{y}(t)$ as shown in equation (??). It also consists of input data $\tilde{u}(t)$, which can be also expected to be corrupted with error, for example due to instrumentation or human error.

For this reasons, measured data for total least squares (TLS) are expected to be generated by

$$\widetilde{y}(t) = \widetilde{\mathbf{h}}(t)^T \mathbf{\theta} + e_y(t) = \left(\overline{\mathbf{h}}(t) + \mathbf{e}_h(t)\right)^T \mathbf{\theta} + e_y(t) \quad ; \ e_y \sim \mathcal{N}(0, \sigma_y) \quad ; \ \mathbf{e}_h \sim \mathcal{N}(0, \Sigma_h), \tag{8}$$

where $\mathbf{\tilde{h}}(t)$ is regression vector $\mathbf{\bar{h}}$ burdened by noise $\mathbf{e}_h \in \mathbb{R}^n$. This noise is vector of normally distributed white noise variables with zero mean. Variance matrix Σ_h will be in this article considered to be diagonal but otherwise unknown. Batch of measured data can be described by

$$\widetilde{\mathbf{y}} = \mathbf{H}\mathbf{\theta} + e_{\mathbf{y}}(t) \tag{9}$$

This equation again defines in general overdetermined system. It was proved in [?], that LS optimization criteria can be generalized to

$$J(\hat{\theta}) = \min_{\hat{\theta}} \frac{\left[\hat{\theta}^{T}, -1\right] \left(\left[\widetilde{\mathbf{H}}, \widetilde{\mathbf{y}}\right] \left[\widetilde{\mathbf{H}}, \widetilde{\mathbf{y}}\right]^{T}\right) \left[\hat{\theta}^{T}, -1\right]^{T}}{\left[\hat{\theta}^{T}, -1\right] \mathbf{D} \left[\hat{\theta}^{T}, -1\right]^{T}} = \min_{\hat{\theta}} \frac{\Theta \left(\Phi^{T} \Phi\right) \Theta^{T}}{\Theta \mathbf{D} \Theta^{T}},$$
(10)

which for diagonal $\mathbf{D} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \in \mathbb{R}^{n+1 \times n+1}$ is equal to equation (??) and for $\mathbf{D} = \mathbf{I}_{n+1}$ defines optimization criteria for TLS.

The minimum of this criteria is, as suggested by the similarity of equation (??) to Reyleigh quotient, first *n* elements of the eigenvector corresponding to the smallest eigenvalue of the Φ normalized such that the last element \underline{v}_{n+1} is equal to -1.

$$\Phi = \mathbf{U}\Sigma\mathbf{V}^T = \mathbf{U}\Sigma[\overline{\mathbf{v}}, \mathbf{v}_2, \dots, \mathbf{v}_n, \underline{\mathbf{v}}]^T$$
(11)

$$\underline{\mathbf{v}} = \begin{bmatrix} \underline{v}_1, \underline{v}_2, \dots, \underline{v}_{n+1} \end{bmatrix}^T$$
(12)

$$\hat{\boldsymbol{\theta}} = \frac{-\left[\underline{\nu}_1, \underline{\nu}_2, \dots, \underline{\nu}_n\right]^T}{\underline{\nu}_{n+1}} \tag{13}$$

5 RECURSIVE LEAST SQUARES

The recursive least squares formulas are simple to derive from equation (??). For detailed description refer to [?]. Derived formulas are directly usable as recursive algorithm

$$\mathbf{P}(t) = \mathbf{P}(t-1) - \frac{\mathbf{P}(t-1)\left(\overline{\mathbf{h}}(t)\overline{\mathbf{h}}(t)^{T}\right)\mathbf{P}(t-1)}{1+\overline{\mathbf{h}}(t)^{T}\mathbf{P}(t-1)\overline{\mathbf{h}}(t)}$$
(14)

$$\mathbf{k}(t) = \mathbf{P}(t)\overline{\mathbf{h}}(t) \tag{15}$$

$$\boldsymbol{\varepsilon}(t) = \widetilde{\boldsymbol{y}}(t) - \overline{\mathbf{h}}(t)^T \hat{\boldsymbol{\theta}}$$
(16)

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t) + \mathbf{k}(t)\boldsymbol{\varepsilon}(t), \tag{17}$$

where $\varepsilon(t)$ is prediction error and K(t) is gain factor. Although this algorithm can be further optimized for less arithmetic operations, this form is best for comparison with recursive TLS algorithm described in next section.

6 RECURSIVE TOTAL LEAST SQUARES

In contrast with RLS, the RTLS is not directly derivable from equations (??)-(??), because of the dependency on SVD. In order to derive recursive total least squares, one can find a way to compute TLS without SVD. This is not a major issue, since SVD is only used to find eigenvector corresponding to the smallest eigenvalue, which can be computed in many different ways. One of such methods is inverse iteration [?]

$$\mathbf{v}(t) = \left(\left[\widetilde{\mathbf{H}}, \widetilde{\mathbf{y}} \right] \left[\widetilde{\mathbf{H}}, \widetilde{\mathbf{y}} \right]^T \right)^{-1} \underline{\mathbf{v}}(t-1) = \mathbf{P}\underline{\mathbf{v}}(t-1)$$
(18)

$$\underline{\mathbf{v}}(t) = \frac{-\mathbf{v}(t)}{v_{n+1}(t)},\tag{19}$$

which is iterative by definition. Although this method computes $\underline{\mathbf{v}}$ for given \mathbf{P} , there are no drawbacks to update \mathbf{P} with new data $\tilde{y}(t)$, $\tilde{\mathbf{h}}(t)$ after each iteration, resulting in recursive algorithm. This alteration combined with extraction of $\hat{\boldsymbol{\theta}}(t)$ from $\underline{\mathbf{v}}(t)$ results in RTLS algorithm.

$$\underline{\mathbf{v}}(t-1) = \left[\hat{\boldsymbol{\theta}}(t-1)^T, -1\right]^T \tag{20}$$

$$\mathbf{P}(t) = \mathbf{P}(t-1) - \frac{\mathbf{P}(t-1) \left(\left[\widetilde{\mathbf{h}}(t)^T, \widetilde{\mathbf{y}}(t) \right]^T \left[\widetilde{\mathbf{h}}(t)^T, \widetilde{\mathbf{y}}(t) \right] \right) \mathbf{P}(t-1)}{1 + \left[\widetilde{\mathbf{h}}(t)^T, \widetilde{\mathbf{y}}(t) \right] \mathbf{P}(t-1) \left[\widetilde{\mathbf{h}}(t)^T, \widetilde{\mathbf{y}}(t) \right]^T}$$
(21)

$$\mathbf{v}(t) = \mathbf{P}(t)\underline{\mathbf{v}}(t-1)$$
(22)

$$\hat{\theta}(t) = \frac{-[v_1(t), v_2(t), \dots, v_n(t)]^T}{v_{n+1}(t)}$$
(23)

7 VALIDATION

Both methods has to be validated and the results compared. To avoid comparison bias by author, Akaike and Bayes information criteria has been selected as cross-validation method

$$AIC = 2n - 2\ln(\hat{L}) \tag{24}$$

$$BIC = n\ln(k) - 2\ln(\hat{L}), \tag{25}$$

where *n* is number of identified parameters, *k* is number of observations and $\hat{L} = p(\tilde{\mathbf{y}}|\hat{\theta}, M)$ is likelihood of the model *M*. Since data is generated by the same model, comparison is also based on the same model. Model from equation (??) has been selected. Log-likelihood is computed as

$$\ln(\hat{L}) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma_y) - \frac{1}{2\sigma_y^2} \sum_{i=1}^k \left(\overline{\mathbf{h}}(i)^T \hat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{y}}(i)\right)^2.$$
(26)

Since log-likelihood formula requires knowledge of variance σ_y , only proportional part $-\sum_{i=1}^{k} \left(\overline{\mathbf{h}}(i)^T \hat{\mathbf{\theta}} - \widetilde{y}(i) \right)^2$ of the formula is used, which has no effect on comparison of the models.

8 EXPERIMENT

Both methods are implemented and tested on hardware in the loop configuration. The computational part is implemented in MATLAB Simulink, inputs and outputs are provided by X20 CP 0484 B&R PLC with 13bit input and output expansion cards. Identified hardware is a configurable RLC array configured to a stable, underdamped linear system.

The experiment consists of four steps. The first step is data acquisition, the second step is data separation, the third step is the identification and the fourth is cross-validation of identified models.

Data were acquired with a 1ms sample rate, the input signal is PRBS. A combination of the input and output quantization effect and nonlinearities of the RLC array is expected to introduce enough noise to the data.

Acquired data are resampled to 20ms, 50ms, and 100ms, and each of the data sets is identified separately. The first 3s of acquired data, shown in figure 2(a), containing part of one transition, is used to find prior $\hat{\theta}_0$ and \mathbf{P}_0 using a noniterative algorithm. This assures, that none of the compared methods is penalized by prior assumptions. The next 150s of the data is used for iterative identification. The remaining 500s of the measured data are used for cross-validation.



Figure 1: Configuration of experiment workplace. The PLC is in the upper part of the picture, and the RLC array is in the lower portion.

9 RESULTS

As noticeable from figure 2(a), both methods resulted in a model satisfactory fitting to measured data. Plotting $\hat{\theta}(t)$ also proved, that RTLS converges faster. This is visualized in 2(b). In theory, the result of TLS identification should be more resilient to input and output noise, which should be proved by AIC and BIC. As apparent from **??**, this is proved to be true, since all of the information criteria are smaller for TLS compared to LS.



Figure 2(a): Fit of identified models on first 3s of **Figure 2(b):** Convergence of 4th element of $\hat{\theta}(t)$ for data for 100ms sample rate.

T[ms]	$AIC_{LS}[-]$	$AIC_{TLS}[-]$	$BIC_{LS}[-]$	$BIC_{TLS}[-]$
100	10.45	10.399	49.731	49.681
50	5.4134	-5.5867	48.856	37.856
20	10.676	7.563	59.617	56.504
10	9.957	9.9184	63.057	63.019

Table 1: Cross-validation results using Akaike and Bayes information criteria.

10 CONCLUSION

Error in variables is a drawback of least squares type of identification methods, which according to theory and presented experiment leads to worse identified model. This is confirmed by both AIC and BIC of identified system, as presented in table **??**. This table is the result of a practical experiment implementing derived batch and recursive identification algorithms in hardware in the loop configuration.

Further research could be dedicated to deriving likelihood function for model defined in equation (??), deriving less computation heavy algorithms or generalizing TLS identification method to models with specific nonlinearities.

ACKNOWLEDGMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

[1] Sabine Van Huffel and Joos Vandewalle. *The total least squares problem: computational aspects and analysis.* SIAM, 1991.

- [2] T Söderström and P Stoica. System Identification. Prentice-Hall, Inc., USA, 1988.
- [3] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.
- [4] RITEI SHIBATA. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, 63(1):117–126, 1976.
- [5] Carlos E. Davila. An Efficient Recursive Total Least Squares Algorithm for FIR Adaptive Filtering. *IEEE Transactions on Signal Processing*, 42(2):268–280, 1994.
- [6] Stephan Rhode, Felix Bleimund, and Frank Gauterin. Recursive Generalized Total Least Squares with Noise Covariance Estimation. Technical report, 2014.

IN-BED POSTURE CLASSIFICATION

Michal Husák

Doctoral Degree Programme (1), FEEC BUT E-mail: xhusak08@stud.feec.vutbr.cz

Supervised by: Václav Kaczmarczyk

E-mail: kaczmarczyk@feec.vutr.cz

Abstract: The growing trend of the population age contributes to the accumulation of patients in social facilities and in-home care, which leads to growing chronic diseases. Modern systems try to improve the effectiveness of health care interventions. Our work aims to create a widely applicable platform that combines the measurement of in-bed position with another's negative states. All these physical influences are mainly the cause of chronic tissue damage (pressure ulcers). Processing of the pressure distribution on the bed is a more dimension problem. The mentioned data are multimodal. Therefore, we used the machine learning (ML) method to obtain the properties.

Keywords: Decubitus, Matrass, Machine Learning, Body posture classification

1 INTRODUCTION

The thesis aims to create a simple and reliable classification system for monitoring the condition of the patient who is lying on the bed. The main reason for the research was the requirement of the modernization and automation medical or rehabilitation devices. IoT devices and modern data processing methods are the only option for maintaining a high quality of life for patients who are prone to pressure ulcer development. In developed countries with an aging population, it is necessary to work on the care of these patients. The number and age of patients in need of after-care facilities are increasing. The high standards that are placed on this care require based on large staff resources. Moreover, the qualification of new paramedics is expensive and requires a lot of time. The lack of paramedics who can care about such aging patients increases. It affects the quality and availability of care provided, forcing hospitals to close parts of their wards, overburdened and exhausted nurses cannot perform all the procedures that would be necessary for patients. Some foreign studies even show that the lack of paramedics increases the likelihood of complications and death of patients. [1] This fact is evident in crises such as today's pandemic situation. Therefore, more and more attention are being paid to technical solutions in these areas of human activity. The overall trend of innovation in health care is desirable for society and necessary in the future.

However, as result of these trends, we must also ask ethical questions. After all, human contact and care are irreplaceable, and no machine can replace them. Despite the shortcomings, other than technical solutions do not yet appear relevant. Pressure ulcers are a major problem in bedridden care. There are several causes of development, and the only effective method of their treatment is prevention. Prevention of the emergence of pressure sores is based on the general experience of nursing and medical practice. It consists of the repetitive activity of paramedics and depends on the not always telling scales. These actions seek to optimize so-called smart sensors, which should be capable of early warnings or even intervention. This work's result is the classification process of sensitizing systems designed for both institutional and home care. The benefit of the work will be optimizing nursing tasks and minimizing the risks of complicating conditions. Another important aspect of this work is how measured data are processed and classified. The result of the pressure image classification will be the input for the data fusion and analysis of the overall state of the monitored entity.

2 PRESSURE ULCERS

Pressure ulcers, also known as decubitus, are a high-priority problem in everyday nursing. This is local damage of the skin, tissue, and at the worst stage, bone structure degradation. The cause of the formation of decubitus is high and stable pressure on the tissue in the areas of bone growth. Among other described side effects are increased humidity or temperature. Significantly affected points include the coccyx, heel, tulle, or scapula area when positioned on the back. The damage has the character of an ulcer with severe inflammation. In case of neglect, there is also a risk of blood poisoning and death. [2]

However, the problem is not only the prerogative of institutional care. In in-home care, the risk depends on the nurse's experience. Decubitus is not only the prerogative of gerontic patients, but younger patients who suffer from chronic pain are also at risk. The consequences can be tragic, and the solution is only an invasive procedure with long-term regeneration. Recovery is painful and time-consuming, so the best solution to a degrading state is to overcome prevention. The overall severity of the decubitus is also the reason for the emergence of therapeutic products that reduce or eliminate the risk of developing it completely. [2]

2.1 **PREVENTION**

The scientific literature lists the procedures and scales of assessment of the occurrence and damage already incurred. Pressure ulcers are classified into four categories (phases). Each phase represents physiological changes in skin, tissue, and bones in areas of adverse effects. [3]

Risk Assessment Scales for Pressure Ulcers Prevention (RAPs) have been introduced as a preventive risk assessment in patients. However, this is an objective evaluation, and there is no consensus among the professional public about the correctness of the individual criteria. All these scales are based on the point assessment of the physical and mental state of the evaluated patient. The result is an overall rating and a scale cut-off of the risk score. We can select three basic scales from the list of the most used scales. [4]

- Braden Scale
- Norton Scale
- Waterlow Scale

Depending on this assessment, the intervals of the treatment are individual. Such a system is highly dependent on the objectivity of its assessors [4]. It is a knowledge-based (vessel) system that could be integrated into the surveillance system. By eliminating the ambiguity of the evaluation, it is possible to optimize the physical actions of paramedics while also increasing the patient's safety.

3 DATA

The subject of this article is not the discussion of data collection, but its processing. However, it is necessary to show what raw data will look like. In addition to external sensors, the most important is the pressure image, which is universally scalable according to the resolution of the pressure image. The entire classification framework is created and tested in the *MATLAB R2020a development environment*. An image resolution data model (36 x 11) pressure point matrix was created. The pressure point is represented as an 8-bit unsigned integer. The image is colour-converted into a heat map, where the greatest pressure is represented by a dark red to black and areas with low or zero pressure are represented by white followed by yellow.

3.1 DATA SIMULATION

A dataset of three test patterns was created for testing in the initial phase. The use of simulated data for learning methods is not yet foreseeable. Data is designed to debug simple decision structures. A real dataset will be created after the sensory part is put into operation. The first test patterns are:

- Position on your back
- Position on your stomach
- Position on the left side

By transforming the columns of the pressure image horizontally, the data are complemented by a fourth position (position on the right side). By augmentation of the original images, the simulated dataset is extended by other images. The pressure matrix is pre-processed by trash hold filtering, and for presentation, values between points are interpolated in (Figure 1).



Figure 1: Heat map projection

4 FEATERUE EXTRACTION

In this chapter, we will deal with the design of pressure image processing. The features detected in the image are forwarded to the master application, which will be the output interface of the system.

4.1 HARD-FEATURES

Properties that can be precisely, mathematically defined from the input pressure image. They do not require a learning process and are not dependent on the training set. These methods are more robust and based on the physical nature of the measured data. The use of the computing capacity of a marginal facility is nowadays very desirable regarding the ecological footprint of the facility. One possible solution in the future appears to be edge devices in standalone working mode.

Energy

Kinetic energy is expressed as the sum of the maximum values of dynamic load change around the pressure point. The result is a time-consuming quantity that represents the patient's activity in the bed. This feature is crucial for us due to the further processing of the pressure image. If the patient is in motion (changing his position), it is not appropriate to classify his position. The activity score (AS) defines it as: [7]

$$AS(t) = \sum Max\{P_i(t), P_i(t-1)\} \cdot \left(\frac{\Delta P_i}{\Delta t}\right)^2$$
(1)

where P_i represents the *i*th pressure point at time *t*. [7]

Histogram projection

Horizontal and vertical projection of pressure image values is an effective way to obtain information about the pressure image. Each pressure point value is added up in the vertical and then horizontal direction. As can be seen in (Figure 1) the result is further used to find out extra information about the position of the center of gravity. If the center of gravity in the x-axis occurs outside the specified operational boundaries, a request for intervention arises. The subject could fall out of bed. Another use of the center of gravity is the segmentation of the pressure image. The gravity center position divides the pressure image into four sub-images. The upper and lower parts are secondarily divided to produce six partial images. These are then used to analyse Body Symmetry Diagnostic (BSD). [5]

4.2 SOFT-FEATURES

Generally, there are methods based on machine learning and artificial intelligence carried out over a larger data volume. The results depend on the quality of the teaching set. An example of such a feature is the recognition of gestures, objects, or the position of a lying figure. In the meantime, we expect to teach the classifier from the first test data. Our next step is a test with real data in real conditions. The accuracy of classification methods can range from 80% to 90% according to the specified method and the quality of learning. [6][7]

5 CLASIFICATION

The raw data after pre-processing are usable only for the presentation of the heat map, but they are not sufficiently descriptive for posture recognition. The task of recognizing the position from this type of data correlates with recognizing the written text. This technique so-called optical character recognition (OCR). In such a case, OCR techniques are used to obtain image properties, which also contain machine learning methods. Some methods could be categorized as pre-processing because they only transform the input space and transfer it clarified to the classifier. A relatively direct method is the use of convolutional neural networks. Convolutional neural networks (CNN) or deep networks also hide transforming convolutional layers. However, the disadvantage is learning methods, rely on a large dataset. Standard CNNs can still converge to similar solutions in OCR, but we do not have many options to moderate them. OCR methods such as histogram-oriented gradient (HOG) serve not only as pre-processing for extracting image properties but also as a sub-scaling method that reduces the size of the input. We can reduce the classifier, such as a standard neural network (NN), or other machine learning methods like (random forest, k-NN, other) [6]. The adopted method deals with segmentation by region of interest (ROI). Fuzzy C-mean clustering, which is based on standard k-means clustering. For example, in an article Hung [7], successfully used this method in combination with a neural classifier. Subsequently, the authors improved it to solve a low-cost infrared sensor. Even with low resolution and two-state output of the measuring point, they were able to achieve sufficient feature separability.



Figure 2: Two models of classification chain

Figure 2 shows two classification strings. The top string provides for the use of OCR methods in combination with FCM and SLA. The resulting properties are separated by a standard neural network. The second string uses symmetric analysis. The properties of the symmetry are separated by the decision tree. The output is a classification vector of four positions, mentioned in chapter (3.1).

6 CONCLUSION

This work has presented the design of on-bed activity classification. Nowadays trend in health care shows the very high priority of technic solutions for complex health condition diagnostic. Two classification methods for posture recognition based on machine learning were selected. V virtual dataset with four main positions was made.

Further work will be focus on experimental testing and system integration. The same performance and stability test will be performed with real data. The classification vector will be extended with new sleeping positions after the real data come. The case study demonstrated the potential and practicability for pressure ulcer prevention and evaluation of fall risk. Another possible usage of application is the detection of sleep apnoea or somnambulism.

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 – "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- AIKEN, Linda H, Douglas M SLOANE, Luk BRUYNEEL, et al. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. *The Lancet*. 2014, **383**(9931), 1824-1830. ISSN 01406736. Available from: doi:10.1016/S0140-6736(13)62631-8
- [2] MIKŠOVÁ, Zdeňka. *Chapters from nursing care*. Praha: Grada, 2006. Nurse (Grada). ISBN 80-247-1442-6.
- [3] Characteristics of bedsores. In: *Dekubity.eu* [online]. Brno: National Center for Nursing and Non-Medical Health Professions, 2020. Available from: https://www.dekubity.eu/informace-pro-verejnost/charakteristika-prolezenin/
- [4] MANDYSOVÁ, Petra, Jana PECHOVÁ a Ehler EDVARD. Using the Braden scale for the prediction of pressure sore risk: inter-rater reliability. *Central European Journal of Nursing and Midwifery*. 2013, **4**(3), 609-613. ISSN 1804-2740.
- [5] LIU, J.J., Wenyao XU, Ming-Chun HUANG, Nabil ALSHURAFA, M. SARRAFZADEH, N. RAUT a B. YADEGAR. A dense pressure sensitive bedsheet design for unobtrusive sleep posture monitoring. 2013/03/01, 207-215. ISBN 978-1-4673-4573-6. Available from: doi:10.1109/PerCom.2013.6526734
- [6] XU, X., F. LIN, A. WANG, C. SONG, Y. HU a W. XU. On-bed sleep posture recognition based on body-earth mover's distance. In: 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS). 2015, 1-4. Available from: doi:10.1109/BioCAS.2015.7348281
- [7] HUNG, Yu-Wei, Yu-Hsien CHIU, Yeong-Chin JOU, Wei-Hao CHEN a Kuo-Sheng CHENG. Bed posture classification based on artificial neural network using fuzzy c-means and latent semantic analysis. *Journal of the Chinese Institute of Engineers* [online]. 2014, 38(4), 415-425. ISSN 0253-3839. Available from: doi:10.1080/02533839.2014.981212

FEATURE SPACE REDUCTION AS DATA PREPROCESSING FOR THE ANOMALY DETECTION

Simon Bilik

Doctoral Degree Programme (2), FEEC BUT E-mail: bilik@feec.vutbr.cz

> Supervised by: Karel Horak E-mail: horak@feec.vutbr.cz

Abstract: In this paper, we present two pipelines in order to reduce the feature space for anomaly detection using the One Class SVM. As a first stage of both pipelines, we compare the performance of three convolutional autoencoders. We use the PCA method together with t-SNE as the first pipeline and the reconstruction errors based method as the second. Both methods have potential for the anomaly detection, but the reconstruction error metrics prove to be more robust for this task. We show that the convolutional autoencoder architecture doesn't have a significant effect for this task and we prove the potential of our approach on the real world dataset.

Keywords: Anomaly detection, Convolutional autoencoder, PCA, t-SNE, CNN, OC-SVM

1 INTRODUCTION

Anomaly detection is a state-of-the-art task focused on the outlying samples recognition. This task is challenging especially because of the large possible variety of the anomalous samples, which are usually not present in the required amount or even not at all in the training data. The other problem, especially with the image data, is the high dimensionality of those samples, where the most of those features are not crucial for the anomaly detection itself. All of this makes the anomaly detectors difficult to use in a proper manner.

For this reason, we investigate several methods, such as convolutional autoencoders (CAE), principal component analysis (PCA) with the t-distributed stochastic neighbor embedding (t-SNE) and the reconstruction metrics in order to reduce the input data dimensionality. To the best of our knowledge, we are the first to examine the feature space created from the error metrics. We suggest the use of the One Class Support Vector Machine (OC-SVM) to automatically process the feature space.

2 RELATED RESEARCH

The papers [1] and [2] define the backgrounds of the One Class SVM. It defines this problem as the SVM model applied on the unlabeled data and it describes the algorithm and its optimization in detail. The implementation of this method can be found besides others in the *Novelty and Outlier detection* scikit-learn Python package. This method is quite efficient, but it often fails on the high dimensional data, as it is described in [3]. Despite this factor, it is usually used only as a classification layer for a reduced feature space, rather than for straight use with image data.

The article [4] describes the CAE architecture principles. The authors describe general applications of the autoencoders, as image denoising, or feature space reduction. They also compare CAEs with fully connected autoencoders and describe the advantages of CAEs. In this article, authors also provide a mathematical background of the CAE and show the importance of the max-pooling layer for the learning consistency. A possibility of initializing the classification CNN with the same topology CAE is explained including its benefits.

The following paper [5] compares performance of the fully connected AEs, denoising AEs and the Gaussian kernel PCA on the artificial multidimensional data and satellite's telemetry data. The authors prove that the AEs architectures can learn to represent normal data and that the reconstruction error could be used as an anomaly score even in the multidimensional data. The AEs based methods clearly outperform the PCA in this research.

The work [6] from the company MVTec presents a CAE based defect segmentation method. The authors compare several reconstruction loss metrics and structure similarity (SSIM) as L2 metric together with variational AE and feature matching AE architectures. The SSIM metric is concerned to bring the best results. The main advantage of this method is that not only the single pixel values are compared, but also the luminance, contrast and other parameters are reflected. The authors also provide an overview of the state-of-the-art CAE based anomaly detection methods and a description of the metrics mentioned above.

The article [7] shows a new training principle in order to extract robust features with the denoising AEs. The main idea of the presented approach is that a noise-damaged sample is processed by the AE as an input and the original sample is considered as a label. For the input corruption, the authors set a fixed number of randomly selected input elements to zero, but they suggest a use of other corrupting noise. This approach should help the AE to select stronger features and to better generalize the inputs.

3 MATERIALS AND METHODS

3.1 DATASET DESCRIPTION

For this experiment, we used the *Cookie Dataset*. It is our internal dataset designed for the anomaly detection task, which captures *Tarallini* biscuits. It contains 1225 samples in four classes with the following structure:

- No defect (474 captions)
- Defect: not complete (465 captions)
- Defect: strange object (158 captions)
- Defect: color defect (128 captions)

All classes can be seen in fig. 1. In order to augmentate the dataset we rotated each sample by 90° for three times, so the augmented dataset contains 4900 captures in total. Augmented images were cropped to the size of their bounding boxes, the source samples remained uncropped in the original dataset.



Figure 1: *Cookie Dataset* class samples

From this data set we created a subset for the anomaly detection, where we consider only the samples within specification (OK class) and randomly selected faulty samples from all faulty classes (NOK

class). Our training dataset contains 1000 OK samples and no NOK samples as the training set, 2000 OK samples and no NOK samples as the validation set and with 200 OK samples and 200 NOK samples as the test set. All images were cropped to their bounding boxes. The sample ratio of the NOK classes in the test set was set as 0,4 : 0,3 : 0,3 (not complete : strange object : color defect).

3.2 EXPERIMENT DESCRIPTION

We use CAE as a first stage of our pipeline. For this purpose, we compare the feature extraction quality of a basic CAE implementation (BAE1) with six convolutional layers¹, our own simplified version of the implementation mentioned above with four convolutional layers (BAE2) and the architecture suggested by MVTec [6] with sixteen convolutional layers.²

As the dimensionality of the input samples is still too high for the anomaly detection task, we compare two methods for the feature space dimensionality reduction. As the first method, we reduce the number of input features to 50 with the PCA algorithm and then to 2 with the t-SNE method as suggested in [9]. The second method we used was the reconstruction error metrics based feature space. For this purpose, we chose to use the L2 metric together with the SSIM suggested in [6] in order to create a two dimensional feature space.

For the PCA, t-SNE and OC-SVM methods, we used the implementations available from the *scikit-learn* Python library [8]. All computations were performed on the Nvidia GTX2080 graphic card. The size of the input images was 256x256 px and the batch size of the CAEs was set to five for training. Scripts used for the experiments were written in Python 3.6 and the Tensorflow 2.4.1 library was used to train the CAEs.

3.3 EXPERIMENT RESULTS

The examined CAE architectures showed similar performances in the image reconstruction and a slightly different results in the image encoding, as can be seen in fig. 2. The best input dimensionality reduction was achieved with the MVTec autoencoder. We tried to use both encoded and decoded samples as a direct input to the OC-SVM classifier, but the similarity of the samples was still too high and this method did not bring us any good results.



Figure 2: Visualisation of the CAE encoded inputs

¹https://blog.keras.io/building-autoencoders-in-keras.html

²https://github.com/cheapthrillandwine/Improving_Unsupervised_Defect_ Segmentation
Results from the PCA and t-SNE based dimensionality reduction can be seen in the left part of fig. 3. We can see, that this method is able to separate one class of the NOK samples (probably the color defects), but the other NOK classes blend with the OK class due to their structural and shape similarity. The best results were obtained with the MVTec CAE, which has the lowest overlap of the OK and NOK class, although the OK class doesn't create a closed cluster. The disadvantage of this approach is that the t-SNE method has a stochastic behaviour and the feature space slightly differs after each run. We can imagine that this method could be used in case of more different OK and NOK classes.



Figure 3: Comparison of the resulting feature spaces

The feature space based on the L2 and SSIM reconstruction error metrics performed better in the OK and NOK classes separation, although the OK class cluster still overlaps with a part of the NOK data. The results can be seen on the right part of figure 3. The overlapping NOK data are probably from the *Not complete* class which is expected to show a lower reconstruction error than the other NOK class samples. Surprisingly and unlike in the paper [6], the SSIM metric shows a low variance between the OK and NOK class, while the L2 metric seems to better distinguish those classes. It is probable, that the class separation might be increased by adding another reconstruction error metrics.

4 DISCUSSION

Based on high dimensionality of the encoded samples, we assume that the convolutional autoencoders are not suitable for the stand-alone use, but they can be used as a part of the anomaly detection pipeline.

Feature space reduction using the encoded samples with PCA and t-SNE methods didn't perform as expected on the given dataset, when only one NOK class created a separated cluster. On the other hand, the results show that this method can be used in the cases where the NOK class differs more significantly from the OK class (eg. color defects rather then the shape, or small strange objects). The improvement of this approach will be an object of the following research.

The best results in the anomalous samples separation in the feature space were achieved using the L2 and SSIM reconstruction error metric, although the metric selection should be further studied and the classes separation is still far from ideal. We believe that better results might be obtained by increasing the feature space dimensionality with other error metrics, or by comparison of the reconstructed sample with the suitably selected and representative samples from the OK class. This approach might also lead to better results of the SSIM metric.

In all three examined cases we tried to use the OC-SVM classifier to determine the boundaries be-

tween the OK and NOK class. This method did not perform as well as expected - we plan to focus on tuning and comparison with other unsupervised classifiers in our future experiments. We also suggest the normalization of the feature space and experimentation with another error metrics because SSIM in particular did not fulfil our expectations.

5 CONCLUSION

In this paper, we compare several approaches how to reduce the feature space for anomaly detection. Because the performance of all studied CAE architectures is similar - based on the lowest encoded space dimensionality, we would recommend the MVTec for the further use. We prove that the error metrics can be used to create feature space for the anomaly detection task and that this space is more suitable for this task than the feature space based on PCA and t-SNE. To the best of our knowledge, we are the first to use the error metric in this manner and not only as an anomaly score.

We also plan to test this approach on another dataset for anomaly detection, because the *Cookie Dataset* shows low variance between the classes and especially the structural defects are a difficult task to distinguish.

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 -"Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- [1] SCHOLKOPF, Bernhard, et al. Estimating the support of a high-dimensional distribution. Neural computation, 2001, 13.7: 1443-1471.
- [2] RATSCH, Gunnar, et al. SVM and boosting: One class. GMD-Forschungszentrum Informationstechnik, 2000.
- [3] ERFANI, Sarah M., et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recognition, 2016, 58: 121-134.
- [4] MASCI, Jonathan, et al. Stacked convolutional auto-encoders for hierarchical feature extraction. In: International conference on artificial neural networks. Springer, Berlin, Heidelberg, 2011. p. 52-59.
- [5] SAKURADA, Mayu; YAIRI, Takehisa. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis. 2014. p. 4-11.
- [6] BERGMANN, Paul, et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. arXiv preprint arXiv:1807.02011, 2018.
- [7] VINCENT, Pascal, et al. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning. 2008. p. 1096-1103.
- [8] PEDREGOSA, Fabian, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 2011, 12: 2825-2830.
- [9] T-distributed Stochastic Neighbor Embedding. Scikit learn [online]. [cit. 2021-03-11]. Available from: https://scikit-learn.org/stable/modules/generated/sklearn. manifold.TSNE.html

SMART HOME WIRED ARCHITECTURE

Tomas Sykora

Doctoral Degree Programme (2), FEEC BUT E-mail: xsykor23@stud.feec.vutbr.cz

Supervised by: Zdenek Bradac

E-mail: bradac@feec.vutbr.cz

Abstract: This article deals with the problem of Smart home systems. Electrical installations in new built and older buildings have recently integrated smart devices. These devices use IoT technology, which needs to communicate with each other. This article discusses the possibilities of replacing most often used wireless communication for a new type of wired based on Industry IoT standards. Using wired technologies will enhance the system reliability, safety and security. What more it involve reduction of the electromagnetic smog, a potentially health-endangering effect generated by devices communicating wirelessly at high frequencies.

Keywords: Home Automation, Smart home system, Industry IoT

1 SMART HOME

Smart Home increase the comfort of living, but also reducing the costs of living. It based on IoT devices located all around the house. These devices creating a comprehensive, safe, comfortable, convenient, energy saving system. Compared to a classic home installation, starting a home installation is more expensive, but the costs will be recouped during use. Nowadays, smart home systems and devices have become more user-friendly and emphasize the overall interaction to the users. Internet and IoT technologies have been used to connect various devices such as lighting systems, temperature regulation, curtain control, security system, audio and video equipment, computer and communication equipment, kitchen appliances, alarm clock and cleaning. Users can control these devices remotely through mobile phones, tablets, and touchscreen panels, super convenient to manage between daily routines, home appliances and other smart home devices. [1]



Figure 1: Smart Home System

There are numerous technologies and applications that can be installed in smart homes. But people often have no idea what do they really need for their homes. Also, installation and high maintenance cost are stopping them to make the first step for intelligent living.

Every year, new manufacturers of smart devices appear on the market, which sees the potential in a new market. This creates a large number of new devices that do not support mutual compatibility. New projects, such as Home Assistant, are being created to enable the interconnection of devices from different manufacturers.

1.1 HOME ASSISTANT

The solution to this problem is Home Assistant. It is a free open-source software for creating and controlling home automation. It is a software brain for a smart home. The main application is written in python and the first version was released in 2013. According to GitHub State of the Octoverse1, in 2020 it became the second largest open-source project according to the number of contributors, which is written in python.



Figure 2: Home Assistant logo [2]

Home assistant also has an app for android, iOS and watchOS. Most commercial IoT and Home automation products have an official add-on to this system, these official add-ons are available through 1700. The biggest focus of the Home Assistant is on safety. It is a local solution, so data is not stored anywhere other than on your device, so it is one of the biggest advantages over cloud services and other proprietary systems. It can also be connected to already known home automation ecosystems such as Google Home, Apple HomeKit, Mi Home, IKEA Smart Home. With an emphasis on security, it is possible to set for each of these services which entities will be available for this service. [2]

2 COMMUNICATION TECHNOLOGY FOR SMART HOME DEVICES

There are a wide variety of technology platforms, or protocols, on which a smart home can be built. Each one is, essentially, its own language. Each language speaks to the various connected devices and instructs them to perform a function.

Choosing a smart home protocol can be difficult. Obviously, you want one that will support a large number of devices, as well as one that offers the best possible ability for devices to talk to each other. But there are also other factors to consider, such as power consumption, bandwidth and, of course, cost.

Following is an overview of some of the most popular home technology platforms on the market.

2.1 ETHERNET

Fast and reliable wired communication, with a range of up to 100 m and low susceptibility to electromagnetic interference. One of the most used wired type of connection between devices. Based on type it can achieve high communication speeds.

2.2 WI-FI

Currently the most used protocol for communication between devices. Fast and reliable wireless communication, with a range of around 25 m. The 2.4 GHz frequency band is most often used for IoT.

2.3 BLUETOOTH

A short-range wireless protocol around 10 m. Often used on phones, headphones, and speakers. Its adaptive frequency hopping system detects existing signals, such as Wi-Fi, and negotiates a channel map for the Bluetooth devices in order to minimize interference.

2.4 THREAD

A wireless protocol developed by a group of companies including Nest, Samsung, QUALCOMM, and OSRAM. Designed to allow the devices in its protocol to communicate even when the Wi-Fi network does not work.

2.5 ZIGBEE

Zigbee is a wireless protocol which operates in a mesh network. It uses a device to relay a signal to other devices, strengthening and expanding the network. Zigbee can be built in dimmers, door locks, thermostats, and more. The advantage is lower energy consumption compared to Wi-Fi. The latest version, ZigBee 3.0, promises better interoperability between devices and versions.

2.6 Z-WAVE

Similar to Zigbee, Z-Wave is an open source mesh network protocol. Technically speaking, the main difference between the two is the data throughput. Z-wave is roughly 6 times slower than Zigbee. Require less energy to cover the same range as Zigbee. Z-Wave runs on the 908.42MHz frequency band.



Figure 3: Overview of protocol logos

3 SYSTEM TOPOLOGY

An important part of every system is the construction of its topology. This is the basis for communication between devices and the entire software architecture. It has an impact on the stability, security and representativeness of the whole system. [3]



Figure 4: Types of topology [3]

3.1 CENTRALIZED SYSTEM

In this system, there is one central control unit, connected via a data bus to other components of the installation. This unit contains software that allows it to receive information from sensors and, based on them, controls and regulates active elements. Main advantage is lower price of used sensors and actuators. A centralized system is easy to set up and can be developed quickly. But this system has an important limitation. If the central point crashes, the system no longer works properly.

3.2 DECENTRALIZED SYSTEM

Decentralized systems do not have one central point. Instead, they use multiple central points. A decentralized system can be just as vulnerable to crashes as a centralized one. However, it is by design more tolerant to faults. That is because when one or more central points fail, the others can continue to provide data access. Resources remain active if at least one of the central points continue to operate. Usually, this means that system can repair faulty points and address any other problems while the system itself continues to run as usual. Crashes in a decentralized system may affect the performance and limit access to some data. But in terms of overall system uptime, this system offers a big improvement over a centralized system. Another advantage of this design is that the access time to the data is often faster. That is because system can create nodes in different regions or areas where activity is high.

3.3 DISTRIBUTED SYSTEM

A distributed system is similar to a decentralized in that it doesn't have a single central point. But going a step further, it eliminates centralization. In a distributed system, users have equal access to data. The best example of distributed system is the internet. Hardware and software resources are also allocated between devices, which in some cases may improve the performance of the system. Distributed systems have evolved as a result of the limitations of the other systems.

4 PROBLEM SOLVING PROPOSAL

The main problem of a smart home is caused by the number of devices that communicate on the same protocols and are often unable to communicate with each other. Considering the current trend towards marked preference of wireless solutions at the expense of the wired technology, however, the novel communication network reference model will have to be designed in a manner reflecting and satisfying the present developments in the field.

My proposal aims at an affordable variant of smart home equipment for new buildings, where the use of this technology is already planned in advance. This solution would lead to significant savings in installation material costs.

My idea is to develop a new kind of device connection using twisted pair. Specifically using only two wires. The terminal would have to be adapted to the way of communication. Therefore, it would have to contain a battery that would provide power to the device at intervals when the line is used for communication. This concept will exploit decentralized communication, with control (da-ta-processing), sensing (service-providing), and enabling (command-executing) devices operating concurrently.

Problems will occur, such adverse effects are likely to include, for example, non-deterministic response to an action or event, communication collisions that may lead to control loop desynchronization, data delays, deadlocks, and/or other issues arising from the actual concurrent-based process. The theory accompanying the proposed approach will consist in a set of regulations for each family of devices (sensing, enabling, and control); the regulations will be created according to the formal models, in such a manner that the problems are prevented completely.

Another aspect enabling us resolve the difficulties generated by decentralized control and distributed communication network rests in, as regards the architecture proposed herein, implementing a state-of-the-art communication protocol outlined through Industry 4.0. This innovative manufacturing concept introduces interactive, data, and topology models ensuring the availability and reliability of all information being transmitted. The principles that find use industrially will be also reflected in the Smart home, namely, they will become generally applicable across convenient disciplines.

5 CONCLUSION

The aim of this article is to improve the overall availability of Smart home systems, with the desired result being an increased number of smart homes practising effective energy management. The proposed solution is to use the principles of Industry IoT in the home. A combination of decentralized control and distributed communication is used, where control units are located in individual rooms. These units serve as the main control point for all devices located in the room, and in addition, the individual devices communicate with each other. In the event of a failure of the central unit of the room, it is possible to replace its functionality with a superior main unit, which is able to maintain its functionality and communicate with devices in the room.

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCE

- Davidoff S., Lee M.K., Yiu C., Zimmerman J., Dey A.K., 2006, Principles of Smart Home Control. In: Dourish P., Friday A. (eds) UbiComp: Ubiquitous Computing. UbiComp 2006. Lecture Notes in Computer Science, vol 4206. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11853565_2
- [2] Home Assistant [online]. 2021 [cit. 2021-2-15]. Dostupné z: https://www.home-assistant.io
- [3] Centralized vs Decentralized vs Distributed Systems [online]. 2021 [cit. 2021-2-20]. Dostupné z: https://berty.tech/blog/decentralized-distributed-centralized

Doktorské projekty

Kybernetika a automatizace II.

INFLUENCE OF PIEZOELECTRIC MATERIAL PROPERTIES ON IMPEDANCE CHARACTERISTIC AND IMPROVEMENTS OF CALIBRATION EQUIPMENT

Jakub Krejčí

Doctoral Degree Programme (4), FEEC BUT E-mail: xkrejc44@stud.feec.vutbr.cz

> Supervised by: Petr Beneš E-mail: benesp@feec.vutbr.cz

Abstract: This paper deals with finite element simulations of impedance characteristics piezoelectic elements utilized in sensors of acoustic emission (AE). Coefficients of piezoelectric material are changed to evaluate the sensitivity to their variation in case they are not known precisely, so to which extent can they influence the sensitivity of the sensor. Result showed sensitivity on coefficient in c-matrix, mainly to coefficients c_{11} and c_{13} . Second investigated topic is issue with test stand for calibration of AE sensors, where output of force sensor was affected by parasite signal. By contactless measurement with laser interferometer was discovered the parasite signal is caused by mechanic of the system and it is no threat for the measuring chain.

Keywords: acoustic emission, simulation, COMSOL, impedance characteristic, laser interferometry

1 PIEZOELECTRIC PARAMETERS

Piezoelectric effect describes transformation of mechanical and electric energy in capable material. Piezoelectric materials are often used in sensors of acoustic emission as a sensing element to transform mechanical wave to electric signal. Key parameters of piezoelectric materials are defined by 4 matrices: [1] [2]

- s matrix of compliance coefficients, dimension 6x6, m^2/N
- c matrix of stiffness coefficients, dimension 6x6, N/m^2
- ε matrix of electric permittivity, dimension 3x3, F/m
- d matrix of piezoelectric coupling coefficients for Strain-Charge form, dimension 3x6, C/N

These matrices have zeroes at some positions, and some of them are symmetrical, therefore the situation is simplified and for example matrices s and c are each defined by 5 parameters, instead of 36.

2 SIMULATIONS OF IMPEDANCE CHARACTERISTICS

Simulations often use the method that noise is added to the input data to verify the robustness of the whole model. In my procedure I will try to achieve a similar effect, because the piezoelectric material is defined by many inputs and its parameters are not always easily measurable, they may differ from the real values. In simulations, the values of material constants will be changed step by step and it will be possible to quantify which parameter changes the simulation significantly and which has less or none effect. In other words, if any of the material constants are measured incorrectly, then the result may differ from the original state.

I used COMSOL Multiphysics to perform simulations of impedance characteristics of sensing element. Shape of the element is truncated cone with diameters 6.24 and 1.50 mm and height 2.512 mm and I choose typical piezoelectric material from PZT-5H from material library. All material constants were changed to the same extent (original value and values increased or decreased by 20 %) and the resulting change will be evaluated. Results are shown in Fig. 1.



Figure 1: Change of different parameters in simulation and their effect on impedance characteristic

As can be seen, impedance characteristic is most sensitive to change of c_{11} , c_{13} and c_{33} . Fluctuations of c_{12} and c_{55} caused only a small difference in comparison with previous ones. Change of damping influences only the amplitude to a small extent. There were no changes in the results when elements of *s*, ε and *d* matrix were changed, therefore these are not shown. We currently use the workplace to calibrate AE sensors, but future plans include this process to also simulate and optimize sensor parameters in advance before they are manufactured. The findings from these partial simulations will be used as a basis for which material parameters we must know exactly, or with how much tolerance, and how they may possibly affect the results.

3 ISSUES WITH AE SENSOR CALIBRATION APARATURE

One of issues we encountered in the calibration using a step function - capillary break [3] (test equipment [4] is shown in Fig. 2) - was that the output signal from the force transducer had a superimposed harmonic component with a frequency of approximately 2 kHz after capillary refraction (Fig. 3). It is not a parameter that significantly affects the measurement of the output signal from the sensor, but it was an interesting topic for diagnosing the cause. One possible reason can be the signal is crosstalk on the measurement path or on the sampling card. However, since it is a frequency that is not usual and, moreover, it appears in the signal only after the refraction, it seems to be false option.

Another possible and very probable source of this signal may be the mechanical side of the whole structure, ie how the force sensor is mounted. After the capillary break, the structure can oscillate and the inertial mass can additionally affect on the force transducer, which would not only be statically relieved, but alternately stressed by periodic resonances. An experiment was performed to verify this possible source. One of the main problems is that it is not possible to perform the measurement typically with accelerometers, because the mounting of the sensor would dampen the possible resonance of the whole system with its inertial mass. Thus, the measurement was performed contactlessly using a laser interferometer Polytec OFV 5000 and the only effect was the adhesion of the reflector tape attached to the measured object, which is negligible.

The measurements were performed in several places, first on the surface of the force transducer, then below and above it. To verify the possible clogging of the mechanical influence, the measurement was performed even at greater distances from the sensor (measurement points are shown in Fig. 2a). Result of one measurement (performed at the location of sensor, which is crucial), is shown in Fig. 4, where the parasite signal is also noticeable. The vibration velocity was also integrated to calculate a maximal unexpected displacement during the process. Results of all measurements are summed up in Table 1.

Results of the measurements show that this is not a crosstalk signal that would be difficult to diagnose,



(a) Calibration stand

(b) Force sensor





Figure 3: Typical waveform of measured force



Figure 4: Velocity measured at the sensor body

but it deals with a mechanical resonance of the system. It is evident the amplitude of examined frequency is decreasing with increasing distance from the sensor. Although the measurement was performed only in 1 axis, the trend is noticeable. There mechanic oscillations could be suppressed by adjusting the mounting of the force transducer, but since the measurement only evaluates the maximum amplitude of the force during capillary break, this is not necessary. Alternatively, it is possible to use a narrowband filter to smooth the signal, as the cause of this issue is known and it is static. Smoothing of the measured characteristic by Butterworth filter, 3rd order with boundary frequencies 1 000 to 3 000 kHz is shown in Fig. 5.

point n.	description	vibration velocity of 2140 Hz (mm/s)	maximal amplitude (µm)
1	sensor body	3.2	20
2	spike below sensor	3.61	20
3	attachment above sensor	0.38	20
4	place near to sensor	0.74	15
5	place farther from the sensor	0.16	6
6	external structure	0.065	2

Table 1: Comparison of measured results



Figure 5: Comparison of original and filtered signal measured by force sensor

4 CONCLUSION

This paper studied finite element simulations of impedance characteristics of piezoelectric elements utilized in sensors of AE. Results showed biggest influence, when elements c_{11} , c_{13} and c_{33} of *c* matrix are changed. When characterizing these materials, it is necessary to pay increased attention to these parameters. The other elements of the matrix caused less modification of the impedance characteristics and as expected change in amplitude, when damping is varied. Based on the simulations, result were the same for changes of elements in matrices *s*, ε and *d*, they probably do not have link to impedance characteristic, or some features were omitted, which will be a topic of my next studies. Secondly, in issue with test stand for calibration of AE sensors was investigated, where output of force sensor was affected by parasite signal. Contactless measurement with laser interferometer allowed to discover the parasite signal is caused by resonance of the fixture and it means no threat for the measuring chain concerning previous measurements or measurements performed in the future.

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- [1] Uchino, K. Advanced piezoelectric materials: Science and technology, Woodhead Publishing, 2017.
- [2] Tichý, J., Erhart, J., Kittinger, E., Privratska, J. Fundamentals of piezoelectric sensorics: mechanical, dielectric, and thermodynamical properties of piezoelectric materials, Springer Science & Business Media, (2010).
- [3] ISO 12713:1998 (E) Non-destructive testing Acoustic emission inspection Primary calibration of transducers
- [4] Keprt, J. Primární kalibrace snímačů akustické emise Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. 171 s. Vedoucí dizertační práce doc. Ing. Petr Beneš, PhD.

COMPARISON OF METHODS FOR IMPULSE RESPONSE COMPUTATION

Vilem Karsky

Doctoral Degree Programme (4), FEEC BUT E-mail: xkarsk010@stud.feec.vutbr.cz

> Supervised by: Pavel Jura E-mail: jura@feec.vutbr.cz

Abstract: This paper deals with obtaining impulse responses from integer order and fractional order transfer functions. There are shown three method how to compute inverse Laplace transform. The first method is based on Mittag-Leffler functions, the second method is formed on generalized Laguerre functions and the third method lays on Fourier transform. These methods are also compared on two examples.

Keywords: Mittag-Leffler functions, Generalized Laguerre functions, Fourierre transform, Fractional order transfer function, Inverse Laplace transform

1 INTRODUCTION

When we deal with systems we often need to obtain impulse response. This isn't huge problem for integer order systems but for fractional order systems it could be quite difficult. In this paper three methods for impulse response computation will be compared.

2 MATHEMATICAL BACKGROUND

In this section some basic concepts will be described. Also in this section methods for computation impulse response will be rely briefly outlined. In following section we will assume that our system is described with transfer function in form

$$F(s) = \frac{\sum_{m=0}^{M} b_m s^{qm}}{\sum_{n=0}^{N} a_n s^{qn}},$$
(1)

where $q \in \mathbb{R}$.

2.1 IMPULSE RESPONSE

Impulse response is system response to dirac impulse. When we have system describer with transfer function, the impulse response could by obtained by calculating inverse Laplace transform of our transfer function. We can write

$$g(t) = \mathscr{L}^{-1}\{F(s)\}.$$
 (2)

2.2 DIRECT COMPUTATION OF IMPULSE RESPONSE

When we have system described with transfer function in form (1) we could split this transfer function into sum of partial fractions and transform each partial fraction according to formula [1]

$$\mathscr{L}\left\{t^{\alpha k+\beta-1}E_{\alpha,\beta}^{(k)}(\pm at^{\alpha})\right\} = \frac{k!s^{\alpha-\beta}}{\left(s^{\alpha}\mp a\right)^{k+1}},\tag{3}$$

where $E_{\alpha;\beta}(z)$ are Mittag-Leffler functions defined by

$$E_{\alpha;\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)} \quad [1].$$
(4)

This method lead to solution in form of quite slow convergent infinite series.

2.3 EMPLOYING GENERALIZED LAGUERRE FUNCTIONS FOR COMPUTATION OF IMPULSE RESPONSE

Generalized Laguerre functions we can define, according to [2], as

$$l_n^{\alpha}(t) = \sqrt{2\lambda} e^{-\lambda t} L_n^{\alpha}(2\lambda t), \qquad (5)$$

where $L_i^{\alpha}(t)$ are generalized Laguerre polynomials. Generalized Laguerre polynomials are defined as

$$L_n^{\alpha}(x) = \frac{\mathrm{e}^x x^{-\alpha}}{n!} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \left(x^{n+\alpha} \mathrm{e}^{-x} \right) \quad [3]. \tag{6}$$

Generalized Laguerre functions generates orthogonal base in time space and also in operator space and spectrum coefficients of function f(t) in generalized Laguerre functions base could be computed according to formula

$$c_n = \langle l_n^{\alpha}(t), f(t) \rangle = \langle \mathscr{L} \{ l_n^{\alpha}(t) \}, \mathscr{L} \{ f(t) \} \rangle, \tag{7}$$

where $\langle f(t), g(t) \rangle$ means scalar product of functions f(t) and g(t).

This property of generalized Laguerre functions could be used for computation inverse Laplace transform. As was shown in [2, 3] we could compute spectrum coefficients using formula

$$c_n^1 = \frac{-\sqrt{2\lambda}}{(n+1)!} \left[\frac{\mathrm{d}^{n+1}}{\mathrm{d}z^{n+1}} F(z) \right]_{z=0},\tag{8}$$

where F(z) is our transfer function after applying bilinear transform in form

$$s = \lambda \frac{1+z}{1-z} \tag{9}$$

on it. So final impulse response would be in form

$$g(t) = \sum_{n=0}^{\infty} c_n^1 l_n^1 = \sqrt{2\lambda} e^{-\lambda t} \sum_{n=0}^{\infty} c_n^1 L_n^1(2\lambda t).$$
(10)

2.4 FOURIER METHOD FOR COMPUTATION OF IMPULSE RESPONSE

In paper [4] was shown method how to compute impulse response from transfer functions using Fourier transform. If we substitute *s* by $j\omega$ in our transfer function we could rewrite our transfer function into

$$F(j\omega) = \frac{M(\omega) + jN(\omega)}{Q(\omega) + jZ(\omega)}.$$
(11)

After that the impulse response could be calculated according to

$$g(t) = \frac{2}{\pi} \int_0^\infty \frac{M(\omega)Q(\omega) + N(\omega)Z(\omega)}{Q^2(\omega) + Z^2(\omega)} \cos(\omega t) dt.$$
(12)

3 COMPARISON OF BOTH METHODS

For comparison of all three methods was chosen two transfer functions (first is integer order transfer function and second is fractional order transfer function).

3.1 INTEGER ORDER SYSTEM

Firs system is defined by this transfer function

$$F(s) = \frac{10}{s^2 + 4s + 8}.$$
(13)

Analytical solution of this transfer function is

$$g_a(t) = 5e^{-2t}\sin(2t).$$
 (14)

This impulse response is plotted in Figure 1 with blue line.

When we use formula (4) we get impulse response in form

$$g_m(t) = \frac{5}{2} \mathbf{j} \left[E_{1;1}(-(2+2\mathbf{j})t) - E_{1;1}(-(2-2\mathbf{j})t) \right].$$
(15)

This impulse response can be modified to equation (14), but it is possible only for integer order system. In Figure 1 the impulse response, which was calculated for first 100 terms, is drawn with red dashed line.

Third way to get impulse response is by using Generalized Laguerre functions as mentioned earlier. This impulse response will be in form (10). For this system was employed only first 7 generalized Laguerre functions with $\lambda = 3.7465$. In Table 1 you can see spectrum coefficients. This impulse response $g_g(t)$ is plotted in Figure 1 with green dash-dotted line.

Table 1: The coefficients' spectrum: integer order system

i	0	1	2	3	4	5	6
c_i^1	1.7199	-1.1594	0.0505	0.2043	-0.0762	-0.0140	0.0199

And finally the fourth way is to employ Fourier transform as was said previously. Final result $g_w(t)$ is shown in Figure 1 with black dotted line.

For comparison of the approximations was calculated three differences $g_a(t) - g_m(t)$, $g_a(t) - g_g(t)$ and $g_a(t) - g_w(t)$. These differences are shown in Figure 2. You can see that MLF better approximate impulse response in first part but, then they diverge. In opposite GLF have some approximation error in the beginning, and then they are converging to g(t). It is worth mention that for MLF was used 100 terms and for GLF was used only 7 terms. But the best solution is obtained by employing Fourier method.

3.2 FRACTIONAL ORDER SYSTEM

For fractional order transfer function was chosen quite similar system. This system is described

$$F(s) = \frac{10}{s^{1.2} + 4s^{0.6} + 8}.$$
(16)

If we use formula (4) to get impulse response, we get

$$g_m(t) = \frac{5}{2}jt^{-0.4} \left[E_{0.6;0.6}(-(2+2j)t^{0.6} - E_{0.6;0.6}(-(2-2j)t^{0.6}) \right].$$
(17)



Figure 1: Impulse responses of the first system

You can see it in Figure 3 with blue line. For computation was used first 200 terms.

But when we employ first 7 GLF with $\lambda = 4.8697$ we get spectrum coefficients, which are in Table 2. It's plotted in Figure 3 with red dashed line.

Table 2: The coefficients' spectrum: fractional order system

i	0	1	2	3	4	5	6
c_i^1	1.4178	-0.2816	0.1962	-0.1093	0.0773	-0.0586	0.0430

And finally solution obtained using Fourier transform is shown in Figure 3 with green dotted line. As you can see in Figure 3 all methods give similar result, but MLF diverge.



Figure 2: Differences from the $g_a(t)$



Figure 3: Impulse response of the second system

4 CONCLUSION

In this paper was shown that all methods are suitable for approximation impulse response. Solution obtained with MLF offers better results in the beginning of the impulse response but it diverges and needs a lot of terms. On the other hand solution using GLF converges and needs only a few terms but the approximation of the beginning of the impulse response is little worse. And solution using Fourier transform offers great results but this method could consume a lot of computation power to compute integrals numerically.

ACKNOWLEDGEMEN)

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- PODLUBNY, I.: Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications. San Diego, Calif.: Academic Press, c1999. Mathematics in science and engineering, v. 198. ISBN 0-12-558840-2.
- [2] MAIONE, G.: Inverting fractional order transfer functions through Laguerre approximation. *Systems and Control Letters*, 52(5), 387–393, 2004.
- [3] KARSKV, V.: An Improved Method for Parameterizing Generalized Laguerre Functions to Compute the Inverse Laplace Transform of Fractional Order Transfer Functions. *In INTERNATIONAL CONFERENCE OF NUMERICAL ANALYSIS AND APPLIED MATHEMATICS ICNAAM 2019.* 2020. p. 340004-1 (340004-4 p.)ISBN: 978-0-7354-4025-8.
- [4] YUCE, A., F. N. DENIZ, N. TAN and D. P. ATHERTON: Obtaining the time response of control systems with fractional order PID from frequency responses. 2015 9th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkey, 2015, pp. 832-836, doi: 10.1109/ELECO.2015.7394522.

SEMI-SUPERVISED APPROACH TO TRAIN CAPTCHA LETTER POSITION DETETOR

Ondrej Bostik

Doctoral Degree Programme (5), FEEC BUT E-mail: bostik@feec.vutbr.cz

> Supervised by: Karel Horak E-mail: horak@feec.vutbr.cz

Abstract:

Common Optical Character Recognition (OCR) methods benefit from the fact, that the text is distributed in images in a predictable pattern. This is not the situation with CAPTCHA systems. Utilizing OCR algorithms to overcome common web anti-abuse CAPTCHA systems is therefore a challenging task. To train a system to overcome any CAPTCHA scheme, an attacker needs a huge dataset of annotated images. And for some methods, the attacker needs not only the right answers but also an exact position of the character in the CAPTCHA image.

Annotate the positions of the object in an image is a time-consuming task. In this paper, we propose a system, which can help to annotate the position of CAPTCHA character with minimal human interaction. After annotating a small sample of targeted CAPTCHA images, a YOLO-based region detection deep network is used to search for the characters' locations.

Keywords: OCR, CAPTCHA, Deep learning, YOLO v2, semi-supervised learning, MATLAB

1 INTRODUCTION

Nothing is most synonymous with an anti-abuse system used on the world-wide-net nowadays as a CAPTCHA (Completely Automated Public Turing Test to tell Computers and Humans Apart). And the major part of the commonly used CAPTCHA system world-wide is based on OCR (Optical Character Recognition) problem.

An image with characters scattered around is sent as a part of almost every web form. Every user who wants automatically interact with this kind of service needs to create a system to overcome CAPTCHA protection [1, 2]. The side effect of creating new systems to overcome text-based CAPTCHAs is therefore an improvement in the field of OCR [3].

Recent CAPTCHA breaking systems are based on convolutional neural networks (CNN). One great example is [4], one deep network is used for the length of the text estimation and the other deep network is using this information for classification estimation. Another similar work [4] utilized CNN for feature extraction and Long Short-Term Memory (LSTM) for recognition. In both cases, a huge annotated dataset is needed for the successful attack.

The work presented in [5] overcomes some of these issues by transfer learning. The initial solver is created with the use of synthetic CAPTCHA images. After that 500 manually annotated CAPTCHA images are used for fine-tuning this solver. Although the base solver has an average success rate close to 0%, after fine-tuning on real examples of targeted CAPTCHA images the success rates increases to 50%-97%.

2 PROPOSED SOLUTION

The work described in this paper is based on our previous work in [6]. In our precedent paper, we propose using the target web-page for generating a sufficient enough dataset without annotating a huge number of data. To create a versatile end-to-end system, we need a huge training dataset containing not only the right answer (which we can obtain by the algorithm presented in [6]) but also the correct positions of the characters in every CAPTCHA image.

Annotate the positions of every character in the CAPTCHA image is a time-consuming task, so we want to automate this task by utilizing a deep learning approach. Every CAPTCHA image is processed by region detector based on YOLO v2 (You Only Look Once) [7, 8] algorithm with deep network based on ResNet 50 architecture in a role of feature extraction unit.

The last part of the proposed solution is the filtering stage. Because the CAPTCHA challenges are generated automatically without human interaction, each implementation takes into account the possibility that the presented CAPTCHA image will not be human-readable. The user can request an alternate image without penalization. Therefore, it is more advantageous for the CAPTCHA breaking algorithm to decide that the image is unreadable and request a new image instead of sending a bad response.

In future work, the automatic validation stage will follow. The extracted region will be fed into the classification stage, where every region is recognized via classic OCR methods (for example another deep network, Tesseract OCR engine, or for example simple k-NN algorithm). If the answer is correct for all the characters, we can save the image with the correct annotation. This method relies on the previously known transcription of the CAPTCHA text, which can be extracted by the methods from [6], or simple human transcription. Either way, this kind of annotation is far quicker, then forcing the human to hand-labelling every character in the image.

In this paper, we substitute the validation engine with a human for greater control over the annotation results.

2.1 INPUT DATASET OVERVIEW

For this experiment, we enlarged the dataset used in our previous work [6]. With the use of the same python web crawler utilizing Selenium Webdriver we automatically obtained 50 000 CAPTCHA images from the domain http://geocheck.org. For this experiment, we annotated 250 CAPTCHA images for training the object detector and we also extract 1 000 more unannotated images for validation. The targeted CAPTCHA scheme consists only of digits with a fixed length of 5 numbers.

The main advantage of this CAPTCHA scheme, in particular, is the verification process, which is located on the client-side. The correct answer is presented in the HTML code sent to the end-user computer in the form of an MD5 hash, which can be broken rapidly utilizing the simple rainbow table. Our automated system can download and break around 4 CAPTCHA images per second exploiting this weakness. Assembling the huge dataset of real-world CAPTCHAs all annotated with 100% correct answers is then a simple task. The sample images from the used dataset are displayed in figure 1.



Figure 1: Sample of input dataset used for the experiments

2.2 IMPLEMENTATION

For the rapid deployment, the experiment was performed in MATLAB computational environment using Deep Learning Toolbox and Computer Vision Toolbox. Proposed experimental setup is visualized on figure 2.



Figure 2: Experiment setup overview (including future work)

Before the experiment started, we annotated 250 CAPTCHAs images via *Image Labeler MATLAB* app to create precise ground-truth data of character location.

On every epoch during the learning stage, the input dataset was augmented with random transformations. Every image undergoes random changes, in contrast, hue, saturation, and brightness, after that the image was randomly scaled with the cropping to preserve the original size.

Images with annotated regions were used to train segmentation deep neural network [9] based on ResNet [10] as a feature extraction segment and YOLO v2 [7, 8] as region proposal segment with layers shown on table 1. Training used Adam Stochastic Optimization algorithm [11], initial learning rate set to 0.001 and mini-batch size of 16 for the total number of 25, 50, or 100 epochs. The size of anchor boxes for YOLO layers was estimated according to [8], the number of anchor boxes was set to 7.

	Layer type	Layer description
1	Convolution	1024pcs 3x3x1024 convolutions with stride 1 and padding 'same'
2	Batch normalisation	Batch normalization with 1024 channels
3	ReLU	ReLU
4	Convolution	1024pcs 3x3x1024 convolutions with stride 1 and padding 'same'
5	Batch normalisation	Batch normalization with 1024 channels
6	ReLU	ReLU
7	Convolution	42pcs 1x1x1024 convolutions with stride 1 and padding [0 0 0 0]
8	YOLO v2 Transform Layer	YOLO v2 Transform Layer with 7 anchors
9	YOLO v2 Output	YOLO v2 Output with 7 anchors

Table 1:	YOLO based ob	ject detector layers
----------	---------------	----------------------

The last part of our model is the filtering stage. The YOLO object detection network for each image outputs a number of bounding boxes. As the evaluated schema contains a fixed number of digits, we can easily filter the results with the incorrect number of regions accordingly to our expectations. The second filtering stage is based on identifying the overlapping segments. For each pair of bounding boxes in the image, the Intersection over Union (IoU) is calculated and when it reaches threshold experimentally set to 40% intersection, the regions are merged into one.

3 RESULTS OF EXPERIMENT

The presented experiment was realized on a laptop computer with Intel Core i5-6300HQ with 4 cores and 2.3GHz frequency and with graphic card NVIDIA GeForce GTX 950M. The presented segmentation network was trained on 200 manually annotated images.

Part of this experiment was evaluating the speed of training on common hardware. The training time per 25 epochs was 52 minutes and grows linearly with more training epochs, with the 100 epochs, the training time was almost 3.5 hours.

For validation 50 more CAPTCHA images were manually annotated. Using these images the validation accuracy for classifier trained for 100-epoch reached 95.90%. Resulting comparison of validation accuracy is depicted in figure 3.



Figure 3: Validation accuracy per training epoch on manually labeled data

To further assess the experiment 1.000 unlabelled images were used. Without the filtering stage, there is a significant difference in the portion of images with the correct number of segments. Images with the correct number of segments were then manually evaluated by a human. After the filtering stage, there is only a marginal difference between 50-epoch and 100-epoch learning, both closed to the 100% region proposal accuracy. The results are summarized in table 2.

Training	Training	Validation	Portion of images with the	Manually evaluated region		
epochs	time	accuracy	correct number of region	proposal accuracy		
			without / with filtering	without / with filtering		
25 epochs	52m	81.38%	27.60% / 79.10%	66.30% / 96.97%		
50 epochs	1h 44m	92.66%	40.00% / 89.80%	91.50% / 99.55%		
100 epochs	3h 25m	95.90%	77.10% / 94.50%	99.09% / 99.79%		

 Table 2:
 Segmenting process success rate per learning epochs

4 CONCLUSION

This paper presents an idea of a semi-supervised system for the position annotation of characters in the CAPTCHA image. In this paper, we present a system requiring only a very small initial dataset

used for classifier training. Utilizing 200 annotated images we achieved the accuracy of 99.09% of character region detection for 3.5hours of learning on a laptop computer without filtering. Even more, with the simple filtering presented in this paper, we can achieve 99.55% accuracy with a detector trained only for half the time.

The following work will focus on adding the validation engine based on OCR recognition of CAPTCHA images. Furthermore, this data can be used to quickly train end-to-end CAPTCHA breaking system without any human interaction.

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- [1] E. Bursztein, J. Aigrain, A. Moscicki, and J. C. Mitchell, "The End is Nigh: Generic Solving of Text-based CAPTCHAs," 2014.
- [2] E. Bursztein, A. Moscicki, C. Fabry, S. Bethard, J. C. Mitchell, and D. Jurafasky, "Easy Does It: More Usable CAPTCHAs," in *CHI '14 Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, (1600 Amphitheatre Pkwy), pp. 2637–2646, 2014.
- [3] K. Kaur and S. Behal, "Designing a Secure Text-based CAPTCHA," in *Procedia Comput. Sci.*, vol. 57, pp. 122–125, Elsevier, 2015.
- [4] M. Tang, H. Gao, Y. Zhang, Y. Liu, P. Zhang, and P. Wang, "Research on Deep Learning Techniques in Breaking Text-Based Captchas and Designing Image-Based Captcha," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, pp. 2522–2537, oct 2018.
- [5] P. Wang, H. Gao, Z. Shi, Z. Yuan, and J. Hu, "Simple and Easy: Transfer Learning-Based Attacks to Text CAPTCHA," *IEEE Access*, vol. 8, pp. 59044–59058, 2020.
- [6] O. Boštík, "SEMI-SUPERVISED DEEP LEARNING APPROACH FOR BREAKING GEO-CACHING CAPTCHAS," in *Proc. II 26th Conf. STUDENT EEICT 2020 - Sel. Pap.*, (Brno), pp. 166–170, Vysoké učené Technické, Fakulta elektrotechniky a komunikačních technologií, Vysoké učené Technické, Fakulta elektrotechniky a komunikačních technologií, 2020.
- [7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conf. Comput. Vis. Pattern Recognit., pp. 779–788, 2016.
- [8] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in 2017 IEEE Conf. Comput. Vis. Pattern Recognit., pp. 6517–6525, 2017.
- [9] K. Horak and R. Sablatnig, "Deep learning concepts and datasets for image recognition: overview 2019," in *Elev. Int. Conf. Digit. Image Process. (ICDIP 2019)*, no. 11179, pp. 484–491, SPIE, 2019.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770–778, 2016.
- [11] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *Int. Conf. Learn. Represent.*, 2014.

MODELING OF MULTIPLE ELECTRODE SYSTEM FOR INHOMOGENEITY DETECTION

Vlastimil Mancl

Doctoral Degree Programme 2, FEEC BUT E-mail: xmancl00@stud.feec.vutbr.cz

> Supervised by: Petr Beneš E-mail: benesp@feec.vutbr.cz

Abstract: Capacitive sensing gained on popularity through the last decades. From proximity sensors through humidity measurement to touch screen matrices. Capacitive sensing offers reliability, simplicity, and low financial cost. One of the crucial parts of the capacity sensor is the setup of the electrodes. The behavior of such setup is often needed to be known before application. Due to the complex nature of electrostatic arrays, the precise analytic solution is often unavailable so the other means to gain solution is needed. This paper focuses on using the simulation for modeling of multiple electrode setup, which could be used for measuring inhomogeneity in an otherwise homogeneous dielectric. The paper describes the problem of sharp edges. The description of the sensing depth of a planar capacitor is also included. Three simulations were created to prove initial assumptions.

Keywords: Ansys Electronics Desktop, capacitive sensing, Maxwell 2D and 3D, Model accuracy, planar capacitor, Simulation

1 INTRODUCTION

Capacitive sensing offers a wide range of solutions in almost every field of industry or science. It is often reliable, simple, and rather cheap. This paper focuses on an important capacity sensors principle, which uses fringe field detection and its usage for proximity detection. An inseparable part of every device is an analysis of its characteristics. Often the analytic solutions are sufficient, but sometimes the analytic solution can be non-existent. In these cases, simulations are the easiest way to provide the desired characterization.

This paper describes the problem of analytic solutions and modeling of a planar capacitor. Then explains the aspect of sensing depth, which is also demonstrated on two simple models, created in Ansys Maxwell 3D. After which the multielectrode set up is proposed, created, and analyzed. During the research, several problems emerged. All of those problems and their solutions are described.

The main goal of this paper is not to design a perfect electrode setup, but rather to offer a reader an insight on what mistakes to avoid when creating a similar model.

2 PLANAR CAPACITOR

Capacitive sensing uses a wast variety of electrode setups. The best-known setup is a parallel plate capacitor, which contains two or three electrodes parallel to each other, with a dielectric in the middle. Capacity between two parallel electrodes is determined by a well-known equation 1. But this equation is valid only if the distance d is much smaller in comparison to area s. The larger the distance, the more noticeable the deviation from reality. The reason for deviations is the fringing effect at the end of the capacitor. The problem becomes much more complex if the electrodes are not parallel. If we open, figuratively speaking, the parallel capacitor, the equation stops to be even remotely valid and the planar capacitor is created.

$$C = \varepsilon_0 \cdot \varepsilon_r \cdot \frac{s}{d} \left[F \right] \tag{1}$$

The planar capacitor uses two or more electrodes as well, but in this case, the plates are in the same plane. So the dielectric lies on the same side of the electrodes rather than in the middle of them. The difference between parallel and planar capacitors can be seen in the figure 1a.



Figure 1: Transformation of a parallel plate capacitor into a planar capacitor and its model.

2.1 MODELING PLANAR CAPACITOR

There is currently no universal analytic solution for planar capacitors due to the complicated nature of electric fields.[2] However three methods exist to determine the analytic model for individual capacitive setups, but needless to say those methods often rely on simplification, such as infinite dimensions or ideal shape. As a result, such models are always an approximation of a real model. In addition, immense knowledge of a mathematics and physics apparatus is a necessity to derive such a model. Those methods are the method of images, the method of conformal mapping, and the solution of Laplace's equation. [3]

The more convenient and less demanding method, to determine expected behavior, is the use of simulation software with an application of the finite element methods (FEM). This method is numerical and therefore relies on the computational power of modern computers. The Ansys Maxwell implements this method and as such was used for the determination of capacity and sensing depth of the multiple electrode setup.

2.2 SENSING DEPTH

The planar capacitor can be described by several attributes, such as penetration depth, imaging resolution, signal strength, or measurement sensitivity. [4] This paper focuses on determining the penetration depth. The rest is a matter of future research.

The penetration depth can be described as the smallest detectable change of capacitance, for moving inhomogeneity away from the electrodes in an otherwise homogeneous dielectric.[1] Based on past research it seems that the sensing depth can be as much as half of a distance between the centroids of electrodes. But this concerns only the elementary shapes of electrodes, such as rectangle or ring. The effective depth is also dependent on the dimensions of electrodes. The penetration depth is demonstrated on two planar capacitor models with rectangular electrodes with different widths. Those models are described in section 3

3 SIMULATION OF A SIMPLE PLANAR CAPACITOR MODEL

For validation of the theory, the two iterations of the same model had been designed. The basis of the model consists of one planar capacitor with two electrodes, one shielding electrode, and polystyrene

block with cut out for distilled water. The model can be seen in figure 1b. The materials were chosen, with future measurements in the mind. The polystyrene was chosen for its accessibility and water for its high permittivity. The width w of electrodes varies for both iterations. The first model has 2 cm wide electrodes, the second one 6 cm. The distance from its centroids is 10 cm for both models. The model was used for parametric simulation with parameter d [mm]. The range of parameter was $\langle 0; 80 \rangle$ with step of 5 mm. The parameter was bound to the distance of the water layer from the planar capacitor. The dimensions of polystyrene also changed accordingly. The total error was set to 0.05%. Each parametric simulation converged and satisfied the requirements. The shielding electrode was placed under the capacitor electrodes, to simulate shielding, which will be needed for real measurement.

From simulated data (figure 3a) it is observable, that the absolute capacity change (Δ_C) near the 50mm is rather small, only in the matter of fF. But in ideal conditions, this could be measurable. The capacitor with wider electrodes has, as expected, higher capacity, and the effective sensing depth is theoretically deeper by at least 5 cm.

4 SIMULATION OF A MULTIPLE ELECTRODE MODEL

Another experiment was brought out to simulate more complex model with a multiple electrode setup. At its core, the model was almost identical to the previous one. However, 10 electrodes were placed into the model instead of two. The detail of the final electrode setup can be seen in figure 2b.

During the creation, many difficulties emerged. First of all, the convergence for this model was severely influenced by the sharp edges of the electrodes. Without filleting the edges, the simulation was unable to converge on an average PC, due to high computational requirements. It is known, that perfect edge creates a singular point that affects the simulated electric intensity E_i . The reason behind this is theoretically infinite E_i near the perfect edge. The finer the mesh the closer are the nodes to the edge, thus the simulation obtains higher intensity. This can be solved by filleting the edges or by bounding the convergence to the capacity, but with a cost of imprecision in E_i . The precise distribution of E_i was desired so the edges were rounded, which created another problem.

As can be seen in figure 2a, when the filleted electrode lays on top of the object, another sharp point is created. This time it does not affects the E_i , but the mesh has to be detailed, which inadvertently increases the computing demands. Also the performed simulation shows, that algorithm, generating the mesh, occasionally fails to create a valid mesh.

This problem was solved by hovering the electrodes above the objects. The distance between the electrodes was 0.2 mm from each side. The error created by this is negligible. Mainly because the expected accuracy of the sensing depth is in a range of a few mm.



(a) The creation of a miniature (b) Electric field around final multi-electrode setup, with detail. (b) Electric field around final multi-electrode setup, with inner electrode excitated.



The measuring depth of the multi-electrode setup depends on combinations of excitated electrodes. Not only the opposing electrodes can be excitated. To strengthen the signal for deeper reach, more than just two electrodes can be used. This paper, however, shows only the excitation of the two opposing electrodes, with rest of the electrodes serving as a shield. So as of now, only five combinations were simulated. For every combination, the parametric simulation was performed. Other combinations are a matter of future research.

The presumption about distribution of capacitance was, that the capacitance will be largest for the middle electrodes and the smallest for outer electrodes.

The first iteration of model was similar to the final model, showed in figure 2b, but the Shields (SHLD) were not present. The shields were added because the first simulations showed, that the electrodes closest to the middle (1.) had unproportionally larger capacitance in comparison to other combinations. It was due to the significant contribution of the space between those electrodes. This problem was solved by inserting a shielding electrode between them. Subsequently, the simulation showed that the outer electrodes (5.) had higher capacitance than the ones situated beside them (4.), therefore failing the initial presumption about capacity distribution. The reason behind this was the fact that outer electrodes were shielded only from one side. This was solved by placing two other shielding electrodes on the outer rims. With this setup only the area on top and bottom of the planar capacitor is now influencing the final capacitance. The last iteration satisfies the initial presumption. The final distribution of capacity can be seen in table 1. The 1. in the table represents the combination of the electrodes closest to middle as seen in figure 2b.

Combination	1.	2.	3.	4.	5.
Capacity [fF]	-898.4808	-98.3735	-33.0574	-17.3396	-13.5462

Table 1: Distribution of capacity. The 1 being the electrodes closest to middle



inhomogeneity

(a) Capacity C and its increment Δ_C vs depth of (b) Dependence of $\delta(x,d)$ on excitation combination and distance of plywood layer.

Figure 3: Acquired dependencies.

5 **DETECTING INHOMOGENITY**

Early results show that by this setup, an inhomogeneity could be detected. By performing the following calculation of sensitivity $\delta(x, d)$.

$$\delta(x,d) = \frac{C(x,d) - C_0(x)}{C(x,d=0) - C_0(x)} \quad [-]$$
⁽²⁾

Where C(x,d) is a capacity of electrode combination x in dependence on distance d of an inhomogeneity (water layer), $C_0(x)$ is a capacity with no inhomogeneity present and C(x,d=0) is the highest capacity when d = 0.

The result $\delta(x, d)$ is a standardized sensitivity value, which describes the reaction of an electrode setup to a distance of inhomogeneity in comparison to the capacity of a set up without inhomogeneity. The resulting function can be seen in figure 3b.

The results indicate, that the inhomogeneous layer could be detected. In the figure 3b it is noticeable, that the combinations of electrodes, that are furthest from each other (5.), react to larger depths of an inhomogeneity. Needless to say, that the capacity differences, in greater distances, are again a matter of fF, but by enlarging the electrodes, the capacity differences should also be more distinguishable.

Whats more interesting is, that by shielding excitated electrodes from both sides the sensitivity is not largest, when the inhomogeneity is nearest to the electrodes, but a bit further away. This phenomenon is best seen in figure 3b for excitation combination 4. This discovery could be essential for future research.

6 CONCLUSION

The principle and sensing depth of a planar capacitor sensing were described and demonstrated. Two models of planar capacitors were created, described, roughly evaluated, and used to confirm initial assumptions about the sensing depth. The results show, that the sensing depth could be as much as half of the distance between the centroids of the electrodes.

The model of multiple electrode planar capacitor was proposed, designed, and used in the simulation. During the creation, several problems occurred, but they were successfully solved. The important conclusion is to avoid the sharp edges when the precise distribution of E_i is required. The results of the simulation of multiple electrodes are promising. It seems that this setup could be used for future inhomogeneity detection, but it is a subject of future research. The comprehensive evaluation of multiple electrode model is planed. Also the verification by measurement is planned to determine the real effective sensing depth.

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- [1] Xiaohui Hu, Wuqiang Yang *Planar capacitive sensors designs and applications*. Manchester, UK, 2010. doi 10.1108/02602281011010772.
- [2] Rashed H. Bhuiyan, Roger A. Dougal. *Proximity Coupled Interdigitated Sensors to Detect Insulation Damage in Power System Cables*. 2007, doi: 10.1109/JSEN.2007.908440
- [3] Y. Ye, C. Zhang, C. He, X. Wang, J. Huang and J. Deng. A Review on Applications of Capacitive Displacement Sensing for Capacitive Proximity Sensor. 2020, doi: 10.1109/AC-CESS.2020.2977716.
- [4] X. B. Li, S. D. Larson, A. S. Zyuzin and A. V. Mamishev. Design of multichannel fringing electric field sensors for imaging. Part I. General design principles. 2004 doi: 10.1109/ELINSL.2004.1380616.

CALIBRATION OF PRECISION ROTARY TABLE USING LASER INTERFEROMETER SYSTEM

Michal Šindelář

Doctoral Degree Programme (4), FEEC BUT E-mail: sindelar@feec.vutbr.cz

Supervised by: Petr Beneš

E-mail: benesp@feec.vutbr.cz

Abstract: Calibration of a rotary axis can be useful to determine angular deviations throughout the measuring range. Specified accuracy presented in the specification list represents the worst case, therefore information about magnitudes of deviations of individual measuring points are unknown. The resulting error map of the device under test allows correction of the indicated value. As a result, the measurement accuracy of the device can be increased. Calibration of a precision rotary table was performed using a Keysight 5530 laser interferometer system. The presented paper briefly describes the fundamentals of this portable calibrating system. Also, the results and accuracies are discussed.

Keywords: Calibration, rotary table, laser interferometer

1 INTRODUCTION

Precision rotary and index tables are devices with accurate angular positioning capabilities. Rotary tables are widely used as measuring devices in many applications, such as machining, welding, manufacturing of ball bearings cages, calibration, or R&D purposes in laboratory environments.

In terms of precision rotary positioning, parameters as absolute deviation, repeatability, and angular resolution are crucial. Usually, manufacturers of rotary tables specify these parameters in a datasheet, but there are two important facts to consider. Firstly, a given absolute deviation of a rotary table represents maximal positioning error throughout a measurement range. Hence, the specified absolute deviation is not assigned to the particular angle and affects all measuring points in the same way. Secondly, a manufacturer guarantees absolute deviation in the determined running conditions. Since the performance of a rotary table may vary due to different running conditions (aging, unbalanced load, bearing condition, different temperature, etc.), it can be useful to provide in situ calibration directly on its application axis [1].

In general, calibration is a process, where the device under test is compared to the more accurate reference standard. Thus, the calibration of a rotary table is based on a comparison of angular values indicated by a rotary table and angular standard. Comparison for i-th measuring point is obtained by the difference between i-th measured value and i-th reference value. As a result, these angular deviations provide an error map that contains information about the magnitude of absolute measurement error related to nominal angular positions. Consequently, known angular deviations allow compensation of indicated values. Hence, measurement accuracy can be significantly increased.

2 CALIBRATION APPROACHES

The calibration process can be carried out in several ways. We can divide calibration techniques into two groups. Firstly, methods referred to as cross-calibration. These techniques are carried out in the way mentioned above, i.e., comparison of the device under test against a reference standard. For example, an autocollimator against precision optical polygon is widely used in metrology laboratories, such as Czech metrology institute [2], or National Institute of Standard and Technology. This

method with optical polygon offers very high accuracy, however, the angular step is limited by the number of polygon faces. Another option is to employ a laser interferometer system with retroreflective optics [3], [4].

Secondly, methods referred to as self-calibration are based on the device capability to determine its deviations. Thus, the external angular standard is not necessary and calibration of the device under test is accomplished by the device itself [1]. To achieve this self-calibration functionality, a feature of a circle closure is applied. The total sum of plane angular intervals around any point inside any closed curve is exactly 2π rad. Hence, self-calibration devices require special design which increases the cost of the device. Also, at least one full rotation run from 0° to 360° is essential for the self-calibration process. As a result, a self-calibration approach is not applicable for rotary axes with reduced operating range. On the other hand, self-calibration techniques have many advantages, such as calibration directly on the application axis under real running conditions, no external standard, zero-cost and time-saving calibration, etc. More information about self-calibrate rotary tables can be found in [5], [6].

3 EXPERIMENTAL SETUP

The device under test is a rotary table SDL 1401 manufactured by RMS. This rotary table allows free turning and provides several operating modes including absolute positioning mode with a resolution of 0.18 arcsec, repeatability of ± 1 arcsec, and absolute accuracy of ± 4 arcsec [7]. The motion controller of a direct drive motor can be controlled by using a touch panel or remotely. One of the main goals of this paper is to verify absolute angular positioning error. Another goal is to obtain an error map to increase angular positioning accuracy by correcting indicated values.



Figure 1: Calibration of a rotary table

The angular standard is represented by Keysight 5530 laser interferometer system cooperating with rotary axis measurement kit E5290K by Cullam Technologies. A rotary axis measurement kit consists of a precision rotary table, which is mounted on a device to be calibrated. The laser system with the rotary axis measurement kit offers total absolute accuracy of 1.0 arcsec, repeatability of 0.8 arcsec, and resolution of 0.02 arcsec [8]. The laser system maintains its accuracy even if it is mounted

on the to be calibrated rotary axis. Typically, the dominant contributor to the total error budget is caused by imperfect installation due to misalignment of the to be calibrated rotary axis and the rotary axis of the reference rotary table. Eccentrics movement is interpreted as corresponding angular displacement. Hence, the indicated value of angular displacement also contains invalid angular displacement acquired by eccentricity movement. The magnitude of eccentricity error negatively affects total accuracy. In general, the more accurate measuring system, the higher requirements for its installation. For the measuring system with an accuracy of 1 arcsec, the eccentricity has to lie within several microns and the centering of the rotary axis becomes challenging. To avoid this, compensation of the eccentricity error is crucial. For this reason, the 5530 laser calibration system employs an angular interferometer, which is sensing the position of the retroreflector mounted on the precision (i.e. reference) rotary table. Change in relative position is measured and used for angular position compensation.

In addition, the accuracy of the laser measurements is affected by environmental conditions, such as air temperature, air pressure, and relative humidity. To accomplish compensation of these error producing conditions, the Keysight 5530 laser system employs a remote environmental sensor. The environmental data are processed in the metrology software E1733A, where a compensation factor is computed and applied for the measurement.

The optical path of the Keysight 5530 laser calibration system is comprised of a laser head 5519B, an angular interferometer, and retroreflective optics. An experimental setup is illustrated in Figure 1. A retroreflector is mounted on the top of the precision calibration rotary table E5290K. An angular interferometer is placed in front of the retroreflector. The laser head is aligned so its beam goes through an angular interferometer and enters the port of the retroreflector and is perpendicular to the port surface.

4 RESULTS

Calibration of the rotary table was carried out at steps of 30° from 0° to 360° . Each angular interval was measured five times in each approach direction (forward and backward). Therefore, ten calibration values per measuring step were acquired, and the mean values of the deviations were calculated as shown in Figure 2. This graph also includes the maximum and minimum envelope obtained by repeated measurements.



Figure 2: Result of the calibration

It can be seen, that the greatest angular deviation reaches 23,5 arcsec and it is related to the nominal angle of 150° . Figure 3 shows the standard deviation of each measured angular position. The maximal value of 1.3 arcsec corresponds to the repeatability of the rotary table. Note that the standard deviation of each nominal position is taken from five measurements by forward and backward approach direction. Thus, the estimated bidirectional standard deviation involves the hysteresis effect. Hysteresis presented in the specification is ± 1 arcsec. Considering the specified repeatability and hysteresis, both of ± 1 arcsec, the estimated standard deviation complies with the specification.

The resulting uncertainty of the calibrating method assumes zero eccentricity error due to laser compensation. Also, compensation for the environmental conditions and proper installation is assumed. Uncertainty of type A was estimated to 0.4 arcsec. Absolute accuracy of ± 1.0 arcsec of the E5290K measurement kit represents an error contributor for the uncertainty of type B, which is estimated to 0.6 arcsec using rectangular distribution. The expanded uncertainty of the reference calibrating system (coverage factor of 2) is evaluated to 1.5 arcsec.



Figure 3: Standard deviation of positioning error

5 CONCLUSIONS

This article has only been able to touch on the most general features in the field of rotary axis calibration. Employing of commercial laser calibrating system Keysight 5530 is straightforward and provides high precision and resolution. The resulting angular deviations from the calibration reveal an angular positioning error map which can be applied to compensate positioning error of the SDL 1401 rotary table. It can be noted that the measured accuracy of the calibrated rotary table does not meet the specified accuracy. The worst case of the measured angular deviation reaches 23.5 arcsec.

The resulting position accuracy is not as good as we expected. On the other hand, the rotary table has been using for many years and the performance of the rotary table could be slightly reduced due to aging and wear. Thus, the parameters of the motion controller may need to be updated. Specification of the E5290K rotary axis measurement kit allows installation with 1 mm geometrical offset between measured and reference axis due to eccentricity compensation by the 5530 laser calibrator system. For this reason, error contributors for the type B uncertainty estimation were reduced to specified accuracy of ± 1.0 arcsec and other contributors were neglected.

In future work, a detailed uncertainty analysis will be realized. Also, an experiment to verify the actual measure of eccentricity compensation of the calibrating system will be carried out. After this, the presented error map in this paper can be verified.

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- [1] WATANABE, Tsukasa. Is an angular standard necessary for rotary encoders? Development of a rotary encoder that enables visualization of angle deviation. Synthesiology English edition [online]. 2009, 1(4), 273-281. DOI: http://doi.org/10.5571/syntheng.1.273
- [2] Státní etalon rovinného úhlu. Státní etalon rovinného úhlu | Český metrologický institut [online]. Brno: ČMI, 2006, https://www.cmi.cz/statni%20etalon%20rovinneho%20uhlu
- [3] KINNANE, Mark N, Lawrence T HUDSON, Albert HENINS a Marcus H MENDENHALL. A simple method for high-precision calibration of long-range errors in an angle encoder using an electronic nulling autocollimator. Metrologia [online]. 2015, 52(2), 244-250. ISSN 00261394. DOI: 10.1088/0026-1394/52/2/244
- [4] ZAHWI, S., M. AMER, M. A. ABDO a A. EL-MELEGY. Calibration of Optical Rotary Tables using Autocollimators and Laser Interferometer Systems. 10th IMEKO TC14 Symposium on Laser Metrology for Precision Measurement and Inspection in Industry [online]. Braunschweig, 2011, https://www.imeko.org/publications/tc14-2011/IMEKO-TC14-2011-43.pdf
- [5] WATANABE, Tsukasa, Masahito KON, Nobuo NABESHIMA a Kayoko TANIGUCHI. An angle encoder for super-high resolution and super-high accuracy using SelfA. Measurement Science and Technology [online]. 2014, 25(6). ISSN 09570233. DOI: 10.1088/0957-0233/25/6/065002
- [6] KOKUYAMA, Wataru, Tsukasa WATANABE, Hideaki NOZATO a Akihiro OTA. Angular velocity calibration system with a self-calibratable rotary encoder. Measurement [online]. 2016, 82(1), 246-253. ISSN 02632241. DOI: 10.1016/j.measurement.2016.01.011
- [7] DYNAMIC MOTION-SIMULATOR WITH TEMPERATURE CHAMBER SDL 05 / SDG10. RMS Dynamic Test Systems [online], http://www.dspmindustria.it/img/cms/sdl05sdg10.pdf
- [8] Cullam Technologies E5290K Rotary Axis Measurement Kit. CULLAM Technologies Co., Ltd.

TECHNIQUES FOR AVOIDING MODEL OVERFITTING ON SMALL DATASET

Lukas Kratochvila

Doctoral Degree Programme (2), FEEC BUT E-mail: kratochvila@feec.vutbr.cz

> Supervised by: Karel Horak E-mail: horak@feec.vutbr.cz

Abstract: Building a deep learning model based on small dataset is difficult, even impossible. To avoiding overfitting, we must constrain model, which we train. Techniques as data augmentation, regularization or data normalization could be crucial. We have created a benchmark with a simple CNN image classifier in order to find the best techniques. As a result, we compare different types of data augmentation and weights regularization and data normalization on a small dataset.

Keywords: Deep Learning, Dataset size, Overfitting, Data Augmentation, Regularization, Image Classification, Batch Normalization, Data Normalization

1 INTRODUCTION

Recent deep learning algorithms deals with several problems. One of them is overfitting on the training dataset. Overfitting is a situation, when the model loss on the validation set (which represents the test set during training) starts increasing, but the loss on the training set still decreases. This situation is called memorising of the training dataset because we lose the generalization property of the model and we are only remembering the training set. The example of the overfitting is shown in Fig. 1.



Figure 1: Example of model overfitting on training dataset.

2 METHODS

There are several methods preventing overfitting. The first one considers dataset and model size. When the model overfits, we probably have too intricate model, too small dataset, or both. This approach refers to Occam's Razor principle. This principle states "Entia non sunt multiplicanda praeter necessitatem"[1]. In deep learning models this principle says that we need the simplest model, which can represent the data. We can use statistics to find the right size of the model. The Eq. 1 represents an a posteriori probability for model M_i . Rasmussen and all.[2] suggest that a posteriori probability can show needed model complexity.

$$P(M_i|Y) = \frac{P(Y|M_i)P(M_i)}{P(Y)}$$
(1)

When we do not have enough data, we can try to create more with data augmentation. Augmentation is a process, which creates new data samples with transformation which preserves annotation. We can use two types of data augmentation. The first type is geometrical transformation, e.g. rotation, cropping, or flipping. The second kind is photometric transformation, e.g. colour shifting, brightness, or contrast adjustment.

The second approach to create more data is regularization. The idea of regularization came from maximum a posteriori (MAP) estimation and therefore minimization of the objective function Eq. 2 as we want to eliminate noise.[3]

$$J(f) = \sum_{i=1}^{N} [y_i - f(x_i)]^2 + \lambda \phi[f]$$
(2)

where λ refers to regularization parameter, y_i is the target and $f(x_i)$ is our model. This parameter controls the trade-off between the noise level and strength of priory assumption.[3] Term $\phi[f]$ refers to energy function and can be expressed as Eq. 3.[4]

$$\phi[f] \Leftrightarrow \sum_{j=1}^{M} |w_j|^q \le \eta \tag{3}$$

The η refers to the constraint and with q we can choose, which regularization we want, i.e. q = 1 means Lasso regularization.

The third important approach is normalization. We consider three methods. The first one is data normalization to the interval [0,1]. The second is global contrast normalization, i.e. subtract mean across pixels and normalize by scale, which is either the standard deviation or L1-norm across pixels. And the third is batch normalization. This is a technique normalizes data in mini-batch. For d-dimensional input $x = (x^{(1)}...x^{(d)})$, the normalization is performed over dimension as show Eq. 4.[5]

$$x_{new}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{Var[x^{(k)}]}}$$
(4)

3 TOOLS AND LIBRARIES

For the purpose of increasing dataset size, when refering to the first approach to deal with model overfitting, we have collected tools and libraries for this task.

The authors in [6] created a CNN and compared some photometric and geometric data augmentations. Namely, authors compare flipping, rotating, cropping, color jittering, edge enhancement, and fancy PCA. Colour jittering and image cropping are the two most successful processes.

In the [7] authors create a rigorous comparison of shallow and deep learning methods. For the boost, they use geometric data augmentation.

Useful Python-base implementation for augmentation is *Imgaug*.[8]

3.1 AUTOMATED AUGMENTATION

AutoAugment [9] is an approach for dealing with small dataset. It uses reinforcement learning for finding the best augmentation policy. The algorithm can work in two ways: direct or transfer. The direct way finds a new policy for the current dataset. On the other side transfer way transfers learned policy to new dataset. Direct AutoAugment could be very slow, therefore Sungbin et al. developed faster AutoAugment named FastAutoAugment.[10]

An alternative could be the algorithm from Ratner et al., which uses General adversarial network for finding ideal augmentation policy.[11]

3.2 **REGULARIZATION**

Some of the regularization techniques are implemented in deep learning frameworks, e.g. in Tensorflow are implemented layer regularizers, which are written for this purpose. In PyTorch is implemented L2-regularization. Different regularizations must be hardcoded.[12]

4 EXPERIMENTAL SETUP

We have constructed a benchmark to consider dataset size, regularizations and preprocessing w.r.t model performance. This benchmark consist of 24 experiments, i.e. we have 6 different approach for 4 datasets. The approches are *normalization*(N), *global constrain normalization*(GC), *L2-regularization*(R2), *batch normalization*(BN), *application of ImageNet Policy*(AA) and *application of CIFAR10 Policy*(AC).

4.1 DATASET

As dataset we have 1225 images with resolution 640x480px divided into 4 classes. The classes are named *in order* with 474 images, *broken* with 465 images, *bad colour* with 158 images and *foreign object* with 128 images. One example of each class is shown on Fig. 2. From this base of the dataset we construct 4 datasets with 100, 200, 300 and 400 samples. In each dataset, the label occurance is balanced, i.e. in the dataset with 400 samples are 100 samples from each class, rest of the images from base dataset are in test dataset. For the validation dataset the first 100 samples from each class of the test dataset were chosen if possible.



Figure 2: From left to right classes examples: in order, broken, bad colour and foreign object.
4.2 ARCHITECTURE

We choose a simple classification network for the evaluation. The network is created in the PyTorch library. The summary is shown in List. 1. Layer *BatchNorm* (3,6,9) is used only when specified. For each convolution and fully connected layer (Linear in the listings) we manually initialize weights.

```
INFO:MY_LeNetBN: Trainable parameters: 266592
INFO:MY_LeNeBNt:MY_LeNetBN(
(1) (conv1):Conv2d(3,32,kernel_size=(5,5),stride=(1,1),padding=(2,2),bias=True)
(2) (pool1):MaxPool2d(kernel_size=2,stride=2,padding=0,dilation=1,ceil_mode=False)
(3) (bn1): BatchNorm2d(32,eps=0.0001,affine=False,track_running_stats=True)
(4) (conv2):Conv2d(32,64,kernel_size=(5,5),stride=(1,1),padding=(2,2),bias=True)
(5) (pool2):MaxPool2d(kernel_size=2,stride=2,padding=0,dilation=1,ceil_mode=False)
(6) (bn2): BatchNorm2d(64,eps=0.0001,affine=False,track_running_stats=True)
(7) (conv3):Conv2d(64,128,kernel_size=(5,5),stride=(1,1),padding=(2,2),bias=True)
(8) (pool3):MaxPool2d(kernel_size=2,stride=2,padding=0,dilation=1,ceil_mode=False)
(9) (bn3): BatchNorm2d(128,eps=0.0001,affine=False,track_running_stats=True)
(10)(fc1): Linear(in_features=2048, out_features=4, bias=True)
```

Listing 1: Model summary

In the input, images are rescaled to 32x32px. The output is the same as the number of classes. Every experiment was trained for 100 epochs after 50 epochs was learning rate multiplied by 0.1. The rest of the hyperparameters are listed in Tab. 1.

Optimizer	Init learning rate	batch size	Activation function	Weight Initialisation
Adam	0.001	128	Leaky ReLu	Xavier

 Table 1: Network hyper-parameters

4.3 **RESULTS**

For evaluation, we chose the average Top-1 accuracy metric. This metric consider only the best result and average throughout all classes. Results on the test dataset are shown in Tab. 2. The best results though experiments are bold.

	Test average Top-1 accuracy [%]			
Experiment	100	200	300	400
baseline	74.22	88.02	87.35	89.94
N	74.49	77.46	78.46	81.45
GC	70.04	83.71	89.08	89.09
BN	72.89	78.44	79.46	82.42
R2	74.09	83.90	87.24	89.33
AA	64.53	68.98	72.86	78.30
AC	63.29	70.54	77.08	76.00

Table 2: Experiment results: columns refer to the dataset size and rows refer to experiments results.

The Fig. 3 shows training and validation loss for all datasets at baseline. We can see strong overfitting on the training dataset, i.g. the training loss decreasing, but the validation loss stagnates or starts to raise. The Fig. 4 shows training and validation losses for all experiments on dataset with 400 samples.

4.4 **DISCUSSION**

Data augmentation methods does not reach the best results, but from comparison of the training and validation loss on the Fig. 4, we can see that the margin between the losses is reduced compare to



Figure 3: Training performance on baseline for all four datasets

the baseline. From this point of view the data augmentation methods improves model generalization ability. For the *batch normalization* (BN), we can see the same reduction as for the data augmentation.

On the Fig. 4 the validation loss for experiment *normalization* (N) and *global constrain normalization* (GC) achieves the best results on validation dataset. This is not seen from Tab. 2 and it could be caused by small size of testing dataset.



Figure 4: Training performance on all experiments for dataset with 400 samples.

5 CONCLUSION

This paper discuss some of the methods for avoiding overfitting on small training dataset. Focus is on three differrent approaches: data augmentation, regularization and data normalization. Necessary

theory is reminded with examples of existing implementations. The main acquisition of this work is benchmark situated for very small dataset. Based on the results, we conclude that data normalization methods are crucial for preventing overfitting on small dataset. Discussion in subsection 4.4, leads us to conclusion that the all methods mentioned above can help with overfitting on the training dataset.

In future work, we would like to extend developed benchmark and develop metric for ranking datasets and their augmentation versions. This metric could amplify evolving models. Next step will be creation of automated library like [9], [10] or [11] for easy use.

6 ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-20-6205 - "Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- [1] W. M. Thorburn, "The myth of occam's razor," Mind, vol. 27, no. 107, pp. 345–353, 1918.
- [2] C. E. Rasmussen and Z. Ghahramani, "Occam's razor," Advances in neural information processing systems, pp. 294–300, 2001.
- [3] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural computation*, vol. 7, no. 2, pp. 219–269, 1995.
- [4] M. Mishra, "Regularization: An important concept in machine learning." https: //towardsdatascience.com/regularization-an-important-conceptin-machine-learning-5891628907ea, 2018. [Online; accessed 10-Mar-2021].
- [5] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, PMLR, 2015.
- [6] L. Taylor and G. Nitschke, "Improving deep learning using generic data augmentation," *CoRR*, vol. abs/1708.06020, 2017.
- [7] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *CoRR*, vol. abs/1405.3531, 2014.
- [8] A. B. Jung, "imgaug." https://github.com/aleju/imgaug, 2021. [Online; accessed 1-Jan-2021].
- [9] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [10] S. Lim, I. Kim, T. Kim, C. Kim, and S. Kim, "Fast autoaugment," *CoRR*, vol. abs/1905.00397, 2019.
- [11] A. J. Ratner, H. R. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré, "Learning to compose domain-specific transformations for data augmentation," *Advances in neural information processing systems*, vol. 30, p. 3239, 2017.
- [12] E. Stevens, L. Antiga, and T. Viehmann, *Deep Learning with PyTorch*. Manning Publications, 2020.

Doktorské projekty

Mikroelektronika a technologie, Elektronika a komunikace I.

GEL POLYMER ELECTROLYTES BASED ON ETHYLENE GLYCOL DIMETHACRYLATE AND THEIR ELECTRO-CHEMICAL CHARACTERIZATION

Iuliia Veselkova

Doctoral Degree Programme (4), FEEC BUT E-mail: xgrach00@vutbr.cz

Supervised by: Marie Sedlaříková

E-mail: sedlara@feec.vutbr.cz

Abstract: The electrochemical properties of cross-linked gel polymer electrolyte based on ethylene glycol dimethacrylate are investigated. Impedance spectroscopy was used for studying the mode of electrical conductivity in prepared gel electrolytes. A study of electrochemical stability was provided using linear sweep voltammetry. The increasing amount of ethylene glycol dimethacrylate shows the decreasing of electrical conductivity. The presence of cross-linking agent affects not only the electrochemical properties but also the mechanical properties.

Keywords: gel electrolyte, polymer, methyl methacrylate, cross linking.

1 INTRODUCTION

Gel polymer electrolytes (GPEs) have attracted a big interest in the field of lithium-ion batteries due to their good features. It is one of the most desirable alternatives among various electrolytes for electrochemical devices [1]. In comparison with liquid electrolytes, GPEs have high ionic conductivity and wettability, which is unique [2]. Also, gel electrolytes eliminate the problem with leakage of electrolytes due to cohesive properties and a problem with fire hazards due to the high flammability of organic solvents [3], [4]. To reach high mobility close to that of the liquid fraction and to reach mechanical cohesion, a cross-linking strategy was used [4].

Crosslinking is a strategy for improving a gel polymer electrolyte with good mechanical strength. The network structure of cross-linked polymers is characterized by the higher density of cross-linking points and the consequently reduced deformability of the chain segments. The cross-linking is formed by covalent chemical bonds or by physical interaction [5], [6].

The main goal of the research was to determine the impact of cross-linker on the electrochemical properties of GPEs. Gel polymer electrolytes based on methyl methacrylate were polymerized by UV irradiation with varying concentrations of ethylene glycol dimethacrylate as a cross-linking agent. Results indicated the temperature dependence of electrical conductivity and electrochemical stability of gel polymer electrolytes, which were prepared.

2 PREPARATION OF GEL POLYMER ELECTROLYTE

Lithium hexafluorophosphate-98% (LiPF₆), ethylene carbonate (EC), diethyl carbonate (DEC), ethylene glycol dimethacrylate (EDMA), methyl methacrylate (MMA) and benzoin ethyl ether (BEE) were chosen for preparation of gel polymer electrolytes. All these materials were purchased from Sigma-Aldrich (Merck) [7].

The preparation of gel polymer electrolytes can be divided into two phases. The first phase is the preparation of a conductive solution (liquid electrolyte) and the second phase is the formation of the gel structure of the prepared solution.

The preparation of the solution involves mixing chemicals. The conductive salt LiPF₆ must be dissolved in an organic solvent EC/DEC (1:1 by weight) along with the monomer MMA, the crosslinking agent EDMA and the initiator of polymerization BEE. After being stirred for 20 minutes all materials were dissolved, and the prepared solution can be a move to a second phase. The second phase can be described as the conversion of the liquid state of a conductive solution into a gel structure. It means that the prepared and mixed solution then be poured into a form, which is placed in a chamber with UV-light and a radiation intensity of 1250 μ W/cm² [7], [8], [9].

3 MEASUREMENT TECHNIQUES

The gel samples were sandwiched between two stainless steel electrodes in the electrochemical testing cell for measurement of electrochemical properties.

Electrical conductivity was measured by electrochemical impedance spectroscopy with a frequency range between 1 MHz and 0.1 Hz. There were 6 steps per decade and the amplitude sinusoidal signal was 10 mV. The value of electrical conductivity γ was calculated from Eq. (1) [7]:

$$\gamma = \frac{1}{R} \cdot \frac{h}{s} \tag{1}$$

where R is the bulk resistance, h is the thickness and S is the surface area of the gel sample.

The electrochemical interface stability of gel samples was performed using linear sweep voltammetry in the voltage range between 0.1 V and 5.1 V with sweep speed 0.5 mV/s. Values of the potential window were calculated for 5 μ A/cm² of limiting current density [7], [8], [9]. The measurements have been provided at room temperature 23 °C.

4 RESULTS AND DISCUSSION

Gel polymer electrolytes with EDMA were prepared during the experiment. The amount of EDMA was changed between 0.35 mol% and 3.5 mol%. Samples for electrochemical characterization were prepared in an argon-filled glove box JACOMEX and measurements taken using a Biologic potentiostat [7].

Electrical conductivity is one of the important factors in the gel polymer electrolyte. Table 1 presents values of electrical conductivity of GPE at room temperature 23 °C.

Amount of EDMA	Electrical conductivity
0.35	3.16
0.75	4.27
1.0	3.21
1.35	3.05
1.75	3.51
2.0	1.81
2.35	2.32
2.75	1.96
3.0	1.95
3.5	1.74

Table 1: Electrical conductivity of gel polymer electrolytes at room temperature

From Table 1 can see that the electrical conductivity generally decreases with an increase in the amount of cross-linking agent EDMA. The cross-linking reaction causes an increase in the resistance for ion transport, which decreases the electrical conductivity with increasing cross-linking density [10]. The highest electrical conductivity has a gel sample with a 0.75 mol% amount of EDMA.

Figure 1 shows the temperature dependence of the electrical conductivity of GPE based on EDMA. The temperature was changed from 30 °C to 60 °C. The Arrhenius plots illustrated an increase of electrical conductivity with increasing temperature due to enhancement of the ionic mobility.



Figure 1: Arrhenius plots of electrical conductivity of GPEs

Values of the potential window were presented in Tab.2. These values describe the electrochemical stability of prepared gel samples.

Amount of EDMA [mol%]	Potential window U [V]
0.35	4.07
0.75	3.61
1.0	2.06
1.35	3.11
1.75	3.48
2.0	3.86
2.35	3.77
2.75	2.80
3.0	3.89
3.5	3.70

 Table 2:
 Potential windows of gel polymer electrolytes based on EDMA

The gel polymer electrolytes with an amount of EDMA of 0.35 mol% has a wider potential window in comparison with other prepared samples. It means that this sample more stable.

Fig. 2 and 3 presents curves of current densities of prepared gel samples. The curves have some peaks, which appear as a potential window, where become oxidation of the surface of gel polymer electrolyte. It may be due to some particles, which stay in a liquid state, some chemical reactions on the electrolyte surface.



Figure 2: Linear sweep voltammograms of gel polymer electrolytes (amount of EDMA 0.35 mol% - 1.75 mol%)



Figure 3: Linear sweep voltammograms of gel polymer electrolytes (amount of EDMA 2.0 mol% - 3.5 mol%)

5 CONCLUSION

Cross-linked gel polymer electrolytes were introduced. The electrochemical properties such as electrical conductivity and electrochemical stability were measured of prepared gel polymer electrolytes. The highest electric conductivity has GPE with the concentration of EDMA of 0.75 mol%. The electrochemical stability of this gel sample was 3.61 V. Comparison of all samples show the decreasing tendency of electrical conductivity with the increasing EDMA concentration. However, electrical conductivity increases with the temperature increase. In contrast with electrochemical properties, the mechanical properties are improved with the increase of a concentration of the crosslinking agent.

ACKNOWLEDGEMENT

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-20-6206.

REFERENCES

- [1] L. Wangyu, P. Ying, L. Jingyuan: A PEO-based gel polymer electrolyte for lithium ion batteries, RSC Advances. 2017, 38, 23494-23501
- [2] Niu Yu-Bin, Yin Ya-Xia, Wang Wen-Peng: In situ copolymerizated gel polymer electrolyte with cross-linked network for sodium-ion batteries, CCS Chemistry. 2020, 589-597
- [3] Shin, Won-kyung, Cho J., Kannan G. A., Lee Y., Kim D.: Cross-linked Composite Gel Polymer Electrolyte using Mesoporous Methacrylate-Functionalized SiO₂ Nanoparticles for Lithium-Ion Polymer Batteries, Scientific Reports. 2016, 6(1)
- [4] Hosseinioun A., Nrnberg P., Schnhoff M., Diddens D., Paillard E.: Improved lithium ion dynamics in crosslinked PMMA gel polymer electrolyte, RSC Advances. 2019, 9(47), 27574-27582
- [5] Maitra J., Shukla V. K.: Cross-linking in Hydrogels A Review, Americal Journal of Polymer Science 4(2) (2014) 25-31
- [6] Wong R., Ashton M., Dodou K.: Effect of crosslinking agent concentration on the properties of unmedicated hydrogels, Pharmaceutics. 2015, 7, 305-319
- [7] Veselkova I., Jahn M., Sedlaříková M., Vondrák J.: An Influence of Cross-linking Agent on Electrochemical Properties of Gel Polymer Electrolytes, Renewable Energy Sources: Engineering, Technology, Innovation. 2019, 1035-1042.
- [8] Veselkova I, Sedlaříková M.: Preparation and electrochemical characterization of gel polymer electrolyte based on tetraalkylammonium tetrafluoroborate for electrochemical power sources, ECS Transaction. 2020, 31-37
- [9] Veselkova I., Sedlaříková M., Fafilek G., Gierl-Mayer C.: Electrochemical and thermal properties of gel polymer electrolytes modified by flame retardants, ECS Transaction. 2019, 47-55
- [10] Yun, Ye, Ji-ae Choi a Dong-won Kim.: Lithium polymer batteries assembled with in situ cross-linked gel polymer electrolytes containing ionic liquid, Macromolecular Research. 2013, 21(1), 49-54

XRD ANALYSIS OF CHEMICAL COMPOSITION ON LEAD SURFACE MICROELECTRODES

Jan Smejkal

Doctoral Degree Programme (2), FEEC BUT E-mail: xsmejk18@stud.feec.vutbr.cz

Supervised by: Ladislav Chladil

E-mail: chladil@feec.vutbr.cz

Abstract: This work is focused on the observation of changes in the chemical composition on the surface of the negative electrode of the lead-acid battery using an XRD diffractometer. The measurement was performed on thin pasted electrodes with a thickness of about 1mm. After the maturation and molding process, the fully charged negative electrode was gradually discharged and charged at a rate of 0.2 C and 0.3 C. The negative electrode was analyzed at 100 %, 80 %, 60 %, 40 % 20 % a 0 % SOC states. Chemical changes occurring on the surface of the negative electrode during discharging and charging of the lead-acid battery were measured using X-ray diffractometer Rigaku MiniFlex HR 600.

Keywords: lead-acid, accumulator, negative electrode, X-ray diffractometer, morphology

1 INTRODUCTION

Although the principle of operation of lead-acid batteries has been known for a very long time, this type of battery is again intensively studied. Lead-acid batteries are widely used due to the good availability of the materials used, their recyclability, high electrochemical efficiency, the high number of cycles, and high rate of voltage around 2 V. These properties together with the low production cost make them very attractive for use in the automotive industry, where they can supply high current for motor starters and also in electromobility. Large arrays of lead batteries are used as backup power sources for data and data centers. Lead-acid batteries are also widely used for the long-term storage of energy produced from renewable sources. Today, more than half of the world's lead production is used to make batteries [1].

The degradation mechanisms of accumulators are closely related to the materials used. If one of the undesirable properties is improved, it usually leads to a deterioration of the other properties. One of the most important degradation mechanisms is the PCL phenomena (Premature Capacity Loss). PCL phenomena were formerly known as antimony free effect mainly because the first cases of this degradation mechanism were observed in batteries in which antimony in the lead alloy (Pb-Sb) positive electrode was replaced by a lead-calcium (Pb-Ca) alloy [2]. We currently know three main PCL phenomena, namely PCL 1, PCL 2, and PCL 3. PCL phenomena cause premature loss of capacity and thus reduce the ability to recharge and charge batteries. At the moment, the research team is trying to suppress these negative phenomena as much as possible.

PLC 3 phenomenon was discovered as the last of the phenomena of premature loss of capacity. At the moment when the negative phenomena on the positive mass of the electrode were suppressed, the attention of the research teams turned to the problems related to the negative electrode. To achieve optimal results for the life of a lead-acid battery, we can only achieve this if the positive and negative plates retain their optimal properties. In PLC 3 phenomenon, the negative electrode is sulfated. This phenomenon is often associated with the selective discharge of the negative electrode, which is the result of a combination of the high rate of hydrogen evolution and high efficiency of oxygen recombination. A key parameter for the long life of the electrodes is an even distri-

bution of redox reactions in the entire volume of the electrode during charging and discharging. Even distribution of these reactions prevents the long-term presence of sulfate in certain parts of the electrode, which leads to gradual sulfation and loss of the ability to supply currents at higher loads [3].

According to Dr. Moseley, there are several possible causes of a significant loss of capacitance affecting the negative electrode. One of the causes is the reduction of the active area of the electrode, which is caused by the loss of organic expanders. Organic expanders are destroyed by oxygen, which reaches the negative electrode during the oxygen cycle [4].

Additives such as carbon, glass fibers, or titanium dioxide can be used to suppress the PLC 3 phenomenon. These additives are added to the negative active mass and can reduce and slow down the sulfation of the negative electrode [5, 6, 7].

2 MEASUREMENT

For the experiment of chemical changes on the surface of the negative electrode, it was necessary to create thin electrodes, which were pasted on a lead substrate of circular shape with a diameter of 13 mm. 24 holes were drilled in the lead substrate. These holes allow better adhesion of the active mass to the surface of the lead substrate. To precisely define the reaction area of the electrode, a thin layer of epoxy adhesive was applied to the part of the electrode that did not participate in the chemical processes. 10 g of the mixture was prepared for the measurement and the exact composition of the active mass of the negative electrode is given in Tab. 1. The electrode was then placed in a holder made of PEEK (semicrystalline thermoplastic). Before the measurement, the electrode was placed in the holder so that there were no movements that could adversely affect the measurement. The electrode matured for 5 days in a controlled high humidity environment. The total weight of the active substance of the electrode after maturation was 0.82 g. After maturation, the electrode was placed in a measuring cell with an electrolyte consisting of an aqueous solution of sulfuric acid at a concentration of 27 %. 20 formation cycles were performed in the measuring cell at a supplied current of 13.11 mA.

Ingredients	Lead powder	Demineralized water	Sulfurid acid	Borosilicate	Barium sulphate
Quantity (g)	8.42	0.92	0.51	0.02	0.13

Table 1: Exact composition of the active mass of the negative electrode

Before measurement, the active substance of the electrode was instilled with an electrolyte, which consists of an aqueous solution of sulfuric acid with a concentration of 33 %. The electrode was then covered with a 12 μ m thick polyamide foil. The foil is attached to the holder with a rubber ring to prevent air from entering the active material of the electrode (see Fig. 2).



Figure 1: On the left there is a picture of the detail of the lead substrate. On the right there is a picture of the complete test electrode covered with polyamide foil.

The analysis of the surface of the active substance took place in a three-electrode circuit. The electrode was immersed in an electrolyte consisting of a 33% aqueous sulfuric acid solution. After the formation cycles, the fully charged electrode was gradually discharged. The discharge was perfor-

med by the galvanostatic method in five steps at a supplied current of 18.86 mA for 1 hour with a potential limitation $E_M = -0.8$ V vs. MSRE. The electrode was discharged at a rate of 0.2 C and subsequently analyzed in XRD at 100 %, 80 %, 60 % 40 % 20 %, and 0 % SOC (State of Charge). Charging was also performed by the galvanostatic method in five steps at a supplied current of 18.86 mA for 1 hour with a potential limitation $E_M = -1.3$ V vs. MSRE. For a speed of 0.3 C, the same setting was used for the experiments, only a current of 31.43 mA was supplied for 31 minutes. Chemical composition on the surface of the negative electrode during charging and discharging was observed using X-ray diffractometer Rigaku MiniFlex HR 600. Diffractograms were measured over a range of 10-100°at step 0.01, IHS 5, divergence slit of 0.625, and k β filter 0.03.

3 RESULTS AND DISCUSSION

During the discharge, in each of the states 100 %, 80 %, 60 %, 40 %, 20 %, and 0 % SOC, the chemical composition on the surface of the active mass of the negative electrode was measured by XRD. The results of the chemical composition can be seen in Fig. 2.



Figure 2: Chemical composition on the surface of the active mass of negative electrode at different SOC at a rate 0.2 C for discharging and charging.

It can be seen that when the electrode is fully charged, the electrode contains 8 mol. % lead sulfate. In the fully discharged state, the electrode contains almost 32 mol. % metallic lead. The composition on the surface of the electrode in the fully charged and discharged state indicates that at almost 40 mol. % of active material does not convert during cycling. It is also possible to observe the prior dissolution of the metal lead when discharging between 80 % and 60 % SOC. There was a decrease in the lead on the surface of the electrode by 18 mol. %, in contrast, in the last part of the discharge from 20 % to 0 % SOC there was a decrease of only 8 mol. %. This half-decrease in lead describes a two-stage discharge. In a two-stage discharge, lead ions first begin to dissolve into the electrody surface.

Subsequently, the electrode was charged at a rate of 0.2 C Charging was also performed by the galvanostatic method in five steps at a supplied current of 11.35 mA for one hour with a potential limitation $E_M = -1.3$ V vs. MSRE. At 0 % SOC during the charge cycle, the electrode contained 30 mol. % lead. This amount of lead correlates with discharge results. The electrode was again measured at 0 %, 20 %, 40 %, 60 %, 80 % a 100 % SOC. Up to 40 % SOC, the electrode was made of lead sulfate. From 60 % SOC, the electrode was already largely lead. At full charge, there was 14 mol. % of lead sulfate on the electrode, which is 6 mol. % more than at 100 % SOC at discharging cycle.

In Fig. 3 we can observe a comparison of the molar proportion of lead and lead sulfate when charging and discharging the electrode at a rate of 0.2 C. When charging and discharging the electrode, there were no significant variations in the proportion of Pb and $PbSO_4$ and the course of representation of elements on the surface of the electrode was almost linear.



Figure 3: Representation of the surface composition of the active substance at different SOC at a rate 0.2 C during charging and discharging.

At a higher discharge rate of 0.3 C, it can be observed that in the fully charged and discharged state there is a significantly higher amount of active material on the electrode surface, which does not convert from lead to sulfate and vice versa when cycling the negative electrode. Even though a smaller amount of active material was used at a higher charging and discharging speed, almost the same charge was delivered at both speeds.



Figure 4: Representation of the surface composition of the active substance at different SOC at a rate of 0.3 C during charging and discharging.

4 CONCLUSION

As can be observed from the measured results, there is almost 40 mol. % on the negative electrode of the active substance in which no material is converted during cycling. The addition of a carbon additive to the active substance could help to reduce the percentage of unused material, and thus

contribute to increasing the capacity of the lead-acid battery. Furthermore, during discharging, there was a negative phenomenon of preferential dissolution of lead at the beginning of the discharging cycle. The preferential dissolution of lead leads to an increase in sulfates on the surface of the electrode and then in the entire volume of active mass. During charging, the conversion of lead to sulfate was almost linear. Only in the range of 60 % - 80 % SOC did the conversion of lead to sulfate accelerate. Furthermore, it was possible to observe that the electrode can absorb an equally large charge at a higher charging and discharging speed, even though there was a lower amount of active material on its surface, which is converted into charged or discharged form.

One step to better understand the operation of the negative electrode in lead-acid batteries could be to compare the behavior and chemical composition on the electrode surface at different battery cycling speeds. The next step could be to analyze the chemical composition of the negative electrode surface with added additives such as glass fibers or carbon. These additives could significantly contribute to the elimination of the negative properties described above occurring in the cycling of the negative electrode.

ACKNOWLEDGEMENT

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-20-6206.

REFERENCES

- [1] MOUBAYED, Nazih, Janine KOUTA, Ali EL-ALI, Hala DERNAYKA a Rachid OUTBIB. Parameter identification of the lead-acid battery model. 2008 33rd IEEE Photovolatic Specialists Conference [online]. IEEE, 2008, 2008(33), 1-6. ISBN 978-1-4244-1640-0. ISSN 0160-8371. Dostupné z: doi:10.1109/PVSC.2008.4922517
- [2] HOLLENKAMP, A.F. When is capacity loss in lead/acid batteries 'premature'?. *Journal of Power Sources* [online]. 1996, **59**(1-2), 87-98 [cit. 2021-02-10]. ISSN 03787753. Dostupné z: doi:10.1016/0378-7753(96)02306-3
- [3] LAM, L.T., H. CEYLAN, N.P. HAIGH, T. LWIN a D.A.J. RAND. Influence of residual elements in lead on oxygen- and hydrogen-gassing rates of lead-acid batteries. *Journal of Power Sources* [online]. 2010, **195**(14), 4494-4512 [cit. 2021-02-11]. ISSN 03787753. Dostupné z: doi:10.1016/j.jpowsour.2009.12.020
- [4] MOSELEY, Patrick. ALABC 2000 the way ahead. *Journal of Power Sources* [online]. USA, 2001, (95), 218-223 [cit. 2021-02-11].
- [5] BODEN, D.P., D.V. LOOSEMORE, M.A. SPENCE a T.D. WOJCINSKI. Optimization studies of carbon additives to negative active material for the purpose of extending the life of VRLA batteries in high-rate partial-state-of-charge operation. *Journal of Power Sources* [online]. 2010, **195**(14), 4470-4493 [cit. 2021-02-25]. ISSN 03787753. Dostupné z: doi:10.1016/j.jpowsour.2009.12.069
- [6] BAČA, P., P. KŘIVÍK a S. VACULÍK. Negative Lead-Acid Battery Electrodes Doped with Glass Fibres. International Journal of Electrochemical Science [online]. 2015, 2206-2219
 [cit. 2021-02-21]. Dostupné z: https://www.researchgate.net/publication/282822951_Negative_Lead-Acid_Battery_Electrodes_Doped_with_Glass_Fibres
- [7] MICKA, K., M. CALÁBEK, P. BAČA, P. KŘIVÁK, R. LÁBUS a R. BILKO. Studies of doped negative valve-regulated lead-acid battery electrodes. *Journal of Power Sources* [online]. 2009, **191**(1), 154-158 [cit. 2021-02-24]. ISSN 03787753. Dostupné z: doi:10.1016/j.jpowsour.2009.01.014

HISTORY OF ELECTROMOBILITY, FUTURE BATTERY TRENDS AND HOW TO MEASURE THEM

Martin Šedina

Doctoral Degree Programme (1.), FEEC BUT E-mail: xsedin00@stud.feec.vutbr.cz

> Supervised by: Tomáš Kazda E-mail: kazda@feec.vutbr.cz

Abstract: This work deals with the description of the development of electromobility from history to the present day. Problems associated with the use of silicon in anodes and a measuring system for testing the volume changes of batteries are also mentioned.

1 INTRODUCTION

Today's world is full of electrification, it is due to the fact that we can relatively easily get and transform electricity from renewable sources. This causes (r)evolution in all directions of our lives. One of the big industries affected by this big change is automotive. The whole industry is trying to adapt and introducing certain degrees of electrification to most models on the market.

2 ELECTROMOBILITY

While electromobility may seem to be a trend of modern years, it certainly is not. Inventors created their prototypes as early as the 19th century. In 1884, the British inventor Thomas Parker began producing the first production electric car in Britain, thus two years ahead of Karl Benz. Electric cars at that time used lead-acid batteries for propulsion, so they had a short range and were heavy, another disadvantage was that in most cases the batteries had to be removed from the vehicle to recharge them. The big advantage, however, was that they achieved much higher performance and speed. Because of the development of internal combustion engines and the complex handling of heavy batteries, the electric drive was abandoned. Occasionally a manufacturer came up with a model powered by electricity as a demonstration of this technology. One of these vehicles is the BMW 1602e, which served in the parade of the Munich Olympics games. You can see the vehicle on the figure 1 [1].



Figure 1: BMW 1602e [2]

The interest in alternative propulsion was started during the oil crisis in the 1970s. At that time, new battery technologies such as Ni-Cd and Ni-MH batteries were already appearing on the market. First, Ni-Cd batteries were used, vehicles with these batteries had a range of about 100 km, later these batteries were replaced by Ni-MH batteries It was because cadmium is toxic and Ni-MH batteries were also able to offer a higher gravimetric energy density. In this days we can find Ni-MH batteries in some hybrid vehicles [1].

Ni-MH battery The anode of the battery is made of an alloy of metals, it is reaction with hydrogen and produce hydrides (MH). The cathode is made of nickel hydrooxide (NiO(OH)). Potassium hydroxide solution (KOH) serves as the electrolyte. Today, they are used mainly as a replacement for disposable cells. Their advantages are a good price/capacity ration, it has no memory effect and does not contain toxic substances, as a minus we can mention self-discharge, but this has been significantly reduced in today's modern batteries [3].

A new era of electric vehicles began in the 1990s when the world start focusing on ecology and also on the fact that fossil fuel reserves are not infinite. That fact forced car manufacturers to produce their own electrified vehicles, but this cars did not attracted lots of customers. It was due to several aspects. That vehicles had a short range, were expensive, had very limited performance, and another limitation was the almost non-existent infrastructure. One of the 1990s representatives is the GM EV1, which is shown on figure 2 [1].



Figure 2: GM EV1 [4]

The modern boom in electric vehicles was unleashed by Tesla with its model Roadster (figure 3), which began production in 2008. It was the first production electric car which overcome 300 km range. Every car manufacturer on the market produces an electric vehicle. It it because of demonstration of technological sophistication and efforts to reduce fleet emissions. All modern electric cars use Li-ion batteries [1].



Figure 3: Tesla Roadster [5]

Li-ion battery It is the latest and most dynamically developed technology in the field of batteries. The technology of lithium-ion batteries allows use of several types of anode and cathode materials. The most commonly used anode material is carbon (C), lithium titanium oxide (LTO) or silicon (Si). In the case of cathode materials, the aim is to reduce the use of cobalt as much as possible and preserve properties. Today, the most common technologies are lithium nickel-manganese-cobalt oxide (NMC), lithium iron-phosphate (LFP) and lithium nickel-cobalt-aluminum oxide (NCA), there are also lithium-cobalt oxide (LCO) and lithium-manganese oxide (LMO). The electrolyte is a lithium salt dissolved in an organic solvent. This is the most widely used type of battery today, and can be found in all types of wearable electronics, electrified vehicles and in energy storage. The biggest advantages of these batteries are large volumetric and gravimetric capacity, low self-discharge, large terminal voltage, etc., the disadvantages include the high cost, the need of battery management and more complex recyclability [3].

2.1 IMPLEMENTATION OF BATTERIES IN AN ELECTRIC CAR

The battery of an electric car is a complex element whose price exceeds the rest of the components in electric vehicles. The battery pack contains cooling, a control unit, battery management system (BMS) and a several battery modules. The battery pack is designed to make the best use of the vehicle's space and to be able to occupy the vehicle with as many batteries as possible, usually the batteries are stored in the floor of the crew cabin, which strengthens the entire structure. One of the parameters that determines the maturity of the carmaker is the gravimetric capacity of the whole pack. The battery has a voltage between 300-400 V.

Battery modules contain individual battery cells, they can contain cooling and heating, or secondary control units and temperature sensors. Figure 4 shows an divided battery.

The BMS is a control circuit that serves to monitor the voltage of individual cells, and also controls the charging and discharging of individual cells so that they do not exceed the permitted voltage limits. It also monitor the temperatures of the batteries and allow to limit the power flow from the batteries. At the same time, it is balancing the voltage of individual cells, because as the batteries age, it begins to happen that the voltage on all batteries is not the same.

Three types of cells can be used in the modules. Each type of cell has some advantages and disadvantages. The first type are roller batteries. These batteries resemble ordinary disposable cells and come in several sizes. Their advantage is easy thermal management, the biggest disadvantage is the low volumetric density of battery modules, because the batteries can not effectively use the entire space. The second type are prismatic cells. These cells have the shape of a block. Thanks to their shape, they allow full use of the space in the pack and, thanks to the solid packaging, they have a relatively easy temperature management, the disadvantage is the smaller volumetric capacity of the batteries. The third type of battery is a cell cell. The batteries can be of any shape, usually rectangular. They use non-conductive foil as packageand. This is the most widely used type of battery today. Their advantage is a large volumetric capacity and the disadvantage is more demanding temperature management [3].



Figure 4: Look inside the Audi A3 battery [3]

2.2 BEHAVIOR OF NEW LI-ION BATTERIES

Following the efforts of the new Li-ion batteries, it is necessary to focus on both cathode and anode technology. Carbon is most common anode material today, it has a theoretical gravimetric density of 375 mAh/g. In this days, in some cases, is anode doped by silicon in range of 5-10 %, but in the future it is expected that the doping will increase up to 40 %. The size of the doping will also affect the growth of the anode capacity. Silicon has a theoretical gravimetric capacity of 4000 mAh/g.

The biggest problem in using a combination of silicon and carbon is that silicon can increase its volume by up to 280 % during cycling. Combined with use in pouch cells, this will lead to battery expansion. It is necessary to verify how will expansion affected other batteries in electric vehicles. It is because of closely stacking in module.[6].

3 MEASUREMENT OF VOLUME CHANGES OF CELLS

I designed the construction of the device in the SolidWorks program which will serve to study the cells at a certain pressure or measure the pressure generated by the battery during expansion. Device could be used for both of these cases.

The device will consist of four metal clamps, which will be used for making pressure, two wooden boards will be used for insulation and pressure distribution from the metal clamps, the top wooden board will be fastened with screws to clamps to reduce the effect of its own weight on the battery. The top plate will be in two versions, the first will serve to create a constant pressure on the battery, in the second case will be piezoelectric sensors attached to the plate to measure the volume changes of the battery cell. The clamp consists of two metal straps, which are connected at the edge with screws and a nut to tighten the joint. A spring will be mounted on the screw to eliminate the effect of the clamp weight. The whole assembled device is shown on the figure 5.



Figure 5: Device for measuring battery volume changes

4 CONCLUSION

As I described in this post, electromobility is a future transport trend. There are still many things that can affect the operation of electric vehicles. This test which was described in this article will serve to Škoda Auto to improve their electric cars. Batteries which will be used for this test are from real electric vehicles. The output should be a verification of how much the pressure load affected the battery performance.

ACKNOWLEDGMENT

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-20-6206.

REFERENCES

- [1] The History of the Electric Car. *Department of Energy* [online]. 2014 [cit. 2021-02-22]. Dostupné z: https://www.energy.gov/articles/history-electric-car
- [2] Concept We Forgot: 1972 BMW 1602e. In: *Motor 1* [online]. Motor 1, 2017 [cit. 2021-02-22]. Dostupné z: https://www.motor1.com/news/133336/concept-we-forgot-bmw-1602e/
- [3] Intelligence in Mobile Battery Applications (D5.1 Desk Research Data Analysis IMBA – Release 1). Brusel, 2020. Dostupné také z: https://www.projectalbatts.eu/Media/Publications/4/Publications_4_20200930_12811.pdf
- [4] General Motors EV1. In: *Wikipedia* [online]. 2011 [cit. 2021-02-22]. Dostupné z: https://en.wikipedia.org/wiki/General_Motors_EV1#/media/File:EV1A014_(1)_cropped.jpg
- [5] Tesla Roadster (2008-2012), the electric car that sparked a revolution. In: *Wheelsjoint.com* [online]. 2020 [cit. 2021-02-22]. Dostupné z: https://www.wheelsjoint.com/tesla-roadster-2008-2012-the-electric-car-that-sparked-a-revolution/
- [6] MCDOWELL, Matthew T., Seok Woo LEE, William D. NIX a Yi CUI. 25th Anniversary Article: Understanding the Lithiation of Silicon and Other Alloying Anodes for Lithium-Ion Batteries. In: *Advanced Materials* [online]. 2013, s. 4966-4985 [cit. 2021-02-22]. ISSN 09359648. Dostupné z: doi:10.1002/adma.201301795

Doktorské projekty

Mikroelektronika a technologie, Elektronika a komunikace II.

IN-SITU SEM CHARACTERIZATION OF LITHIUM-ION BATTERY CREATED ON MEMS CHIP

Ondřej Klvač

Doctoral Degree Programme (1), FEEC BUT E-mail: xklvac02@vutbr.cz

Supervised by: Tomáš Kazda

E-mail: kazda@feec.vutbr.cz

Abstract: The aim of the work is to develop an in-situ method for electrode nanostructure investigation by scanning electron microscopy. In the first part, the preparation of a model battery using FIB-SEM system is described, which consists of lithium titanate oxide (LTO) and metallic lithium material cutting by the focused ion beam, transport to the MEMS chip, and contacting by carbon deposition. After that the electrodes were soaked by the electrolyte, the voltage changes were measured on two created battery systems. Finally, a current was applied to the cell. However, cycling failed due to technological problems. Challenges connected with battery preparation are described and possible solutions for future testing are proposed. The work is evolved in cooperation with Thermo Fisher Scientific Brno.

Keywords: SEM, FIB, MEMS, in-situ, Li-ion, LTO, battery characterization

1 INTRODUCTION

Lithium-ion batteries (LIBs) are currently the most commonly used cells for high-performance applications. There are several reasons, they have a high energy density, low self-discharge, high number of charge/discharge cycles, low maintenance, no "memory effect", and high nominal voltage. This makes them a suitable candidate for future applications such as energy storage for renewable electricity sources or electric vehicle propulsion [1,2,3].

However, there are still some issues that need to be solved before they can be spread on a massive scale. $LiCoO_2$, a widely used cathode material in the past, was gradually replaced, due to its high cost and poor stability, by $LiNi_xMn_yCo_zO_2$ (NMC) [4,5]. Achieving higher capacity while maintaining stability has been the subject of research in recent years [6,7]. Today, this material reaches its limits and although there are other promising and cheaper materials (sulfur, silicon), they are not very widespread due to technological problems (solubility, poor conductivity, cracking, and delamination induced by high expansion) [8,9].

During the investigation by electrochemical methods, only spatial averages can be derived. Scanning electron microscopy (SEM) allows seeing mentioned phenomena in detail directly which is essential for modern nanostructures [10]. In most cases, the battery is discharged after a defined number of cycles, disassembled, and the electrode surface or cross-section is scanned [11,12]. However, for safety reasons (manipulation with charged cell), charging phenomena such as lithium plating cannot be observed [13]. In-situ SEM characterization can be helpful, but it needs to prevent electrolyte evaporation and material degradation caused by air exposure. The ionic liquid is often used in combination with special cells for transfer and cycling [14].

This work follows the diploma thesis of Ing. Hujňák [15], where the whole battery is created inside the SEM chamber from active electrode materials only. The materials are observed from the first soaking with the electrolyte and the presented method is universally applicable for various materials. Because this procedure is new a stable LTO and lithium were used for the first steps.

2 EXPERIMENTAL

The battery was connected to biasing contacts of a MEMS chip [16] designed for in-situ heating and biasing experiments in Thermo Fisher Scientific tools [17] (**Figure 1**).



Figure 1: Used MEMS chip

In standard conditions, the chip is contacted in a holder on the SEM stage and connected to the controller at the air side of the vacuum feedthrough via a 25 pin CANON connector. For this experiment, the controller was used for connectivity test only, then it was dismounted and pins connected directly with bias contacts were used for measurements.

Firstly, a chunk of approx. 50x50x20 µm was cut from an electrode with LTO active material coated on aluminum foil produced by CUSTOMCELLS[®] using xenon plasma focused ion beam (PFIB) system Helios Hydra 5 UX DualBeam (Thermo Fisher Scientific).

The sample was not cut completely, a small "bridge" for fixation was left (**Figure 2**a). Then, the EasyLift micromanipulation needle was attached on the top-left chunk corner, glued by carbon deposition using MultiChem Gas Delivery System and the "bridge" was removed (**Figure 2**b). Afterward, the chunk was lifted-up (**Figure 2**c), aligned with bias pad on MEMS, fixed by carbon deposition and the needle was cut-off finally (**Figure 2**d).

In the case of air-sensitive lithium, a small piece from 99.9% pure ribbon (SigmaAldrich) was cut with a knife in Ar-filled glovebox, enclosed to a glass vial, and the cap was tightened with PARA-FILM[®]. To minimize the air exposure time, the vial was opened once everything inside the Dual-Beam chamber was ready. After opening, the lithium was placed on a sample stub, the chamber was closed and pumped immediately. The chamber was then kept closed through the rest of the experiment.

For lithium chunk preparation, Helios 5 UC DualBeam (Thermo Fisher Scientific) with gallium FIB was used similarly. However, lithium has a lower melting point than LTO, it is very soft, and strong redeposition occurs during ion milling. Moreover, carbon is deposited slowly and has poor adhesion. Tungsten deposition was tested too, it is faster, but the deposited layer was nested into lithium after a short time and a chemical reaction probably occurred. Therefore, it was problematic to prepare the chunk with the same dimensions precisely. After tuning appropriate conditions of the FIB sample preparation, a 30x50x20 Li sample was placed on the second biasing contact of the MEMS chip.

As an ionic liquid, 1-Ethyl-3-methylimidazolium tetrafluoroborate (EMIMBF \geq 98%, SigmaAldrich) in a mixture with LiBF₆ salt in 0.5M concentration was used. The liquid was applied to the MEMS chip near the LTO in an amount of 0.5 microliters with a pipette. Once both active materials were connected to the biasing contacts, the electrolyte was moved between them using stage tilt (**Figure 3**).



Figure 2: Cut chunk from electrode (a); contacting with EasyLift needle (b); chunk lift out (c); connection with MEMS chip by carbon deposition, both electrodes (d)



Figure 3: Soaking of electrodes by electrolyte, first (left) and second (right) battery system

For electrochemical characterization, BioLogic SP-150 potentiostat which allows setting currents below nano amperes was used. For instance, the capacity of used LTO given by datasheet is 1mAh/cm², and measured active material thickness is around 30 um. Then, the calculated capacity of the created battery is approximately 16 nAh. The device was connected with a shared reference electrode (RE) and the counter electrode (CE) to the lithium, LTO was on the working electrode. The ground electrode was on the microscope frame.

3 RESULTS AND DISCUSSION

Before the IL was moved between electrodes, an open-circuit voltage was measured with a 0.5 s period. Since the electrodes were not connected, the voltage was zero (only noise in microvolts was present). Once the electrolyte soaked both electrodes, the voltage increased to 302.3 mV within 4.5 s. However, it immediately began to decline, after 2.5 s it was 150 mV, 23 mV after 1 min, and 5.7 mV after 2 min (**Figure 4** – right).

Many battery recovery attempts have been performed with different currents in galvanostatic mode. At 2.5 nA, the voltage fluctuated around 5 mV, but charging was not visible. At 100 nA, the voltage was increased to stable 16 mV, at 1 mA it was stable 167 mV, and 1.5 V at 10 mA. As soon as the current source was disconnected the voltage dropped to units of millivolts immediately. At 20 mA there was a visible increasing voltage trend, however, these currents are in the order of millions of C-rate and it is probably polarization only (**Figure 4** – left).

It is clear that the current flows here, but the battery is not charged. The IL was on the surface of the chip for 8 days and solid particles were floating on its surface. MEMS surface could be dissolved and might become conductive for electrons. It is also possible that during cutting of the needle, redeposition of tungsten occurred and paths were partially connected. During the experiment, the LTO chunk extended by 13%.



Figure 4: Open circuit voltage measurement (left); applied current and voltage response (right)

In the next experiment, it was tried to pass a current of 100 nA through the circuit before the electrolyte soaked the electrodes. The current flowed here and a voltage drop of 1.5 V was measured. It turned out that a very thin layer of material is redeposited between the contacts, which forms a parallel resistance of about 15 M Ω . The layer was not visible, but during the milling of the area between the contacts, the voltage drop gradually increased to 10 V, and then it was not possible to maintain the current. After wetting the electrodes with the electrolyte, the voltage rose to about 0.65 V, which was stable for the 20 s (**Figure 5**). Then, the electrodes were irradiated with an electron beam and the polarity was reversed.

Charging and local lithium deposition induced by electron beam was described for a system where the lithium was connected with the sample stage (ground) [13], however in this case, although the voltage began to return to normal, the battery was destroyed completely and the cycling failed. Only strong noise was present even after turning off the beams and low voltages could not be measured. Obviously, it is necessary to modify the connection.



Figure 5: Open circuit voltage measurement affected by ion and electron beam

4 CONCLUSION

An experimental setup for in-situ research of electrode materials by scanning electron microscopy has been described and will be used for the characterization of nanostructures during doctoral study.

Two batteries from commercially available LTO electrode material, metallic lithium, and ionic liquid were created on a MEMS chip inside the SEM chamber. In the first case, the battery had a voltage of 0.3 V when both electrodes were soaked by the electrolyte. However, it was found, that conductive materials can be redeposited between electrodes during preparation and a parallel load is created. Therefore, the battery was discharged very fast. In the case of the second battery, the load was removed by ion cleaning. There was a stable voltage of 0.65 V for 20 sec, then the battery was damaged

by an electron beam and cycling was not possible. It has been shown that the electron and probably also the ion beam can strongly influence the result and damage the battery.

During the work, it is planned to create a best practice document describing individual steps of the battery preparation, which consists of FIB cut-out of electroactive samples, transfer of these samples on micromanipulation needle, and connection of the samples with the MEMS pads by carbon deposition. The experiment is still in progress, the usage of platinum deposition for better contacting, another ionic liquid, and modification of measurement connection is planned. The described procedure can be used for the characterization of sulfur electrodes in the future.

ACKNOWLEDGEMENTS

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-20-6206. The work was developed in cooperation with Thermo Fisher scientific Brno. Yakub Fam, Libor Novák and Tomáš Kazda participated in the preparation of the experiment.

REFERENCE

- [1] Y. Liang et al., InfoMat, 1, 6-32 (2019).
- [2] H. -J. Kim et al., Electronics, 9 (2020).
- [3] B. Scrosati, Angewandte Chemie International Edition, 50, 5254-5255 (2011).
- [4] M. Li and J. Lu, Science, 367, 979-980 (2020).
- [5] X. Chen, W. Shen, T. T. Vo, Z. Cao, and A. Kapoor, in 2012 10th International Power & Energy Conference (IPEC), p. 230-235, IEEE (2012)
- [6] G. Houchins and V. Viswanathan, Journal of The Electrochemical Society, 167 (2020).
- [7] H. Cha et al., Advanced Materials, 32 (2020).
- [8] L. Borchardt, M. Oschatz, and S. Kaskel, Chemistry A European Journal, 22, 7324-7351 (2016)
- [9] H. -J. Kim et al., Electronics, 9 (2020)
- [10] H. Liu et al., Journal of Power Sources, 306, 300-308 (2016)
- [11] Kovachev et al., Batteries, 5 (2019)
- [12] T. Waldmann et al., Journal of The Electrochemical Society, 163, A2149-A2164 (2016)
- [13] T. Waldmann, M. Wilka, M. Kasper, M. Fleischhammer, and M. Wohlfahrt-Mehrens, Journal of Power Sources, 262, 129-135 (2014)
- [14] D. Chen, S. Indris, M. Schulz, B. Gamer, and R. Mönig, Journal of Power Sources, 196, 6382-6387 (2011)
- [15] J. Hujňák, thesis, Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií, Brno (2020)
- [16] L. Mele et al., Microscopy Research and Technique, 79, 239-250 (2016)
- [17] L. Novák, J. Stárek, T. Vystavěl, and L. Mele, Microscopy and Microanalysis, 22, 184-185 (2016)

STUDY OF DIELECTRIC BEHAVIOR OF POLYMERS USING A COMPLEX ELECTRIC MODULUS

Luděk Horák

Doctoral Degree Programme, FEEC BUT E-mail: xhorak58@stud.feec.vutbr.cz

Supervised by: Jiří Kazelle

E-mail: kazelle@feec.vutbr.cz

Abstract: The complex electric modulus is mathematically defined as the inverse value of the complex permittivity. It is particularly useful in evaluating the dielectric properties of polymeric insulating materials, especially at low frequencies and at high temperatures. Presented thesis is focused on mathematical interpretation, practical example and evaluation of electrical module. In our case, the measurements are performed on a composite material based on epoxy resin.

Keywords: Complex electric modulus, complex permittivity, dielectric relaxation, composite materials, electrical insulation materials, epoxy resin

1 INTRODUCTION

The literature states that the electric modulus was first introduced by McCrum et al. [1] Based on their knowledge and theoretical knowledge of Macedo et al. they used an electric modulus to study the phenomena of electrical relaxation in glass ion conductors in practice. [2] Over time, the complex electric modulus proved to be a suitable tool for the analysis of dielectric properties of polymeric insulating materials, in our case for the observation of modified epoxy resins.

2 THEORY

The complex electric modulus M^* is defined as the inverse value of the complex permittivity ε^* . If it is true that

$$\varepsilon^*(\omega) = \varepsilon'(\omega) - j\varepsilon''(\omega), \tag{1}$$

then M^* is expressed by the equation:

$$M^{*}(\omega) = \frac{1}{\varepsilon^{*}} = M' + jM''$$
(2)

$$M^{*} = \frac{\varepsilon'}{(\varepsilon')^{2} + (\varepsilon'')^{2}} + j \frac{\varepsilon''}{(\varepsilon')^{2} + (\varepsilon'')^{2}},$$
(3)

where M' is the real part of the complex electric modulus, M'' is the imaginary part of the complex electric modulus, ε' is the real part of the complex permittivity (relative permittivity) and ε'' is the imaginary part of the complex permittivity (loss number). [3][4][5]

For a Cole-Cole type relaxation, ε^* can be expressed as:

$$\varepsilon^* = \varepsilon_{\infty} + \frac{\varepsilon_s - \varepsilon_{\infty}}{1 + (j\omega\tau)^{\beta}}, \quad [3][4]$$

where ε_s is the static relative permittivity at a frequency approaching zero, ε_{∞} is the optical relative permittivity at a frequency approaching infinity, τ is the temperature-dependent relaxation time and β is the arc shape parameter (taking values from 0 to 1). [3][4]

When we combine the previous equations (1-4), after several adjustments we get:

$$\begin{split} M^{*}(\omega) &= \frac{1}{\varepsilon^{*}} = \frac{1}{\varepsilon_{\infty} + \frac{\varepsilon_{s} - \varepsilon_{\infty}}{1 + (j\omega\tau)^{\beta}}} = \frac{1}{\varepsilon_{\infty}} \cdot \frac{1}{1 + \frac{\varepsilon_{s} - 1}{1 + \frac{\varepsilon_{s}}{\varepsilon_{\infty}} - 1}} = \\ &= \frac{1}{\varepsilon_{\infty}} \cdot \left(\frac{1 + \frac{\varepsilon_{s}}{\varepsilon_{\infty}} - 1}{1 + (j\omega\tau)^{\beta}} - \frac{\varepsilon_{s}}{\varepsilon_{\infty}} - 1}{1 + (j\omega\tau)^{\beta}} \right) = \\ &= \frac{1}{\varepsilon_{\infty}} \cdot \left(1 - \frac{\frac{\varepsilon_{s}}{\varepsilon_{\infty}} - 1}{1 + (j\omega\tau)^{\beta} + \frac{\varepsilon_{s}}{\varepsilon_{\infty}} - 1}} \right) = \\ &= \frac{1}{\varepsilon_{\infty}} - \frac{\frac{\varepsilon_{s}}{\varepsilon_{\infty}^{2}} - \frac{1}{\varepsilon_{\infty}}}{1 + (j\omega\tau)^{\beta} + \frac{\varepsilon_{s}}{\varepsilon_{\infty}} - 1}} = \frac{1}{\varepsilon_{\infty}} - \frac{\frac{1}{\varepsilon_{\infty}} - \frac{1}{\varepsilon_{s}}}{1 + (j\omega\tau)^{\beta} \cdot \frac{\varepsilon_{\infty}}{\varepsilon_{s}}}} = \frac{1}{\varepsilon_{\infty}} - \frac{\frac{1}{\varepsilon_{\infty}} - \frac{1}{\varepsilon_{s}}}{1 + (j\omega\tau'_{M})^{\beta}} \end{split}$$
(5)

$$M^* = M_{\infty} - \frac{M_{\infty} - M_s}{1 + (j\omega\tau'_M)^{\beta}},$$
[3] (6)

where $M_s = 1/\varepsilon_s$, $M_{\infty} = 1/\varepsilon_{\infty}$ and relaxation time τ'_M is given by the relation:

$$\tau'_{M} = \left(\tau \cdot \left(\frac{\varepsilon_{\infty}}{\varepsilon_{s}}\right)^{\frac{1}{\beta}}\right) = \tau_{M} \cdot \left(\left(\frac{\varepsilon_{\infty}}{\varepsilon_{s}}\right)^{\frac{1}{\beta}-1}\right),$$
[3][4]

where $\tau_M = \tau(\varepsilon_{\infty}/\varepsilon_s)$. If ε_s is larger then ε_{∞} , τ_M or τ'_M is always smaller than τ . This means that the relaxation in the modular spectrum moves to a higher frequencies compared to the permittivity spectrum. In addition, because both ε' and ε'' appear in the denominators, the numerical value of M' and M'' decreases with increasing ε' and ε'' . Therefore, in the spectrum of the complex electric modulus, the tendency that large values of ε' and ε'' exceed the relaxation peaks is completely eliminated. Furthermore, the equations show that the parameter β remains unchanged regardless of whether the arc is in permittivity mode or electric modulus mode. In the case of both Cole-Cole arcs of ε^* and M^* with a smaller parameter β , there is a wider distribution of relaxation time around τ , τ_M or τ'_M at each respective center. [3][4]

Another problem is described in the article [3]. The accumulation of space charge near the electrodes (electrode polarization) causes a high permittivity in place, where are the low frequencies. Therefore, the dielectric behavior of the system, especially at high temperatures and low frequencies, would be largely obscured in the permittivity spectrum. In this regard, the complex electric modulus M^* is a powerful tool for the analysis of dielectric relaxation processes. [3]

3 SAMPLES FOR MEASUREMENT

Samples for the study of dielectric behavior were made in the laboratories of SYNPO a.s. Pardubice. It is an internal system called "Sadurit" with increased burning resistance and glass transition temperature T_g in the range of 100 - 120 °C. The system is multicomponent, most often filled with different kinds of micro-ground siliceous sand. Liquid anhydride, REACH free, is used as a hardener.

Epoxy samples have a thickness of approximately 2 mm and a diameter of 35 mm. They are cleaned in a detergent solution, degreased with isopropyl alcohol and then placed in a desiccator filled with desiccant for at least 72 hours, where they are freed of any moisture. Resins can absorb water up to several percent of their own weight, they are very sensitive to moisture. The absorption of water causes changes in the material and leads to a deterioration of the dielectric properties, the phenomenon is described in more detail in article [6].

4 EXPERIMENT

The measurement was performed on a dielectric impedance analyzer from company Novocontrol Technologies. The samples were measured in the frequency range from 10^{-3} Hz to 10^{6} Hz, in the temperature range 25 °C - 200 °C and at an operating voltage of 1 V. The measurement is very time consuming, it took several days.

It can be seen from the graphs that with decreasing frequency and increasing temperature, the values of relative permittivity (Fig. 1) and loss figure (Fig. 2) increase significantly. When displaying graphs of the whole permittivity spectrum, information about the investigated material is lost, so we must plot the dependencies in a detailed view - only for specific values of the components of complex permittivity (not part of this work). By transforming the values of complex permittivity according to equations (2-3), we obtain a complex electric modulus, with which we can better describe the properties of the material (Fig. 3 and Fig. 4).



Figure 1: Dependence of relative permittivity on frequency



Figure 2: Dependence of loss number on frequency

Relaxation peaks (α -polarization) can be observed in the spectrum (Fig. 4) in the temperature range 100 °C - 200 °C, they are accompanied by a gradual transition from low to high values in the spectrum (Fig. 3). β -polarization and γ -polarization are not observable, they need to be measured in the negative temperature range. Relaxation peaks are observable above the glass transition temperature (T_g). Above the temperature T_g, ions from the hardener begin to release (jump), there is an increase in conductivity and associated losses. The charge carriers move (jump) over long distances at frequencies below the relaxation peak, and conversely at frequencies higher than the maximum (e.g. for 200 °C ~ 10⁻³ Hz) they are only movable over short distances.



Figure 3: Dependence of real part of the complex electric modulus on frequency



Figure 4: Dependence of imaginary part of the complex electric modulus on frequency

5 CONCLUSION

In the case of imaging the entire permittivity spectrum, the components of complex permittivity increase significantly, especially at high temperatures and low frequencies. As a result, information about the material under investigation is lost. A powerful tool for the analysis of dielectric properties is a complex electric modulus, we can use it to detect dielectric relaxation and charge transport in the material.

ACKNOWLEDGEMENT

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-20-6206.

REFERENCES

- [1] N. G. MCCRUM, B. E. READ, G. WILLIAMS. Anelastic and Dielectric Effects in Polymeric Solids, Wiley, London, NewYork, pp. 108-111, 1967. Dostupné z: <u>https://www.sciencedirect.com/science/article/abs/pii/0032386168900700</u>
- [2] P. B. MACEDO, C. T. MOYNIHAN, R. BOSE. The role of ionic diffusion in polarization in vitreous ionic conductors, Phys. Chem. Glasses, vol. 13, pp. 171-179, 1972.
- [3] F. TIAN and Y. OHKI. Electric modulus powerful tool for analyzing dielectric behavior [online]. *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 21, no. 3, June 2014. [cit. 2021-03-01]. ISSN: 1558-4135. Dostupné z: https://ieeexplore.ieee.org/document/6832233
- [4] TSANGARIS, G.M., PSARRAS, G.C., & KOULOUMBI, N. Electric modulus and interfacial polarization in composite polymeric systems [online]. Journal of Materials Science, 33, 1998. DOI: 10.1023/A:1004398514901. Dostupné z: https://link.springer.com/article/10.1023/A:1004398514901
- [5] S. BOUKHEIR, Z. SAMIR, R. BELHIMRIA, L. KREIT, M. E. ACHOUR, N. ÉBER, L.C. COSTA, A. OUERIAGLI & A. OUTZOURHIT. Electric Modulus Spectroscopic Studies of the Dielectric Properties of Carbon Nanotubes/Epoxy Polymer Composite Materials [online]. Journal of Macromolecular Science, Part B, 57(3), 2018. DOI:

 10.1080/00222348.2018.1439243.
 ISSN
 0022-2348.
 Dostupné
 z:

 https://www.tandfonline.com/doi/full/10.1080/00222348.2018.1439243

[6] HARVÁNEK, L., TOMÁŠKOVÁ, T., SVOBODA, M., MENTLÍK, V. Composites with nanosilica [online]. *IEEE 11th International Conference on the Properties and Applications* of Dielectric Materials (ICPADM), Sydney, Australia, 2015. [cit. 2020-05-30]. ISBN: 978-1-4799-8903-4. ISSN: 2160-9241. Dostupné z: <u>https://ieeexplore.ieee.org/document/7295404</u>

MEASUREMENT-BASED 60 GHZ TAPPED-DELAY-LINE CHANNEL MODEL

Radek Zavorka

Doctoral Degree Programme (1), FEEC BUT E-mail: xzavor03@stud.feec.vutbr.cz

> Supervised by: Ales Prokes E-mail: prokes@feec.vutbr.cz

Abstract: With the development of communication technology a new electromagnetic spectrum is necessary. It is the reason for the research of millimeter waves (MMW) about 60 GHz. This paper is focused on the evaluation impact of multipath components on receive signals. It depends on the radiation pattern of used antennas and the direction of the receiving antenna related to the line of sight (LOS) component. The multipath components (MPCs) are received from surrounding points around the center of the transmitter. To model the MPCs propagation a basic Tap Delay Line (TDL) model is established including individual tap indexes and important points are compared mutually. Research is also focused on the determination of the reflectivity of the surrounding objects.

Keywords: channel measurement, mmWave, Power Delay Profile, Tapped Delay Line

1 INTRODUCTION

Millimeter waves are defined as the electromagnetic spectrum from 30 to 300 GHz. Due to the growing requirements on communication speed, it was a logical step to move to the higher frequency. Band of mmWave offers a wide range of possibilities, like a wider communication spectrum and a higher frequency, which allow the rise of the communication speed. Another reason was the abrupt increase of the number of communication devices, because they need individual bands for transmitting signals. There are a few unlicensed industrial, scientific, and medical (ISM) frequency bands specially designated for high speed, indoor and outdoor connectivity.

The 60 GHz mmWave is a perspective ISM band and the International Telecommunication Union (ITU) suggested it for unlicensed operation. Many researchers study the propagation of this wave for various environments and applications. The physical limitations are predominantly caused by the extremely high propagation path loss reducing the propagation range, since the path loss increases with the square of frequency. A potential problem for longer propagation distances can also be caused by rain attenuation and absorption of MMW energy by the resonance of oxygen and water vapor molecules in the atmosphere. Heavy rain can cause attenuation of about 12 dB/km at 60 GHz, while attenuation about 15 dB/km is excited by oxygen in the frequency band 57-64 GHz. Modern equipment and laboratories are necessary for the demonstrative approach because common measuring devices aren't sufficient. The paper [1] presents results obtained from a vehicle-to-vehicle channel measurement campaign carried out in the millimeter-wave band around a 60 GHz. In [2] you can see the study of propagation mmWave inside a bus. In the article [3] authors measured and modeled mmWave channel for outdoor microcellular environment and a very large waiting hall at a railway station at 26 and 28 GHz.

2 MEASUREMENT SCENARIO

The measurement simulates a static vehicle to infrastructure communication. It was performed in the Brno University of Technology campus between the building at Technicka 12 and VW CC car park on the far side of the road as shown in Fig. 1a. A transmit omnidirectional antenna together with the power amplifier and cooler were placed on the roof of the car. And a receiver with a horn antenna was situated in a window on the 6th floor of the university building. Receiving antenna scanned the area around the car and measured multipath components (MPCs), see points in Fig. 1b. A motorized Sky-Watcher AllView mount controlled by PC and LabView was used to direct the antenna. From Fig. 1a it is obvious that the height of the transmitting (TX) and receiving (RX) antennas above the ground is h = 1.55 m and H = 14.5 m respectively. Because the distance between the TX antenna and the building is D = 28.5 m, the propagation distance between the antennas is $L = \sqrt{(H-h)^2 + D^2} = 31.3$ m.



a)

b)

Figure 1: Position of the transmitter relative to the receiver a), measurement points b). [4]

3 MEASUREMENT SETUP

The channel measurement was carried out using the 60 GHz time-domain channel sounder described in detail in [4]. The transmitter is based on an Anritsu MP1800A Signal Quality Analyzer working as a pseudorandom binary sequence (PRBS) generator. The receiver is created using a Tektronix MSO72004C (20 GHz, 50 GS/s) Mixed Signal Oscilloscope working as a very fast analog-to-digital converter. The baseband PRBS signal is converted into the MMW band and back using SiversIma FC1000V series V-band up/down converters [5]. The channel sounder bandwidth is 8 GHz, the number of samples per measured channel impulse response (CIR) is set to N_{Sa} = 8092 and the number of saved CIRs per measurement is N_{CIR} = 932. Due to the correlation gain and averaging, the sounder dynamic range is about 55 dB. The transmitter was equipped with an omnidirectional SIW slot antenna described in [6], whose radiation pattern related to the car is shown in Fig. 2. The MMW signal was received using a directional horn antenna with a dielectric lens. The antenna gain depends on the angle which is shown in Fig. 3. All measured values mentioned below are related to the low noise MMW preamplifier output. After that, the down converter SiversIma increases them by a gain of about 15-20 dB.

4 CHANNEL CHARACTERIZATION

In this work, the static channel is analyzed which means that the position of all objects is stable in time. The channel is often modeled by the linear time invariant filter (LTI), where the impulse



Figure 2: E-plane (left) and H-plane (right) measured radiation pattern of double-sided SIW slot antenna at 55 GHz, 60 GHz, and 65 GHz. [4]



Figure 3: E-plane a) and H-plane b) measured gain of horn antenna. [4]

response of the channel is marked h(t), and the output is presented as a convolution of a useful signal with an impulse response and then noise n(t) is added.

$$r(t) = s(t) * h(t) + n(t)$$
 (1)

The main focus is to analyze a power delay profile (PDP) for the line of sight and nonLOS (NLOS) components. Special attention will be paid to the evaluation of the influence of the surrounding objects to the power received from different directions. Fig. 4 shows the received power from all points.



Figure 4: Power of received signal.

The red line means LOS power, black line is the sum of individual components of the channel impulse responses (CIRs) for NLOS and the blue line is the total received power from each point. It is evident that the strongest component is received from the direct path (point 13). Also the nearest points have relatively strong components in LOS and NLOS, but most significantly it is in a vertical direction from the centre for points 8, 18 and also 7, 9, 14, 17, 19. It is due to the radiation pattern of both antennas, Fig. 2 and 3.

From this measurement it is also possible to estimate the reflectivity of individual objects or surfaces. As is evident from the graph in Fig. 4 a satisfactory reflection off the surface of the road or sidewalk was achieved. It can be caused by very smooth sections of the road, because the length of waves is extremely short, about 5 millimeters. There is a very good reflectivity of the building behind the car. NLOS components have a higher intensity, about 6-12 dB, than their LOS. On the other hand, the grass surface and clay exhibit higher losses, it is about 30 dB in comparison with the direct path from the center point 13. Of course, branches and leaves of a tree cause a significant attenuation of the signal due to the fading effect. A metal box (point 22 in Fig. 1b) doesn't cause such important losses in the MPCs. This might be caused by the ray refraction on the edge.

4.1 TDL MODEL

Tapped delay line model for every point from Fig. 1b is analyzed from the performed measurement. This model corresponds to the Finite Impulse Response (FIR) filter and the general model is shown in Fig. 5. Equation (2) describes the output of TDL and corresponds with the general model:

$$y(\tau) = \sum_{n=1}^{N} s(\tau - nT)g_n(\tau), \qquad (2)$$

where $s(\tau)$ is a delay input signal and $g_n(\tau)$ are TDL tap gains. Every tap index is obtained from a 10 ns cluster like max value, which allows to distinguish 3 meters differences in the length of the signal path. The reason why the clusters are established is to create a simplified model. To evaluate normalized tap gain, the max value of CIRs is taken from the cluster, because there are sometimes very weak components, which isn't significant.

In Fig. 6, there are important points from Fig. 1b where normalized tap gains are shown in the bar chart depending on the delay and normalized in the view of LOS. The value of LOS is in the title of the graph. Strength of the received power depends on the angle of the rotation of the receiving antenna and reflectivity surface, as described above. Differences between propagation path lengths of NLOS components are described above the dominant components.



Figure 5: General model of Tapped Delay Line.

When a signal is received from a direct path from point 13, there is only one tap index whose value considerably exceeds other components, but it is significantly attenuated in comparison with LOS. It is due to the radiation pattern of the receiving antenna, because the incoming signal that is skewed five degrees from the direct path is significantly attenuated. Point 8 has an important index 4, which shows a reflected signal with a longer way, about 12 m. It corresponds with the distance between the car and the sidewalk near the building behind the car. Other components for this point are much less

significant. Points 17, 18, 19 are interesting due to the first tap index, which contains a reflected signal with a distance of about 3 m from the LOS. It is the same case as point 13, the antenna is heading to the road and the edge of the lobe is receiving a signal from the center point. It is obvious that the surface of the road has a very good reflectivity. TDL models for these three points have another tap index with a delay between 80-90 ns, which corresponds with the difference of the distances of about 24-27 m. The component at the tap 8/9 is strongly attenuated due to the large path loss and also due to the multi-reflection. The signal is probably reflected from the building behind the car and from the surrounding car body, whose surface reflects the signal very well.



Figure 6: Normalized tap gain for select points.

5 CONCLUSION

This paper offers general information about the mmWave propagation and provides instructions for channel measurement. The main goal of this work is to analyze the multipath components between the transmitter located on the roof of the car and the receiver situated in the window of the building. It is found that the strongest power was received from the direct path and from the nearest points around the transmitter. MPCs have a significant influence when they are reflected from the road and sidewalk because the delayed received signal has a considerably greater amplitude than LOS (for a given antenna direction). Also, it is found that the grass surface attenuated the signal very well (more than 30 dB in relation to the center point). TDL model is defined for every point and the most significant components are described and shown with the value of delay time and distance.

REFERENCES

- [1] J. Blumenstein et al., *Vehicle-to-Vehicle Millimeter-Wave Channel Measurements at 56-64 GHz*, IEEE 90th VTC, Honolulu, HI, USA, 2019, pp. 1-5.
- [2] A. Chandra et al., 60 GHz millimeter wave propagation inside bus: Measurement, modeling, simulation, and performance analysis, IEEE Access, vol. 7, no. 2019, pp. 97815-97826.
- [3] X. Zhao et al., *Neural network and GBSM based time-varying and stochastic channel modeling for 5G millimeter wave communications*, in China Communications, vol. 16, no. 6, pp. 80-90, June 2019, doi: 10.23919/JCC.2019.06.007.
- [4] A. Prokes et al., Multipath Propagation Analysis for Vehicle-to-Infrastructure Communication at 60 GHz, 2019 IEEE Vehicular Networking Conference (VNC), Los Angeles, CA, USA, 2019, pp. 1-8, doi: 10.1109/VNC48660.2019.9062771.
- [5] *FC1005V/00 V-band Converter with LO*. [Online]. [Accessed: 16-Feb-2021]. Available: https://www.ecmstockroom.com/writable/items/pdf_files/fc1005v00-data-sheet.pdf
- [6] T. Mikulasek, J. Lacik, and Z. Raida, *SIW slot antennas utilized for 60- GHz channel characterization*, Microw Opt. Tech. Lett., vol. 57, no. 6, pp. 1365-1370, 2015.
COMPARISON OF REMOTE AND SELF-POSITIONING APPROACHES FOR INDOOR BLE-BASED LOCALIZATION

Stanislav Rozum

Doctoral Degree Programme (4), FEEC BUT E-mail: rozum@feec.vutbr.cz

Supervised by: Ladislav Polak E-mail: polakl@feec.vutbr.cz

Abstract: This paper deals with Received Signal Strength (RSS) positioning based on Bluetooth Low Energy technology. Specifically, it compares two concepts, remote and self-positioning, using their design properties and the RSS maps. The fundamental difference between both modes lies in units' tasks with the known and unknown positions. In our experiment, the fixed units - anchors exploit multiple antennas, whereas the movable device uses a classic single antenna. The system can operate in either mode, thus eliminating the influence of different hardware.

Keywords: remote-positioning, self-positioning, BLE, localization, RSS

1 INTRODUCTION

Positioning is a basic building block for navigation and tracking. It is important to minimize the error during positioning phase, because the inaccuracies are carried from the positioning up, where it is hard to mitigate them. To this end, in the following, we focus on means of positioning.

In an indoor environment, the satellite signal is either highly attenuated or several times reflected, making it hard to estimate position with high accuracy. Therefore, different technologies are used, such as Ultra-Wideband, Wi-Fi, or Bluetooth Low Energy (BLE) based systems [1].

This work compares two main principles of indoor Received Signal Strength (RSS) BLE-based positioning, namely remote-positioning and self-positioning [1]. The following section briefly describes the differences between these modes. Next, the measurement setup and the obtained results are presented. Finally, the conclusion offers a short discussion of the results.

2 BLE POSITIONING CONCEPTS

In indoor positioning systems, there are typically two types of devices, i.e., anchors and tags. The anchors are a part of the fixed infrastructure of the system, and their position is known. On the other hand, tags are moveable, and it is their position we are estimating during the positioning process.

From the hardware perspective, a device can be configured to work in either remote or self-positioning mode. While in remote-positioning, the anchors only listen (are passive), and tags periodically transmit advertisement reports; in the self-positioning mode, the roles are swapped. There are two additional modes called indirect remote-positioning and indirect self-positioning, where a backchannel is introduced to get the resolved position to the counterpart [1].

2.1 **Remote-positioning**

In the remote-positioning mode, the fixed anchors are listening for reports, and movable tags transmit advertisements (packets). This approach benefits from the option to have more than one receiving

antenna because space limitation and power consumption are not significant issues for grid-powered devices. Hence, the anchor can collect multiple RSS samples for a single advertisement. Because a single BLE advertisement consists of three packets sent on all three advertisement channels [2], each antenna (connected to a dedicated transceiver) can tune on a specific channel. One advertisement then provides a sample per transceiver. Moreover, the spatial diversity may be exploited [3]. The multiple uncorrelated samples offer more ways to improve the accuracy of a positioning system, especially for moving tags.

A remote positioning system enables to track not only smartphones but also very simple tags, for example, embroidered to cloths [4]. The need for tags to be actively transmitting can be considered a drawback because it might compromise privacy as we can track the users who are using our system for navigation. The BLE standards support random addresses [2], which can periodically change, so the tracking system's reach would be limited. High demands (high data rate, low latency) on connections between anchors and processing unit and backchannel established for smartphones navigation are downsides of remote-positioning.

2.2 Self-positioning

On the other hand, in the self-positioning mode, where fixed anchors transmit advertisements and tags receive them, the system communication layer may not be required, and the anchors can operate on their own. Further, the backchannel is not needed for navigation.

Because a common BLE-enabled device has a single transceiver (antenna), it can receive on a single channel at a time. Additionally, in the case where two anchors transmit reports simultaneously, there would be a collision at the receiver side. That means a significantly lower RSS sample count per time, which is problematic, especially for moving tags.

3 MEASUREMENT SETUP

The comparison measurement of RSS samples has been carried out in the Brno University of Technology (BUT), at the Department of Radio Electronics (DREL) in a narrow corridor 44 m long, 1.8 m wide, and 2.7 m high located on the top (seventh) floor. Fig. 1 illustrates the floor plan; the basic properties of the corridor are: the floor is made from concrete, walls are lined with plasterboard, and the ceiling material is mineral wool with paint finish. Wooden cabinets and metal chairs are located on the right side of the corridor, whereas the left side is without obstructions.

During the measurement, the tag was placed in a grid with 0.5 m spacing from 0 to 35 m on the x-axis and from 0.4 to 1.4 m on the y-axis with [0, 0] coordinates put under the Anchor 0 (in Fig. 1, the Anchor 0 is the one closest to X and Y intersection).



Figure 1: Measurement scenario and floor plan. Green boxes denote the anchors at the height of 2.2 m above the floor level. Tags were placed in the 0.5×0.5 m grid at 1.0 m above floor. The scale of the x-axis and y-axis is different. Note x-axis crosses y-axis in 0.9 m

For the purpose of this paper, we have used the system previously introduced in [5], with additional enhancements. In contrast to [5], the Anchor's firmware was upgraded to support both remote and self-positioning [3]. The configuration can be done online via Ethernet command. Fig. 2 displays block diagram of the used system. The BL652 modules capture BLE advertisements, the RSS samples are then passed to the Raspberry Pi B+ via ZigBee network and then to the laptop through Ethernet. In the upgraded version the Tag 1 was introduced. It is now connected to the ZigBee network to relay measured samples to the laptop when the system operates in the self-positioning mode.



Figure 2: Block schematic of positioning system [3].

4 MEASUREMENT RESULTS

From the measurement scenario grid, we have 213 discreet points where the RSS samples were taken. In each position, we have collected approximately 700 advertisements in about 70 seconds for remote-positioning and 700 advertisements in about 140 seconds for self-positioning. One advertisement sample includes RSS measurements from four anchors, each with four antennas, in total 16 RSS samples. During the two days long measurement for each positioning mode, there was minimal movement of any kind. Firstly, the remote-positioning campaign was performed. Two months later, the system has been switched to self-positioning, and the measurement has been repeated.

Fig. 3 shows the averaged RSS over four antennas and channels of Anchor 0 at a position projected to the corridor for remote-positioning mode. The map slope corresponds with placement of Anchor 0 (at [0, 0] m). There does not appear to be a clear connection between the peaks and the furniture. Sometimes the higher RSS is in the middle of the corridor and sometimes on the sides. There are several points where the RSS in adjacent positions (0.5 apart) differs by about 10 dB.

Because the RSS map for self-positioning mode is very similar to Fig. 3, Fig. 4 shows only the difference between measurement campaigns. The differences range from -5 to 5 dB. Because the RSS significantly varies even for small offsets around the position and the tag was placed on the measuring positions manually, we can assume that the positioning concept does not affect the RSS fluctuation. For reference, Fig. 5 depicts RSS samples collected by four antennas (with 3 cm spacing) of Anchor 0. Although it captures the farthest tag position, the differences between RSS samples are similar throughout the corridor.

On the contrary, for the moving targets, a tag in self-positioning mode receives RSS samples on a single channel at a time (the channel can be periodically changed, e.g., each 200 ms), and the advertisements from anchors are received sequentially. Therefore, the samples from all anchors cannot be captured at a single time and single position. However, the remote position systems benefit from multiple receivers, where it receives four samples per anchor per advertisement from the tag. Moreover, each antenna can receive on a different channel. That significantly increases the amount of information to the positioning processor; thus, it increases the accuracy.



Figure 3: Remote-positioning RSS map projected to the corridor floor plan for Anchor 0.



Figure 4: Difference between remote and self-positioning maps for Anchor 0.



Figure 5: RSS fluctuations on four antennas of Anchor 0. Tag in [35, 1.4] m.

5 CONCLUSION

Intuitively, there should be no difference between remote and self-positioning since both uses the same band and principles. That is experimentally verified by Fig. 4, where the differences between RSS maps are less than RSS fluctuations on antennas of an anchor. Table 1 lists statistical comparison for the RSS maps of each anchor.

However, when introducing multiple antennas on the anchor side, they can be efficiently used only in remote-positioning mode. We can see that it took twice as much time for self-positioning to collect the same number of samples as in remote-positioning. The channel switch tuning and collisions of advertisements from multiple antennas and anchors are to blame. The slower sampling rate and limited diversity make the self-positioning mode less favorable for moving tags.

	A0	A1	A2	A3
MSE(remote - self) [dB2]	6.613	7.695	6.314	7.174
Mean(Abs(remote - self)) [dB]	2.045	2.158	1.999	2.124

 Table 1:
 Mean squared error and mean of the difference between remote and self-pos. RSS maps

ACKNOWLEDGEMENT

This work was supported by the Brno University of Technology Internal Grant Agency under project no. FEKT-S-20-6325.

REFERENCES

- [1] Liu H., et al.: Survey of wireless indoor positioning techniques and systems, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 37 (6) (2007) 1067–1080
- [2] Townsend K., et al.: Getting started with Bluetooth Low Energy: tools and techniques for low-power networking, Revised First Edition. Sebastopol, CA: O'Reilly, 2014.
- [3] Polak L., et al.: Received Signal Strength Fingerprinting-Based Indoor Location Estimation Employing Machine Learning, 2021 Ad Hoc Networks (Under Review, ADHOC-D-21-00093)
- [4] Komai K., et al.: Elderly person monitoring in day care center using Bluetooth Low Energy, 2016 10th International Symposium on Medical Information and Communication Technology (ISMICT), Worcester, MA, USA, 2016, pp. 1-5, doi: 10.1109/ISMICT.2016.7498897.
- [5] Rozum S., Kufa J. and Polak L.: Bluetooth Low Power Portable Indoor Positioning System Using SIMO Approach, 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 2019, pp. 228-231, doi: 10.1109/TSP.2019.8769114.

Doktorské projekty

Silnoproudá elektrotechnika a elektroenergetika I.

ISSUE OF MEASURING PARAMETERS OF DIELECTRIC MATERIALS

Luděk Pelikán

Doctoral Degree Programme (3), FEEC BUT E-mail: xpelik15@stud.feec.vutbr.cz

Supervised by: Michal Krbal

E-mail: krbal@feec.vutbr.cz

Abstract: This article deals with the problems arising in the electrical measurement of various dielectric materials used in power engineering. The article presents the basic quantities and their measurement principles. It focuses mainly on parameters measurable in non-destructive measurements, such as volume resistivity, relative permittivity, and dissipation factor. The article draws attention to the problems that may arise during the measurement and thus guides the reader on how to measure these parameters correctly. These problems and findings were gathered during several different measurements with different types of dielectrics.

Keywords: Insulation, Dielectric, DC Resistivity, Polarization, Permittivity, Dissipation Factor, Electronic Bridge.

1 INTRODUCTION

The dielectric parameters of dielectrics used in the power industry are important properties that can greatly affect the properties of the entire device. In the case of dielectric insulation parameters for devices that are connected to an electrical grid (for example electrical cables or transformers), these parameters affect e.g., the efficiency of electricity transmission or the short-circuit properties in this grid.

These dielectric parameters can change over time, not only with normal increasing insulation temperature, but also in terms of material aging. From the beginning of use the device is subject to thermal, electrical, and mechanical stresses, foreign particle ingress and variations in temperature and humidity. This deteriorates the insulation parameters and degrades the material. This process then tends to be exponential.

Therefore, it is important to know these parameters, not only when designing these devices, but it is advisable to determine them during the operation of the device, for their continuous diagnosis or at the end of the life of the device, for possible modifications to the design of new devices generations.

Of course, there are many other measurable parameters that indicate the condition of the insulation material, whether measurable by mechanical tests or breakdown tests. However, the result is usually determined by the destruction of the dielectric.

2 DIELECTRIC PARAMETERS OF MATERIALS

The dielectric parameters of materials are very closely related to the processes that take place when an electric field is applied to a dielectric. These processes are called dielectric polarization. These are movements at the level of molecules, atoms, and their particles. Accordingly, the polarization can be simply divided into fast and slow, while from a practical point of view, mostly slow polarizations are measured, which are more indicative of the state of the material. And above all, these types of polarization tend to be energy loss. Dielectric parameters can be divided in several ways. For example, the division according to the type of applied voltage when AC and DC tests can be performed. Using an AC voltage, parameters such as dissipation factor or relative permittivity of the dielectric can be measured. Using DC voltage, it is possible to measure parameters related to dielectric absorption, i.e., polarization and depolarization currents or polarization indexes or resistivity derived from them.

Electrical conductivity and different types of polarizations are dependent on many external factors. Therefore, it is possible to create various dependences of parameters on other factors by measurement, while one of the most important is the dependence of a given parameter on temperature.

Another possibility that can often be encountered when measuring dielectrics is diagnostics with variable frequency measurement. For entire devices, such as transformers, this diagnostic is called frequency response area (FRA). The so-called dielectric spectroscopy deals with this when examining materials. It can be a measurement of a given parameter in a very wide frequency range.

2.1 **RELATIVE PERMITTIVITY**

Since all electrical equipment is composed of metal and any kind of insulation, they therefore have capacitive properties. The relative permittivity of the dielectric, also sometimes referred to as the dielectric constant, determines the magnitude of the electrical capacitance. This parameter is very closely related to the dielectric polarization. In general, with a greater value of relative permittivity, the capacity increases and with a greater insulation thickness, the capacity decreases. The relative permittivity of a specific material indicates a multiple of the vacuum permittivity value ($\epsilon_0 = 8,8542 \cdot 10^{-12}$ F/m). It is thus a dimensionless quantity. Typical relative permittivity values for solid dielectrics are in the range of 2 to 5.

2.2 DISSIPATION FACTOR

Dissipation factor (DF) also sometimes referred to as tg δ or loss factor is a parameter, which describes the quality of the dielectric material. Compared to permittivity, this is a more complex parameter, which also takes into account the conductivity of the dielectric. It practically indicates how much the dielectric differs from the ideal one, through which flows only capacitive current. In practice, however, such an ideal dielectric material does not exist.





Instead, electrical measuring on each material can be described according to Figure 1, where the current flowing through the resistor represents the current caused by the lossy polarizations and conductivity of the dielectric. The loss factor then determines the value of the ratio of the current flowing through the resistor to the capacitive current, i.e., the tangent of the loss angle δ . Typical values for solid dielectrics can vary widely, usually in the order of 10^{-5} to 10^{-1} . It is thus a dimensionless quantity, but the value can also be given in a percentage format, i.e., multiplied by 100.

You may encounter a quantity used primarily in the United States called "Power Factor". This is the value of $\cos \varphi$, i.e., the ratio of the current flowing through the resistor to the current *I*.

2.3 VOLUME RESISTIVITY

In practice, the measurement of volume resistivity involves the measurement of a current with a known magnitude of the electric field, the thickness of the material and the area of the measuring electrode. In contrast to the relative permittivity and dissipation factor, which are measured in a variable electric field, the volume resistivity is measured using a direct current electric field. The unit of volume resistivity is $\Omega \cdot m$.



Figure 2: Principal scheme of volume resistivity measurement

During this measurement, there is practically dielectric absorption of the material. The point is that after applying an electric field to the dielectric, a charging current begins to flow, which initially has an absolute maximum limited only by the resistance of the source, the leads to the dielectric test sample, their sum is shown as R_0 , and the magnitude of the voltage *U*. Charging current is given by the immediate increase of the charge charging the geometric capacity of the dielectric, to which can be added an electric charge which flows to the electrodes of the capacitor due to fast polarizations. After charging the geometric capacitance, the current theoretically decreases exponentially with a time constant given by a multiple of the capacitance C, mainly caused by the inserted dielectric between the electrodes and the resistor R_0 . In practice, other polarization currents caused by other types of polarizations in the dielectric will also contribute to the current. The course of the current will not be exponential but will decrease more slowly. The value of the steady current is given by the conductivity of the dielectric. The volume resistivity is the value, read from the meter 60 seconds after applying a DC voltage. In the same way, it would be possible to measure the current when the capacitance C is short-circuited and thus to measure the so-called resorption process in the dielectric, while it is not possible to measure the volume resistivity.

2.4 OTHER ELECTRICAL PARAMETERS

Among the measured parameters, it is appropriate to mention the surface resistivity of the dielectric (Ω), which is important for certain surface dielectrics. Its measurement is performed similarly to the volume resistivity, the arrangement of the electrodes is different.

When using a high voltage dielectric, the electrical strength of the material (V/m) is also very important. It is measured by impulse, AC or DC electric field, while statistical evaluation of the result is necessary.

3 PRACTICAL MEASUREMENT

Methods of measuring electrical parameters vary depending on the measuring equipment, often on the material being measured.

3.1 MEASURING SYSTEM

This article deals with the method of measurement using the Tettex 2830/2831 device and Tettex 2914 solid dielectric cells. It is a device that can be used to perform very accurate measurements.



Figure 3: Tettex 2830/2831 device and Tettex 2914 solid dielectric cells

It is a device that can be used to perform very accurate measurements, with a built-in AC and DC power supply, a measuring bridge, and a sensitive DC current meter.

The accuracy and limitations of such a measuring system are given by the maximum value of the measuring voltage, the sensitivity of the current meter, the quality of the comparator normal capacitor and the size of the measuring electrode.

3.2 ISSUES OF MEASURING DIELECTRIC SAMPLES

The basis of successful measurement is good creation and preparation of samples. Dimensionally, the samples should ideally extend beyond the edge of the protective electrode, which prevents unwanted stray capacitance measurements.

3.3 THICKNESS OF DIELECTRIC SAMPLES

It is very important to measure the thickness of the dielectric because it is a parameter according to which the volume resistivity and permittivity of the material are automatically measured by an instrument into which the thickness data is entered. According to standards dealing with dielectric measurements, thickness measurements should be made at several different locations and for accurate measurements the differences should not be greater than 1 %. This rule applies to the measurement of alternating parameters, or to a dielectric soft enough for proper contact between the electrode and the dielectric. Unfortunately, when it comes to measuring the volume resistivity of, for example, hard polyethylene samples, which in themselves have a very high volume resistivity, its measurement is problematic, often unrealistic. Even a small asymmetry of the sample can make the measurement impossible. An increase in pressure may help, but even that may not be enough. Therefore, I recommend choosing a very small thickness when creating samples with expected very high resistivity.

3.4 CONDITIONING OF DIELECTRIC SAMPLES

The measurement can also be influenced by the way the sample is stored before the measurement, their so-called conditioning. Standards for some measurements and materials require that the ambient temperature and humidity of the samples be maintained for some time before the measurement begins, taking into account the ability of the material to absorb moisture. It is often convenient to remove the electric charge by grounding the sample surfaces with zero potential. Otherwise, the charge may interfere with the DC tests. At the very least, it is advisable to carry out the charge removal before starting a new measurement, at least for the previous period of application of the DC field on the sample. If the charge is not removed, an incorrectly measured volume resistivity value may appear to be higher.

3.5 TEMPERATURE OF DIELECTRIC SAMPLES

Maintaining the sample temperature during measurement is also crucial. As already mentioned, the measured parameters are very temperature dependent. Problems arise especially when measuring at higher temperatures, which are common in polymeric materials.

The Tettex device can regulate the temperature with an accuracy of approx. 2 °C by means of temperature sensors and direct heating of the electrodes. However, reading the measured values in this temperature range can lead to a significant error in the measurement. The main problem is the reading of volume resistivity in 60 seconds, when due to the temperature it is difficult to schedule the reading.

4 CONCLUSION

The main text contains information that should be known to the reader who is going to measure the electrical parameters of dielectric materials. These possible measurement errors, which can usually be partially avoided by careful preparation for the measurement, only supplement other possible measurement uncertainties. These can form, for example, air bubbles in the material, or inhomogeneously distributed impurities or other impurities.

Therefore, for accurate measurements, several samples of identical material should be prepared and the results arithmetically averaged and the standard deviation determined. Usually there is a sufficient number of five samples, which is also the recommendation of the standard.

ACKNOWLEDGEMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE) in the research infrastructure CVVOZEPowerLab with support of the project LM2015092. Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports of the Czech Republic.

REFERENCES

- PELIKÁN, L.; KRBAL, M.; ORSÁGOVÁ, J. Diagnosis of Dielectric Parameters of PE and PVC Insulations. In Proceedings of the 2019 20th International Scientific Conference on Electric Power Engineering (EPE). Ostrava: VSB - Technical University of Ostrava, 2019. p. 294-297. ISBN: 978-1-7281-1333-3.
- [2] KRBAL, M.; ŠTĚPÁNEK, J.; PELIKÁN, L.; ORSÁGOVÁ, J.; KOLCUNOVÁ, I. Comparison of Electric and Radiometric Methods for Liquid Dielectric Diagnostic. Przeglad Elektrotechniczny, 2019, vol. 6/2019, no. 04, p. 15-22. ISSN: 0033-2097.
- [3] Tettex, Operating Instructions, 2830/2831 Precision Oil and Solid Dielectric Analyzer, Version 2.0, HAEFELY TEST AG

[4] ČSN IEC 250 (346466) A Doporučené postupy ke stanovení permitivity a ztrátového činitele elektroizolačních materiálů při průmyslových, akustických a rozhlasových kmitočtech včetně metrových vlnových délek. Praha: Český normalizační institut, 1998. http://csnonline.agentura-cas.cz/

IMPLEMENTATION OF ENERGY MANAGEMENT IN THE INDUSTRIAL SECTORS

Carina Beermann

Doctoral Degree Programme (1st), FEEC BUT E-mail: carina.beermann@vut.cz

Supervised by: Petr Mastny

E-mail: masty@vut.cz

Abstract: Global climate change, scarce energy resources and the increase in energy consumption make action urgently necessary. [1] For everyday life, suggestions and legal regulations have been created to deal with the scarcity of resources and sustainability. In industry, guidelines and laws regulate the handling of resources such as electrical energy. Controlled energy policy and "voluntary" reduction of energy demand, especially in energy-intensive companies, play a decisive role and offer the approach for a possible resolution of this tension through the application of an energy management system. In order to achieve the goal of efficient energy management, a constant individual review of the systems have to take place, for example with the help of the Plan-Do-Check-Act cycle at short intervals, in order to offer continuous.

Keywords: Energy management, Energy policy, PDCA-Cycle, Renewable energy sources

1 INTRODUCTION

This paper presents the requirements for energy management from the perspective of the International Organisation for Standardisation, European Union and the international and current European legal situation. Based on guidelines, studies, regulations, directives and legal texts, the added value of energy management in connection with the PDCA cycle is analysed. The focus of this work is mainly on the implementation of energy management in the industrial sectors. The strategy and procedure for implementation are exemplified by the PDCA cycle. A large part of society is dependent on resources such as electricity, oil, gas, and water. Without these resources, daily life would be inconceivable. First, energy policy goals oblige and encourage industry to operate an energy management system. In the meantime, European law has developed into a very important source of law, because the primacy of application of European law applies to matters relating to the Union or directly affecting Europe. In this context, regulations and directives are of particular interest within the framework of secondary law. Regulations have direct effect for citizens and member states (Art. 288 TFEU). Directives must be transposed into national, Member State law in conformity with the directive by certain deadlines.

In contrast to a regulation, they only specify the objective, but not the means. They are therefore addressed to the member states, usually with the requirement to transpose them into national law - laws or regulations. Therefore, the objectives of directives are also reflected in concrete form in national laws. The overriding motivation is to raise awareness in dealing with valuable resources; in the industrial environment as well as in private households. A communication-capable compact circuit breaker (Siemens 3VA2) is used as an object in the energy distribution system as an example for the implementation of energy management in the industrial environment. With the help of this switch, it is possible to implement energy management and/or energy monitoring in the field level (Figure 2) and/or the production level (Figure 2).

2 LEGAL FRAMEWORK FOR THE ENFORCEMENT OF ENERGY EFFICIENCY

2.1 CURRENT EUROPEAN LEGAL ACTS

To implement the Paris Climate Agreement of 2015, the European Commission presented its legislative package "Clean Energy for All Europeans" on 31 November 2016 as the so-called "Winter Package" with new climate and energy policy targets for the period after 2020 and the framework for the implementation of the Energy Union until 2030 [2].

The three main objectives of this "winter package" are firstly to reduce greenhouse gas emissions by at least 40 % compared to 1990, secondly to increase the share of renewable energy sources to at least 32 % of final energy consumption, and thirdly to increase energy efficiency by reducing primary energy consumption by at least 32.5 % compared to the reference scenario (1990) [2]. In the meantime, the European Parliament has adopted the amendments to the Renewable Energy Directive (EERL II) [3], the Energy Efficiency Directive (EED) [4], as well as the Regulation on the Governance of the Energy Union¹ [5]. The revised Energy Efficiency Directive, which is of particular interest here, provides for a reduction of primary energy consumption within the European Union by 32.5 % by 2030 compared to an underlying reference development (Art. 1 I EED). The central implementation element, the "final energy savings obligation", was extended and strengthened beyond 2020. For the first time, annual real savings of 0.8 % of energy sales in each Member State were agreed (Art. 7 I EED).

2.2 ENERGY MANAGEMENT ACCORDING TO ISO 50001 IN CONNECTION WITH THE PDCA-CY-CLE

Every company can decide how it wants to organize its energy management, but it must adhere to the requirements of ISO 50001 to ensure that subsequent certification is successful. ISO 50001 is a globally valid standard for the implementation of an energy management system and contains measures to increase the energy efficiency of companies. It describes many criteria, the realization of which can lead to high energy savings potentials and thus reduce costs, but also reduce the influence on the environment.

The ISO certification according to ISO 50001:2011-12 is valid for a period of three years but have to be verified by annual audits. The aim is to monitor compliance with the defined indicators, but also the implementation of further energy saving measures, and to ensure a continuous process, which is a strategic task of the management level of a company. If a company decides to be certified according to ISO 50001, it has entered a so-called PDCA cycle. This includes the several areas like Plan, Do, Check and Act and have to be carried out annually. It is derived from ISO 50004 (2014-12): Energy management systems - Guidance for implementing, maintaining, and improving an energy management system. In this way, continuous optimisation and a permanent reduction in energy consumption are to be achieved. After three years, a renewed certification have to take place as a recertification. For the implementation of an energy management system, an energy assessment of the company/organisation has to be first take place. Therefore, a collection of all relevant data must be carried out by the respective persons in charge beforehand. This data includes, among other things, information on past and current energy use and energy consumption and on the factors that influence energy consumption. Facts about current energy efficiency should also be collected. In a three stage analysis, energy consumption and energy use are presented in de-tail, the areas with the highest consumption are identified, and subsequently possibilities for permanent improvement of the energy balance as well as operational goals are determined ("Plan" in PDCA cycle). Once the planning has been completed, the next step is to implement the defined measures ("Do" in the PDCA cycle). In any case, the employees should be involved in the implementation of the measures at this point at the latest. They work in the respective areas on a daily basis and can often provide additional information on further savings opportunities.

¹ Governance of the Energy Union (GovEnU)

Then those responsible monitor the measures taken and the achievement of the goals ("Check" in the PDCA cycle). For example, the current energy consumption is measured and this data is aligned with the targets. Realtime data is a profitable means of drawing timely conclusions and making comparisons. In addition, internal audits have to be carried out regularly so that system conformity can be checked and any corrections can be initiated. The documentation of the data should be transparent so that it is available in the desired form for future certification. By monitoring the processes, possible changes are possible, which can or have to be made by those responsible for the process, i.e. the management level ("Act" in PDCA cycle) - **Figure 1** [6]. The management review is also particularly important as a strategic task, because the responsible management is responsible for monitoring all measures introduced so far and for checking their effectiveness. If necessary, this process can be followed by further measures to improve the energy balance or new goals can be decided upon. For this purpose, special action plans are drawn up that build on each other.

2.3 **OBJECTIVES OF THE EUROPEAN UNION**

Based on the EU's strategic goals up to the year [7]:

- Reduction of greenhouse gas emissions by 40 % compared to 1999.
- Increase the share of renewable energy sources to at least 32 % of final energy consumption.
- Reduction of primary energy consumption by at least 32.5 % compared to 2007.

These targets are an important driving force for technical innovations in energy management.

3 AIM OF IMPLEMENTING AN ENERGY MANAGEMENT SYSTEM

3.1 CREATE INCENTIVES

In order to achieve the goal of efficient energy management, future plants have to be planned according to the requirements and existing plants must be modernised, if necessary from the ground up. Directives, laws, and standards primarily determine the behaviour of society and industry with regard to access to electronic energy and energy consumption. Legal requirements are obligations and must therefore be respected and complied with. At the same time, however, incentives are created that have a motivating influence to possibly invest in a future-oriented manner beyond the legal basis. This means that digital evaluation, visualisation and storage of measurement data should take place in the area of energy management. With the right components, energy savings, cost savings, quality of supply, security of supply and energy efficiency can be guaranteed.



Figure 2: Example for an Energymanagement-System [8]

This illustration from Siemens shows an example of how an energy management system can be implemented (**Figure 2**). All levels from management to field level are brought together in one flow. Data for energy monitoring can be recorded using energy measuring components such as a communication-capable measuring device, a circuit breaker, a soft starter, a frequency converter or a motor. Furthermore, processing and visualisation are possible with the help of a PLC and a human machine interface. Data selection and evaluation can be realised with the help of software at the management level. There are many different application possibilities for implementing energy management, but the management level, the production level and the field level should be taken into account as a guide. As an additional option, the data can be uploaded to a cloud, which allows access from many different locations. The investments should be in the area of communicationcapable measuring devices with a wide range of measuring possibilities, the energy management software as well as the specialised personnel. As a result, energy monitoring is necessary. The continuous evaluation and recording of relevant measurement data helps, for example, to reduce the effort required for the prescribed regulations and logged energy measurement data.

3.2 GUIDELINES FOR IMPLEMENTATION

To implement these goals, the following important guidelines and laws have been issued or are currently being adapted as you can see in **Table 1**. The table lists the European Union directives that must be transposed into national law. As mentioned above, implementation is mandatory for each member state.

EU
Energy Services Directive (ESD) of 2006 (Directive 2012/27/EU of 25 October 2012) [9]
Eco-design Directive and Energy Using Products Act (EuP Directive) (Directive 2005/35/EC of 6 July 2005) [10]
Emissions allowance trading [11]
Building Efficiency Directive (EU) 2018/ 844 of 30.05.2018 [12]
Energy Efficiency Directive (EED Directive 2018/2001 of 11.12.2018) [4]

Table 1: Basic rules for implementation of energy management

Energy management and the promotion of additional energy storage technologies offer the potential to make optimal use of additional energy sources and to reduce energy demand. The implementation of energy efficiency measures also contributes to independence from energy imports. The introduction of an energy management system should help to determine the exact energy consumption of a company and show possibilities to save resources such as electricity, heating and water. The introduction of an energy management system is not just a one-time assessment and inventory control but must be repeated continuously in cycles. The measures have to be carefully planned - in relation to the respective company/organisation - regularly monitored and the data collected and compared.

3.3 IMPLEMENTATION OF ENERGY MANAGEMENT USING THE EXAMPLE OF THE **3VA2** COM-PACT CIRCUIT-BREAKER (SIEMENS)

Compliance with the regulations for valuable energy monitoring of energy data such as current, voltage, power, phase position would be realisable with a commercial measuring device and current transformer. In an existing installation, there may be a lack of space in the energy distribution system, making it impossible to implement a suitable and compliant energy management system on site.

For such challenges, the 3VA2 compact circuit-breaker offers a space-saving and optimal solution. The flexible communication options via Profinet, Profibus, Modbus TCP and Modbus RTU provide a high level of transparency.

By reporting the system statuses and measured values to higher-level management systems, the system utilisation can be optimised, savings potential can be identified and energy efficiency can be increased (transparency of the status of the energy distribution and its energy flows). The 3VA2 compact circuit-breaker can thus support the company's energy management.

The energy operating data, service and maintenance information from the switch are read out and projected on a human machine interface. In addition, the data is stored e.g. via an internal memory of the human machine interface, which, depending on the setting, is evaluated after one working day with the help of a program on the computer and then made available to the person responsible for energy management for further processing. Relevant energy data include, as already mentioned above, current, voltage, active line, apparent power, current frequency, fundamental oscillation power factor; service and maintenance information includes, for example, the number of trips (due to a fault), mechanical clearances and the reason for the last trip.

A circuit breaker is used in almost every industrial building today. If the company is willing to use an energy management system, the communication-capable compact circuit-breaker helps to record all relevant data for an energy management system right from the field level. The recorded data is then evaluated with a programme and compared with the previous measurement data according to the PDCA cycle in order to have reached the target agreement after successful measures. A communication-enabled component such as the compact circuit-breaker is an entry point for companies that are obliged to operate an energy management system or do so voluntarily. An extension with energyefficient and data-reporting loads or further compact circuit-breakers and PLCs then provides energy management at all levels of the automation pyramid [13].

4 CONCLUSION

Efficiency and savings in the field of energy play a key role today and in the future. With the trend towards renewable energies, the legislative and directive process is being intensified by the European Union in order to save energy. If a company uses the optimal components for an energy monitoring system, a basis is created for operational energy management in accordance with ISO 50001. With maximum plant safety, highest plant availability, short amortisation time of the required equipment, permanent cost reduction (energy savings with plant and cost savings) through e.g. avoidance of financial sanctions such as fines, the company benefits in many ways. However, the optimum can only be achieved with a constantly individually adapted energy management system; this means that a review of the systems takes place at short intervals in order to offer continuous improvement. The implementation of an energy monitoring system is of fundamental importance in the context of the energy audit, but especially in the context of an energy management system. In summary, it can be stated that with EN16247, ISO 50001:2011-12 and its successor ISO 50001:2018, a global standard is available that sustainably supports organisations/companies in driving energy efficiency measures forward. The next work will relate to the international and European standards, examining the relationship between the standards and their impact on energy management. In addition, an investigation of the implementation of the European directives and regulations into national German law will be undertaken in the near future.

ACKNOWLEDGEMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports of the Czech Republic under Brno University of Technology specific research programme (project No. FEKT-S-20-6449).

REFERENCES

[1] Siemens AG (2014): Zuverlässig, nachhaltig und effizient. TÜV-geprüftes Energiemonitoringsystem gemäß ISO50001. pdf: Industry Online Support Siemens (SIOS) - 16.01.2021.

- [2] European Commission (2017): 2030 climate & energy framework. 2030 climate and energy framework- existing ambition. https://ec.europa.eu/clima/policies/strategies/2030_en European Commission 16.01.2021.
- [3] Official Journal of the European Union (21.12.2018): DIRECTIVE (EU) 2018/2001 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, https://eur-lex.europa.eu/legal-con-tent/EN/TXT/PDF/?uri=CELEX:32018L2001&from=DE EUR-Lex Access to European Union law 07.01.2021.
- [4] Official Journal of the European Union (21.12.2018): DIRECTIVE (EU) 2018/2002 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L2002&from=EN - EUR-Lex Access to European Union law 07.01.2021.
- [5] Official Journal of the European Union (21.12.2018): REGULATION (EU) 2018/1999 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1999&from=EN - EUR-Lex Access to European Union law 07.01.2021.
- [6] eQM-System (21.09.2020): Energiemanagement. 4. Der PDC-Zyklus in einem Energiemanagementsystem. https://www.eqm-system.de/pdca-zyklus-energiemanagementsystem/ eQM-System 02.01.2021.
- [7] Official Journal of the European Union (21.12.2018): Regulation (EU) 2018/1999 of the European Parliament and of the Council, Erneuerbare Energie. 31.01.2019. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:4372645 EUR-Lex Access to European Union law 07.01.2021.
- [8] Siemens AG: Image database. https://www.automation.siemens.com/bilddb/search.aspx?akt-prim=0&nodeid=10204319&lang=de&usestructure=2.05.3.2020.
- [9] Official Journal of the European Union (14.11.2012): DIRECTIVE 2012/27/EU OF THE EU-ROPEAN PARLIAMENT AND OF THE COUNCIL of 25 October on energy efficiency, amending Directives 2009/125/EC and 2012/30/EU and repealing Directives 2004/8/EC and 2006/32/EC https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32012L0027 - pp. L351/1 EUR-Lex Access to European Union law 02.01.2020
- [10] Institute for European Environmental Policy (April 2009): A Report on the Implementation of Directive 2002/96/EC on Waste Electrical and Electronic Equipment (WEEE), p.13. https://ec.europa.eu/environment/archives/waste/reporting/pdf/WEEE_Directive.pdf-02.01.2021
- [11] European Commission (2015): European Commission Energy Climate change, Environment -Climate Change - EU Action - EU Emissions Trading System (EU ETS) https://ec.europa.eu/clima/policies/ets_en - European Commission - 02.01.2021.
- [12] Official Journal of the European Union (19.06.2018): DIRECTIVE (EU) 2018/844 OF THE EU-ROPEAN PARLIAMENT AND OF THE COUNCIL of 30 May 2018 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency, L156/75. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L0844&from=EN - EUR-Lex Access to European 02.01.2021
- [13] Information from own investigation from the bachelor-thesis "SENTRON Kompaktleistungsschalter 3VA2" -Einsatz in der Industrie 4.0 - Carina Beermann 01.08.2017

ASSESSMENT OF INVESMENTS IN LOCAL DISTRIBUTION SYSTEM USING COST-BENEFIT ANALYSIS

Matěj Vrtal

Doctoral Degree Programme (1), FEEC BUT E-mail: xvrtal00@stud.feec.vutbr.cz

Supervised by: Petr Toman

E-mail: toman@feec.vutbr.cz

Abstract: This paper demonstrates the possibility of using cost-benefit analysis to evaluate investments in the development of local distribution systems, specifically the modernization of distribution transformers. The article interprets the results of the analysis of three investment variants of a specific company operating a local distribution system in the Czech Republic.

Keywords: LDS, CBA, transformer, investments, assessment.

1 INTRODUCTION

The local distribution system (LDS) is being used to connect end-users to the electricity network. Through this system, the supply of electricity for the customer and his delivery point is ensured. All rules for the operation of LDS are set out in the Energy Act and the relevant implementing regulations. The minimum technical, planning, operational and information requirements for the connection of users to distribution networks are set out in the Distribution System Operation Rules, which are approved by the Department of Energy.

This paper presents the principle of assessing investment in the construction of a new transformer station in LDS. To demonstrate the method, the real intention of the company operating the LDS is used. The economic side of a new project is always the main indicator of the suitability of a variant solution, but not the most important. Non-monetary pros and cons can have a greater impact on decisions. Both of these aspects are taken into account in a cost-benefit analysis (CBA), supplemented by static and dynamic monetization methods.

2 CURRENT STATUS OF SOLVED LDS

Two 22/0.4 kV air-cooled transformers, each with an apparent power of 630 kVA, are located in the low-voltage substation of LDS. One of these transformers is permanently maintained as a back-up. The second of the transformers is operated with only a third load, so the current state of LDS is on the verge of profitability. The aim of the intended investment is to reduce the operating costs of the LDS while keeping a sufficient power reserve for the implementation of other investments proposed in [1] - in particular, the construction of charging stations for electric vehicles.

3 DESCRIPTION OF INVESTMENT OPTIONS

The reference variant (state without project implementation) in this case is to keep the current state as it is today. Of the two 630 kVA transformers, only one will continue to operate, both of which will remain at the current site of the current substation. The substation will continue to operate without modernization of the equipment.

3.1 OPTION 1

Investment variant 1 assumes the installation of a new oil transformer with a nominal output of 630 kVA in a new kiosk station. The kiosk will be located outside the existing substation.

3.2 OPTION 2

Investment variant 2 assumes the installation of a new oil transformer with a nominal output of 400 kVA in a new kiosk station. The same placement as for investment variant 1 is assumed, so the only different factor between investment variant 1 and 2 is the rated power of the transformers used.

3.3 OPTION 3

The third investment option involves the construction of two kiosk transformer stations, each at one of the considered locations specified in more detail in [1]. The transformer located on the northeast corner of the area would have a rated output of 250 kVA and would serve to cover its own consumption inside the complex. The second of the transformers, located in the southwest corner of the complex, near the building of the existing substation, would be used to power the land outside the company's premises. This kiosk would also be used as a substation for possible photovoltaic power plant proposed in [1].

4 COST-BENEFIT ANALYSIS

The essence of the method is to analyze the investment impact on actors, quantification of the identified effects, and conversion to a common numeric (ideally financial) unit. Then we can use the indicators of criteria of net present value, internal rate of return of economic CF, profitability index, and payback period.

The advantage of CBA is that the benefits, but also the costs do not need to be expressed in monetary terms necessarily. It can be alternatively expressed in other ways (for example, in socioeconomic, environmental, or qualitative terms). However, it is important that it always needs to be expressed in a meaningfully measurable way.

4.1 CONVERSION OF QUANTIFIED BENEFITS INTO CASH FLOWS

The evacuation of the premises of the current substation will bring a usable area of approximately 200 m^2 , which can be used directly for the purposes of the company or offered for rent. After analyzing the market for renting production premises, we can consider the rental price for 1 m^2 of approximately 100 CZK per month. The stated price is the median of the prices offered for production premises, taking into account the concerned locality. The freed-up area would therefore bring savings of approximately 20000 CZK per month, on which we can further calculate. Compared to the need to provide production space differently, the savings would be significantly higher.

We will quantify the lower no-load and short-circuit losses using data on the annual electricity consumption and the nameplate values of the new and current transformer. The total electricity losses for a given period are determined according to the relation [2]:

$$\Delta W_T = \Delta W_0 + \Delta W_z = (\Delta P_0 + k_\Delta \cdot \Delta Q_0) \cdot T + (\Delta P_k + k_\Delta \cdot \Delta Q_k) \cdot \left(\frac{P_{max}}{S_n \cdot \cos\varphi}\right)^2 \cdot T_\Delta \tag{1}$$

where ΔW_z are the annual losses of electric energy due to the transformer load (kWh), ΔP_k are the active short-term transformer losses (kW), ΔQ_k are the short-circuit reactive losses of the transformer (kVAr), S_n is the apparent transformer power (kVA). ΔW_0 are the annual no-load losses of the transformer (kWh), ΔP_0 are the active no-load losses of the transformer (kW), ΔQ_0 are the no-

load losses of the transformer no-load (kVAr), T is the time the transformer is in operation for the given period (h).

A detailed calculation is realized in [1]. The total electricity losses of the existing transformer are $\Delta W_T = 25571.25$ kWh, for the new transformer with an apparent power of 630 kVA, $\Delta W_T = 17140.86$ kWh and for the new transformer with the apparent power of 400 kVA $\Delta W_T = 7753.39$ kWh.

To economically quantify the losses, we will use the actual price of the electricity consumed. In general, we divide the price of electricity into two parts. The first part is the price for power electricity. It consists of the amount of active electricity consumption and electricity tax. The second part is the payment for distribution. The distribution price consists of fixed charges which are derived from the reserved power input and do not change with the amount of energy consumed, and charges which are derived from the amount of energy consumed. The total variable costs (sum of the price of power electricity and the variable part of the price for distribution) per 1 MWh of electricity for the given company were set in [1] at 1594.68 CZK. The cost of electricity loss N_{dT} for the first variant is then 40777.96 CZK per year, for the second variant 27334.19 CZK per year, and for the third variant 12364.18 CZK per year.

4.2 ECONOMIC LOAD OF THE TRANSFORMER REGARDING LOSSES

The average annual value of the load β of the currently used transformer is approximately 6.3%. The economical load of the transformer is the load at which the total losses of active power caused by the transformer are minimal. To determine the optimal load of the transformer, specific losses are used, i.e., losses per unit load. To calculate the economic load, we use the relation derived in [2].

$$S_{h} = S_{n} \cdot \sqrt{\frac{\Delta P_{0} + k_{\Delta} \cdot \Delta Q_{0}}{\Delta P_{k} + k_{\Delta} \cdot \Delta Q_{k}}}$$
(2)

The calculations show that the economical apparent power that would be appropriate to load a 630 kVA transformer is 286.58 kVA. The economical apparent power with which it would be appropriate to load this 400 kVA transformer is 142.21 kVA. This value is close to the maximum quarter-hour maximum for 2019 in the solved LDS, which was 123 kW. Neglecting the possibility of a significant increase in consumption in the future, we can consider it more appropriate from the point of view of the economy to use a transformer with a lower apparent power.

4.3 ECONOMIC LOAD OF THE TRANSFORMER

The procedure of the previous subchapter is limited to losses only and does not take into account the purchase price of the transformer and its depreciation. From an economic point of view, it is therefore more accurate to speak of annual production costs, which include both losses, expressed in terms of loss costs, and costs derived from the cost of the transformer, using the relation [2]:

$$S_{he} = S_n \cdot \sqrt{\frac{N_{iT} + (\Delta P_0 + k_\Delta \cdot \Delta Q_0) \cdot n_\Delta^0}{(\Delta P_k + k_\Delta \cdot \Delta Q_k) \cdot n_\Delta}}$$
(3)

The calculated value of the economic load from the economic point of view ($S_{he} = 1144.28$ kVA) is well above the rated power of the transformer ($S_n = 630$ kVA). This value is, of course, nonreal. The result can be interpreted as the chosen transformer is very good for this use.

4.4 DETERMINATION OF THE DISCOUNT RATE

Determining of the discount rate is important for calculation by dynamic methods. In our case, we will apply only the dynamic method for evaluating investments – the net present value of *NPV*. The discount rate can be considered 2 % p.a., which is the inflation targeting of the Czech National Bank for this year. If the maximum social permissible rate is up to 5.5 %, we can consider a discount rate of 5 % p.a. This value will be used in the calculations.

4.5 CALCULATION OF ECONOMIC INDICATORS

• Payback time

The payback period of an investment is an important and frequently used investment valuation indicator that gives a rough idea of the time after which the returns on the initial investment will exceed the value of the investment itself. The method is static and is an expression of the so-called simple payback period. The calculation of the simple payback period is based on a formula [3]:

$$TN_{P} = \frac{IN}{CF} \tag{4}$$

where *IN* is the cost of investment (investment expenditure) (CZK), *CF* is the annual cash flow (annual income - cost savings due to the investment) (CZK).

• Return on investment

The return on investment (ROI) compares the net accounting profit to the size of the investment, or the volume of total assets and liabilities. Simply put, we can define *ROI* as the ratio of money earned to money invested. It therefore indicates the return on the amount spent as a percentage. We calculate it according to the relation [3]:

$$ROI = \frac{return}{investment} \cdot 100 \tag{5}$$

• Net present value method

The last method that will be used to assess individual investment options will be the *NPV* method. The main advantage of this method is the consideration of the time factor. We calculate the net present value according to the formula [3]:

$$NPV = \sum_{n=0}^{20} \frac{CF}{(i+i)^n}$$
(6)

where CF is the cash flow (CZK), *i* is the interest rate (-), *n* is the number of years we have to wait for income (-), *IN* is the initial investment (CZK).

Calculations of economic indicators for individual investment variants are summarized in Table 1.

Investment option	Payback period (years)	ROI (%)	NPV (CZK)
1	7.10	281.60	1 665 369.79
2	6.41	312.11	1 974 362.60
3	11.92	167.83	380 739.24

Table 1: Economic indicators for particular investment variants

5 CONCLUSION

Compared to the reference variant, all three considered investment variants are preferably considered an economic comparison. It should be mentioned that in the calculation of economic returns and other indicators, the possibilities of selling the current equipment of the substation and transformers were not considered, which would increase the profitability of the investment. After making a comparison from a noneconomic point of view, we identified this investment option as the least meaningful, and we were able to move on to deciding between options 1 and 2. Figure 1 shows comparison between the reference variant and investment options 1 and 2. The third investment option is on the verge of profitability, as we can see in Table 1.



Figure 1: Cost comparison taking into account the growth of electricity prices

In general, we can say that while maintaining the current state, it pays off to invest in a transformer with a nominal output of 400 kVA. With increased consumption in the future according to the planned intention, it is a reasonable choice to invest in a higher rated power of the transformer, despite the slightly higher purchase price and higher operating costs.

REFERENCES

- [1] VRTAL, M: Conceptual development of the Local Distribution System Prototypa, a.s. [online]. Brno, 2020 [cit. 2021-03-10]. Master thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Electrical Power Engineering. Supervised by Lukáš Radil.
- [2] CHMELA, M. Economics and Management. Brno, 2007. Brno University of Technology.
- [3] SOUKOPOVÁ, J. Cost-Benefit Analysis: Public Orders and Public Projects Evaluation [cit. 2021-03-10]. Masaryk University.

HARDWARE-IN-THE-LOOP SIMULATION OF A DISTRIBU-TION NETWORK WITH AN ON-LOAD TAP CHANGER CONTROLLED BY AN INTELLIGENT ELECTRONIC DEVICE VIA DIGITAL COMMUNICATION

Viktor Jurák

Doctoral Degree Programme (2), FEEC BUT E-mail: xjurak04@stud.feec.vutbr.cz

Supervised by: Jaroslava Orságová

E-mail: orsagova@feec.vutbr.cz

Abstract: This paper deals with hardware-in-the-loop simulation of the distribution network, which is equipped with the on-load tap changer on the HV / MV transformer. This tap changer in the simulation is controlled from a real physical external device - the multi-functional intelligent electronic device REX 640 from ABB. This device receives and processes the sampled measured values of voltages and currents. Based on these values, an instruction to switch the transformer tap is then given back into the simulation. All communication between the simulator and the external device is provided using the IEC 61850 industrial communication protocol. The paper summarizes both the benefits of hardware-in-the-loop simulations and the possibilities of using intelligent electronic devices and data communication in distribution systems.

Keywords: HIL simulation, Real-time simulator, IED, Voltage regulation, SMV, GOOSE

1 INTRODUCTION

The way they operate distribution networks is changing. Increasing distributed generation sources, which leads primarily to change the direction of power flow. Therefore, there is an effort to create complex control systems with advanced digital communication, which will help eliminate the negative effects of these sources and will turn them into positives (for example, smart management of these resources).

The increasing complexity of control systems leads to the need to use methods of testing their functionality, which are more sophisticated. One option is real-time simulator (RTS) testing. The RTS is a device that is able to perform EMT (Electromagnetic Transient) simulations in real time. Analog and digital input and output peripherals can also be connected to the simulator. Through these peripherals, it is possible to connect an external device and integrate it into the simulation. Therefore, RTS calculates the network states and sends the results to the external device via the output peripherals. The device responds to these results and sends feedback via input peripherals back to the simulator, which reacts again to these stimuli. In this case, it is a HIL (Hardware-in-the-loop) simulation. For example, the source [1] deals with these simulations.

The paper describes a real-time simulation of a simple HV / MV substation and one section of MV overhead lines. An intelligent electronic device (IED) is connected to the simulator via digital communication. The voltage regulation function is set in the IED. This function controls the on-load tap changer (OLTC) of the HV / MV transformer. A similar test, but with a simpler communication structure and without RTS, is discussed, for example, in article [2].

2 DESCRIPTION OF THE SIMULATION

The simulation was designed in RSCAD in the draft module. The scheme of the simulated network is shown in Figure 1. The network is supplied from an ideal voltage source at a voltage level of 110

kV. A three-winding transformer with a nominal apparent power of 25 MVA with OLTC on the primary side, provides the transformation from HV to MV. The transformer has ± 8 taps, each with a value of 2% of the nominal voltage. Taps can be switched within the simulation, both manually and externally (green arrow), using IEC 61850 GOOSE (Generic Objected Oriented Substation Event). The tertiary winding is delta-connected and is loaded with a fixed value of active power of 1 MW. The secondary winding of the transformer is connected to the AlFe 110/22 type overhead line in a side-by-side arrangement, which is 20 km long. The line was modelled in the RSCAD program in the Tline module, based on the physical parameters of the line. These parameters were taken from sources [3] and [4]. The resulting value of the series impedance is $\overline{Z}_k = (R_k + jX_k) = (0.257 + j0.340) \Omega \cdot \text{km}^{-1}$. A load is connected at the end of this line. The voltage (blue arrow) and current (red arrow) measurements are located on the secondary side of the transformer and are routed to a module that provides communication according to IEC 61850 SMV (Sampled Measured Values).



Figure 1: Simulated network diagram

The external connection is also shown in Fig.1. A computer (PC) is connected to the Ethernet interface of the simulator, through which the simulation was designed and which is used to parameterize and run the simulation. There are also Ethernet ports on the simulator of the GTNETx2 network card, which has two modules for communication via the IEC 61850 protocol. Another port has a GTSYNC card for time synchronization according to IEEE 1588. All these ports are connected to one switch, to which the IED REX 640 is also connected. the synchronization source is the internal clock of the IED.

2.1 SETTING AND PARAMETERIZATION OF THE IED

The nominal value of the voltage U_n was set to 22 kV (L-L, RMS), i.e. to the same value as the nominal voltage of the system. The rated current I_n has been set to 300 A with respect to the maximum value of the overhead line load (318 A) to the value of the next lower rated value.

The OL5ATCC function block was used for voltage regulation. According to the manual for the IED [5], the function works so that if the average measured RMS value of all three phases U_{meas} for the time delay t_1 exceeds the range of the set voltage band U_{band} and does not fall below the hysteresis value, the tap is switched. The hysteresis value is 10% of the U_{band} value and is located at the

lower and upper limits within the U_{band} range. The U_{band} range is symmetrical according to the set value $U_{\text{band_center}}$. This centre value is therefore the required voltage value. According to [5], it is recommended to set the U_{band} range to a value around twice the step of one tap, but never below this value. For simulation purposes, this value was chosen as twice the step of one tap. If the voltage exceeds the limit by a larger value and it would not be sufficient to switch only one tap, the timer t_2 is activated instead of the timer t_1 . Timer t_1 is set to 1 s, timer t_2 to 2 s. In practice, these values are set to a significantly larger value. According to [5], the recommended value for t_1 is 60 s and for t_2 30 s. The low set value of the timers is chosen to speed up the performed tests.

In the settings, it is also possible to compensate the voltage drop on the line. This function is called LDC (Line Drop Compensation). If the LDC function is switched on, it is necessary to set the percentage values of the voltage drop on the resistance of the line U_r and on the reactance of the line U_x . These values are calculated according to equations (1 and 2), which are taken from [5].

$$U_r = \frac{\sqrt{3} \cdot I_n \cdot R}{U_n} \cdot 100 , U_x = \frac{\sqrt{3} \cdot I_n \cdot X}{U_n} \cdot 100$$
(1, 2)

3 SIMULATION RESULTS

Two tests were performed as part of the simulations. In test a), the load LO1 was set to a constant value and the voltage at the source changes. Test b) consisted of verifying the LDC function. Here, the voltage of the source remained at the nominal value and the value of the load at the end of the overhead line changes. The measurement in test b) remains at the same place as in test a), but now the IED calculates the voltage on the load based on the measured current and the set line impedance.

3.1 TEST A)

Simulation results a) are shown in Figure 2. The time from the beginning of the simulation in seconds is plotted on the common x-axis. The first part of the figure shows the set voltage value on the primary side of the transformer u_{T110} . The voltage is plotted in per units, where the reference value is the nominal value of the voltage $U_{nHV} = 110$ kV. At the beginning of the simulation, this voltage is at the nominal value. At time 1 s, the voltage begins to rise to a peak of $1.05U_{nHV}$, which occurs at time 11 s. Then the voltage drops sharply back to the nominal value. In the second part of the figure, the voltage on the secondary side of the transformer is plotted. This voltage is also in per units. The reference value is the nominal voltage $U_{nMV} = 22$ kV. In this part of the graph, both the voltage from the simulation u_{T22} (black curve) and the voltage u_{T22_GOOSE} (red curve), which measures the IED and sends its value back to the simulator using GOOSE messages. When comparing these two waveforms, it can be seen that they are similar.

The voltage profile of the u_{T22_GOOSSE} is stepped, which is because the IED sends the measured value only every 500 ms. This value is sent in FLOAT32 format and is a period value. This value is calculated directly within the OL5ATCC function block and is used to check whether the function calculates the value correctly. Another way to send values from the IED is using SV, but the network card of the simulator does not allow receiving and sending data in SV format at the same time within one module. However, for data received in SV format, it would not be possible to verify the correctness of the calculation of the average three-phase RMS value.

At the beginning of the simulation, it can be seen that while the voltage on the primary side of the transformer is nominal, the voltage on the secondary side is below the nominal value. This difference is due to the voltage drop at the transformer. The voltage u_{T22} shows the correct function of the voltage regulation. When this voltage is above the value of one tap, i.e. above $1.02U_{nMV}$ ($U_{band_center} + 0.5U_{band}$), for 1 s (time set by timer t_1), the tap is switched downwards. This fact is also evident from the third part of the graph, where the binary signals received from the IED via GOOSE are plotted. The *G_TIMER* signal indicates an active timer and the *G_DOWN* and *G_UP*

signals are commands to decrease / increase the tap. In the last part of the figure, the actual tap position is plotted, which is in the basic position at the beginning of the simulation. It can be seen from the diagram that after the sixth second of the simulation the timer is activated, then after the seventh second the signal from the IED G_DOWN decreases the tap. The timer is reactivated before the 11th second, but before it can finish, the voltage drops sharply. The timer is then activated again, but now counts down the time until the tap increases, which occurs after the 12th second.



Figure 2: Results of simulation a)

3.2 TEST B)

The results of the simulation are evident from Fig.3. The graph is in a similar format as in the previous case. In the first part of the graph, the set load S_{LO1} of the load is plotted.



Figure 3: Results of simulation a)

In the second part of the graph, the voltage on the secondary side of the transformer u_{T22} (red curve) and the voltage at the end of the line U_{LO1} (green curve) are plotted. The difference between these curves is due to the voltage drop on the line. As the load increases, so does the difference between these voltages. The third and fourth parts of the graph are in the same format as in test a). It can be seen from the diagram that the default state is now the first tap. As soon as the voltage on the load falls below the value of one tap (0.98 U_{nMV}), the timer is activated and after time t_1 , the tap is changed. Therefore, the controller now does not regulate to the value of the voltage behind the transformer, but to the value of the voltage at the load node. It can be seen from the picture that the tap will increase three times. After a sharp load reduction, the tap will decrease.

4 CONCLUSION

HIL simulations using a real-time simulator are a suitable tool for testing devices for grid control and protection. With the help of a real-time simulator, it is possible to test the devices in almost the same conditions to which they are exposed during real operation. In this experiment, the possibility of using a multifunctional IED for voltage regulation by switching transformer taps using IEC 61850 SMV and GOOSE was successfully verified on a simulated part of the distribution network. This concept, where one universal device is used, instead of several single-purpose ones, will already be the standard. The paper also outlined another possibility to improve voltage quality by using a LDC function, which may help, but which needs to be approached in a reserved manner and based on judgment based on knowledge of the character of the network.

ACKNOWLEDGEMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Author gratefully acknowledge financial support from the Ministry of Education, Youth and Sports of the Czech Republic under BUT specific research programme (project No. FEKT-S-20-6449).

Author gratefully acknowledge financial support from the Technology Agency of the Czech Republic (project No. TK01030094).

REFERENCES

- A. S. Makhzani, M. Zarghami, B. Falahati and M. Vaziri, "Hardware-in-the-loop testing of protection relays in distribution feeders with high penetration of DGs," 2017 North American Power Symposium (NAPS), Morgantown, WV, 2017, pp. 1-6, doi: 10.1109/NAPS.2017.8107191.
- [2] N. Sichwart, A. Eltom and G. Kobet, "Transformer Load Tap Changer control using IEC 61850 GOOSE messaging," 2013 IEEE Power & Energy Society General Meeting, Vancouver, BC, Canada, 2013, pp. 1-5, doi: 10.1109/PESMG.2013.6672628.
- [3] PNE 34 7509 Z1: Holé vodiče pro venkovní vedení ze slaněných kruhových drátů.
- [4] PNE 34 8601: Součásti venkovních vedení distribučního vedení VN 45 kV Příloha 01
- [5] REX 640: Technical Manual [online]. Vaasa, Finland, 2020 [cit. 2021-03-04]. Dostupné z: https://searchext.abb.com/library/Download.aspx?DocumentID=1MRS758407&LanguageCode=en&Doc umentPartId=&Action=Launch

Doktorské projekty

Silnoproudá elektrotechnika a elektroenergetika II.

BACK-COMMUTATION AND STICKING REDUCTION METHODS IN DC CONTACTOR

Jakub Piska

Doctoral Degree Programme (1), FEEC BUT E-mail: xpiska02@stud.feec.vutbr.cz

> Supervised by: Jiri Valenta E-mail: valentaj@feec.vutbr.cz

Abstract: The back-commutation and sticking are phenomena which cause an increase in time delay between arc ignition and arc reaching the arc chute. These effects cause increase in total switching time and contact and arc runner wear. This paper gives a comparison of several approaches to reduction of these effects in a medium voltage contactor for traction usage. The three chosen approaches were: changes of the shape of the arcrunners, changes of the pole plates of the magnetic circuit and increase of the ventilation cross section. Metrics on how to compare the performance of these different alterations were given, based on switching times. From the three tested methods the ventilation increase yielded the best results as it both reduced the switching time and time delay caused by the back-commutation and reduced the variation of these characteristics.

Keywords: Contactor, breaking, back-commutation, sticking, reignition

1 INTRODUCTION

The arc extinguishing mechanism in the majority of DC contactors is magnetic blow-out. This method uses magnetic field to push the arc into an arc chute which by the means of cooling and arc prolongation increase the arc voltage and thus help the breaking process. This magnetic field can be created by several means - shape of the current path, addition of coils or permanent magnets. The goal is to transfer the arc into the arc chute as quickly as possible. This is because without the arc voltage exceeding the source voltage the arc cannot be distinguished in a DC circuit.

However, the transfer process is often hindered by effects like back-commutation and sticking. This is because the arc motion is not determined only by the external forces like magnetic force, pressure of surrounding gasses and fluid resistance. The arc tries to assume a position with the lowest arc voltage possible and often stays in places where the increase in voltage would be significant. These places are most commonly located on original contacts, just before arc chute and in places where arc-runners diverge. The arc staying in this position is called *sticking*. In case the arc moves past these places another phenomenon called *back-commutation* can occur. This is because the arc leaves behind it a space filled with ionised gas and after the arc voltage increases, again due to increasing its length, another parallel arc is ignited in the ionised vapours. This new arc is shorter and thus has lower voltage and as they are parallel, the original arc decays. After that we are left with a new arc in a position further from arc chute. Both of these effects significantly increase the total switching time and thus increase the thermal strain on the contactor which lowers the expected service life of the contactor.[1] There have been proposed many ways to alter a switching device to decrease the influence of these reignition effects. Some promising ones were tested in this paper.

2 METHODOLOGY

A way to quantify the results of the breaking tests was needed. A system of three times is used $-t_1, t_2, t_d$. t_1 is the total switching time, measured from the first increase in arc voltage to current zero. t_2 is travel time to the arc chute, measured from first increase in arc voltage to the first peak, which coincides with arc entering the arc chute. This fact was determined by comparing the arc voltage data and fast camera recordings. Lastly t_d is the sum of time delays caused by arc stagnation and back-commutations. These delays were defined as the time needed for the arc voltage after back-commutation to reach its original value before the back-commutation as shown in fig.(1). The software used disregarded any delay that was shorter than 100 μ s. This was done because these delays of shorter length are a normal part of any arc voltage signal and bear no connection to back-commutation or sticking.



Figure 1: Back-commutation delay

Figure 2: Measurement setup

3 EXPERIMENTAL SETUP AND TESTED DEVICE

The test circuit is shown in fig.(2). All tests were conducted with parameters 2250 V, 960 A, time constant 8.6 ms. These parameters were chosen as they coincide with the limits of switching capability of the tested contactor. Each sequence of tests was conducted as a series of 10 test with one alteration. There was always a 5 minute break between each test. This was done, so the contactor would not overheat. After each series the ceramic plates from arc chute were replaced as their degradation is the main reason for worsening of the contactor switching performance.

4 ALTERATIONS

Other authors have presented several possible solution to the reignition problems. [2] and [3] have investigated the influence of contact materials, [2] and [4] have tried altering the width and material of the contact chamber and [5] have swapped the splitter plates for a narrow slot with a great success. All of these solutions, while somewhat effective are unsuitable for alteration of already existing device.

One of easily realisable implementations is given by [3] and [6] who investigated the influence of venting condition on arc movement by opening or choking the exhausts which can be implemented in three ways - covering the exhausts, changing the shape of the ceramic plates or by removing some of the ceramic plates. The last method unfortunately leads to lower arc voltage which is detrimental to breaking process as a whole. Both authors measured the lowest travel time with maximally open vents. The second mentioned option was chosen for further tests. The ventilation area was increased from 880 mm² to 1540 mm². This is shown in fig.(3)

[7] gives another alteration which leads both to shortening the travel time and reduction of the amount of back-commutations - changing the shape of arcrunners. [7] tested several arcrunners which opening

angles from 60 to 150° and the number of back-commutations only increased with the opening angle. As can be seen from fig.(5) the original construction has a sharp increase of arcrunner distance as the arc passes from lower arcrunner to the upper. This was changed so the arcrunner distance is more gradual.

-1 1-2
1
11
11
1000
1

Figure 3: Ventilation alteration



Figure 4: Pole plate alterations



Figure 5: Arcrunner alterations

And as could be expected, changing the magnetic field so it is increased in the places where arc stagnation and back-commutation occur leads to lowering of the total travel time as shown by [8]. For this the pole plates of the magnetic circuit were altered so the magnetic field between the contacts was increased from average magnetic flux denstity 16.4 μ T to 17.2 μ T. This was done by adding two iron parts to the original pole plates as can be seen in fig.(4). This change can lead to a negative effect of decreasing the magnetic field in the arc chute and thus lowering the arc voltage. This would in turn prolong the total switching time. The values were calculated in Ansys as an average magnetic flux in the volume between fully opened contacts. The calculation was done as magnetostatic with permanet magnets as the only source of magnetic field.

5 RESULTS AND DISCUSSION

To determine whether there was any significant change in behaviour of the contactor independent two sample t-tests at significance level of 5% were done on the average values and variations of the results. The comparison was always done against the results of the measurement with no changes. The t-tests for altered arc runner were done as t-tests with unequal variances. T-tests for pole plate and ventilation alterations were done as t-tests with similar variances.

The only set of data which was not found significantly different was the total time t_1 for pole plate alteration. After establishing that the differences in data are not due to fluctuations, the results can be compared.

No change			Arc runner			Pole plate			Ventilation		
t ₁ [ms]	$t_2[ms]$	t _d [ms]	t ₁ [ms]	t ₂ [ms]	t _d [ms]	t ₁ [ms]	t ₂ [ms]	t _d [ms]	t ₁ [ms]	t ₂ [ms]	t _d [ms]
21.70	7.36	3.32	19.42	6.14	2.14	21.38	6.62	2.44	20.08	6.62	2.86
22.33	7.04	2.54	21.20	6.60	2.40	21.90	6.66	2.40	20.66	6.76	2.68
23.16	8.22	3.54	20.72	5.92	2.32	21.70	6.94	2.54	21.14	5.96	2.06
22.20	6.46	2.96	22.02	7.00	2.44	22.52	7.20	2.96	21.58	6.04	2.10
23.16	8.04	3.36	22.54	8.28	4.10	23.00	6.58	2.62	21.38	6.38	2.46
23.44	7.48	3.22	22.04	6.28	1.80	22.52	6.06	1.84	21.20	6.04	1.88
22.98	6.84	3.04	21.86	6.66	2.58	22.88	6.38	2.20	21.94	6.90	2.50
22.90	6.98	2.40	22.66	7.24	2.60	26.46	7.94	3.62	21.72	6.00	2.00
24.86	7.56	3.10	22.36	6.84	2.46	23.26	6.67	2.62	21.82	6.34	2.24
22.74	7.00	2.54	22.42	7.26	3.06	23.84	7.34	3.00	21.70	6.12	1.96

Table 1: Measured data

	No change			inge Arc runner		Pole plate			Ventilation			
	$t_1[ms]$	$t_2[ms]$	t _d [ms]	$t_1[ms]$	$t_2[ms]$	t _d [ms]	$t_1[ms]$	$t_2[ms]$	t _d [ms]	$t_1[ms]$	$t_2[ms]$	t _d [ms]
E(X)	22.95	7.30	3.00	21.72	6.82	2.59	22.95	6.84	2.62	21.32	6.32	2.27
s(X)	0.65	0.27	0.14	0.92	0.42	0.35	1.87	0.26	0.21	0.30	0.11	0.10

Table 2: Average values and standard deviations

	$\Delta t_1[\%]$			$\Delta t_2[\%]$			$\Delta t_{\rm d}$ [%]		
	AR	PP	VEN	AR	PP	VEN	AR	PP	VEN
E(X)	-5.3	0.0	-7.1	-6.5	-6.3	-13.5	-13.7	-12.6	-24.3
s(X)	41.2	186.5	-53.6	56.0	-2.9	-60.6	154.0	55.8	-26.9

Table 3: Changes in switching times t_1 , t_2 , t_d

As can be seen from tab.(3) all of the alterations lower the back-commutation and sticking effects. In case of arc runner and ventilation this leads to the lowering of total switching time. In case of pole plates this does not occur due to the lower magnetic field at the ceramic plates which leads to a smaller prolongation of arc and lower arc voltage.

The most significant effect was that of the increasing of the ventilation cross section. This alteration even positively affected the variation of the switching times which increased with the other two types alterations.

6 CONCLUSION

- 1. The mechanisms of back-commutation and sticking in switching devices were described.
- 2. Three metrics were suggested for comparing different switching operations: total switching time, arc travel time and time delay caused by back-commutations.
- 3. Three method how to reduce the back-commutation and sticking suggested by other authors were tested on a middle voltage contactor for traction usage: Shaping the arcrunners so there are no sudden increases in arc length. Shaping the pole plates of magnetic circuit so the magnetic field is increased between the original contacts. Increasing the ventilation cross section by changing the shape of ceramic plates in the arc chute.
- 4. Series of 10 tests were done on each alteration and were compared to the original construction.
- 5. All of the alteration yielded positive results as all of the metrics only decreased. The most promising results were of the increasing of ventilation cross section. This has both decreased the back-commutation delay by 24 % and also lowered the variation of the measured times.

7 ACKNOWLEDGMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-20-6379).

REFERENCES

- [1] SLADE, Paul G. *Electrical contacts: principles and applications*. Second edition. Boca Raton: CRC Press, Taylor Francis Group, [2014]. ISBN 1439881308.
- [2] GAUSTER, Ewald a Werner RIEDER. Arc restrikes yielding back-commutations in the contact gap of low voltage interrupters. IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A [online]. 1998, 21(4), 549-555 [cit. 2021-01-29]. ISSN 1070-9886. Available from: doi:10.1109/95.740047
- [3] MCBRIDGE, J.W., K. PECHRACH a P.M. WEAVER. Arc root commutation from moving contacts in low voltage devices. In: Electrical Contacts - 2000. Proceedings of the Forty-Sixth IEEE Holm Conference on Electrical Contacts (Cat. No.00CB37081) [online]. IEEE, 2000, pp. 130-138 [cit. 2021-01-29]. ISBN 0-7803-5960-7. Available from: doi:10.1109/HOLM.2000.889922
- [4] MA, Ruiguang, Mingzhe RONG, Fei YANG, Yi WU, Hao SUN, Duanlei YUAN, Haiyan WANG a Chunping NIU. Investigation on Arc Behavior During Arc Motion in Air DC Circuit Breaker. IEEE Transactions on Plasma Science [online]. 2013, 41(9), 2551-2560 [cit. 2021-01-29]. ISSN 0093-3813. Available from: doi:10.1109/TPS.2013.2273832
- [5] CHEN DEGUI, CHEN YONG a YUAN HAI WEN. Investigation of back commutation phenomena for narrow slot arc quenching chamber in current limiting circuit breaker. In: Electrical Contacts - 1996. Proceedings of the Forty-Second IEEE Holm Conference on Electrical Contacts. Joint with the 18th International Conference on Electrical Contacts [online]. IEEE, 1996, pp. 121-128 [cit. 2021-01-29]. ISBN 0-7803-3578-3. Available from: doi:10.1109/HOLM.1996.557188
- [6] ZELLER, P.R. a W.F. RIEDER. Arc structure, arc motion, and gas pressure between laterally enclosed arc runners. IEEE Transactions on Components and Packaging Technologies [online]. 24(3), 337-341 [cit. 2021-01-29]. ISSN 15213331. Available from: doi:10.1109/6144.946476
- [7] GAUSTER, E. a W. RIEDER. Arc lengthening between divergent runners: influence of arc current, geometry, and materials of runners and walls. IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A [online]. 21(1), 82-95 [cit. 2021-01-29]. ISSN 10709886. Available from: doi:10.1109/95.679037
- [8] REN, Zhigang, Ruiguang MA, Hao SUN, Jiaqi NING, Zhexin CHEN a Chunping NIU. Experimental investigation of arc characteristics in medium-voltage DC circuit breaker. In: 2013 IEEE International Conference of IEEE Region 10 (TENCON 2013) [online]. IEEE, 2013, 2013, pp. 1-4 [cit. 2021-01-29]. ISBN 978-1-4799-2827-9. Available from: doi:10.1109/TENCON.2013.6718987

DESIGN OF A HIGH-SPEED GENERATOR FOR A HELIUM EXPANSION TURBINE

Daniel Pribulla

Doctoral Degree Programme (1), FEEC BUT E-mail: xpribu02@feec.vutbr.com

> Supervised by: Ondřej Vítek E-mail: viteko@feec.vutbr.com

Abstract: This paper presents design of an ultra-high-speed generator, propelled by a helium expansion turbine. The power output of the presented generator is 5 kW at 160 000 rpm. Very high efficiency, over 95%, was achieved thanks to the use of permanent magnet synchronous machine type and proper material selection. Special care was taken during design of the machine rotor to meet both mechanical and electromagnetic requirements. Only high-strength materials can withstand the load. Stator sheets have to be very thin, to suppress the skin effect and eddy current losses, especially with the 4 pole configuration. It was shown that the design process for standard, low-speed machines can be used with the awareness of it's limitations. Weak points of the design, which should be further investigated, were pointed out.

Keywords: PMSM, high-speed, generator, turbine, design

1 INTRODUCTION

Due to the current demand for reduction of energy consumption and improvement of the overall efficiency among all industry sectors, new applications for high speed electric machines are being sought. One of them is an expansion turbine generator as a part of a Brayton cycle helium cryocooler. Vast amount of energy is lost during expansion phase of the cycle in conventional device. By expansion through the high-speed turbine this energy can be harvested. The gearless coupling between electric machine and turbine on a single shaft promises efficient and reliable operation of the device. Various high speed rotating machinery for different applications have been developed for long time. Based on the overview in [1], the achievable size and speed of particular machine type can be estimated using parameter RPM. \sqrt{kW} . The design specification in this application is 5 kW at 160000 min⁻¹ corresponding with $3.5 \times 10^5 \text{min}^{-1}$. \sqrt{kW} . Only permanent magnet synchronous machine (PMSM) and induction machine (IM) can reach this value, according to [1]. PM machine was selected since it has generally higher overall efficiency.

2 MECHANICAL DESIGN

Proper rotor dimensioning is a complex task as it has to meet both mechanical and electromagnetic loads. Tangential stress constant σ_{Ftan} was used to determine the required rotor volume per given deign speed ω and power output *P*.

$$\sigma_{Ftan} = \frac{2P}{\omega \pi D_r^2 l'} \tag{1}$$

A value of 10 kPa was used after review of previously designed machines with similar parameters in [2][3]. This is below the typical range of tangential stress value for normal machines found in [5]. It is mainly cased by lower air gap flux density as well as increased windage and core losses of the high-speed machines.

The structure of the designed rotor is shown in Fig. 1. A solid, magnetic shaft is suspended between two bearings. The surface-mounted permanent magnets are secured in place by pressed-in sleeve to withstand high centrifugal forces. The sleeve is the most mechanically loaded component since



Figure 1: A cross-section view of the PMSM generator rotor.

the centrifugal force rises with square of diameter. It was modeled as rotating cylinder, loaded with pressure of PMs on the inner surface. Radial and tangential stress within the sleeve is expressed by equations

$$\sigma_r = A - \frac{B}{r^2} - \frac{3+\mu}{8} \rho r^2 \omega^2, \quad \sigma_t = A + \frac{B}{r^2} - \frac{1+\mu}{8} \rho r^2 \omega^2, \tag{2}$$

where μ is Poisson's ratio and ρ is material density. Constants A and B are derived from boundary conditions

$$r = \frac{D_r}{2} - h_{sl} : \sigma_r = -p_{PM} = 2\pi \rho_{PM} h_{PM} \omega^2 \left(r_{in} - \frac{h_{PM}}{2} \right)^2; \quad r = \frac{D_r}{2} : \sigma_r = 0 , \qquad (3)$$

By solving equations (2) and (3), maximum stress within the sleeve for a given diameter, sleeve and PM thickness was obtained. With 1 mm thick Inconel 718 sleeve material and 2 mm of PM, yield strength of the sleeve (Tab. 1) is reached at 29 mm diameter. This calculation assumes that the radial stress between the shaft and PMs caused by the interference fit disappears due to the centrifugal forces at the nominal speed. Some radial stress is needed to transmit the torque, thus the sleeve diameter has to be smaller. Detailed calculation of the required sleeve interference can be found in [6].

Material	Density p	Yield strength	Young's modulus E	Poisson's ratio μ	Use
Inconel 718	8190 kg/m ³	1100 Mpa	205 GPa	0.31	sleeve
41CrMo4	7800 kg/m ³	800 Mpa	205 GPa	0.29	shaft
Recoma 35E	8300 kg/m ³	N/A	140 GPa	N/A	PM

Table 1: Mechanical properties of used rotor materials.

Another phenomenon limiting the length of the rotor is mechanical resonance. A critical speed occurs when the rotation of shaft excites its first normal mode of vibration. The speed of the machine usually has to be kept below this limit to avoid catastrophic failure. An empirical estimation of the maximum length of rotor to avoid resonance for a given design speed was described by Wiarth [4]

$$l_{max}^{\prime 2} = \frac{\pi}{k\omega} \sqrt{\frac{EJ}{\rho S}} , \qquad (4)$$

where J is the quadratic moment of the shaft's cross-section area. For 41CrMo4 steel shaft with a diameter of 20 mm the maximum length is 69 mm. Since the shaft has complex structure, FEA analysis should be carried out, including the effect of additional weight of the impeller and bearings.
Rotor diameter D_r [mm]	28
Rotor lenght l' [mm]	25
Sleeve thickness h_{sl} [mm]	1
PM thickness <i>h_{PM}</i> [mm]	2

Table 2: Parameters of designed PMSM generator rotor.

3 ELECTROMAGNETIC DESIGN

The design process accommodated from [5] starts by defining the air gap δ and the air gap magnetic flux density B_{δ} as these imply the rest of magnetic circuit dimensions. Since relative permeability of Inconel 718 material, used for the retaining sleeve is close to 1, it's increasing the apparent thickness of the air gap. Increasing the thickness of PMs to gain sufficient magnetomotive force results in higher centrifugal forces acting on the sleeve. Thus the magnetic circuit design has to be carried out iteratively, together with the mechanical.

A simplified calculation can be used as an initial guess of the required PM thickness by omitting the magnetic voltages over the majority of reluctance is caused by the air gap.

$$h_{PM} = \frac{\frac{B_{\delta}}{\mu_0} \delta}{H_c - \frac{H_c}{B_*} B_{PM}}$$
(5)

The machine was designed as four-pole, with 12 stator slots, giving 1 slot per pole and phase. A minimum number of slots was chosen, because of small diameter of the stator bore. With 24 slots and 2 slots per pole and phase, more sinusoidal EMF would be obtained, but very narrow stator teeth are problematic to manufacture. Two-pole machine would lead to a higher stator yoke and copper volume, increasing the losses.

To determine necessary stator tooth width and height, the winding slot cross-section area has to be evaluated.

$$N = \frac{e_{PM}}{p \,\omega_m \,k_w \,\alpha_{PM} \,B_\delta \,\tau_p \,l'} \,, \tag{6}$$

where e_{PM} is effective value of induced voltage, k_w is winding factor, and α_{PM} relative effective PM width. The cross section of a single wire is determined from the mean power output requirement and the permitted current density J_s

$$I_s = \frac{P}{m\eta U_s \cos\varphi}; \quad S_{Cs} = \frac{I_s}{aJ_s}, \tag{7}$$

where *m* is number of phases and *a* number of parallel branches. High current density has to be selected to obtain reasonable shape of the stator teeth. Widening the slots leads to thin teeth, resulting in high tooth flux density and increased losses. Tall slot compromises mechanical rigidity of the teeth and increases its magnetic voltage. The minimal cross-section area of the stator slot S_{Cus} is obtained by multiplying the number of conductors per slot z_Q by required cross section of a single conductor S_{cs} .

$$z_Q = 2 a m \frac{N}{Q}, \quad S_{Cus} = \frac{z_Q S_{cs}}{k_{Cu}} \tag{8}$$

Proposed dimensions and parameters of the machine stator are shown in Table 3. A FEMM software was used to verify the calculations. Figure 2 shows flux density distribution within the magnetic circuit at nominal load angle and current (Tab. 4). The current is supplied by a current sources into the q-axis of the machine.

Number of coil turns N		32	Stator tooth height h_d	[mm]	4
Number of slot conductors z_Q		16	Stator tooth flux density B_d	[T]	1.4
Slots per pole per phase q		1	Stator yoke height h_{ys}	[mm]	8
Single wire cross section S_{cs}	[mm ²]	1.18	Stator yoke flux density B_{ys}	[T]	0.8
Winding space factor k_{Cu}		0.5	Stator outer diameter D_{re}	[mm]	70
Stator winding connection		Star	Stator material		NO10
Linear current density J	$[A/mm^2]$	4.0	Rotor flux density B_r	[T]	0.57
Stator tooth width b_d	[mm]	12	Tangential stress σ_{Ftan}	[kPa]	10

Table 3: Electromagnetic design of the PMSM generator.

It is important to take the skin effect into account, when choosing materials for the machine rotor. Penetration depth at 5333 Hz is approximately 0.14 mm for electrical steel and 1.26 mm for copper. As a result, winding has to be made of multiple parallel branches or Litz wire. Very common M230-35A electrical steel with 0.35 mm thickness is not suitable choice. Only advanced, thin electrical steel, like NO10-1270N can be used in order to achieve satisfactory performance and losses.



Figure 2: Magnetic flux density distribution within machine cross-section geometry at nominal load.

4 RMXPRT SIMULATION RESULTS

Properties of the designed machine were evaluated by Ansys RMxprt software. The most important parameters are listed in Table 4. The machine has large power reserve and reaches the nominal power with relatively low load angle. With proper cooling, it may be possible to reduce the machine volume. Winding resistance and inductance are both very low and should not cause problems with the control of the machine. The windage losses P_{ρ} were calculated by equations (9.8) – (9.18) in [5], assuming air at 100 °C as surrounding medium and surface roughness 1.2. The efficiency of the machine is very high, but the bearing mechanical losses as well as eddy current losses in the sleeve and PMs were not taken into account.

5 CONCLUSION

Permanent magnet synchronous machine is a suitable solution for the helium expansion turbine generator. It is possible to meet required power output and speed of the turbine. Proposed machine can withstand the extreme mechanical loading and operate with reasonable efficiency. This can be

Max. line voltage U	[V]	231
RMS phase current <i>I</i> _s	[A]	19.2
Armature resistance <i>R</i>	$[m\Omega]$	9.5
Synchronous inductance $L_d = L_q$	$[\mu H]$	33.4
Torque angle δ_{load}		19.5°
Windage loss P_{ρ}	[W]	68
Iron core loss P_{Fe}	[W]	125
Copper loss P_{Cu}	[W]	11
Efficiency η	[%]	96
Peak output power P_{max}	[W]	23 400

 Table 4: Rated performance parameters of the designed PMSM generator.

achieved only by the use of high-end materials. It was shown that the standard, analytic design process used in low-speed applications can be used with limited accuracy. Further FEM modeling is necessary to guarantee reliable operation. Rotor dynamics analysis should be carried out to prevent mechanical failure. The effect of slot permeance harmonics, inducing eddy currents into retaining sleeve and PM, should be analyzed to avoid problems with overheating of the rotor.

ACKNOWLEDGEMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-20-6379).

REFERENCES

- GERADA, David, Abdeslam MEBARKI, Neil L. BROWN, Chris GERADA and Aldo BOGLI-ETTI. 2014. High-Speed Electrical Machines: Technologies, Trends, and Developments. In: *IEEE Transactions on Industrial Electronics*. 61(6), p. 2946-2959.
- [2] CHO, Han-Wook, Kyoung-Jin KO, Jang-Young CHOI, and Hyun-Jae SHIN. 2011. Rotor Natural Frequency in High-Speed Permanent-Magnet Synchronous Motor for Turbo-Compressor Application. In: *IEEE Transactions on Magnetics*. 47(10), p. 4258-4261.
- [3] UZHEGOV, Nikita, Jan BARTA, Jiri KURFURST, Cestmir ONDRUSEK and Juha PYRHONEN. 2017. Comparison of High-Speed Electrical Motors for a Turbo Circulator Application. In: *IEEE Transactions on Industry Applications*. 53(5), p. 4308-4317.
- [4] WIART, Albert. 1982. New high-speed high-power machines with converter power supply. In: *Proceedings of the International MOTORCON Conference*. Geneva, p. 354–365.
- [5] PYRHÖNEN, Juha, Tapani JOKINEN and Valéria HRABOVCOVÁ. 2014. *Design of rotating electrical machines*. 2nd edition. Chichester: Wiley.
- [6] UZHEGOV, Nikita, Emil KURVINEN, Janne NERG, Juha PYRHONEN, Jussi T. SOPANEN and Sergey SHIRINSKII. 2016. Multidisciplinary Design Process of a 6-Slot 2-Pole High-Speed Permanent-Magnet Synchronous Machine. In: *IEEE Transactions on Industrial Electronics*. 63(2), p. 784-795.

COPPER LOSSES IN PARALLEL WIRES AND LITZ WIRE IN PERMANENT MAGNET SYNCHRONNOUS MACHINE

Petr Klima

Doctoral Degree Programme (1), FEEC BUT E-mail: xklima28@vutbr.cz

Supervised by: Cestmir Ondrusek

E-mail: ondrusek@feec.vutbr.cz

Abstract: The paper is focused on comparing the effects of a winding made of the solid wire and the Litz wire to the copper losses in a 3kW 180 000 RPM high-speed permanent magnet synchronous motor. It shows the influence of skin and proximity effects on a current density and copper losses in the wires. The paper uses data from results of FEM simulations.

Keywords: High-speed machine, synchronous machine, proximity effect, winding

1 INTRODUCTION

The high-speed permanent magnet synchronous machines have an advantage of a high power density. It is achieved by a high operating frequency of the machine. With that comes an increased demand on the stator steel and the stator winding. The eddy effects are one of the main reasons for the increased losses in both of those cases. To mitigate those effects, the best available solution is the usage of thinner and superior quality steels and the Litz wire. [1]

The Litz wire is a special type of wire consisting of many parallel strands of a small diameter. The strands are twisted so that every strand comes through every position in the whole Litz to decrease the induced voltage between the strands. The strands can be organized into branches to allow better manufacturability and flexibility. The strands in the branch are twisted and the branch is then twisted with other branches. One of the disadvantages of the Litz wire is a higher prize and a lower fill-factor. An ideal Litz wire has the same AC and DC resistance. [2]

It is generally advised to use a winding with wires of smaller diameter than the depth of penetration to reduce the skin effect [3]. The proximity effect is caused mainly by a leakage flux in the stator slot and needs to be carefully addressed by the placement of parallel wires when using them. When placed incorrectly, the induced circulating currents will cause additional losses. It is therefore advised to place the parallel strands into approximately the same radial position in the slot [4].

A similar subject as this paper was studied in [5]. The studied machine in that case was an induction motor of the same nominal speed and power as the synchronous machine in this paper. The results showed the increase of copper losses for the solid wire in the active part of a machine up to 3.5 the copper losses for a Litz wire.

2 STUDIED MACHINE

The machine in which the skin and proximity effects were studied was a 3 kW 180 000 RPM permanent magnet synchronous motor. The motor is planned to be used in a compressor application. The stator core is made of NO10 steel. The permanent magnet in the rotor is mechanically supported by a carbon fibre sleeve. The winding is a double-layer tooth coil winding to reduce the length of the end winding. The basic parameters of the motor are in Table 1.

Rated RPM	(\min^{-1})	180 000
Rated Output Power	(kW)	3
Number of Poles	(-)	2
Rated Amplitude of Phase Voltage	(V)	300
Stator Winding Connection	(-)	Star
Number of Turns in Series of the Stator Winding	(-)	80
Rated Stator Frequency	(Hz)	3000
Number of Stator Slots	(-)	6
Stator Outer Diameter	(mm)	125
Rotor Outer Diameter	(mm)	22
Air Gap Length	(mm)	3
Active Length of Motor	(mm)	25

Table 1: Parameters of the studied synchronous motor

To properly analyse the eddy current effects in different types of the winding, a FEM software was used. In the first step, the machine model was created with a winding made of the Litz wire. The feasibility of the model was verified by a 2D transient analysis. In the next step, the winding of one phase was modelled using individual stranded wires. The transient analysis was done with a step of 1/800 of the period. According to [6], the simulation step for high-speed machines is to be at least 1/400 of the period, which was satisfied.

According to [3], it is sufficient to model only one phase of the machine to be able to find out the accurate copper losses. The rest of the phases are modelled as the areas of the same cross section as the sum of the wires in the according half slot. The eddy effects are not calculated in these phases and the copper losses are obtained only from the stranded winding. The simulation time is therefore reduced. The next thing which reduces simulation time is a usage of symmetry of the machine. The windings were supplied by an ideal three-phase current source with an effective current 7.4 A to achieve its nominal power.



Figure 1: Cross-section of the studied motor

The studied winding wire combinations are the solid wire, 2 and 3 parallel strands wire and the Litz wire. There are no other changes in the machine between the simulations to show only the effect of the winding on the copper losses. The external circuit was used to properly connect individual strands of the stranded winding. Each strand included its end winding resistance and inductance which were computed analytically according to [6]. The resistance of the end winding area was assumed to be

without the skin and proximity effects as it is difficult to assess their influence in this area. The calculation of the inductance is based on the calculation of the inductance of an air-cored solenoid and it is shown in equation (1).

$$L_{\text{end}} = \frac{Q_{\text{s}}}{m} \mu_0 \mu_{\text{env}} \frac{\left(\frac{z_{\text{q}}}{2}\right)^2 \pi (l_{\text{ew}})^2}{h} \tag{1}$$

where Q_s is the number of stator slots, *m* is the number of phases, μ_0 is the permeability of vacuum, μ_{env} is the relative permeability of environment in the end winding area ($\mu_{env} = 2$ was used in the calculation), z_q is the number of wires in the slot, l_{ew} is the radius of end winding and *h* is the height of the end winding.

An example of the external circuit used in simulation of winding with 3 parallel strands is shown in Figure 2. In this case the resistance and inductance for individual strands are 3 times the values for whole phase as they are connected parallel and their combination must make the total value for the phase.



Figure 2: External circuit used in simulations of the 3x0.95mm wire



Figure 3: Stranded winding of the model, conductors of the same turn have the same colour A) 1.6mm solid wire B) 2x1.12mm wire C) 3x0.95mm wire

3 SIMULATION RESULTS

Table 2 shows the copper losses for each type of the wire. The losses in the table are the total losses for all three phases. The computed copper loss in the stranded winding was simply multiplied by the number of phases. The resistance factor is used to easily compare the influence of skin and proximity effects. It is a ratio of losses with skin and proximity effects and losses without those effects.

Wire type		1x1.6mm	2x1.12mm	3x0.95mm	435x0.08mm Litz
Total copper area in half of the slot	(mm ²)	80.38	78.78	85.02	83.24
Resistance factor	(-)	3.11	1.86	1.73	1.00
Copper losses in ac- tive part of motor	(W)	23.0	14.0	12.1	7.1
Total copper losses	(W)	34.6	25.8	23.1	18.3

Table 2:

Results of the simulations for different wires

Figure 4 shows the current density in the winding. The skin and proximity effects cause the peak current density in some strands to locally reach values up to 24 A/mm^2 for 1.6 mm solid wire and 16.3 A/mm^2 for 3x0.95 mm wires. These values are 4.64 and 3.33 times higher than the nominal peak current density.



Figure 4: Current density A) 1.6mm solid wire B) 2x1.12mm wire C) 3x0.95mm wire

4 CONCLUSION

The results of the simulations show a substantial increase of copper losses in the synchronous machine when using a solid wire and wires consisting of 2 and 3 parallel strands instead of a Litz wire. One of the main reasons for the increase is believed to be the proximity effect, because the highest current density in the winding is located toward the opening of the slot and toward the other half of the slot.

The resistance factor for the solid 1.6mm wire is 3.11 times the Litz wire. For the parallel 3x0.95mm the factor reaches 1.73. It is assumed that with a bigger number of smaller parallel strands the resistance factor would be even lower but the placement of the strands would need to be carefully assessed as the induced circular currents could cause an increase in the losses.

ACKNOWLEDGEMENT

"This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support, BUT specific research programme (project No. FEKT-S-20-6379) and from the Technology Agency of the Czech Republic (project No. TK02020168)."

REFERENCES

- D. Gerada, A. Mebarki, N. L. Brown, C. Gerada, A. Cavagnino, and A. Boglietti, "High-Speed Electrical Machines: Technologies, Trends, and Developments", *IEEE Transactions on Industrial Electronics*, vol. 61, no. 6, pp. 2946-2959, 2014.
- [2] Bartoli, M., N. Noferi, and A. Reatti, "Modeling Litz-Wire Winding Losses in High Frequency Power Inductors", 27th IEEE Pow. Elecs. Spec. Conf., Jun 1996, vol. 2, pgs. 1690-1696.
- [3] F. Birnkammer, J. Chen, D. Bachinski Pinhal, and D. Gerling, "Influence of the Modeling Depth and Voltage Level on the AC Losses in Parallel Conductors of a Permanent Magnet Synchronous Machine", *IEEE Transactions on Applied Superconductivity*, vol. 28, no. 3, pp. 1-5, 2018.
- [4] D. Bauer, P. Mamuschkin, H.-C. Reuss, and E. Nolle, "Influence of parallel wire placement on the ac copper losses in electrical machines," in *Proc. IEEE International Electric Machines* & *Drives Conference (IEMDC)*, Coeur d'Alene, ID, May 2015, pp. 1247-1253.
- [5] P. Klíma, "Comparison of copper losses of litz wire and parallel wires in high speed induction motor", *PROCEEDINGS I OF THE 26TH STUDENT EEICT 2020 GENERAL PAPERS*. Brno, Czech Republic: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2020. p. 408-412. ISBN: 978-80-214-5867-3.
- [6] J. Pyrhönen, T. Jokinen, and V. Hrabovcová, *Design of rotating electrical machines*, 2nd edition. Chichester: Wiley, 2014.

TOPOLOGY AND BREAKING PROCESS OF HYBRID DC CIRCUIT BREAKER

Jakub Kaser

Doctoral Degree Programme, FEEC BUT E-mail: xkaser00@vutbr.cz

Supervised by: Jiří Valenta

E-mail: valentaj@vutbr.cz

Abstract: This study is focused on a general function of hybrid direct current circuit breakers. The one of topics is design topology of these circuit breakers with all necessary branches. Mainly, the study deal with their principles of interruption of fault direct current. The description of issue is in general and then for the specific interesting proposals. The main purpose of these protections is continuous reducing the fault current to the current-zero outside the mechanical breaker contacts which conduct solely the nominal current.

Keywords: Hybrid DC circuit breaker, interruption, fault direct current, proposals

1 INTRODUCTION

The products of electrical engineering are constantly evolved. Many developments and explorations are carried out for reaching the phenomenal attributes and reliable controlling. Therefore, the protections of electrical circuits must be competitive, too and have a front view.

The circuit breakers (CBs) of high voltage direct current grid (HVDC) have been developed continuously and even increasingly in the recent years. The main purpose is to make proposals to reduce or eliminate an electric arc which occurs on the mechanical contacts during short-circuit current interruption. Although existing mechanical circuit breakers (MCBs) offer low conduction loss with negligible resistance, long breaking time and arc extinguishing wear out the mechanical break contacts which makes to decrease their service life. Hybrid circuit breaker (HCB) commutates the fault direct current (DC) to the specific parallel branch consists of power semiconductors. The solid-state switches [2] have a faster breaking time and are controlled by system with high reliability. However, the main limiting factors are on-state resistance and heating. Thereby, it should not utilise for normal conditions in the HCB, solely for break and reclose process. Thus, the hybrid DC circuit breaker increases the fault current interruption capability.

2 HYBRID DC CIRCUIT BREAKER

2.1 STRUCTURE OF CIRCUIT

The circuit topology of hybrid direct current circuit breaker (HCB) often consists of three parallel connected parts. A conductive branch, the main breaker branch of power semiconductors and absorption branch with snubber components.

The conductive branch consists of an ultra-fast disconnector (UFD) to conduct the load current during normal process. The UFD is often combined with a load commutation IGBT switch (LCS) to help the commutation of fault current. In the main breaker branch, power thyristors are very often used for their high current breaking capability and low cost [1], and IGBTs for a fast and reliable full control. The absorption part is employed to next commutate energy and reduce the fault current to a current zero. The rest of accumulated excess energy is shifted to a MOV (metal oxide varistor).

2.2 GENERAL BREAK PROCESS

1. Commutation from the conductive branch to the main breaker branch. The power semiconductors are turned on. The fault current starts to flow in this branch and di/dt significantly decreases across the UFD.

2. Current zero in the conduction branch. The UFD starts to break without the electric arc and a dielectric strength is sequentially increased through the mechanical contacts.

3. Commutation to the absorption branch after the UFD is fully opened. The solid-state switches are usually turned off. Subsequently, all energy commutates to the absorption part and the snubber capacitors are charged.

4. Commutation from the snubber capacitors to the MOV or other absorption components. For instance, the capacitor is charged to the voltage value of MOV and all excess energy transfers outside the HCB. In this article, the reclose and rebreak process after interruption is not presented.

3 HYBRID DC CIRCUIT BREAKER BASED ON SERIES CONNECTION OF THYRIS-TORS AND IGBT HALF-BRIDGE SUBMODULES

This proposal [1] has all branches described in Chapter 2. The UFD is combined with the LCS for its effectiveness and simplicity. A main breaker has the branch of series connected thyristors and IGBT half bridge submodules with full-controlled lower Q_L and upper arm Q_U , as shown in Fig. 1. The thyristors reduce the number of IGBTs which makes to lower overall costs. Besides, the thyristors withstand the major part of turn-off surge voltage instead of IGBTs. These solid-state switches are series connected to protect themselves against a high DC voltage and surge voltage. The snubber circuit consist of snubber capacitor and parallel connected diodes and resistor. The resistor is used for following reclosing. MOV absorb the excess energy.



Figure 1: Circuit topology and waveform during break process [1].

The break process of the proposed HCD includes 8 stages, as shown in Fig. 1. At time t1, the fault occurs and load current i_L starts to rise linearly through the conductive branch, as shown in Fig. 1. At time t2 starts the commutation from the LCS to the thyristors. First, turn-on drive signal will be sent to the thyristors (T_1-T_N) and the lower arm IGBTs ($Q_{L1}-Q_{LM}$) in half-bridges. Then, the LCS will be turn off and afterwards the load current i_T will be transfer to the thyristor branch. At time t3, the UFD starts to break without the arc and has a low breaking speed. So, the dielectric strength across it will be built slowly in several milliseconds. After the time delay, at time t4 are charged

the submodule capacitors of IGBT half bridges for thyristors break. The submodule capacitors will be gradually charged to obtain the voltage value of one submodule capacitor (U_m) and to maintain relatively low du/dt across the mechanical contacts. First, the lower IGBT QLM is turn off while remain QL are conducted. Same way will be for the rest of submodule capacitors until the charge process is complete. At time t5 is commutation to the snubber circuit. The UFD is full opened and the Q_L are turned off. After the short dead time, the Q_U will be turned on. The generated reverse voltage from submodule capacitors will be applied to force the interruption of thyristors and will need to last long enough to fully turn off these thyristors. At time t6, the thyristors and then the Q_U are completely turn off. The snubber capacitor is sequentially charged by the fault current rapidly and its capacity C should be high to limit the maximum du/dt of proposed thyristors. At time t7, after the voltage across C reaches the break voltage of MOV ($U_{MOV,1}$), the load fault current will transfer to this MOV. At time t8, the fault current starts to decrease at the peak MOV voltage ($U_{MOV,2}$) and at t9, it reaches the current zero. This peak MOV voltage is commonly 1,2-2 times of UDC. The bidirectional system is proposed but it is not presented in this article. [1]

4 CASCADED COMMUTATION CIRCUIT OF HYBRID DC CIRCUIT BREAKER

This proposal [2] is different, especially, without the snubber circuit. The conduction branch is without the LCS to reduce the power semiconductors. The commutation circuit is the main branch of HCB and consists of the series connected capacitors (C_C) which are associated with IGBT solid state switch (T_n) in each cascaded voltage stage. The last part of circuit is commutation coil L_C to produce the resonant counter current. The MOV is used in the absorption stage. Its clamping voltage ($U_{MOV,2}$) can be twice the DC source and IGBTs is chosen with respect to this voltage.



Figure 2: Circuit topology (the UFD is replaced by the vacuum circuit breaker VCB) [2]

When the fault occurs, the T_3 will be turned on and the stored energy of C_{bank} will be transmitted to the R_{FAULT}. Thus, the process of preparation begins. After that, the C_C capacitors start charging the reverse voltage by turning on the IGBT T₂ and turning off the T₁ and this voltage across C_C is equal (V_{C1} = V_{C2} = V_{C3}). Then the electrodes start to open with arcing and after the specific distance of contacts, T₄ will be turned on and the resonant counter current ($i_{LC} = i_{C3}$), produced by resonant energy, will be rise and force the fault current though the VCB to decrease ($i_{VCB} = i_{FAULT} - i_{LC}$). Afterwards, the T₅ is closed and the T₄ is opened. The resonant loop is changed (C_{C1,2} - T₅ - L_C -VCB) and i_{LC} is still increased. Then, the T₅ is opened, the T₆ is closed and the fault current reaches the first current zero. After the current zero, the capacitor voltage is solely V_{C1} to limit the rise rate of transient recovery voltage du/dt through the VCB to maintain the low dielectric strength and to prevent the arcing. The voltage of capacitors C_C is discharged to the voltage zero by sequential switching T_6 , T_5 and T_4 . When the $C_{C1,2,3}$ is discharged and the VCB is fully opened (T_4 is closed), the all capacitors will be charged by the DC source and the current will flow through R_{FAULT} . After discharging, the i_{VCB} will produce the second current zero, too. The second failure is deemed to interruption failure of breaker. The excess energy will be absorbed in the MOV part, and the fault current will break completely. [2]

The solid-state switches are controlled by sequential time intervals, overlapping for a short duration $(1 \ \mu s)$ to ensure commutation current continuity. After interruption, the capacitors need to maintain the DC voltage. Therefore, the required auxiliary circuits (shunt connected resistors) will discharge the capacitor energy when the repair or the maintenance of system is solely necessary. [2]

5 T-TYPE HYBRID DC CIRCUIT BREAKER

This proposal [3] has similar procedures and design as the HCB [1] in Section 2. In Fig. 3 is a bidirectional blocking ability with D1 - D4. The main branch is designed as cascaded full-bridge submodules which consist of the IGBTs and capacitors. The main difference is that these capacitors are already used for charging by the fault current and the subsequent transfer of excess energy to the MOV of individual submodules. When the normal current flows, these capacitors will be precharged by the DC source. During pre-charging, one of IGBTs is closed and the n-1 capacitors are in the process. After that, the sum of n capacitors voltage is higher than the nominal DC voltage. Therefore, this prevent the main branch from connection with the system during normal process as the diode D1 or D2 will stop to conduct.



Figure 3: Structure of the T-type with bidirectional system [3]

When the fault current is detected, the IGBTs will be conducted and the LCS will be blocked. In the conduction branch, the current drops to the zero and then the UFD will be full opened under the zero-voltage condition due to pre-charging. It will take no more than 2 ms. Afterwards, the IGBTs can be blocked. The fault current transfer to the capacitors and decrease immediately due to the capacitors has been already pre-charged. After reaching the peak MOV voltage, this current shifts to the absorption branch and the fault is cleared. [3]

6 CONCLUSION

In this article, the interesting topologies of HCB are described and compared. The thyristors and IGBT half-bridge submodule in series [1] in Section 3 have several advancements and advantages. The thyristors withstand the main part of overvoltage and reduce the required number of IGBTs and thus, the overall costs are decreased. The number of thyristors (N) is defined according to the peak voltage of MOV, thyristors and IGBT half-bridge submodule. The submodule capacitors are sequential charged during the UFD breaking. Therefore, the total breaking time is not risen significantly, and the auxiliary supplies are not required. The proposal is verified on high voltage (up to 100 kV) and the equations are simple for verification. The HCB of ABB [4] has a faster breaking time but consists of solely the IGBTs. In Section 4, the commutation circuit [2] produce the resonant counter current to the mechanical contacts. In each commutation branch is solely one IGBT device to reduce the power loss but in high voltage application (>15 kV) will result in more cascaded voltage stages. However, controlling dv/dt through the VCB is increased. The main issue is that the VCB starts to separate the contacts with the electric arc. Therefore, the probability of interruption depends on the first and second current zero through the VCB. The preparation on fault current, charging and discharging capacitors during breaking, and the arc extinguishing prolong the total breaking time, too. The next challenge are difficult resonant equations and controls. The HCB with T-bridge [3] in Section 5 has similar advancements as the HCB in Section 3. The main benefit, the submodule capacitors are pre-charged in normal process and adopt the features of snubber capacitor. Thus, the fault blocking process is shorter compared with the HCB in Section 2 and the peak fault current is lower. The T-bridge structure avoids the dynamic voltage balancing due to the synchronous opening of IGBT. However, the number of submodules and MOV is increased. Thus, the implementation and the calculation will be reflected in overall costs.

The total breaking time of HCB is dependent on the opening mechanical contacts (several ms). The solid-state circuit breaker can achieve the shorter interruption (<1 ms) with lesser components. [2] However, this breaker has significant drawbacks. The physical disconnection of the current path is more reliable with mechanical contacts and the main limiting factors are on-state power losses and heating during normal conditions. The rapid decline di/dt of fault current to produce the induced inductive voltage. Thus, the semiconductors in series must sustain this overvoltage.

ACKNOWLEDGEMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-20-6379).

REFERENCES

- Novel Hybrid DC Circuit Breaker Based on Series Connection of Thyristors and IGBT Half-Bridge Submodules. *IEEE Xplore* [online]. Feb. 2021, 36(2), s. 1506-1518 [cit. 2021-02-14]. Source: doi:10.1109/TPEL.2020.3010965
- [2] Cascaded commutation circuit for a hybrid DC breaker with dynamic voltage control. *IET Power Electronics* [online]. May 2017, 10(6), s. 676-686 [cit. 2020-12-05]. Source: doi:10.1049/iet-pel.2016.0472
- [3] Basic Topology, Modelling and Evaluation of a Novel Hybrid DC Breaker for HVDC Grid (T-Bridge topology). *IEEE Xplore* [online]. Oct. 2020, s. 1-1 [cit. 2021-02-14]. Source: doi:10.1109/TPWRD.2020.3031671
- [4] M. Callavik et al., "Breakthrough! ABB's hybrid hvdc breaker, an innovation breakthrough enabling reliable hvdc grids," ABB Tech., Nov. 2012.

ACCURACY ANALYSIS OF ROTOR FREQUENCY CALCULA-TION FOR INDUCTION MOTOR DRIVE

Tomáš Lažek

Doctoral Degree Programme (2), FEEC BUT E-mail: xlazek00@stud.feec.vutbr.cz

Supervised by: Ivo Pazdera E-mail: pazderai@feec.vutbr.cz

Abstract: This paper deals with the accuracy analysis of rotor frequency calculations for an analytical formula of the optimal linkage flux. First, an equivalent circuit of the induction machine and a loss model are described. Furthermore, the calculation of the rotor frequency is performed for two cases, which are compared. The results show that the rotor frequency can be calculated in a simplified form without a large difference in accuracy.

Keywords: accuracy analysis, induction motor, loss minimization, variable speed drive

1 INTRODUCTION

In recent years, there has been an emphasis on reducing electricity consumption. The largest consumers of electricity are variable speed electric drives. Here is a great opportunity to reduce the consumption of electric drives by using an algorithm to increase the efficiency of the electric drive.

The efficiency of the drive can be controlled by an adaptive regulator, which reduces the value of the linkage flux, thus reducing losses. One of the commonly used strategies is model-based method which requires the loss model of the induction motor [1].

The goal of this paper is to determine the difference between the two methods of calculating rotor frequency. Both calculation methods are compared with each other. The influence of other parameters such as stator currents and power losses are compared. The conclusion evaluates the influence of both calculations on the parameters accuracy necessary for the operation of the regulator.

2 INDUCTION MACHINE MODEL

An equivalent circuit in the form of a commonly used gamma network is used. The resistor representing the iron resistance is connected parallel to the magnetization inductance. Detailed development of the model can be found [2]. The steady-state model shown in Figure 1 is defined in the rotating (d,q) stator flux frame. Thus $\Psi_{sd} = \Psi_s$ and $\Psi_{sg} = 0$.



Figure 1: Steady-state induction machine equivalent circuit in: a) d-axis and b) q-axis.

In the steady-state model, the coils can be thought of as a short circuit. The stator voltage u_{sd} and u_{sq} is given as:

$$u_{sd} = R_s \cdot i_{sd} \tag{1}$$

$$u_{sq} = R_s i_{sq} + \omega_s \psi_s \qquad \qquad \omega_s \psi_s = \omega_r L_\sigma i_{rd} + R_r i_{rq} + \omega \psi_s \qquad (2a, b)$$

where ω_r is the difference between the synchronous frequency ω_s and the mechanical speed ω . The synchronous frequency ω_s can be expressed as the mechanical speed ω multiplied by the number of pole-pairs p. The currents i_{sd} and i_{sq} in the model are described as follows:

$$i_{sd} = i_u + i_{rd}$$
 $i_u = \frac{\psi_s}{L_m}$ $i_{rd} = \frac{\omega_r L_\sigma i_{sq}}{R_r}$ (3a, b, c)

$$i_{rq} = i_{sq} - i_{Feq} = i_{sq} \qquad \qquad i_{sq} = \frac{2T}{3p\psi_s}$$
(4a, b)

The current i_{Feq} and i_{Fed} can be considered as zero, because the resistance R_{Fe} is many times greater than the resistance R_r . It is advantageous to express the current i_{sq} in terms of torque and stator linkage flux. Furthermore, from the Figure 1 follows that the voltage at R_{Fe} is determined by the voltage $\omega_s \cdot \psi_s$.

3 LOSS MODEL

The loss model of an induction motor is advantageous to express depending on the mechanical speed, torque, and linkage flux. The aim is to obtain an analytical expression for the optimal linkage flux at a given torque and known shaft speed. However, the formula compilation is not within the scope of this article.

The parameter R_{Fe} must not be constant due to the loss model accuracy. The iron losses consist of losses by eddy currents and hysteresis losses. According to [2], both resistance values can be combined to create a simple linear dependence of the total iron resistance on frequency.

$$R_{Fe} = R_{Fe0} \frac{\omega_s}{\omega_{s0}} \tag{5}$$

The iron resistance R_{Fe0} must be determined at the specified frequency ω_{s0} .

Determining stator Joule losses P_{js} in the substitution model in d-q axis is very simple. The resistance of the stator winding R_s is multiplied by the square of the current flowing through it. It can be seen from Figure 2 that both i_{sd} and i_{sq} flow through the resistor R_s . The determination of Joule losses in the rotor P_{jr} is similar to the difference that the rotor currents i_{rd} and i_{rq} flow through the resistor R_r . As mentioned above, $i_{rq} = i_{sq}$. Thus:

$$P_{js} = R_s(i_{sd}^2 + i_{sq}^2) \qquad P_{jr} = R_r(i_{rd}^2 + i_{sq}^2) \qquad (6a, b)$$

4 CALCULATION OF ROTOR FREQUENCY

For calculation current i_{rd} , it is necessary to know rotor frequency ω_r . The rotor frequency can be expressed from equation (2b) as follows:

$$\omega_r \psi_s = \omega_r L_\sigma i_{rd} + R_r i_{sq} \tag{7}$$

To simplify the expression, the $L_{\sigma} \cdot \omega_s \cdot i_{rd}$ can be neglected due to the small rotor current in the d-axis. Then ω_r with aid of (4b) may be written as:

$$\omega_{rA} = \frac{2R_r T}{3p\psi_s^2} \tag{8}$$

Then the currents in the d-axis can be expressed:

$$i_{rdA} = \frac{4T^2 L_{\sigma}}{9p^2 \psi_s^3} \qquad \qquad i_{sdA} = \frac{\psi_s}{L_m} + \frac{4T^2 L_{\sigma}}{9p^2 \psi_s^3}$$
(9a, b)

The natural solution of equation (7) leads to a quadratic equation:

$$\omega_{rB} = \frac{3p\psi_s^2 \pm \sqrt{9p^2\psi_s^4 + 16p^2L_{\sigma}^2}}{4T\,L_{\sigma}^2/R_r} \tag{10}$$

Only the negative sign can be taken into account. A positive sign represents operate with high slip before the pull-out torque of the torque characteristic, which is undesirable.

Then the currents in the d-axis can be expressed as follows:

$$i_{rdB} = \frac{2T}{3p\psi_s} \frac{3p\psi_s^2 - \sqrt{9p^2\psi_s^4 + 16p^2L_{\sigma}^2}}{4TL_{\sigma}} \qquad \qquad i_{sdB} = \frac{\psi_s}{L_m} + \frac{2T}{3p\psi_s} \frac{3p\psi_s^2 - \sqrt{9p^2\psi_s^4 + 16p^2L_{\sigma}^2}}{4TL_{\sigma}} \qquad (11a, b)$$

5 ACCURACY VERIFICATION OF ROTOR FREQUENCY CALCULATIONS

The aim of this chapter is to verify the accuracy of the calculation of currents and losses in the motor when considering the simplified expression for ω_r (8) and when considering the natural expression for ω_r (10) for

different values of linkage flux. It should be noted that at reduced flux, no permanent load with nominal torque is expected. The value of ω_r according to the equation (10) is limited only to the value of the pullout torque, which can be expressed from the condition for the square root of the expression (10). It can be seen from this equation that the pull-out torque decreases with decreasing linkage flux.

Verification was performed in MATLAB environment with the parameters of the real induction motor. Since the determination of parameters is not the subject of this article, the values of the required rated parameters of induction motor ATAS T22VR512 are: torque $T_n = 2$ Nm, speed $n_n = 2380 \text{ min}^{-1}$, linkage flux (peak) $\Psi_n = 1$ Vs, number of pole-pairs p = 1; stator resistance $R_s = 11.8 \Omega$, rotor resistance $R_r = 9.2 \Omega$, iron loss resistance R_{Fe} (50 Hz) = 4900 Ω , leakage inductance $L_{\sigma} = 90$ mH. Magnetization inductance is a function of the linkage flux, and its values are: $L_m (\Psi = 1 \text{ Vs}) = 0.9 \text{ H}$, $L_m (\Psi = 1.1 \text{ Vs}) = 0.7 \text{ H}$, $L_m (\Psi = 0.75 \text{ Vs}) = 1.07 \text{ H}$, $L_m (\Psi = 0.5 \text{ Vs}) = 1.2 \text{ H}$. Details of the parameters of the induction motor are given in [3].

Figure 2 shows the difference of the values of ω_r according to equations (8) and (10) as a function of the torque for four linkage flux values. It can be seen that at rated linkage flux $\psi = 1$ Vs the difference between the values ω_{rA} and ω_{rB} is negligible to the rated torque. This difference increases at higher torques than the



Figure 2: Dependence of rotor frequency on torque at different linkage flux values.

rated. At the linkage flux $\psi = 1.1$ Vs the difference is even smaller. On the other hand, when the linkage flux decreases to $\psi = 0.75$ Vs and $\psi = 0.5$ Vs, the difference increases.

Figure 3 shows the rotor current in the d-axis according to the equations (9a) i_{rdA} and (11a) i_{rdB} and the stator current in the q-axis i_{sq} as a function of the torque for four linkage flux values. It can be seen that all rotor d-axis currents i_{rdA} and i_{rdB} and stator q-axis current i_{sq} increase as the linkage flux decreases. The difference between the value of i_{rdA} and i_{rdB} increases with increasing torque. A significant difference occurs at reduced linkage flux, especially at $\psi = 0.5$ Vs. Furthermore, the stator q-axis current i_{sq} is higher than the rotor d-axis current i_{rdA} and i_{rdB} .

In Figure 4, the stator currents in the d-axis i_{sdA} and i_{sdB} are plotted as a function of torque at four different linkage fluxes. In particular, the magnetizing current i_u is plotted here using equation (3b). The ellipse corresponding to the rated operation condition of the induction machine i_{max} is also plotted in the graph. In this area, the induction motor can operate continuously. For the linkage flux $\psi = 1$ Vs and $\psi = 1.1$ Vs it can be seen that the difference between the currents i_{sdA} and i_{sdB} is very small. For lower linkage flux $\psi = 0.75$ Vs and $\psi = 0.5$, the difference increases, especially at higher torque. The currents i_{sdA} and i_{sdB} are also affected





by the magnetizing current i_u , which increases with increasing linkage flux. Only at $\psi = 0.5$, the rated torque line passes only with i_{sdB} and not with i_{sdA} .

Figure 5 shows the iron loss P_{Fe} , Joule losses in the rotor winding (P_{jrA} and P_{jrB}) and stator winding (P_{jsA} and P_{jsB}), depending on the torque at different linkage flux values. Losses were calculated by considering



Figure 4: Dependence of stator currents on torque at different linkage flux values.



Figure 3: Dependence of power losses on torque at different linkage flux values.

two current values according to equations (9a, b) and (11a, b). It can be seen that the difference between P_{jrA} and P_{jrB} and P_{jsA} and P_{jsB} , respectively, occurs at low value of linkage flux. Furthermore, it can be seen how the iron loss P_{Fe} increases with increasing linkage flux.

6 CONCLUSION

The aim of this paper was to calculate ω_r according to two different formulas and to verify its influence on quantities in the equivalent circuit of the induction machine, especially on current and power losses. It can be seen from Figures 2-5 that at higher linkage flux value, the difference between ω_{rA} (8) and ω_{rB} (10) is negligible. At low flux values, the difference between ω_{rA} and ω_{rB} increases, especially especially at a higher torque value. However, the motor is not expected to operate at rated torque at reduced linkage flux. Thus, the simplified formula for calculating ω_{rA} (8) can be considered sufficient and can be replaced by the natural formula ω_{rB} (10) for the analytical expression of the optimal linkage flux.

ACKNOWLEDGEMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-20-6379).

REFERENCES

- M. N. Uddin and S. W. Nam, "New Online Loss-Minimization-Based Control of an Induction Motor Drive," in *IEEE Transactions on Power Electronics*, vol. 23, no. 2, pp. 926-933, March 2008, doi: 10.1109/TPEL.2007.915029.
- [2] N. P. Quang and J. A. Dittrich, *Vector Control of Three-Phase AC Machines System Development in the Practice*. Berlin, Germany: Springer-Verlag, 2008.
- [3] M. Toman, R. Cipin, P. Vorel and M. Mach, "Algorithm for IM Optimal Flux Determination Respecting Nonlinearities and Thermal Influences," 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Palermo, 2018, pp. 1-5, doi: 10.1109/EEEIC.2018.8493953.

ANALYSIS OF EQUIVALENT THERMAL CONDUCTIVITY OF WINDING USING FEM-BASED MODEL

Marek Toman

Doctoral Degree Programme (6), FEEC BUT E-mail: marek.toman@vutbr.cz

> Supervised by: Pavel Vorel E-mail: vorel@fecc.vutbr.cz

Abstract: This paper deals with analysis of equivalent thermal conductivity of winding. Analytical method from literature is described first. Then, the FEM-based model is described, which was created in order to verify the functionality of the analytical method. Results of the FEM simulations showed that the analytical method works well.

Keywords: Electric machine, fill factor, slot, thermal analysis, thermal conductivity, winding

1 INTRODUCTION

From a temperature point of view, the most critical part of an electrical machine is usually a winding. The high temperature in the winding can significantly affect a winding insulation and thus a motor lifetime. Each thermal class of winding insulation has prescribed maximum permitted temperature rise (not a hot spot temperature) for which an insulation lifetime is declared (usually 20 000 h) [1]. Exceeding this temperature rise by approximately 8-10 K yields to shortening the insulation lifetime by half, see [1].

Winding temperature affects not only the insulation lifetime but it has also impact on an electric resistance of the winding. It is important to know the winding electric resistance (and thus its temperature) in many control strategies of electric machines, e.g., in an induction motor control strategy achieving maximum efficiency in a wide range of speed and torque, see [2]. For these reasons, it is necessary to know the winding temperature both, i.e., during a designing process of the machine and during machine operation.

Thermal modeling of winding can be quite difficult because it is composed of several parts with different thermal properties, namely of conductors with very good thermal conductivity (TC), conductor insulation with very low TC and impregnation which also has very low TC, see the Fig. 1. In thermal models of electrical machines, the winding is usually replaced by a homogeneous space with an equivalent thermal properties in individual directions of coordinate system, i.e. the equivalent homogenous material is orthotropic. Analytical equations for the calculating of an equivalent thermal conductivity (ETC) of the winding can be found in the literature, see [3]. The goal of this paper is to compare the analytical method with a numerical solution. There was created a thermal model based on the finite element method (FEM) in the Ansys software. This model is described in the paper.

2 ANALYTICAL METHOD FOR CALCULATING THE EQUIVALENT THERMAL CON-DUCTIVITY OF WINDING

As previously mentioned, the resulting ETC of the winding depends on the thermal conductivities of the individual parts of the winding. But the resulting ETC also strongly depends on a fill factor of the slot. There are different definitions of the fill factor. Papers dealing with calculation of dependence of



Figure 1: (a) Random distribution of wires in the slot and definition of slot dimensions. (b) Aligned distribution of wires for the FEM model. (c) Homogenized winding and equivalent thermal conductivities in individual directions. (d) Detail of the aligned distribution.

the ETC on the fill factor unfortunately do not usually define how the fill factor should be calculated. In this article, the fill factor is taken as ratio of the area of conductors (without their insulation) to the slot area in which the winding is concentrated, which is mathematically written as follows

$$K_{\rm f,Cu} = \frac{\pi d_{\rm c}^2 N_{\rm c}}{8A_{\rm 1s,win}},\tag{1}$$

where $K_{f,Cu}$ is the fill factor of winding, d_c is diameter of conductor (without its insulation), N_c is number of conductors in one slot, and $A_{1s,win}$ is a part of one slot area in which the winding is concentrated, which can be calculated according to approximate equation

$$A_{1s,\text{win}} \approx A_{1s} - p_s t_{si},\tag{2}$$

where A_{1s} is total area of one slot, p_s is perimeter of the slot, and t_{si} is thickness of a slot insulation. The parameters A_{1s} and p_s can be calculated according to slot dimensions, see Fig. 1 (a), as follows

$$A_{1s} = h_{2s} \frac{b_{1s} + b_{2s}}{2} + \frac{\pi b_{2s}^2}{8},\tag{3}$$

$$p_{\rm s} = b_{1\rm s} + \frac{\pi b_{2\rm s}}{2} + 2\sqrt{\frac{(b_{2\rm s} - b_{1\rm s})^2}{4} + h_{2\rm s}^2}.$$
(4)

The analytical calculation of the ETC of winding is divided into two steps [3]. First, the equivalent thermal conductivity of homogenised conductor and conductor insulation k_c^* is calculated according to the equation [3]

$$k_{\rm c}^* = k_{\rm ci} \frac{k_{\rm c}(1+\chi_{\rm c}) + k_{\rm ci}(1-\chi_{\rm c})}{k_{\rm c}(1-\chi_{\rm c}) + k_{\rm ci}(1+\chi_{\rm c})},\tag{5}$$

where k_c is the TC of conductor, k_{ci} is the TC of conductor insulation, and χ_c is the area ratio of the conductor cross-section to wire cross-section (including conductor insulation) calculated according to the equation [3]

$$\chi_{\rm c} = \left(\frac{d_{\rm c}}{d_{\rm ci}}\right)^2,\tag{6}$$

where d_{ci} is outer diameter of conductor insulation.

Subsequently, the required ETC of homogenised winding $k_{win,x,y}$ can be calculated according to the equation [3]

$$k_{\rm win,x,y} = k_{\rm imp} \frac{k_{\rm c}^*(1 + K_{\rm f,Cu}) + k_{\rm imp}(1 - K_{\rm f,Cu})}{k_{\rm c}^*(1 - K_{\rm f,Cu}) + k_{\rm imp}(1 + K_{\rm f,Cu})},\tag{7}$$

where k_{imp} is the TC of winding impregnation. The symbols "x" and "y" in the subscript indicate that the calculated ETC corresponds to the *x*- and *y*-direction, see the Fig. 1 (c). Graphical interpretation of the equation (7) can be seen in the Fig. 3 (b) where the results are also compared with the FEM-based model results.

Calculation of the ETC of winding in z-direction is much easier than in the x or y ones. Assuming that $k_c \gg k_{ci}$ and $k_c \gg k_{imp}$, the ETC of homogenised winding in z-direction can by calculated simply as follows

$$k_{\text{win,z}} = k_{\text{c}} K_{\text{f,Cu}} \frac{A_{1\text{s,win}}}{A_{1\text{s}}}.$$
(8)

The equation (8) assumes that the cross-sectional area of the homogenised winding in XY-plane equals to A_{1s} .

3 FEM-BASED MODEL FOR CALCULATING THE EQUIVALENT THERMAL CONDUC-TIVITY OF WINDING

To verify the analytical method presented in the previous section, a thermal model based on the finite element method was created, which is described in this section. The model was created in the well-known Ansys software.

3.1 MODEL DESCRIPTION

Individual wires are in produced motor usually distributed randomly as shown in the Fig. 1 (a). Parametrization of such an arrangement would be very complicated, therefore in the FEM-based model, it is assumed that the wires are distributed in rhombus alignment, see the Fig. 1 (b), (d). The advantage is that this layout can be fully parametrizable. This approach is used for example in [4].

Created geometry can be seen in the Fig. 2. There are totally $15 \times 15 = 225$ wires in this geometry. The geometry has outer dimensions L_x , L_y , L_z in the respective directions. According to the Fig. 1 (d), there were derived equations for calculating the distances between the individual wires depending on the fill factor. In this case, the fill factor is taken as the ratio of the conductor area to the rhombus area, i.e., $K_{f,Cu} = (\pi d_c^2)/(8\delta_{c,x}\delta_{c,y})$. The distance of the conductors can be then calculated according to the following equations

$$\delta_{\rm c,y} = d_{\rm c} \sqrt{\frac{\pi \tan(\beta_{\rm c})}{8K_{\rm f,Cu}}},\tag{9}$$

$$\delta_{c,x} = \frac{\delta_{c,y}}{\tan(\beta_c)},\tag{10}$$

where $\delta_{c,x}$ is the distance of adjacent conductors in the *x*-direction, $\delta_{c,y}$ is the distance of adjacent conductors in the *y*-direction, and β_c is the angle between the centres of adjacent conductors. The

outer dimensions of the created geometry are calculated according to the equations

$$L_{\mathbf{x}} = \boldsymbol{\delta}_{\mathbf{c},\mathbf{x}}(N_{\mathbf{c},\mathbf{x}}+1),\tag{11}$$

$$L_{\rm y} = \delta_{\rm c,y}(N_{\rm c,y} + 1), \tag{12}$$

where $N_{c,x}$ is the number of conductors in x-direction and $N_{c,y}$ is the number of conductors in ydirection. The dimension L_z should be as small as possible to ensure a small number of elements. It is obvious according to the Fig. 2 that created geometry does not meet conditions of a consistent fill factor near the boundaries, but with a sufficiently large number of conductors, this effect is negligible.

According to the Fig. 1 (d), it is also possible to derive the maximum theoretical fill factor

$$K_{\rm f,Cu,max} = \frac{\pi}{8\sin(\beta_{\rm c})\cos(\beta_{\rm c})} \left(\frac{d_{\rm c}}{d_{\rm ci}}\right)^2.$$
 (13)

Graphical interpretation of the equation (13) can be seen in the Fig. 3 (a).

3.2 BOUNDARY CONDITIONS AND SIMULATION RESULTS

Using the FEM model, the ETCs $k_{win,x}$ and $k_{win,y}$ were evaluated. The boundary conditions that were used in the simulations are shown in the Fig. 2. The individual ETCs were determined separately, so the shown boundary conditions were not used simultaneously (the solutions are one-dimensional).



Figure 2: The geometry of the FEM-based model and definition of boundary conditions (during the simulations, the boundary conditions for *x*- and *y*-direction are not applied simultaneously).

By applying mentioned boundary conditions, the maximum temperature ϑ_{max} is at the side where the heat flow (P_x or P_y) enters. Then it possible to determine the required ETCs according to the equations

$$R_{\rm x} = \frac{L_{\rm x}}{L_{\rm y}L_{\rm z}k_{\rm win,x}} = \frac{\vartheta_{\rm max}}{P_{\rm x}} \Rightarrow \left[k_{\rm win,x} = \frac{L_{\rm x}}{L_{\rm y}L_{\rm z}} \frac{P_{\rm x}}{\vartheta_{\rm max}} \right],\tag{14}$$

$$R_{\rm y} = \frac{L_{\rm y}}{L_{\rm x}L_{\rm z}k_{\rm win,y}} = \frac{\vartheta_{\rm max}}{P_{\rm y}} \Rightarrow \boxed{k_{\rm win,y} = \frac{L_{\rm y}}{L_{\rm x}L_{\rm z}}\frac{P_{\rm y}}{\vartheta_{\rm max}}}.$$
(15)

A total of 75 simulations with different values of the parameters $K_{f,Cu}$, β_c , and k_{imp} were performed (dimensions of the FEM geometry were recalculated according to the equations (9)-(12)). For the all simulations: $d_c = 1 \text{ mm}$, $d_{ci} = 1.07 \text{ mm}$, $k_c = 380 \text{ W/(m·K)}$, and $k_{ci} = 0.25 \text{ W/(m·K)}$. The FEM model results are shown in the Fig. 3 (b).



Figure 3: (a) Maximum theoretical fill factor depending on the angle between the centres of adjacent conductors. (b) Comparison of analytical calculations with FEM simulation results.

4 CONCLUSION

The aim of the paper was to verify the analytical method for calculating the equivalent thermal conductivity of winding. The FEM-based model was created for this purpose. The comparison of the analytical and numerical results can be seen in the Fig. 3 (b). The differences in the results are very small. It was also found that the equivalent thermal conductivities in different directions do not differ much depending on the arrangement of the conductors.

ACKNOWLEDGEMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-20-6379).

REFERENCES

- PYRHONEN, Juha, Tapani JOKINEN and Valeria HRABOVCOVA. Design of rotating electrical machines. Second edition. Chichester, West Sussex, United Kingdom: Wiley, 2014. ISBN 978-1118581575.
- [2] TOMAN, Marek, Radoslav CIPIN, Pavel VOREL and Martin MACH. Algorithm for IM Optimal Flux Determination Respecting Nonlinearities and Thermal Influences. In: *Conference EEEIC/I&CPS Europe*. IEEE, 2018, 2018, pp. 1-5. ISBN 978-1-5386-5186-5. doi:10.1109/EEEIC.2018.8493953
- [3] LIU, Haipeng, Sabrina AYAT, Rafal WROBEL a Chengning ZHANG. Comparative study of thermal properties of electrical windings impregnated with alternative varnish materials. *The Journal* of Engineering. 2019, 17. ISSN 2051-3305. doi:10.1049/joe.2018.8198
- [4] OŠLEJŠEK, Oldřich. Tepelná vodivost svazku vodičů kruhového průřezu. *Technika elektrických strojů*. Brno: Výzkumný a vývojový ústav elektrických strojů točivých, 1972, **17**, 65-80.

CONTROL UNIT FOR ELECTROPORATING GENERATOR

Martin Folprecht

Doctoral Degree Programme (4), FEEC BUT E-mail: xfolpr01@stud.feec.vutbr.cz

Supervised by: Dalibor Cervinka E-mail: cervinka@feec.vutbr.cz

Abstract: This contribution deals with a design and improvement of a logic unit, which controls a HV generator for AC cell electroporation. The generator produces short bursts of bipolar high-voltage pulses. Its safety is provided by appropriate topology of a power part and by safety circuits at the control unit.

Keywords: AC cell electroporation, DC-AC inverter, HV generator, Schmitt trigger, voltage bursts

1 INTRODUCTION

Demands on safety of electronic devices used in medicine are strict and approval of new devices for clinical use takes a long time [1]. Thus, new devices are used for experiments. One of these new devices is the HV generator for AC cell electroporation developed at FEEC BUT. This contribution is focused on its control unit, which provides a safety of the whole device together with appropriately chosen topology of a power part.

2 HV GENERATOR FOR AC CELL ELECTROPORATION

AC cell electroporation is based on the use of short bursts of bipolar rectangular pulses. The generator (Figure 1) of these bursts works as a DC-AC inverter with pulse transformer. Transformers are replaceable, so two different output peak voltage values (1.3 kV or 2.5 kV) can be chosen. Delivered peak power 27.5 kW in both cases is the same, so two maximal output current values (21 A resp. 11 A) are obtained. This topology is not widely used compared to topology with capacitor bank. The topology with capacitor bank is advantageous with respect to shape of output pulses, which can be purely rectangular [2]. Energy accumulation at capacitor bank also takes a less space.



(a) Front panel

(b) Internal arrangement

Figure 1: HV generator for AC electroporation

On the other hand, malfunction of a semiconductor switch, which connects the capacitor bank to application electrodes, can cause uncontrolled discharge of the capacitor into the tissue. For this reason, safety switch, which grounds the electrodes, must be used [3]. This problem does not occur, when the topology with pulse transformer is used. Eventual transistor malfunction leads to magnetizing current increase and core saturation [4]. Similar problem can occur in control unit, when signals are affected by interference. For this reason, control unit must contain some protective circuits.

3 CONTROL UNIT

The control unit allows the operator to set time properties of the output voltage: a length of bursts, a length of space between bursts and frequency of pulses. Output bursts can be synchronized with ECG signals, that is important during operation near myocardium. A built-in overcurrent protection reacts, when a primary current of the transformer exceeds its peak value. A block diagram of the unit can be seen in Figure 2a and typical shape of the output voltage in Figure 2b.



(a) Block diagram of the control unit

(b) Typical shape of the burst

Figure 2: Control unit of the generator

High-frequency pulses are generated by tunable CMOS 4047 oscillator, their frequency can be changed from 60 to 470 kHz. Output signals Q and \overline{Q} are led to inputs IN A and IN B of an integrated gate driver UCC 27524, whose outputs are boosted by MOSFET transistors. Gate signals for switching transistors in power part are isolated by pulse transformer. The oscillator, which is also a source of a clock signal (CLK) for a D-flip-flop (CMOS 4013), runs permanently. Bursts are created by cutting of output signals Q and \overline{Q} . It is provided by integrated driver in cooperation with D-flip-flop. Its input S is grounded (S = 0) and signal from burst length generator (Figure 3) is led to its input D.



Figure 3: Burst length generator

Logic state from input D is transferred to output Q with rising edge of the clock signal. Thus, when output signal from the burst length generator D = 1, integrated driver is unblocked by its enable inputs EN A and EN B. Burst length can be varied from 40 to 120 μ s by potentiometer R_4 . The generator is triggered on a rising edge of a signal brought to its input. A source of the input signal can be a space length generator (AUTO mode) or a synchronization circuit (ECG mode). Working mode is selected by toggle switch. Especially in ECG mode, when the input signal can be affected by interference, a safety flip-flop circuit (IC_{2A}, IC_{2B}) is needed. It does not allow a creation of the new burst less than 0.3 s after the end of the previous one. This situation is captured in Figure 4.



Figure 4: Input and blocking signals at the burst length generator

Blocking signal is triggered on a falling edge of the input signal brought to the mode switch. A normal situation can be seen in Figure 4a. Rising edge of the input signal comes more than 0.3 s after the falling edge, so power part is unblocked with each rising edge of the input signal for a period set by potentiometer R_4 . An error situation is in Figure 4b, when the rising edge of the input signal comes less than 0.3 s after the falling edge and the power part is permanently blocked. The space length generator (Figure 5a) works as a relaxation oscillator. Its period can be set by potentiometer R_5 from 0.5 to 1.5 s. A diode D₁ with resistor R_6 provide, that a duty cycle is lower than 0.5. Thus, the safety flip-flop circuit IC_{2A}, IC_{2B} (Figure 3) can not block the power part.



Figure 5: Space length sources

A signal brought to the ECG input (Figure 5b) is isolated by optocoupler D_1 , T_1 , which also works as a signal shaper. Shaped signal (ECG) is led to the mode switch and to a flip-flop circuit, which extends the signal to synchronize an external oscilloscope used for burst visualization. Oscilloscope output is isolated by the second optocoupler D_2 , T_2 , which works also as a signal inverter. Mentioned signals in one sequence are captured in Figure 6. It is obvious that the burst length generator is triggered on the rising edge of the space length signal and the power part is unblocked with rising edge of the burst length signal. The first and the last pulse is shortened and the whole burst is delayed because of used D-flip-flop. When the rising edge of the burst length signal appears at the D-input of CMOS 4013 (Figure 2a), the flip-flop waits for the rising edge of the CLK signal and after that, the rising edge from the D-input is transferred to its Q-output. The power part is unblocked. A similar situation becomes with the falling edge of the burst length signal. The last pulse is created, but the power part is blocked with the next CLK rising edge and the pulse is cut off.



Figure 6: Sequence of control signals

Primary current of the pulse transformer in the power part is sensed by a pulse transformer with transfer ratio 1:1500 (Figure 7) and rectified. When the primary current exceeds its nominal value 96 A, voltage drop on shunt resistors R_7 , R_8 reach to 8.5 V and RESET signal created by Schmitt triggers IC_{1C}, IC_{1D} is transferred to the R-input of CMOS 4013 and power part is blocked. Unblocking is provided by mains switch. Undervoltage protection, which is based on TL 431 regulator and Schmitt triggers IC_{1A}, IC_{1B}, is also needed. The control unit is powered from auxiliary supply + 15 V DC and in the case of voltage drop, power transistors can be switched incorrectly. To suppress this problem, gate signals are blocked, when the supply voltage decreases under 13 V.



Figure 7: Undervolatge and overcurrent protection

4 BURST LENGTH GENERATOR IMPROVEMENT

Burst length generator can be improved by protective flip-flop circuit (IC_{3A} , IC_{3B}), which does not allow a creation of longer burst than 150 μ s. A cutout of current circuit diagram can be seen in Figure 8a. Improved version is shown in Figure 8b, added components are red. When an error in flip-flop (IC_{1B} , IC_{1C}) occurs, added protective circuit cuts the burst after 150 μ s, so the amount of energy delivered to the tissue cannot be increased. Other part, that can be improved, is the synchronization circuit. Actual version allows to synchronize the oscilloscope only in ECG mode. But it is advantageous to synchronize the oscilloscope also in AUTO mode, because it simplifies the work with devices at workplace. And finally, control unit based on Schmitt triggers is simple, but changes of current parameters or addition of new functions is associated with component replacements or new PCB design. For this reason, new sample of AC electroporating generator will be controlled by MCU or FPGA.



Figure 8: Protective circuit in burst length generator

5 CONCLUSION

This contribution deals with control unit of HV generator for AC electroporation. The unit, which is based on Schmitt triggers, produces gate signals for switching transistors, protects these transistors against the overrcurrent and allows the user to set time properties of output voltage bursts. Additional function is ECG synchronization possibility. Safety of the whole device is provided by chosen topology of the power part and by safety circuits as parts of the control unit. The topology with pulse transformer is more safe compared to the topology with capacitor banks, because the risk of uncontrolled capacitor discharge is eliminated and a body of the patient is isolated from AC mains by the pulse transformer. The control unit contains mentioned overcurrent protection, undervoltage protection and flip-flop circuit, which provides a minimal delay between two bursts. Next sample of the unit will be supplemented by other flip-flop, which will determine maximal length of the burst. Finally, the whole device will by controlled by MCU or FPGA.

ACKNOWLEDGMENT

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-20-6379).

REFERENCES

- [1] European Standard EN60601-1 Medical Electrical Equipment Part I: General Requirements for Safety, 2nd ed., 1998.
- [2] N. F. Kasri, M. A. M. Piah, and Z. Adzis, "Compact high-volatge pulse generator for pulsed electric field applications: Lab scale development," Journal of Electrical and Computer Engineering, vol. 2020, DOI 10.1155/2020/6525483, pp. 1–12, Sep. 2020.
- [3] V. Stankevic, P. Simonis, N. Zurauskiene, A. Stirke, A. Dervinis, V. Bleizgys, S. Kersulis, and S. Balevicius, "Compact square-wave pulse electroporator with controlled electroporation efficiency and cell viability," Symmetry, vol. 12, DOI 10.3390/sym.12030412, no. 3, pp. 1–14, Mar. 2020.
- [4] D. Cervinka and V. Novotna, "High-voltage pulse source for cell electroporation," Advances in Intelligent Systems and Computing, DOI 10.1007/978-3-319-65960-2-11, pp. 1–6, Sep. 2018.

Doktorské projekty

Teoretická elektrotechnika, fyzika a matematika

SIMULATING A UAV SWARM - OVERVIEW AND PROPOSAL

Petr Raichl

Doctoral Degree Programme (1), FEEC BUT E-mail: xraich02@vutbr.cz

> Supervised by: Petr Marcon E-mail: marcon@vutbr.cz

Abstract: The paper is aimed at simulations of a UAV swarm. The first part of the paper presents a UAV swarm overview, application areas, and two main kinds of UAVs swarm control, i.e. with/without a base station. The second part of this paper discusses possibilities and problems in UAVs swarm simulations. The main part of this paper offers our own proposal on how to simulate a UAV swarm. The proposal is architecture, which contains 3 main parts - Our own logic unit, which is responsible for data processing, decisions, and actions of a device, Gazebo as a realistic simulation environment, and ROS as a bridge between Gazebo and our logic unit. Additionally, a simple simulation of multiple UAVs was done.

Keywords: UAV, swarm, swarm simulation, ROS, Gazebo

1 INTRODUCTION

A Swarm of Unmanned Aerial Vehicles (UAVs) is a group of UAVs that work together to accomplish a goal of specific mission. The idea of swarm cooperation is that a swarm of robots can achieve a goal of specific mission more effectively in comparison of individual robot itself.

Before deployment of UAV swarm system, a lot of algorithms have to be tested, i.e., localization, detection, formation, swarming, cooperation algorithms. Testing of mentioned algorithms with real UAVs could be financially inefficient, so there is need of simulations. This paper presents review of existed options to simulate UAV swarms and presents another proposal.

1.1 APPLICATION AREAS

The UAVs are usually equipped by different sensors to sense information about environment in which it is located. Then these collected information can be processed and transmitted to control center or some kind of specific action can be triggered.

Nowadays, single UAVs can be deployed in areas such as: Monitoring of disasters, monitoring of city traffic, agriculture monitoring, search and rescue missions, attacks/defense of guarded object. [1]

UAV swarms can be deployed in the same application areas as single drones, however the control of swarm is needed. The main control approaches are discussed below.

1.2 A BASE STATION DRIVEN SWARM

Nowadays, most of UAVs swarm studies, [2] - [4], are aimed at a swarm controlled from a base station, sometimes called Ground Control Station (GCS). In this type of UAV swarm control, every single UAV communicates with a base station. The base station is typically high-performance work-station. This workstation receives data from UAV, processes this data, and sends control data to UAV. It means that the control station is responsible for every UAVs actions. [3]

Disadvantages of this type of UAV swarm control can be clearly seen. The performance of the base station limits the maximum number of UAVs in the swarm. The need for secured, fast and stable communication between UAVs and base station on long distances is another disadvantage.

1.3 AUTONOMOUS UAV SWARM

In this type of UAV swarm control, every single UAV has some level of autonomy. It means that this UAV can trigger some actions independently on a base station. It can trigger actions based on information from sensors or based on information from sensors of some swarm unit or information received from a base station. However, it doesn?t mean that the base control is not present or that UAVs can not communicate with the base control.

If communication with the base station exists in this type of UAV swarm control, it is limited to elementary instructions. These limited instructions can inform UAVs about the goals of the actual mission, inform about important facts. In the opposite direction of communication, i.e. from UAVs to the base station, it can be results reports or issues.

2 POSSIBILITIES AND PROBLEMS OF UAV SWARM SIMULATION

Generally, two main kinds of UAVs simulations can be done. The first kind is Software-In-The-Loop (SITL) and the second one is Hardware-In-The-Loop (HITL). Both of these kinds have some advantages and disadvantages. The main advantage of SITL is no need for real hardware (HW), so in this kind of simulation, theoretically unlimited number of UAVs in a swarm could be simulated. Another advantage is, that software changes can be done easily without the need for reprogramming of HW. Speed of simulation can be faster than in real-time too. On the other side, characteristics of HW, such as sensor noises and time delays, can be counted on it in HITL. So, HITL simulations are more realistic, however, in UAVs swarm HITL simulation, the structure of multiple UAVs can be much bulkier and the task very challenging.

Naturally, with the growing popularity of UAVs, a number of UAVs simulators have been created. Nevertheless, these simulators were primarily created for simulation of single UAVs, and simulation of UAVs swarm is problematic or impossible. There are simulators such as X-Plane and FlightGear where each UAV is simulated on a single machine and visualizations are displayed together on a monitor, however, this solution is not cost and place effective. Another type of simulator enables to run several instances on a single simulation computer and integrates visualization to one. It is clear that the number of instances is limited by computing resources.

Nevertheless, none of these simulators enables flexible simulations based on own written algorithms. So, this paper presents a proposal of UAVs swarm simulation software, which could be used for realistic and computational effective simulation of algorithms used in UAV swarms.

3 PROPOSED SOLUTION

This section contains our proposal of SW architecture for UAVs swarm simulation and a quick introduction of modules used in this proposal.

3.1 ROBOT OPERATING SYSTEM (ROS)

ROS is an open-source collection of software libraries and tools for building robot applications. ROS enables solving problems such as real-time data collection of data from cyber-physical systems sensors or management of commands from a user to actuate related actions [5].

ROS implements the publish/subscribe model. It means the components of the robots are represented by nodes, which communicates one to another via the anonymous and asynchronous publish/subscribe mechanism [5].

3.2 GAZEBO

A Gazebo is an open-source simulator that enables effective and realistic simulation of the environment faced by robots in operation. The main advantage of Gazebo is possible integration between ROS and Gazebo. This integration is provided by a set of Gazebo plugins that support many existing robots and sensors. A gazebo user can write ROS nodes that are compatible with simulation and hardware. Another advantage is that it is possible to develop an application directly in a simulation environment and then deploy the application to a physical robot with little or no changes [5].



Figure 1: SW architecture proposal

3.3 ARCHITECTURE PROPOSAL

Our proposal of software architecture is shown in fig. 1. This architecture can solve most of the problems discussed above: simulation of multiple UAVs, simulation of own algorithms, simulation of algorithms in simulation environment with easy adaption to the real world.

The architecture contains three units: Our SW unit where all logic(algorithms, data processing, etc.) is done, Gazebo for realistic simulation environment, ROS as a bridge between our SW unit and Gazebo.

For enabling the simulation of UAVs we ported the popular Hector quadrotor [6], developed in TU Darmstadt, to ROS Noetic and the newest Ubuntu 20.04. For the simulation of multiple UAVs in a single environment, we use different namespaces for every UAV. The idea is, that generation of multiple namespaces will be done automatically based on configuration file, which will be edited by the user. Meantime, namespaces can be generated manually.

The logic unit contains 5 main blocks: Simulation controller, Perception block, detection, navigation, and device control. These modules were designed based on the typical pyramid of autonomy devices presented in [7]. Information from the Gazebo environment, i.e. information from UAVs sensors goes through ROS to the Perception module. This module processes sensors data. It means localization in the environment, localization in the swarm, pose estimation, and eventually obstacle avoidance. The Detection module is responsible for the detection of objects of interest. So, this module contains an object detector. The Navigation module is responsible for path planning based on information from Perception and eventually the Simulation control module. The Navigation module transmits instructions to the device control module, which is responsible for controlling the UAV. Eventually, when deploying the application to a real device, only the device control module, and format of sensors information which entries to Perception module will be changed. The Simulation control module is responsible for logging simulation data, evaluation of simulation - for example detection of collisions, evaluation of obstacle avoidance algorithms, etc.



Figure 2: Simulation example

3.4 SIMULATOR EVALUATION

Based on the architecture discussed above, simulation was done for evaluation of the suitability of the proposed solution. A few of Hector's quadrotors were placed in the Gazebo simulation environment by editing the configuration file. Then C++ program, which communicates with The Gazebo environment through ROS bridge was written. This program sends commands to single UAVs with coordinates where they shall move in regular intervals. It means that this program is connected to specific ROS topics which are responsible for UAV moving. So, this way simple path planning was made, based on knowledge of the environment. The described simulation illustrates fig. 2.

Additionally, we displayed the sensor data that are published by specific Hector's quadrotors ROS topics in RVIZ. RVIZ is a simple program that enables the viewing of sensor data from ROS, for example, Camera data, Lidar data. The fact that we are able to connect to these ROS topics by RVIZ and view the data means, that we are going to be able to work with this data later in our program and algorithms.

4 CONCLUSION

Firstly, the overview of the UAVs swarm was presented. Secondly, two main kinds of UAVs swarm control were discussed - with/without a base station. The main disadvantages of controlling with the base station are the limited performance of the base station, the need for fast, stable communication on long distances. However, controlling with the base station can be easier, because the need for level autonomy of the robot is lower than when controlling without the base station. Then possibilities and problems of UAVs swarm simulations were presented. Based on this section, our own proposal of a UAVs swarm simulation was presented. This proposal is based on SW architecture, see fig. 1, and two open-source SW - ROS and Gazebo, which are mutually compatible. Simple simulation of multiple UAVs was described, for results illustration see fig. 2. By this simulation, we evaluated that the proposed solution is suitable for described needs. In future work, we are going to implement and test the proposed architecture.

ACKNOWLEDGEMENT

This paper was funded from the general student development project at Brno University of Technology.

REFERENCES

- A. Tahir, J. Böling, M. Haghbayan, H. T. Toivonen, J. Plosila: Swarms of Unmanned Aerial Vehicles - A Survey, In: Journal of Industrial Information Integration, Volume 16, 2019, ISSN 2452-414X
- [2] P. Tosato, D. Facinelli, M. Prada, L. Gemma, M. Rossi, D. Brunelli: An Autonomous Swarm of Drones for Industrial Gas Sensing Applications, In: 2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Washington, DC, USA, 2019, pp. 1-6
- [3] M. Campion, P. Ranganathan, S. Faruque: UAV Swarm Communication and Control Architectures: A Review. In: 2018 Journal of Unmanned Vehicle Systems
- [4] I. Bekmezci, O. K. Sahingoz, S. Temel: Flying Ad-Hoc Networks (FANETs): A survey. In: Ad Hoc Networks, Volume 11, Issue 3, 2013, Pages 1254-1270
- [5] C. Bernardeschi, A. Fagiolini, M. Palmieri, G. Scrima, F. Sofia: ROS/Gazebo Based Simulation of Co-operative UAVs, 2019, ISBN 978-3-030-13735-9
- [6] M. Johannes, A. Sendobry, S. Kohlbrecher, U. Klingauf, O. Von Stryk: Comprehensive Simulation of Quadrotor UAVs Using ROS and Gazebo, 2012, ISBN 978-3-642-34326-1
- [7] S. J. Plathottam, P. Ranganathan: Next generation distributed and networked autonomous vehicles: Review. In: 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, India, 2018, pp. 577-582

DESIGN OF THE PULSE GENERATOR FOR TESTING AND VALIDATING PARTIAL DISCHARGE MEASURING SYSTEMS

Tomáš Hejtmánek

Doctoral Degree Programme (3), FEEC BUT E-mail: xhejtm07@stud.feec.vutbr.cz

Supervised by: Petr Drexler

E-mail: drexler@feec.vutbr.cz

Abstract: This article deals with the design and implementation of a pulse generator serving as a calibration generator for setting up measuring systems for monitoring the activity of partial discharges (PD) in the high-voltage power transformers or gas insulated switchgears (GIS). After installing the measuring device, it is advisable to test the entire measuring chain and, if necessary, set parameters that would affect the correct detection of PD signals. Calibration pulse generators are used for setting and validating measuring chain. These are usually used in systems that sense the electromagnetic activity of PD. Correct and timely detection of PD can avert a transformer accident and thus contribute to higher safety in power plants. The proposed generator should serve as the output part of a programmable generator, which did not reach sufficient parameters of output pulses. The goal was to achieve a rising time around 100 ps and an amplitude of more than 6 V using modern and very fast operational amplifiers and step recovery diodes as a sharpener of the input pulse. The measurements show waveform of the output voltage pulse.

Keywords: Pulse generator, partial discharge, UHF band, step recovery diode

1 INTRODUCTION

Pulse generators are used in the field of partial discharge measurement to obtain the most similar signal of the partial discharge. An important element of any measuring system for PD is also a pulse test generator, which verifies the functionality and settings of the entire measuring chain. Many different approaches are used to generate very short and steep pulses, these are described in the following chapters. The article will mainly deal with the design of a generator with a step recovery diode (SRD), which is the most suitable for verification and validating PD measuring chain in terms of parameters and required properties.

2 WHOLE DESIGN OF IMPULSE GENERATOR

The following chapter will describe the design of individual parts of the generator. A programmable generator will be described in more detail, which will serve as an exciter for the SRD. The pulse SRD generator itself will form the output part of the programmable generator, which, however, did not reach sufficient parameters of output pulses for the UHF band and was therefore unsuitable for PD simulations. In fig. 1 you can see the programmable generator fictional name "G1". The generator has three high frequency (HF) and ultra-high frequency (UHF) outputs. It is powered from the mains and can be programmed via USB cable using software on a PC. The software allows you to program different signal patterns and place them in a defined part of the mains frequency period. It is also possible to adjust the polarity of the signal.



Figure 1: Programmable generator "G1".

Predefined patterns can be activated using the buttons on the generator. Unfortunately, a closer examination of the UHF output revealed that the output pulses did not reach the parameters of the PD pulses. The times of rising and falling edges of pulses were 5-7 ns, for the purpose of calibration of measuring systems for PD the output voltage was also small, maximum 2,8 V. The aim was to improve the existing G1 so as to maintain its main advantage - programmability but output pulses were as similar as possible to PD. Fig. 2 and 3 show the actual waveforms from UHF and HF output G1.



Figure 2: Chosen pattern replayed to a 50 Hz line-triggered oscilloscope with persistence display enabled (500 mV/div; 2ms/div).



Figure 3: The individual pulses on the 'HF' output channel have variable amplitude (shown here with display persistence enabled, 500 mV/div). Pulse duration is about 150 - 200 ns (200 ns/div).
The individual parts of the SRD generator connection will now be described. Fig. 4 is a block diagram of the proposed generator.



Figure 4: Block diagram of generator.

2.1 PREAMPLIFIER WITH ULTRA FAST OPERATIONAL AMPLIFIERS (OA)

To obtain the highest possible excitation voltage, a preamplifier from OA was included behind the generator G1. It is a two-stage non-inverting amplifier. The minimum required slew rate to transmit a signal with an amplitude of 8 V and a leading edge of 5 ns is

$$SR = \Delta V_{out} / \Delta t \tag{1}$$

where *SR* is the slew rate, ΔV_{out} is the output voltage, Δt is the rise time, after calculating we get 1600 V/µs. The leading edge of 5 ns corresponds to a frequency of 200 MHz both of these parameters were important for the selection of a suitable OA. Very fast operational amplifiers from Texas Instruments were chosen for the amplifier, which achieve a slew rate of up to 8000 V/µs and their bandwidth is 350 to 900 MHz. This ensures that the input signal from G1 is not distorted. Specifically, it is a type THS3491, which meets all required parameters with sufficient reserve. According to the manufacturer's recommendations [1], suitable sizes of feedback resistors were calculated and chosen. The first stage of the amplifier has a gain of 4x and the second stage 2x. The total gain is therefore 8x. Capacitors C14 and C19 have been added to limit the frequency band and prevent amplifier oscillation. Schematic is in figure 5.



Figure 5: Schematic of preamplifier.

2.2 STEP RECOVERY DIODE PULSE GENERATOR

The second part of the design was a circuit with SRD. The SRD actually acts as a signal edge sharpener here. A pulse-forming network is also included in the circuit, which ensures a narrow

Gaussian pulse. We used the literature [2] and [3] to design the generator. The designs of the generator with SRD and their possible modifications and improvements are very well described here, [4], [5], [6]. The SRD connection contains a Schottky diode and delay line, which allows the generation of narrow Gaussian pulses, which increases the bandwidth of the generator. Another advantage of this involvement is the large ringing reduction. Schematic of SRD generator is on fig. 6.



Figure 6: Schematic of SRD pulse generator.

Of course, the generator also includes a power supply. A laboratory symmetrical source was used for operational amplifiers and a linear stabilizer of the LM337 series, which is located directly on the PCB, was used to set the DC bias current of the SRD. The power supply of the operational amplifiers is \pm 15 V. The fabricated and equipped PCB is shown in fig. 7.



Figure 7: Complete PCB with SRD generator.

3 MEASURING ON IMPULSE GENERATOR

This chapter deals with measurements of output pulses from the generator. Both the laboratory generator and the G1 generator for which the proposed generator is intended were used for the measurement. A Tektronix TDS 694C oscilloscope (3 GHz, 10 GS/s) was used to measure the output waveforms. Unfortunately, the manufacturer did not state the time of the leading edge of the oscilloscope, but according to the sampling rate, we assume that it is around 90 ps. For completely correct results, it would therefore be appropriate to subtract this time from the measured leading edge.

In the first tests, the SRD generator was excited from the laboratory generator by rectangular pulses. The leading edge of the pulse was set to 7 ns and the repetition frequency was 3,15 MHz. After amplification, the control pulse had 12V. This corresponded to an output pulse of about 8,8 V and a leading edge without correction of 149 ps as seen in fig. 8 left part.

The last measurement was performed with the driving generator G1 to show that the output pulses from the SRD are dependent on the size of the control pulses. This confirmed the measurement, the pulses have different sizes according to the G1 programming, depending on the repetition frequency and the shape of the control pulses, the rise times of the output pulses also change slightly. The captured waveform is shown in fig. 8 right part.



Figure 8: Measured output impulse of SRD generator (left), screenshot of SRD generator output impulse driving by G1 (right).

4 CONCLUSION

In this paper, we presented the design of an impulse generator as an additional part of the programmable generator G1, which is used to set and calibrate sensors for GIS or PD. The generator consists of a preamplifier with very fast operational amplifiers in a non-inverting circuit. The second part consists of a pulse generator with SRD and pulse-forming network. The whole device was realized and subjected to measurements. The measurement results show a maximum pulse amplitude of 11,5 V and a leading edge time after correction of the oscilloscope's own leading edge below 100 ps. The graph shows a slight but undesirable ringing, which may be due to the PCB material used or the imperfect PCB design. We have thus successfully improved the G1 programmable generator according to our ideas. Subsequently, further testing will take place during the calibration of GIS or PD sensors.

ACKNOWLEDGEMENT

The preparation of this paper was assisted by the general student development project BD FEKT-S-20-6360 in progress at Brno University of Technology.

REFERENCES

- [1] THS3491 High-Power Output Current Feedback Amplifier, Texas Instruments, Datasheet, Available https://www.ti.com/lit/ds/symlink/ths3491.pdf?ts=1611669014773&ref_url=https%253A%2 52F%252Fwww.google.com%252F
- [2] Hewlett-Packard Application Note AN918: Pulse and Waveform Generation with Step Recovery Diodes, Hewlett-Packard, Santa Clara, CA, 1984
- [3] Wei, Y. Y.: (2006), Design a simple, low-cost UWB source, p. 68-74
- [4] Protiva, P., Mrkvica J., Machac, J.: (2008), Universal Generator of Ultra-Wideband Pulses, Radioengineering, p. 17
- [5] Protiva, P., Mrkvica J., Machac, J.: (2009), A compact step recovery diode subnanosecond pulse generator, Microwave and Optical Technology Letters, 52. 438 - 440. 10.1002/mop.24945
- [6] Ceramic Packaged Step recovery diode 5082-0253, Hewlett-Packard, Technical Data, Available at https://www.rf-microwave.com/resources/products_attachments/5a4386f25e561.pdf

Doktorské projekty

Zpracování signálů, obrazu a dat

MULTILINGUAL ANALYSIS OF HYPOKINETIC DYSARTHRIA IN PATIENTS WITH PARKINSON'S DISEASE

Daniel Kováč

Doctoral Degree Programme (1st year), FEEC BUT E-mail: xkovac41@stud.feec.vutbr.cz

> Supervised by: Jiří Mekyska E-mail: mekyska@feec.vutbr.cz

Abstract: This article deals with the multilingual analysis of hypokinetic dysarthria (HD) in patients with Parkinson's disease (PD). The goal is to identify acoustic features that have high discrimination power and that are independent of the language of a speaker. The speech corpus contains 59 PD patients and 44 healthy controls (HC) speaking in Czech (cs) and American English (en-US). Based on non-parametric statistical tests and logistic regression, we observed the best discrimination power has the speech index of rhythmicity (extracted from a reading text) and harmonic-to-noise ratio (extracted from a sustained vowel). We were able to identify PD with 67% sensitivity and 79% specificity in the Czech corpus and with 78% sensitivity and 67% specificity in the English one. The performance of the model was significantly lower when combining both datasets, thus suggesting language plays a significant role during the automatic assessment of HD.

Keywords: Acoustic analysis, Parkinson's disease, hypokinetic dysarthria, classification

1 INTRODUCTION

Parkinson's disease (PD) is a chronic neurodegenerative disorder characterized by progressive degeneration of neurons in the substantia nigra pars compacta (black substance) [1]. The risk of developing PD increases with the age. Other significant factors are mainly genetic predisposition, traumatic brain injury and gender. It is also obvious that the number of patients increases with population and life expectancy [2]. The disease was first described in 1817 by the English surgeon James Parkinson [3]. Although science has progressed since this time, we still do not know the true cause of PD and are not able to cure it. However, it is possible to alleviate its course. The non-invasive way is for example, using levodopa, an amino acid, which is converted to dopamine in the brain. The riskier option is surgery [4]. Therefore, early detection and initiation of treatment have a major impact on the further course of the disease. Hoehn and Yahr defined five stages of PD, with speech impairments occurring already in the second early phase of the disease [5]. These speech impairments are caused by deteriorative phonation, articulation, prosody, and respiration and are collectively referred to as hypokinetic dysarthria (HD), which affect up to 90 % of patients with PD [6]. Acoustic analysis of speech and voice signal is one way to an automated diagnosis of PD. Many studies aim to find acoustic features with high discrimination power, but only a few are multilingual. There is no multilingual study dealing with Czech and American English. Finding features that have high discrimination power and are independent of language is key to the global automated diagnosis of PD based on acoustic analysis of speech.

2 METODOLOGY

Speech recording took place in the silent room using Audacity software and a condenser microphone with the cardioid polar pattern. The result was a dataset of mono signals with a 16 kHz sampling frequency obtained from each participant.

2.1 DATASET

The database contains subjects speaking Czech and American English. Table 1 represents the numbers of males and females in combination with language and diagnosis. The mean age of Czech speakers is 73.9 ± 15.4 (PD) and 66.1 ± 5.0 (HC). For American English, it is 67.7 ± 4.1 (PD) and 70.4 ± 8.6 (HC). The Czech participants were enrolled at the St. Annes University Hospital Brno and the American ones at the Department of Neurology, College of Medicine, University of Arizona. All subjects signed informed consent. The study was approved by local ethic committees.

Table 1: Demographic data.							
		cs			en-US		- 11
	male	female	together	male	female	together	an
PD	29	19	48	8	3	11	59
HC	9	17	26	5	13	18	44
together	38	36	74	13	16	29	103

2.2 SPEECH TASKS

To analyze all dimensions of HD, several speech tasks have been employed (see Table 2). Tasks are designed to examine phonation, articulation and prosody.

Label	Speech task	Description
TSK1	Monologue	Monolog, at least 90 s long without interruption of a clinician. The participants were
		instructed to speak about their hobbies, family, job, actual date activity, etc.
TSK2	Reading	Reading a short text. The patient could read the text for her-/himself in advance.
TSK3	Sustained phonation	Approximately 3-s (not longer than 5 s) sustained vowel of [a] at a comfortable
		pitch and loudness. Performed on one breath.
TSK4	Sustained phonation	Approximately 3-s (not longer than 5 s) sustained vowel of [i] at a comfortable
		pitch and loudness. Performed on one breath.
TSK5	Sustained phonation	Approximately 3-s (not longer than 5 s) sustained vowel of [u] at a comfortable
		pitch and loudness. Performed on one breath.
TSK6	Sustained phonation	Sustained phonation of [a] at a comfortable pitch and loudness as constant
		and long as possible, at least 5 s. Performed on one breath.
TSK7	Diadochokinetic task	Rapid steady [pa]-[ta]-[ka] syllables repetition as constant and long as possible,
		repeated at least 5 times. Performed on one breath.

Table 2: Speech tasks

2.3 ACOUSTIC FEATURES

Selected acoustic features were calculated from different speech tasks for each participant. A list of all parameters together with the disorders they investigate is given in Table 3. Acoustic features depend on sex and age, therefore these confounding factors were removed using linear regression.

2.4 STATISTICAL ANALYSIS

For testing the normal distribution of features, the Kolmogorov–Smirnov test was applied. Because many features did not fit the normal distribution, for subsequent statistical analyses, Mann–Whitney U test was used. The first test investigates the dependence of the feature on the disease (3 models: cs, en-US, all) and the second the dependence on the language (2 models: PD, HC). We additionally performed a multiple testing correction via the false discovery rate (FDR) approach.

2.5 MULTIVARIATE ANALYSIS

All acoustic features were used as input independent variables for training the machine learning models (3 models: cs, en-US, all). Logistic regression was chosen as the classifier. Training data was

		Ta	ble 3: Acoustic features
Speech	Acoustic	Specific	Feature
task	feature	disorder	definition
			Phonation
TSK6	MPT	Airflow	Maximum phonation time, aerodynamic efficiency of the vocal tract
		insufficiency	measured as the maximum duration of the prolonged vowel.
TSK3–6	relF0SD	Irregular pitch	Standard deviation of fundamental frequency relative to its mean,
		fluctuations	variation in frequency of vocal fold vibration.
TSK3-6	Jitter	Microperturbations	Frequency perturbation, extent of variation of the voice range. Jitter is
	(PPO)	in frequency	defined as the variability of the F0 of speech from one cycle to the next.
TSK3–6	Shimmer	Microperturbations	Amplitude perturbation, representing rough speech. Shimmer is defined
	(APO)	in amplitude	as the sequence of maximum extent of the signal amplitude within each
	(in ampricae	vocal cycle
TSK3 6	HNR	Increased noise	Harmonics to noise ratio, the amount of noise in the speech signal
1513-0	IIIII	mereased noise	mainly due to incomplete vocal fold closure. HNR is defined as the
			amplitude of poise relative to topal components in speech
TSV2 6	DIW	Apariadiaity	Degree of unuclead segments, the fraction of nitch frames marked
13K3-0	DUV	Apenodicity	Degree of unvolced segments, the fraction of pitch frames marked
marza (IELOD	T	as unvoiced.
TSK3-6	relFISD,	Tremor of Jaw	Standard deviation of first (F1) and second (F2) formant relative to its
	relF2SD		mean. Formants are related to resonances of the oro-naso-pharyngeal
			tract and are modified by position of tongue and jaw.
			Articulation
TSK1-5	VAI	Decreased tongue	Vowel articulation index, based on formant centralization, defined as
		movement	VAI = (F1a + F2i)/(F1i + F1u + F2a + F2u).
TSK1-2	relF1SD,	Rigidity of	Standard deviation of first (F1) and second (F2) formant
	relF2SD	tongue and jaw	relative to its mean.
TSK2	#lndmrk	Imprecise	Number of speech landmarks representing local energy maxima
		articulation	characterized by harmonic power. Landmark patterns are identified
			by comparison between "coarse" and "fine" spectral detail.
TSK7	PR	Slow alternating	Pace rate, representing the number of syllable vocalizations per second.
		motion rate	Considering first 30 syllables
TSK7	COV	Instability	Coefficient of variation defined as the ratio of the standard deviation
1011/	001	of diadochokinetic pace	of the duration of the fourth to tenth DDK cycles to the average
		of diadoenokinetic pace	duration of the first three cycles
TSV7	DI	Instability	Dhuthm instability, defined as sum of absolute deviations
13K/	KI	of diadaahakinatia paaa	from a regression line modelling each DDV quale duration
		of diadochokiletic pace	noin a regression line modelling each DDK cycle duration,
mouzz	DA	A 1 /	weighted to the total DDK performance time.
15K/	PA	Acceleration	Pace acceleration, defined as $PA = 100 \text{ x} (avCycDur4-6 - avCycDur7-9)/$
		of diadochokinetic pace	avCycDur1–3, where $avCycDurX-Y$ is average duration of cycles X–Y.
TSK7	RA	Acceleration	Rhythm acceleration, defined as gradient of regression line modelling
		of diadochokinetic pace	DDK cycle durations (positive values mean acceleration).
			Prosody
TSK1-2	relSEOSD	Monoloudness	Speech loudness variation, defined as a standard deviation of intensity
			contour relative to its mean after removing silences exceeding 50 ms.
TSK1-2	relF0SD	Monopitch	Pitch variation, defined as a standard deviation of F0 contour
			relative to its mean.
TSK2	SPIR	Inappropriate silences	Number of pauses relative to total speech time after removing periods
			of silence lasting less than 50 ms.
TSK2	PPR	Higher proportion	Percentual pause ratio, defined as total duration of silences (longer than
		of silence time	50 ms)/total duration of speech.
TSK2	DurMED	Longer duration	Median duration of silences longer than 50 ms.
		of silences	
TSK2	DurMAD	Higher variability	Median absolute deviation of silence duration (longer than 50 ms)
10112	Dunnin	of silence duration	fredam absolute de factori of shenee daration (ronger anal 20 ms).
TSK2	ΔR	Unnatural speech rate	Number of speech sounds produced per second after pauses longer
10112	1 111	Cimatarai speech fate	than 50 ms were removed
TSKO	NCT	Uighar proportion	Not speech time
13K2	101	of silonos time	net specen unie.
TEVO	TDT	Uigher properties	Total names time (names longer than 50 ms were removed)
15K2	111	nigher proportion	total pause time (pauses longer than 50 ms were removed).
TOUC	mam	of stience time	Tetal and the first
15K2	151	Higher proportion	Iotal speech time.
mar		of silence time	
TSK2	EEVOL	Unstable loudness	Energy evolution, defined as the slope of intensity.

Table 3: Acoustic features

balanced by Synthetic Minority Oversampling Technique (SMOTE). Hyperparameter tuning focused on the best accuracy was led by grid search and 10-fold cross-validation with 10 repetitions. The leave-one-out method was used for validating the models.

3 RESULTS

3.1 STATISTICAL ANALYSIS

Significant combinations (level of significance $\alpha = 0.05$) of acoustic features and tasks are listed in Table 4. P-values of Test 1 are associated with the difference between PD and HC. P-values of Test 2 are linked with the differences between languages.

Iable	Tuble 4. Digitile and combinations of acoustic reature and specen task.								
Speech task	Acoustic feature	Language	Test 1: p-value	Test 2: p-value	Change with PD				
TSK2	SPIR	all	0.019	HC: 0.992 PD: 0.359	\bar{x} : $\downarrow Me$: $\downarrow s$: \downarrow				
TSK3	mean HNR	all	0.019	HC: 0.185 PD: 0.102	\bar{x} : $\uparrow Me$: $\uparrow s$: \downarrow				
TSK3	median HNR	all	0.019	HC: 0.136 PD: 0.087	\bar{x} : $\uparrow Me$: $\uparrow s$: \downarrow				
TSK3	relF1SD	all	0.041	HC: 0.875 PD: 0.441	\bar{x} : $\downarrow Me$: $\downarrow s$: \downarrow				
TSK3	relF2SD	all	0.019	HC: 0.986 PD: 0.990	\bar{x} : $\downarrow Me$: $\downarrow s$: \uparrow				

Table 4: Significant combinations of acoustic feature and speech task.

Arrows represent the increase or decrease of mean \bar{x} , median Me and standard deviation s.

3.2 MULTIVARIATE ANALYSIS

The performance of the logistic regression model is given in Table 5 and evaluated by the following metrics: Area Under The Curve (AUC), Matthews Correlation Coefficient (MCC), Accuracy (ACC), Sensitivity (SEN) and Specificity (SPE).

 Table 5: Results of logistic regression.

 threshold
 AUC
 MCC
 ACC [%]
 SEN [%]
 SPE

	threshold	AUC	MCC	ACC [%]	SEN [%]	SPE [%]
cs	0.51	0.68	0.46	73	67	79
en-US	0.66	0.73	0.45	72	78	67
all	0.51	0.53	0.34	67	64	69

ROC curves show how SEN and SPE vary by changing the classification threshold. The importance of features is given by the coefficients of logistic regression (see Figure 1).



Figure 1: ROC curves and important features (cs: left, en-US: middle, all: right).

4 DISCUSSION

A total of 52 combinations of acoustic features and speech tasks were statistically tested. After correction of p-values, only five remained significant, and only for the model combining both languages. These features are: SPIR of TSK2 and mean HNR, median HNR, relF1SD, relF2SD of TSK3. Therefore, sustained phonation of vowel [a] and reading text played a significant role during the multilingual discrimination of PD and HC. According to the logistic regression results, it is obvious that for different languages, different features have high discrimination power. Five features are important for the model combining both languages including the earlier mentioned SPIR and median HNR. Because these two features are also significant according to statistical tests, they are highly promising for the global automated diagnosis of PD based on acoustic analysis of speech. Classification accuracy and overall performance of the machine learning model are higher when focusing on individual languages than both combined.

5 CONCLUSION

This article deals with the automated diagnosis of hypokinetic dysarthria and related Parkinson's disease focused on the dependence of acoustic features on language. Especially phonatory features and measures of pausing seem to play a significant role in this sense. To verify these features, it is desirable to expand the testing dataset, especially with other languages.

ACKNOWLEDGEMENT

This study has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 734718 (CoBeN), and from grant no. NU20-04-00294 of the Czech Ministry of Health.

REFERENCES

- HORNYKIEWICZ, O. Biochemical aspects of Parkinson's disease. Neurology [online]. 1998, 51(2, Supplement 2), S2-S9 [cit. 2020-11-17]. ISSN 0028-3878. Dostupné z: doi:10.1212/WNL.51.2_Suppl_2.S2
- [2] DeMaagd G, Philip A. Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis. P T. 2015 Aug;40(8):504-32. PMID: 26236139; PMCID: PMC4517533.
- [3] PARKINSON, James. An Essay on the Shaking Palsy. The Journal of Neuropsychiatry and Clinical Neurosciences [online]. 2002, 14(2), 223-236 [cit. 2020-11-17]. ISSN 0895-0172. Dostupné z: doi:10.1176/jnp.14.2.223
- [4] GOETZ, C. G. The History of Parkinson's Disease: Early Clinical Descriptions and Neurological Therapies. Cold Spring Harbor Perspectives in Medicine [online]. 2011, 1(1), a008862-a008862 [cit. 2020-11-17]. ISSN 2157-1422. Dostupné z: doi:10.1101/cshperspect.a008862
- [5] OWNWARD, Emily. What Are the Stages of Parkinson's Disease? [online]. [cit. 2019-12-20]. Dostupné z: https://parkinsonsdisease.net/basics/stages/
- [6] Ho AK, Iansek R, Marigliani C, Bradshaw JL, Gates S. Speech impairment in a large sample of patients with Parkinson's disease. Behav Neurol. 1999 Jan 1;11(3):131-7. PMID: 22387592.

VISION-BASED AUTONOMOUS NAVIGATION OF AN UNMANNED AERIAL VEHICLE

Jiří Janoušek

Doctoral Degree Programme (4), FEEC BUT E-mail: xjanou09@vutbr.cz

Supervised by: Petr Marcoň

E-mail: marcon@feec.vutbr.cz

Abstract: This article is focused on the real-time detection and recognition of the objects by using a camera located on Unmanned Aerial Vehicle (UAV). The detected objects in the image carry information about the next points of the flight plan according to which the drone performs an autonomous flight. The drone can recognize the information encoded in the patterns on the ground by the field of view of the camera, therefore, it can perform flight manoeuvres consequently. The detection and information recognition is performed by a single-board computer and it is based on comparing images with a predefined Haar cascade. The single-board computer on the drone recognizes a size and spacing of the individual contrast points which determine the partial information about the next flight path. Afterward, single-board computer transmits the detected information to the Pixhawk control unit by using the Mavlink protocol in real-time. The aircraft control unit ensures the accurate positioning of the UAV by sequential execution of the commands.

Keywords: Image processing, Haar, UAV, Autonomous, Mavlink

1 INTRODUCTION

Pilot's intervention is not necessary as the Unmanned Aerial Vehicle UAVs flight path is controlled by the real-time image detection where variability and various number of detected waypoints need to be processed. This thesis verifies the possibilities of using the system for autonomous flight. It brings verification of limit values of distance from detected objects, their size or flight speed of an unmanned aircraft. It verifies if is possible to perform a completely autonomous flight from take-off of the aircraft to landing. If successful detection is performed, a set of flight instructions will be found. Various speed and altitude may be set during the flight and additional accessories may be controlled by the drone. The use of real-time image detection on an autonomous flight of the UAV can serve for package delivery, mapping of large areas, or for military missions and other security forces.

2 MATHERIALS AND METHODS

Figure 1 shows a block diagram of proposed solution of an UAV. The entire autonomous system consists of the drone with a control unit Pixhawk, battery power supply, single-board computer Raspberry Pi 4 for image processing and FullHD camera. The drone can have various builds; it can be a multicopter or Vertical Take-Off and Landing (VTOL) drone. The control unit of the drone – Pix-hawk uses Ardupilot firmware that is open autopilot software for drones and other autonomous systems. The control unit is connected thru 2,4 GHz Wi-Fi telemetric module with the ground station and it serves for the control of the flight and transferring information about the drone's status. The control over the navigation and the flightpath is supplied by the processing Raspberry Pi 4 that is connected with the control unit by the serial line. Single-board computer Raspberry Pi 4 was chosen as the best fit because of the small size and sufficient power. The direct placement on the UAV eliminates the need for communication with the ground and the drone is capable of processing and

real-time image detection which allows an autonomous flight in the places where the service provider coverage would be limited [1-3].



Figure 1: Block diagram of UAV

2.1 **OBJECT DETECTION**

The detection and information recognition from the image is performed by a single-board computer Raspberry. In the program language Python, the comparison of the images with predefine Haar cascade is executed by the OpenCV library. The cascade function is trained by the positive and negative images. Haar cascades is based on the various brightness in the image allowing to use weak classifiers. The principle is the difference of the total number of pixels from two re-gions of the identical shape and size that are adjacent either vertically or horizontally. The strongest attributes are identified from the image by the method AdaBoost, which works on the principle of adaptive amplification. The created cascade of the classifiers is then imple-mented by the script on the acquired images. The final classifier consists of a cascade of many simple classifiers. The order in which the classifier then goes through is based on com-plexity of image, which may save computation time. The result is a positive finding of the images identical with the training set [4–6].

2.2 MAVLINK - MICRO AERIAL VEHICLE LINK

Mavlink is a standard communication tool for UAV and it is used for processing of flight commands. It is a simple stateless protocol for communication thru devices with low throughput. The protocol defines a set of defined messages for checking on the drone's status, sending control commands and setting up parameters. The firmware of the control unit generates telemetric messages Mavlink and allows them to be sent to another devices. In this case, the Mavlink package is sent thru connected serial line from the microprocessor in Pixhawk control unit and also in a form of the status messages back. Maylink serves for communication between Raspberry and Pixhawk control unit but it also provides transmission of the telemetric data to the ground station. The data transfer is carried out wirelessly or by USB. Thanks to the Mavlink protocol, it is possible to find out the vehicle status, the status of all connected sensors and to ensure about the command line of the flight. Based on the coordinates, it allows to plan flightpath, altitude and speed of the flight, reaction of the drone on the change of the wind conditions and a deviation from the planned path. It is possible to command where and in what situation it should land and how to react on the emergency situations. The created message is always encoded and sent in a package. The service function is initiated on the receiving site and it executes the relevant actions. A response to the sender follows after sending certain types of messages [5–7].

Each package contains heading of the message, body and a control sum, see Tab. 1. The maximum length of the packages is 263 bytes. Mavlink creates the package body. Each message is identified by the field *Message ID* in the heading of the package. This identifier sets the definition of the data structure in a message. The definition of a single message is saved in XML document that is part of Mavlink library [7].

Byte index	Content	Value	Explanation
0	Packet start sign	0xFE	Protocol-specific start-of-text (STX) marker used
	i uenet start sign		to indicate the beginning of a new packet.
1	Payload length	0 - 255	Indicates length of the following payload section.
2	Packet sequence	0 255	Used to detect packet loss. Components increment
2	I acket sequence	0 - 233	value for each message sent.
2	Sustan ID	1 - 255	ID of system (vehicle) sending the message. Used
5	System ID		to differentiate systems on network.
			ID of component sending the message. Used to
4	Component ID	1 - 255	differentiate components in a system (e.g. autopi-
			lot and a camera).
5	Massaga ID	0 255	ID of message type in payload. Used to decode
5	Message ID	0 - 255	data back into message object.
6 to $(n \mid 6)$	Data		Message data. Depends on message type (i.e.
0 10 (N+6)	Dala		Message ID) and contents.
(n+6) to (n+7)	Checksum		Checksum for error detection.

Table 1: Structure of Mavlink packet [7]

3 EXPERIMENT DESCRIPTION

For testing of an autonomous flight, cascades for five various images in a form of simplified Quick Response (QR) codes were created see Fig. 3. For each image 40 positive and 500 negative photos were used, examples see Fig. 2. The size of all QR codes is $18 \text{ cm} \times 18 \text{ cm}$ and they are made of a chessboard of distinguishing points 3,6 cm \times 3,6 cm. The QR codes were used intentionally because of their high contrast differences.

The training algorithm selects a set of positive images that contain the searched object and negative images that contain different backgrounds. Various types of pre/sumed background (on which the detected object is located) like different surfaces, houses, roads and meadows are used on the negative photos. It then generates all Haar flags and the AdaBoost algorithm selects weak linear classifiers, the combination of which creates a strong nonlinear classifier representing the degree of the cascade.



Figure 2: Selected examples of positive images for training

Images are loaded along with a trained cascade of classifiers to detect objects. The detection process is performed by selecting the size of the image detection window. A cascade of classifiers is applied to this window. The detection window is gradually shifted over the entire image and in each iteration it is magnified and shifted again. The detection process is thus applied by merging the detected windows of the same object and finding the most telling detection, which is then highlighted on the input images.



Figure 3: QR codes for detection

The primary initiations of the flight executed by the manual script initiation on the single-board computer. The command *def arm_and_takeoff(altitude)="true"* are armed motors and drone is ready to take off. Then, the drone takes off to the setup altitude. The command *vehicle.mode = Vehicle-Mode("GUIDED")* sets the drone into a mode when the drone is controlled by the series of commands with the exact position thru Raspberry. The single-board computer on the UAV recognizes the QR codes in its view field based on the images from the camera, see Fig. 4. The points are located on the certain GPS position for creating the exact path of the flight. The created script assigns the real time recognized images to the flight command database saved in the configuration set of the single-board computer. Each QR code contains specific set of instructions for another flight move on the GPS position of another point. If the match with one of the learned cascades is found the program creates a command *"vehicle.simple_goto(LocationGlobalRelative(16.5591343, 49.2365885, 20)"* thru Py-thon library Pymavlink and send it via the serial line directly to Pixhawk. UAV carries out the autonomous flight until the point with assigned change to flight mode *"Return to launch"* is reached [8],[9].



Figure 4: Real-Time Processing of Images and assigning an instruction set for the next flight.

4 **RESULTS**

Variable point distribution and diverse characteristics of the flight envelope of UAV was tested. For higher accuracy, it was necessary to enlarge the size of the QR codes on the ground. It would be beneficial to create a cascade of positive images directly from the flight environment. During a flight, different speed and altitude can be set or another additional devices on the UAV can be controlled; for example the drone can stop for accurate recognition and detection of the QR code. Also, a stop-over for automatic battery charging can be performed. Variable set up of the flight commands and point distribution supports autonomous flights that are based on the detection of the points acquired and processed directly on the UAV board.

Images with a size of 1920 x 1080 pixels captured by a camera with an angle of view of 78.8 $^{\circ}$ were used for image processing and detection. The size of the detected QR codes at an image distance of 15 m were about 100 x 100 pixels. At this camera resolution, it was possible to achieve a maximum processing speed of 5 frames per second due to limited hardware performance of the Raspberry Pi 4. The power consumption of Raspberry during processing can be up to 15 W. The average power

consumption for the flight is 660 W. For smaller drones with lower energy consumption, the consumption of Raspberry could significantly reduce the flight time. Flights tests shows that 15 m distance was the limit for successful detection when the drone approached the position determined by GPS coordinates. At higher distances, no detection occurred due to small QR code captured at detection frame or due to high camera shaking. The independent control of the drone was also tested during the flight over the monitored object. In this case, at a height of 10 m above the ground, the speed limit was 5 m/s for successful detection, at higher speeds there was no detection or there was a problem of capturing the nonblurred image.

5 CONCLUSIONS

Real time detection of known images during a flight allows to build autonomous system and possibility to set up the flightpath. This thesis tested the possibilities of using simple QR codes for autonomous flying. Limit values of height, distance or speed were found at which the successful detection of the required QR codes with defined hardware resources was ensured.

Distributed QR codes can be changed on the screens, therefore, the flightpath is created based on the requirements of the actual changes in the situation. The real time image detection for autonomous flightpath of the UAV can serve for packages delivery, mapping of large areas or for military missions and another safety service. The entire system allows complete real time autonomous flight from take-off to landing with various variability and number of detected waypoints.

The result is fully functioning autonomous drone controlled by the single-board computer with the minimal need for the operator thru the ground station. In the next steps, the robustness of the proposed system will be tested, a larger training set will be created. An accurate test measurement will also be developed to determine the success of the detection of the QR codes used and the percentage success of a set of classifiers at different altitudes and at different aircraft speeds.

ACKNOWLEDGEMENT

This paper was funded from the general student development project at Brno University of Technology.

REFERENCES

- [1] H. Y. Chao, Y. C. Cao, and Y. Q. Chen, "Autopilots for small unmanned aerial vehicles: A survey", International Journal of Control, Automation and Systems, vol. 8, pp. 36-44, 2010.
- [2] J. Janousek and P. Marcon, "Precision landing options in unmanned aerial vehicles", in 2018 International Interdisciplinary PhD Workshop (IIPhDW), 2018, pp. 58-60.
- [3] P. Marcon, J. Janousek, and R. Kadlec, "Vision-Based and Differential Global Positioning System to Ensure Precise Autonomous Landing of UAVs", in 2018 Progress in Electromagnetics Research Symposium (PIERS-Toyama), 2018, pp. 542-546.
- [4] Viola, P.; Jones, M. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR 2001, Vol. 1, pp. 511-518.
- [5] Forsyth, D.; Ponce, J. Computer Vision: A Modern Approach. Prentice Hall; 2 edition (2011).
- [6] D. Safadinho, J. Ramos, R. Ribeiro, V. Filipe, J. Barroso, and A. Pereira, "UAV Landing Using Computer Vision Techniques for Human Detection", Sensors, vol. 20, no. 3, 2020.
- [7] Synchronized control of group of helicopters using direct communication Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Structure-of-MAVLinkpacket-Source_fig2_283013996
- [8] M. Galimov, R. Fedorenko, and A. Klimchik, "UAV Positioning Mechanisms in Landing Stations: Classification and Engineering Design Review", Sensors, vol. 20, no. 13, 2020.
- [9] J. Janousek, P. Marcon, J. Pokorny, and J. Mikulka, "Detection and Tracking of Moving UAVs", in 2019 PhotonIcs & Electromagnetics *Research Symposium - Spring (PIERS-Spring)*, 2019, pp. 2759-2763.

DEMYSTIFYING BAND-LIMITED EXTRAPOLATION

Ondrej Mihálik

Doctoral Degree Programme (2), FEEC BUT E-mail: xmihal06@stud.feec.vutbr.cz

Supervised by: Petr Fiedler

E-mail: fiedlerp@feec.vutbr.cz

Abstract: Extrapolation of band-limited signals gained scientific attention over the last 60 years. The famous methods: Gerchberg-Papoulis algorithm, Prolate spheroidal wave functions (PSWFs), and sinc interpolation—they all promise excellent results. But when it comes to their practical implementation, users may find themselves struggling with many unanswered questions. Especially PSWFs became viewed as mysterious. They are hard to compute and even harder to apply. In theory they promise excellent extrapolation capabilities—something which is contrary to our intuition. This paradox is resolved if we admit that the real-world data contain noise. In this paper we review the abovementioned methods and try to provide a brief assessment of their capabilities by considering the effects of noise and the length of signal observation.

Keywords: Slepian sequences, extrapolation, band-limited signal, reconstruction, regularisation.

1 INTRODUCTION

The signals being processed in real life are always measured in a finite time interval. Band-limited signal extrapolation refers to the process of obtaining signal waveform outside this interval. For this purpose, we make use of the known samples and an information about the band-limit of the signal.

There had been several important milestones in the history of band-limited extrapolation. One of the first algorithms was the famous Gerchberg-Papoulis (GP) iterative process [1], which alternates between band-limiting and time-limiting operations. This was preceded by a series of papers from Slepian and Landau, beginning with the introduction of PSWFs [2], and followed by their discrete versions—Discrete prolate spheroidal sequences (DPSS's) [3]—which seem to be the ultimate tool for the signal recovery from its discrete samples. However, they received relatively small attention over the last decades. The majority of researchers drifted back towards the iterative methods, such as in the paper [4].

This paper aims to offer a good insight into the band-limited extrapolation by avoiding unnecessarily complicated notation. Hopefully, it provides a foothold for easy implementation, e.g., in MATLAB. The article is divided as follows. At first, in Section 2 we recall some of the fundamental facts concerning the band-limited extrapolation. We also discuss the differences between continuous and discrete formulations of the problem, as they have a major impact on the subsequent implementation of the reconstruction algorithm and its computational complexity. Section 3 is devoted to numerical experiments. We show examples which demonstrate how both the length of the observed data, and the level of noise affect the quality of the extrapolated signal. From this comparison it becomes clear that the excellent results, such as [5], are rather exaggerated—they do not translate into practical digital signal processing (DSP), since they were achieved in the absence of noise.

2 THEORY

A continuous signal f(t) is related to its spectrum $F(\omega)$ by the Fourier transform

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt,$$
 (1)

where t denotes time and ω stands for the angular frequency. Signals whose spectrum is of the form

$$F(\omega) = 0, \quad \text{for} \quad |\omega| > \Omega,$$
 (2)

are called band-limited, with band-limit Ω . According to the famous Shannon-Nyquist sampling theorem, any band-limited signal f(t) can be perfectly represented by its discrete samples for all values of the continuous time, t, using the formula

$$f(t) = \sum_{k=-\infty}^{\infty} f(kT_s) \frac{\sin \pi (t/T_s - k)}{\pi (t/T_s - k)}, \quad \text{for} \quad 0 < T_s < \frac{\pi}{\Omega}.$$
 (3)

Nonetheless, if we have only a finite number of samples, perfect reconstruction is no longer possible. The situation in which we sample more densely than with $T_s = \pi/\Omega$ is referred to as oversampling. As a result, the samples exhibit redundancy and contain hidden information invaluable for the band-limited extrapolation. Without oversampling, this form of extrapolation would not be possible [6].

2.1 CONTINUOUS VS. DISCRETE FORMULATION OF THE PROBLEM

We understand that a new researcher might be confused with the spectrum of proposed methods and formulations of the problem which occur in the literature. Some scientists are purely interested in the analysis of continuous problems—thought, it is hard to imagine how could these ideas be reliably employed in the practical DSP. Ignoring this, they devise methods to solve the continuous problem and discretize the resulting algorithms only at the point when it comes to its implementation. A common practice is to select a number of functions (e.g., complex exponentials, spherical Bessel functions, PSWFs, etc.) and use them for the interpolation of the known samples. Furthermore, these models require some additional variables—the number of functions used and their parameters, which are often selected by the rule of thumb.

Even our attention had been initially deflected from the discrete statement of the problem, since the continuous formulation is prevalent. The use of the aforementioned functions lacks its justification as their orthogonality does not hold for a finite number of their discrete samples. We believe that the most important step towards a successful implementation is the correct statement of the problem being solved. We start by showing that the problem we are dealing with becomes finite-dimensional naturally, and we implement the algorithm without the artificial selection of interpolating functions.

2.2 GERCHBERG-PAPOULIS ITERATIVE METHOD FOR RECONSTRUCTION

The paper [1] in which Papoulis analyses Gerchberg's algorithm defines it using time-limiting and band-limiting operators, D and B, respectively. To demonstrate how the problem reduces to finite-dimensional, we rewrite the algorithm explicitly. The iteration starts with the discrete signal which is represented by the sequence of Dirac pulses

$$g_0(t) = \sum_{k=-K}^{K} f(kT_{\rm s})\delta(t - kT_{\rm s}).$$
(4)

(It is expected here that N = 2K + 1 samples are known.) The band-limiting operation, *B*, smooths Dirac pulses into sinc functions. Hence, after the first iteration the signal is of the form

$$g_1(t) = \sum_{k=-K}^{K} f(kT_s) \frac{\sin \Omega(t - kT_s)}{\pi(t - kT_s)}.$$
 (5)

Then further the iteration follows

$$g_m(t) = g_{m-1}(t) + \sum_{k=-K}^{K} [f(kT_s) - g_{m-1}(kT_s)] \frac{\sin\Omega(t - kT_s)}{\pi(t - kT_s)}.$$
 (6)

At each iteration, a difference is computed from the known samples of the original signal and its approximation. Then a band-limiting operation is applied to this differential signal—this all is represented by the sum on the right-hand side of (6). It becomes clear that, by induction, each GP iteration is of the form

$$g_m(t) = \sum_{k=-K}^{K} b_m(k) \frac{\sin \Omega(t - kT_s)}{\pi(t - kT_s)}.$$
 (7)

This means that the amplitudes of N sinc functions change, but their positions in time do not. This is vital for practical implementation, because the problem which originally seemed to be infinite-dimensional is indeed of finite dimensions. We were originally seeking a continuous signal which extrapolates our data, but now we only have to find N numbers stored in the vector

$$\vec{b} = [b(-K) \ b(1-K) \ b(2-K) \ \cdots \ b(K)]^{\mathrm{T}},$$
 (8)

where the 'T' stands for the matrix transpose. As the iteration number goes to infinity, $m \to \infty$, we end up with the amplitudes, \vec{b} , which satisfy

$$\vec{g}_0 = \mathbf{S}\vec{b}.\tag{9}$$

Here $\vec{g}_0 = [f(-KT_s) \ f(T_s - KT_s) \ f(2T_s - KT_s) \ \cdots \ f(KT_s)]^T$ denotes the vector of known signal samples, and the elements of the matrix **S**, in *k*th row and *l*th column, are given by

$$s_{k,l} = \frac{\sin \Omega(k-l)T_{\rm s}}{\pi(k-l)T_{\rm s}}, \qquad k,l = -K, 1-K, 2-K, \dots, K.$$
(10)

The iterative algorithm requires large computational efforts—as we get closer to the solution, the rate of convergence decreases. As demonstrated in [7], the algorithm is in fact nothing more than the method of steepest descend, which is notorious for its poor convergence due to "zig-zagging" when minimizing the error functional. The problem may be solved via accelerated GP algorithm [8], or by matrix inversion.

2.3 RECONSTRUCTION BY INVERSION OF SINC MATRIX

Matrix inversion is the direct approach of solving (9). The calculations become numerically instable as the number of samples, N, increase. To make the inverse feasible, application of Tikhonov regularization (ridge regression) [9] is vital. Note that our approach is less extensive than the typical implementation of one-step GP algorithm [10] as we do not use complex exponentials. By induction in Section 2.2 we effectively avoided the computations which would otherwise involve complex numbers.

2.4 RECONSTRUCTION USING SLEPIAN SEQUENCES

We saw that the two previous methods solve a matrix inversion. If we have many samples, the matrix is large. Therefore, both methods pose a large computational burden. A method of solving the inversion which avoids extensive computations relies on DPSS's—also referred to as Slepian sequences [3]. They are defined as the eigenvectors of the above-mentioned matrix **S** containing sinc functions. Having these sequences precomputed in the memory, we can use them to recover the extrapolated signal needing only two matrix multiplications—one to obtain the vector of their coefficients, \vec{a} , and a second one to use them for the evaluation of the extrapolated signal at desired time instants. This also simplifies the application of Tikhonov regularisation. Having limited space, we refer the interested reader to our previous work [11], which deals with the details of this approach.

2.5 **REGULARISATION**

If a form of noise is present, all three above-mentioned methods should avoid perfect interpolation in order to achieve a stable extrapolation. In some scientific field this would be referred to as avoiding the "overfitting" or the "blow-up problem". To accomplish regularisation, we may use the noise variance σ^2 to stop the GP process once the quadratic error evaluated over the known samples drops to

$$\left(\vec{g}_0 - \mathbf{S}\vec{b}\right)^{\mathrm{T}} \cdot \left(\vec{g}_0 - \mathbf{S}\vec{b}\right) \le K\sigma^2.$$
⁽¹¹⁾

This is known as the Morozov discrepancy principle [7]. It may be used for Tikhonov regularisation of the matrix inversion—or of the method using DPSS's—and we do so in the following section.

3 NUMERICAL EXPERIMENT

We already demonstrated the importance of band-limit in paper [11]; now we want to focus on the effects of different noise levels and lengths of the observation interval. The signal (12) was selected in accordance with the paper of Devasia [5], see Figure 1 (blue).

$$f(t) = \sin(\pi t)\cos(3\pi t)e^{-2t^2+3t} + 0.5[\sin(5\pi t) - \cos(7\pi t)]$$
(12)

In a strictly mathematical sense, the signal (12) is not band-limited; it contains two modulated Gaussian functions. It is, however, *essentially* band-limited for $\Omega = 8\pi$ rad/s. The signal was sampled with sampling period $T_s = 0.1$ s in the intervals denoted by the dashed black lines in Figure 1. Gaussian noise with standard deviation σ of values 10^{-2} , 10^{-4} and 10^{-6} was added during the sampling. This simulates various sources of uncertainty, such as the errors induced by the A/D converters. Selected values roughly cover the most widely used converters—the range between 6 bits and 20 bits.



Figure 1: The signal and its reconstruction at different noise levels and lengths of the observed interval (intervals are denoted by the black dashed lines).

Figure 1 shows something that is not typically seen in the literature. If a signal is too noisy, observing it over a longer period of time provides negligible enhancement of its extrapolation—all extrapolations constructed at $\sigma = 0.1$ (red) are essentially the same if we speak about their right-hand side.

The situation becomes somewhat more favourable as we decrease the noise level. All three figures show improvements for $\sigma = 10^{-4}$ (yellow), and a noticeable dependence on the length of observed data occurs when σ decreases to 10^{-6} (magenta). In such a case, longer time of observation leads to better extrapolability. The extrapolated part of a signal may be conveniently used, e.g., for evaluation of the signal derivatives at the edge of the observation interval. Normally, they would have to be computed using the less accurate finite differences of the last few samples.

4 CONCLUSION

We have briefly summarized the most common implementations of band-limited signal extrapolation. We alluded to the continuous formulation of the extrapolation with its disadvantages and focused mainly on the discrete formulation. We conclude that it is more suitable for implementation in the DSP. To avoid the algorithm's susceptibility to noise, we have applied Tikhonov regularization with its parameter selected according to Morozov's discrepancy principle.

Using a numerical example we demonstrated the results originally predicted by Landau [6]. The most important conclusion is the inevitable fact that if we are extrapolating signal from its noisy samples, then there is a limit to which we can improve the extrapolation of a signal by increasing the length of its observation. For higher noise levels, the size of the matrix to be inverted can be significantly reduced. This is safely achieved by dropping off the past samples and incorporating only those which are in a reasonable proximity of the point from which we are aiming to extrapolate. This is vital for the optimization of the extrapolation algorithm, should it run in a real time application.

ACKNOWLEDGEMENT

We wish to thank Pavel Jura for very helpful suggestions.

The completion of this paper was made possible by the grant No. FEKT-S-20-6205—"Research in Automation, Cybernetics and Artificial Intelligence within Industry 4.0" financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- A. Papoulis, 'A new algorithm in spectral analysis and band-limited extrapolation', *IEEE Transactions on Circuits and Systems*, vol. 22, no. 9, pp. 735–742, Sep. 1975, doi: 10.1109/TCS.1975.1084118.
- [2] D. Slepian and H. O. Pollak, 'Prolate spheroidal wave functions, fourier analysis and uncertainty — I', *The Bell System Technical Journal*, vol. 40, no. 1, pp. 43–63, Jan. 1961, doi: 10.1002/j.1538-7305.1961.tb03976.x.
- [3] D. Slepian, 'Prolate spheroidal wave functions, fourier analysis, and uncertainty V: the discrete case', *The Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, May 1978, doi: 10.1002/j.1538-7305.1978.tb02104.x.
- [4] M. Chen, G. Qu, L. Shi, and S. Gao, 'The reweighted Landweber scheme for the extrapolation of band-limited signals', *Signal Process.*, vol. 164, pp. 340–344, Nov. 2019, doi: 10.1016/j.sigpro.2019.06.022.
- [5] A. Devasia and M. Cada, 'High Precision Numerical Implementation of Bandlimited Signal Extrapolation Using Prolate Spheroidal Wave Functions', *IAENG International Journal of Computer Science*, vol. 40, no. 4, pp. 291–300, 2014, doi: 10.1007/978-94-017-9115-1 30.
- [6] H. Landau, 'Extrapolating a band-limited function from its samples taken in a finite interval', *IEEE Transactions on Information Theory*, vol. 32, no. 4, pp. 464–470, Jul. 1986, doi: 10.1109/TIT.1986.1057205.
- [7] G. Villiers, *The Limits of Resolution*. Boca Raton: CRC Press, 2016.
- [8] B. G. Salomon and H. Ur, 'Accelerated iterative band-limited extrapolation algorithms', *IEEE Signal Processing Letters*, vol. 11, no. 11, pp. 871–874, Nov. 2004, doi: 10.1109/LSP.2004.836950.
- [9] J. N. FRANKLIN, 'On Tikhonov's method for ill-posed problems', 889–907, vol. 28, no. 128, 1974, doi: 10.1090/S0025-5718-1974-0375817-5.
- [10] M. Sabri and W. Steenaart, 'An approach to band-limited signal extrapolation: The extrapolation matrix', *IEEE Transactions on Circuits and Systems*, vol. 25, no. 2, pp. 74–78, Feb. 1978, doi: 10.1109/TCS.1978.1084442.
- [11] O. Mihálik, D. Michalík, P. Jura, and P. Fiedler, 'Spectral Super-resolution and Band-limited Extrapolation Using Slepian Series', *IFAC-PapersOnLine*, vol. 52, no. 27, pp. 388–393, Jan. 2019, doi: 10.1016/j.ifacol.2019.12.691.

NEURAL NETWORKS WITH DILATED CONVOLUTIONS FOR SOUND EVENT RECOGNITION

Stepan Miklanek

Doctoral Degree Programme (2), FEEC BUT E-mail: xmikla12@stud.feec.vutbr.cz

Supervised by: Jiri Schimmel E-mail: schimmel@feec.vutbr.cz

Abstract: Convolutional neural networks, most commonly deployed in image classification tasks, typically use square-shaped convolutional kernels, which are well suited for feature extraction from two-dimensional data. This study explores the effect of utilizing spectrally aware dilated convolutions specialized for sound event recognition. By extending the base kernels in the time or the frequency dimension, the features extracted from the spectral audio representations should, in theory, better capture the temporal and timbral information of different sound events. The baseline neural network model with squared kernels was compared against three models, which used an increasing dilation factor in the subsequent convolutional layers. The three models were purposefully tuned to focus towards the frequency and time feature extraction. The results have shown that the models with dilated convolutions performed noticeably better in comparison with the baseline model.

Keywords: sound event recognition; convolutional neural networks; dilated convolution

1 INTRODUCTION

In recent years, numerous studies presented exceptional results of utilizing convolutional neural networks (CNNs) in image classification or object detection tasks. The neural network models specifically tailored for these problems can also be used when dealing with audio [1]. The main drawback of using these models is that they are usually overly complex for tasks such as sound event recognition or music auto-tagging. The CNNs used in image processing are designed to extract spatial features of various objects present in images. Analogically, the majority of audio-motivated CNN architectures are using spectral transformations, which can also be thought of as image-like representations [2]. In image processing, squared convolutional filters of 3×3 or 12×12 are widely used [3]. Note that the audio processing filter dimensions do not correspond to spatial information. The spectrogrambased filter dimensions, in fact, correspond to time and frequency. In the audio realm, wider filters are capable of learning longer temporal dependencies, while filters with larger height are capable of learning timbral features as shown in [4]. The larger frequency and time receptive fields also result in more complex models as the convolutional kernels get bigger. The main focus of this paper is to build lightweight classification models that make use of dilated convolutions to capture spectral properties and dependencies of varying sound events. The second question is, whether the models with spectrally aware layers can outperform the baseline model with common squared kernels.

1.1 RELATED WORK

The CNNs are well established tool for sound event recognition. The results presented in the previous papers show that various neural network architectures can perform well on different sound event datasets [4, 5, 6]. Usually, these neural network models have up to tens of millions of parameters. Conversely, all the models presented in this study have less than 100k parameters and achieve similar classification accuracy.

2 METHODS

2.1 DILATED CONVOLUTIONS

Dilated convolutional kernels were used instead of rectangular or one-dimensional kernels proposed in earlier literature [4]. The basic square-shaped kernel can be dilated in the time or the frequency axis by inserting zeros between the kernel values as shown in Figure 1.



Figure 1: Frequency and time dilated convolutional kernels.

The main advantage of the dilated convolutions is that the number of trainable weights remains constant when increasing the dilation factor. Larger dilation factor results in extended receptive field in the selected spectrogram dimension. This should adapt the conventional convolutional neural networks towards the timbral and temporal feature extraction needed for accurate sound event recognition.

2.2 DATASET

The *UrbanSound8K* dataset was chosen for the model performance comparison [6]. This dataset contains 8732 labeled sound excerpts, most of which are 4 seconds long. The dataset includes manual annotations of 10 low-level classes—air conditioner, car horn, children playing, dog bark, drilling, engine idling, gun shot, jackhammer, siren and street music. The number of audio clips is evenly distributed across all the categories, except for the car horn and the gun shot class.

All the neural network models had input layers that accepted 88200 samples of audio, which is equal to 4 seconds at a sampling rate of 22050 Hz. In the case of shorter excerpts, the rest of the audio arrays was padded with zeros to match the required input size of the models. The dataset was divided into three subsets, where 80% of the data was used for training and the remaining 20% was split in half for validation and testing. The partitioning was done in pseudo-random fashion so that the classification accuracy of all the models could be directly compared.

2.3 NEURAL NETWORK MODELS

The raw input waveforms were transformed to mel-spectrogram representations by a custom layer from the Kapre module for Python [7]. The mel-spectrograms were normalized across the last dimension by a batch normalization layer. Each model consisted of four convolutional layers with 3×3 kernels activated by a rectified linear (ReLU) function, followed by a dropout layers to prevent overfitting. The max pooling layers were used to downsample the results of the convolutions. The number of convolutional filters was set to 4, 8, 16, and 32 for the last layer. The feature maps of the last convolutional layers were flattened to vectors and followed by a final dropout layer. The feature vectors were fed into fully connected layer with 100 neurons. At the end of each model, there was a second fully connected layer with 10 neurons activated by a softmax function to get the final predictions. The structure of the models is shown in Table 1.

The models were built so that they had the same number of layers and roughly equal amount of trainable variables. The baseline Conv2D model had square-shaped kernels in every convolutional layer.

	Conv2	2D	ConvI	DF	ConvI	DT	ConvE)B
Layer Type	Shape	Dilation	Shape	Dilation	Shape	Dilation	Shape	Dilation
Input Waveform	(88200, 1)		(88200, 1)		(88200, 1)		(88200, 1)	
Mel-spectrogram	(345, 120, 1)		(345, 120, 1)		(345, 120, 1)		(345, 120, 1)	
Batch Normalization	(345, 120, 1)		(345, 120, 1)		(345, 120, 1)		(345, 120, 1)	
Convolution	(343, 118, 4)	(1, 1)	(343, 118, 4)	(1, 1)	(343, 118, 4)	(1, 1)	(343, 118, 4)	(1, 1)
Dropout	(343, 118, 4)		(343, 118, 4)		(343, 118, 4)		(343, 118, 4)	
Max Pooling	(171, 59, 4)		(171, 59, 4)		(171, 59, 4)		(171, 118, 4)	
Convolution	(169, 57, 8)	(1, 1)	(169, 55, 8)	(1, 2)	(167, 57, 8)	(2, 1)	(167, 114, 8)	(2, 2)
Dropout	(169, 57, 8)		(169, 55, 8)		(167, 57, 8)		(167, 114, 8)	
Max Pooling	(84, 28, 8)		(84, 27, 8)		(83, 28, 8)		(83, 57, 8)	
Convolution	(82, 26, 16)	(1, 1)	(82, 19, 16)	(1, 4)	(75, 26, 16)	(4, 1)	(75, 49, 16)	(4, 4)
Dropout	(82, 26, 16)		(82, 19, 16)		(75, 26, 16)		(75, 49, 16)	
Max Pooling	(41, 13, 16)		(41, 19, 16)		(37, 13, 16)		(37, 24, 16)	
Convolution	(39, 11, 32)	(1, 1)	(39, 3, 32)	(1, 8)	(21, 11, 32)	(8, 1)	(21, 8, 32)	(8, 8)
Dropout	(39, 11, 32)		(39, 3, 32)		(21, 11, 32)		(21, 8, 32)	
Max Pooling	(9, 3, 32)		(9, 3, 32)		(5, 5, 32)		(7, 4, 32)	
Flatten	864		864		800		896	
Dropout	864		864		800		896	
Fully Connected	100		100		100		100	
Fully Connected	10		10		10		10	
Trainable Parameters	93,656		93,656		87,258		96,856	

Table 1: Neural network architectures.

The ConvDF and ConvDT models used dilated convolutions in the frequency and the time dimension, respectively. The ConvDB utilized dilated convolutions in both spectrogram dimensions. The dilated convolutions were described in more detail in Section 2.1. The exact dilation patterns are present in Table 1.

2.4 EVALUATION METRICS

To evaluate the performance of each model, following objective metrics were chosen. The accuracy (ACC) is defined as the fraction of correct predictions over *n* samples. Precision and recall metrics were also used. Precision is the ability of the classifier not to label as positive a sample that is negative, and recall is the ability of the classifier to find all the positive samples. Another used metric is the *F-score*, sometimes also called the *F-measure*. This metric can be defined as a weighted harmonic mean of the precision and recall. The binary metrics can be extended to multi-class problems by treating the data as a collection of binary problems, one for each class. In order to compute the multi-class metrics, macro-, or micro-averaging must be used. Macro-averaging is computed as a mean of the binary metrics, given equal weight to each class. On the other hand, micro-averaging gives each sample-class pair an equal contribution to the overall metric.

3 RESULTS

Note that the purpose of this study was not to create a perfect classifier, but rather investigate on the advantages of using dilated convolutions. Although related works already used dilated convolutions for audio classification [8], the models proposed in this study are substantially less complex with comparable results.

3.1 TRAINING

The models were trained by minimizing the categorical cross-entropy loss, which is the most common choice when dealing with multi-class problems. The training process ran for 200 epochs on the same training and validation sets. At the end of every epoch, the training set was randomly shuffled.

It has been found that the models with dilated convolutions outperform the baseline Conv2D model by a significant margin. The ConvDF model with dilations in the frequency dimension had the second

highest validation loss and the second lowest validation accuracy. The ConvDB model had the lowest validation loss and highest accuracy, making it the best performing out of the four trained models. The ConvDT was the intermediate model of the models with dilated convolutions.

3.2 OBJECTIVE RESULTS

The trained models with the lowest validation loss were further validated on a separate subset of 874 audio examples that were not used in the process of training. This was done to check if the models are good at generalizing of the prediction on unseen data. The metrics presented in Section 2.4 were used to compare the models. The evaluation metrics computed on a test set are shown in Table 2.

Model	ACC	Pmacro	R _{macro}	<i>F</i> -score _{macro}
Conv2D	0.847	0.874	0.858	0.857
ConvDF	0.882	0.897	0.889	0.888
ConvDT	0.891	0.906	0.897	0.896
ConvDB	0.914	0.921	0.917	0.917

 Table 2: Performance of the models on a test set.

The results presented in the previous section were confirmed by computing the test ACC and additional classification metrics. The models with dilated convolutions performed better than the baseline model across all the computed metrics. The ACC can be misleading when dealing with unbalanced datasets. Fortunately, this was not the case because the *UrbanSound8K* is a fairly balanced dataset. Nevertheless, the results were also backed by computing the *P*, *R*, and *F*-score, which verified the proposals made in the previous sections of the paper. The fact that the additional metrics did not significantly deviate from the ACC values means that all three models generalized well. The test ACC of the ConvDB model was 6.7% higher than the baseline Conv2D model. The ConvDT model performed slightly worse with ACC 2.3% lower than the ConvDB model. The ConvDF model had the worst performance out of the models with dilated convolutions. Still, its ACC was 3.5% higher than the baseline model. The normalized confusion matrix of the best performing model is shown in Figure 2.



Figure 2: Normalized confusion matrix of the ConvDB model.

According to the confusion matrix, the car horn, the dog bark, and the street music classes were the hardest to predict. The confusions were probably caused by the prevailing background noise, which is present in some of the sound excerpts. Furthermore, as stated in the Section 2.2, numerous sound samples were shorter than the 4 second input of the neural network models. This could also contribute to the number of false predictions made by the models.

4 CONCLUSION

This study presented relevant architectural choices of building deep learning models for sound event recognition. The design strategy of utilizing spectrally aware layers for audio classification purposes can further improve the results obtained with the common feature extraction techniques used in image processing. The deep learning approaches are often criticized for the difficulty in understanding the hidden relationships that neural networks are learning. Many authors are using unnecessary brute force methods without rationalizing the elementary principles behind the data they are trying to classify. Of course, this work is not the universal answer for successful audio classification, but at least it tries to pinpoint the possible future directions. The proposed methods are yet to be further validated on other datasets. Also, the presented neural network models can be further modified and improved.

REFERENCES

- [1] PONS, Jordi, Oriol NIETO, Matthew PROCKUP, Erik SCHMIDT, Andreas EHMANN and Xavier SERRA. End-to-end learning for music audio tagging at scale. In: Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR2018). Paris, France, 2018, pp. 637–643. ISBN 978-2-9540351-2-3.
- [2] CHOI, Keunwoo, George FAZEKAS and Mark SANDLER. Automatic tagging using deep convolutional neural networks. In: Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR2016). New York, USA, 2016, pp. 805–825. ISBN 978-0-692-75506-8.
- [3] HE, Kaiming, Xiangyu ZHANG, Shaoqing REN and Jian SUN. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [online]. Las Vegas, USA, 2016, pp. 770–778 [retrieved 2021-03-08]. ISBN 978-1-4673-8851-1. Available: doi:10.1109/CVPR.2016.90
- [4] PONS, Jordi. *Deep Neural Networks for Music and Audio Tagging*. Barcelona, Spain, 2019. Dissertation. Universitat Pompeu Fabra. Supervised by Xavier Serra.
- [5] LUZ, Jederson S., Myllena C. OLIVIERA, Flavio H.D. ARAUJO and Deborah M.V. MAG-ALHAES. Ensemble of handcrafted and deep features for urban sound classification. *Applied Acoustics*. 2021, **2021**(175).
- [6] SALAMON, Justin, Christopher JACOBY and Juan Pablo BELLO. A Dataset and Taxonomy for Urban Sound Research. In: 22nd ACM International Conference on Multimedia. Orlando, USA, 2014.
- [7] CHOI, Keunwoo, Deokjin JOO and Juho KIM. Kapre: On-GPU Audio Preprocessing Layers for a Quick Implementation of Deep Neural Network Models with Keras. In: Machine Learning for Music Discovery Workshop at 34th International Conference on Machine Learning. Sydney, Australia, 2017.
- [8] CHEN, Yan, Qian GUO, Xinyan LIANG, Jiang WANG and Yuhua QIAN. Environmental sound classification with dilated convolutions. *Applied Acoustics*. 2019, **2019**(148), 123–132.

MULTI-CLASS WEATHER CLASSIFICATION FROM SINGLE IMAGES WITH CONVOLUTIONAL NEURAL NETWORKS ON EMBEDDED HARDWARE

Tomáš Bravenec

Doctoral Degree Programme (2), FEEC BUT E-mail: xbrave01@stud.feec.vutbr.cz

> Supervised by: Tomáš Frýza E-mail: fryza@vut.cz

Abstract: The paper is focused on creating a lightweight machine learning solution for classification of weather conditions from input images, that can process the input data in real time on embedded devices. The approach to the classification uses deep convolutional neural networks architecture with focus on lightweight design and fast inference, while providing high accuracy results. The focus on creating lightweight convolutional neural network architecture capable of classification of weather conditions also enables usage of the network in real time applications at the edge.

Keywords: deep learning, neural networks, computer vision, weather classification, machine learning, parallel computing, inference on edge, reduced precision computing

1 INTRODUCTION

The correct weather analysis is crucial for autonomous machines, as various weather conditions can disrupt the proper function of such machines. Current systems use mostly a sensor arrays or cameras to predict the weather conditions the machine is located in. This issue does not affect only autonomous machines but transportation infrastructure too. Also various traffic accidents could be prevented by better adjustment of driving to the current weather conditions with either not allowing to start autonomous driving system in cars, auto pilot in airplanes etc, requiring full attention from the driver or pilot while the autonomous systems are active or using different constraints for autonomous movement. As the weather can be quite diverse, the approach using classic computer vision techniques is not the best approach here, and the usage of deep neural networks is more suitable for learning the weather patterns directly out of training data.

As weather is not generally changing too quickly and latency is not an issue, it is possible to offload the computation from the device to servers in the cloud. There are situations the use of cloud computing is not an option, either due to lack of connectivity or privacy and security reasons. In situations like these, the inference has to be done on the same or closely connected device that captured the data. This process of doing the computation, usually on embedded devices close to the sensors acquiring the data is called a computing at the edge. For the usage of neural networks on edge devices, the neural networks are trained on a wide training set of data representing all the required situations either using a powerful server or desktop computer, due to the compute limitations of the embedded devices, this approach saves a lot of time as training neural networks on embedded hardware would be extremely time consuming and ineffective.

A lot of the related works use a binary classification, which means those solutions are capable of classifying only two weather conditions [1,2], being either sunny or cloudy, sunny or raining etc. There are also works doing multi-class weather classification [3,4] which mostly work with 3 classification labels, adding either snowing or foggy conditions into the classification.



Figure 1: Example of images used for extension of the dataset containing rainy and snowy conditions

2 DATASET

The dataset used for training and testing the designed neural network architecture was created by modifying existing multi-class weather dataset [5], which contains 1530 images, with another 1189 images. The original dataset contained 5 weather conditions, each represented by 200 to 300 images. The weather conditions represented in the original dataset are cloudy, rainy, foggy, clear weather, sunrise or sunset.

The author's modification of the dataset was necessary as the original dataset did not include any images including snowy conditions, so the dataset was extended with a new class with 619 training images containing snowy weather. Example of snowy and rainy condition images used for the dataset expansion are shown in Figure 1. To make the dataset more robust, the original classes were also extended with about 100 to 200 additional images under Creative Commons licence scraped from the internet.

3 IMPLEMENTATION

Due to the requirement of real time processing on embedded devices, the neural network was designed and trained from scratch with embedded devices in mind. This means the deep neural network uses relatively low amount of layers with small amount of trainable parameters. Bigger networks with more parameters and layers were also tested but the result improvement at at most 1 % was not worth the increase in size and complexity of the network. The architecture of the designed network is in Table 1.

The training of the network was done using Quantization Aware Training (QAT) method, which adds another constraint to the weights of the network, to allow them to be converted from the standard 32-bit floating point format into 8-bit fixed point precision format without significant loss in accuracy of the network.

For the training itself the framework for deep learning applications TensorFlow 2 [6] was used, as it allows conversion of the trained model into TensorFlow Lite model. It is a highly optimized version of TensorFlow specifically designed to be ran in mobile applications and on embedded devices with low compute performance.

4 EVALUATION AND TESTING

The testing of the designed convolutional neural architecture was done on a single board computer NVIDIA Jetson Nano [7]. The 4-core CPU of NVIDIA Jetson Nano are built on an Arm A57 architecture and they are clocked at frequency of 1.5 GHz. The NVIDIA Jetson series of single board computers also contain desktop grade GPU cores even though in smaller numbers than their desktop counterparts. The GPU is based on NVIDIA Maxwell architecture, and contains 128 CUDA cores running at frequency of 0.9 GHz. The NVIDIA Jetson Nano contains 4 GB of DDR4 memory, which

Layer	Input Shape	Output shape	Activation	Parameters
Convolutional 2D	224x224x3	222x222x16	ReLU	448
Max pooling	222x222x16	111x111x16		0
Convolutional 2D	111x111x16	109x109x8	ReLU	1160
Max pooling	109x109x8	54x54x8		0
Convolutional 2D	54x54x8	52x52x8	ReLU	584
Max pooling	52x52x8	26x26x8		0
Convolutional 2D	26x26x8	24x24x8	ReLU	584
Convolutional 2D	24x24x8	22x22x4	ReLU	292
Flatten	22x22x4	484x1		0
Fully connected	484x1	40x1	ReLU	19400
Fully connected	40x1	40x1	ReLU	1640
Fully connected	40x1	40x1	ReLU	1640
Fully connected	40x1	6x1	SoftMax	246

Table 1: Designed convolutional neural network architecture



Figure 2: Comparison of data type influence on the size of weights

is dynamically assigned to either the CPU or the GPU. The whole system is relatively low power and consumes at most 10 W which makes it perfect compute system for applied computer vision projects.

Since the NVIDIA Jetson Nano has desktop class NVIDIA CUDA cores usable for general purpose computing and parallel data processing acceleration, but TensorFlow lite does not support it, the quantized model could not be tested on the Jetson Nano's GPU. The model using standard 32-bit floating point arithmetic was tested both using the CPU and the GPU.

When it comes to the size of the weights of the designed neural network, the usage of 8-bit quantized integer weights made the model a lot smaller than while using the standard 32-bit floating point data type, to be exact $3.6 \times$ smaller, which is expected as almost every weight is quarter of the original size. The comparison of the weights size using reduced precision data type is in Figure 2 while the comparison of accuracy as a whole and inference performance using standard 32-bit floating point and quantized integer weights is in Figure 3.

4.1 STANDARD 32-BIT ARITHMETIC

From the accuracy point of view at the designed network, the results are quite reasonable for the small size and efficiency of the network. The correct weather conditions with the network using 32-bit



Figure 3: Comparison of data type influence on the inference performance

floating point arithmetic were predicted correctly in 81.67 % of all cases. The reason for almost 20 % incorrect classifications is the variability of weather. There can be clouds in the sky even on a sunny day, or snowflakes can be mistaken for rain drops and vice versa. The confusion matrix showing the ground truth and predictions of the designed network are in Figure 4 a).

The inference speed of the standard model was not great at only 5.7 frames per second using CPU, and 7.2 frames per second using GPU. The relatively small increase in performance using the GPU acceleration is due to the small size of the designed neural network. In this network, memory transfers (copy of the input data in and out of the memory space assigned to the GPU) take majority of the time required for the processing which in the end does not give that big performance improvement.

4.2 QUANTIZED 8-BIT ARITHMETIC

The network using quantized 8-bit arithmetic, provides very similar results in terms of accuracy to the non quantized network at 80.83 % of all predictions being correct, which is just 0.84 % worse than using the non-quantized model. The confusion matrix of the quantized network is in Figure 4 b).

The performance gained by converting the model using TensorFlow Lite to use quantized integer arithmetic provided massive bump in inference performance, by managing to get frame rate of 71.2 frames per second which is a lot higher than required for real-time processing.

5 CONCLUSION

The main focus of this paper was to design a system for weather classification, that would be simple and lightweight enough to run on an embedded hardware. For this the training dataset was expanded to contain almost double the amount of images. The designed neural network created using Tensor-Flow is small and provides accurate results. While using traditional 32-bit floating point arithmetic, the performance of the designed neural network was not great, even while using the integrated GPU of the NVIDIA Jetson Nano, managing at most 7.5 frames per second. The usage of weights quantization and fixed point arithmetic with TensorFlow Lite provided a lot better results as the neural network managed to achieve better than real-time inference performance at 71.2 frames per second. The designed neural network could possibly be expanded in the future to be capable of classification of more weather conditions, like thunderstorm, sandstorm, hurricanes and more.



Figure 4: Confusion matrix of weather classification CNN: a) with 32-bit floating point arithmetic, b) with 8-bit quantized integer arithmetic

ACKNOWLEDGEMENT

The research published in this submission was financially supported by the Brno University of Technology Internal Grant Agency under project no. FEKT-S-20-6325.

REFERENCES

- M. Elhoseiny, S. Huang, and A. Elgammal, "Weather classification with deep convolutional neural networks," in 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 3349–3353.
- [2] C. Lu, D. Lin, J. Jia, and C. Tang, "Two-class weather classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2510–2524, 2017.
- [3] Z. Zhang and H. Ma, "Multi-class weather classification on single images," in 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 4396–4400.
- [4] S. Jabeen, A. Malkana, A. Farooq, and U. Ghani Khan, "Weather classification on roads for drivers assistance using deep transferred features," in 2019 International Conference on Frontiers of Information Technology (FIT), 2019, pp. 221–2215.
- [5] V. Gupta, "Weather dataset," https://www.kaggle.com/vijaygiitk/multiclass-weather-dataset, 2020, [Online; accessed 15-February-2021].
- [6] 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: https://www.tensorflow.org/
- [7] Embedded Linux Wiki , "Jetson," https://elinux.org/index.php?title=Jetson&oldid=543051, 2021, [Online; accessed 29-January-2021].

RIGHT CONVOLUTIONAL NEURAL NETWORK FOR CLASSIFICATION ILLUSTRATIONS IN ARTWORKS

Pavel Sikora

Doctoral Degree Programme (third year), FEEC BUT E-mail: xsikor14@stud.feec.vutbr.cz

> Supervised by: Kamil Riha E-mail: rihak@feec.vutbr.cz

Abstract: This paper deals with the image classification problem in the field of artworks. The article uses a custom dataset from artworks with eight classes of some not common objects and illustrations. This dataset is used to train three convolutional neural networks for classification. All classification results are well discussed and evaluated with an example on the images from a dataset.

Keywords: artwork, convolutional neural network, deep learning, image classification, keras, machine learning

1 INTRODUCTION

Nowadays in the world is experiencing huge benefits of neural networks, fall into today's so-called category AI (Artificial Intelligence). Since 2014 very deep convolutional networks have been used more widely [1]. Nowadays AI is used in many industries from medicine to automatic robots in halls to automatizing tasks in mobile phones, which is used by the vast majority of people. Thanks to the rapid trend of increasing the performance of computer technology and GPUs, researchers can create more complex and deeper networks. Thanks to that improvements, AI technology is more accurate and can be used in the next industries. There are also uncharted ends of this technology, one of them is works of art. [1, 2, 3]

This paper focuses on artworks, where can be applied the classification technique. The main interest is objects or illustrations that do not occur in commonly available datasets. For that, is used custom dataset which is made from artworks created by the Vasulka's couple [4].

2 METHODOLOGY

Today's classification tasks greatly solving CNN (Convolutional Neural Network). As the name suggests, CNN is DNN (Deep Neural Network) which among others uses convolutional layers. Nowadays exist a lot of image classification DNN models. In the end, three convolutional networks were chosen, namely VGG [3], DenseNet (Dense convolutional Network) [2] and Inception [1]. All these models handle very well with image classification problematics.

A F1-score is used to evaluate all results. This metric is calculated from precision and recall. The calculation formula is [5]:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{1}$$

2.1 DATASET PREPARATION

This article focuses on the artwork dataset contains these objects and illustrations: air, fire, image processing keying, machine vision (fisheye), woody, stripes, interior, landscape. The dataset



Figure 1: Two images from a dataset [4]

contains images with the sky, for which is a dedicated class *air*. Artworks sometimes contain scenes depicting the fire, for that is class *fire*. *Image processing keying* is on images where technique keying is applied. *Machine vision (fisheye)* occurred in images that are recorded uses the fisheye lens. Class *woody* suggests images where is Woody, one of the images dataset authors. *Stripes* represent specific lines. An example of *stripes* is on the left side of the figure 1. The second image in figure 1 shows the class *interior*, which classifies images that are captured in interiors. Images where occur landscapes are classified with *landscape* class.

At the beginning of the whole process, the dataset must be prepared into a suitable form for training. Dataset has been separated into folders, where folder names are class names that occur in images in that folder. Although the model is trained with a multi-label technique. Thanks to this technique trained model can classify all trained classes in one image, not only one class per image. This hierarchy is loaded to memory, shuffled, and saved to *CSV file*. Next, this file is used to create a data generator in Keras.

All used CNNs have been modified to train for eight classes. For that classification parts of every CNN model were replaced. Firstly with 2D global average pooling and dropout layer. Dropout layer use 40 % rate. Directly behind these layers, eight dense layers were added. Every dense layer is responsible for evaluating one of the eight classes [6]. All these dense layers use the sigmoid activation function. Binary cross-entropy is used to compute the loss function. The training process uses optimizer Adam with learning rate 1^{-5} , with adjustment during training to a minimum 1^{-6} .

2.2 VGG

The network input is an RGB image with a size of 224×224 pixels. The convolution layers used in the VGG model use very small receptive fields of size 3×3 with step one. The use of such small convolution filters allows the VGG architecture to use a larger number of layers, as the authors have researched [3]. They also use convolutional filters of size 1×1 , which are used to create a linear transformation of the input, followed by the use of ReLU (Rectified Linear Unit). These filters are used to create a crucial function more non-linearly without changing the receptive fields. [3, 7]

In this paper is used VGG 19, which consist of most layers and could be most accurate.

2.3 DENSENET

DenseNet comes with direct connections between any layers consists of the same feature-map size, and as the name suggests, the model uses a dense connectivity pattern. This model naturally scales to hundreds of layers without any difficulties in optimization. DenseNet is distinguished by less parameters and computational complexity to achieve great performance. The input of the network is an RGB image with size 224×224 pixels. This work uses DenseNet 201. [2]

2.4 INCEPTION

The inception model passed two improvements. This paper uses the third version of that model. The authors claim that the model is computationally efficient. Inception V3 works with factorizing convolutions into smaller convolutions, spatial factorizing, auxiliary classifiers, and efficient grid size reduction. The network input is an RGB image with a size of 229×229 pixels. [1]

3 EXPERIMENT AND RESULTS

Training environment consists of Keras with TensorFlow backend [8]. The whole environment uses a middle-class workstation but with GPU Nvidia 2080 Ti for accelerating all training processes. For this paper, F1-Score is computed with scikit learn library [5]. The custom dataset is not balanced. Because of this fact, we used a technique using class weight. For each class, a weight is calculated at the beginning of the training. Weights prevent the degradation of classes with low image representation.



Classification example of one model can be seen in figure 2.

Figure 2: Two classified images

On the left image is an example of stripes, where the model great classify stripes with 98,6% probability. The image on the right side of the figure shows a mirror ball in the interior. Model right classifies interior with great probability 100%, but the model also classifies fisheye lens, which could be not exact. Probability is 99,8%. When is it compared with images captured with a fisheye lens, images are very similar. Images have also a warped scene, but the fisheye lens warps the whole image, not the only center of the image as shows this image.

		Table	2: Classes statistic	S
		Class name	Average F1 [%]	Number of images
Table 1: A	verage F1-score	Air	72,8	3048
for all classe	s per model	Fire	58,4	789
Model	Average [%]	Image processing keying	60,0	3595
VGG	69,4	Interior	88,3	10561
DenseNet	79,9	Landscape	57,7	4921
Inception	74,7	Machine vision (fisheye)	88,9	1834
		Stripes	83,3	2057
		Woody	87,8	3161

This experiment shows that the best model for classifying illustrations in artworks is DenseNet 201, which has the best average F1-score 79,9 % of tested models. Average F1-scores are in table 1. Second place has Inception with a score of 5,2 % lower than the best model. The worst model is VGG which gets a score of 10,5 % lower than the best model.

Best average F1-score get class Machine vision (fisheye), which get 88,9%, as shows table 2. The worst classified class in order of average F1-score is a landscape with 57,7%. As mentioned above, the dataset is not balanced, as can be seen in table 2. With the class balance technique, this fact is not so significant. Most images, specifically 10561, contain class interior. On the other hand, at least images contain the class fire, with 789 images.

In a deeper look into the classification of particular classes, also DenseNet get highest F1-score. Model get 94,5 % in classification *woody* class, as shows figure 3. Inception model get second place as in average score, and also *woody* class get best value, which is 93,4 %. VGG classify *woody* class as second best in order of F1-score with 75,4 %. Best classified class with VGG model is *interior*, which get 89,2 %. *Interior* class is on third place classifying by model DenseNet with 89,7 % and fourth place classifying by model Inception with 86,1 %.

VGG and DenseNet classify the worst *landscape* class, with 46 %, respectively 63,2 %. Inception classify the worst *fire* with 49 %. From these results, can be said, that *woody* class is the most accurate classifiable. But with regard to the average F1-score of classes, Machine vision (fisheye) get 88,9 %, which is best. The worst classifiable class is fire as is the same according to the average F1-score of classes.



Figure 3: F1-scores of every model for every class

The Inception V3 model is very robust and has very good accuracy. Comparing these three models on ImageNet dataset [9], different order will get. Best is the Inception model with TOP-1 accuracy 78,8 %, but on the artwork dataset, the Inception model gets second place. On ImageNet, the second is DenseNet 201 with 77,42 % TOP-1 accuracy, but in the artwork dataset, DenseNet took first place. The third place is the same and is taken by VGG 19 with 74,5 % TOP1-accuracy on ImageNet. [10]

4 CONCLUSION

This paper proposed right deep CNN for classification artworks, which is DenseNet 201 as is shown on test results. This model get 79,9% F1-score. At all three CNN models has been trained for classification custom artwork dataset. Article not used common dataset, but prepared custom dataset from artworks with specific classes to classify. The article also finds out that although works of art contain materials similar to those found in common datasets, the resulting accuracy may vary and it may not be sufficient to compare models only on general datasets.

ACKNOWLEDGEMENT

This work was supported by the Technology Agency of the Czech Republic under project TL02000270 Media Art Live Archive: Intelligent Interface for Interactive Mediation of Cultural Heritage.

REFERENCES

- C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [2] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [4] W. Vasulka and V. S., "Binary lives," Source: Archiv Vasulka Kitchen Brno, 1999.
- [5] 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.
- [6] V. J, "Multi-label image classification tutorial with keras imagedatagenerator," Mar 2020. [Online]. Available: https://vijayabhaskar96.medium.com/multi-label-image-classificationtutorial-with-keras-imagedatagenerator-cd541f8eaf24
- [7] J. Wei, "Vgg neural networks: The next step after alexnet," Jul 2019. [Online]. Available: https://towardsdatascience.com/vgg-neural-networks-the-next-step-after-alexnet-3f91fa9ffe2c
- [8] "Tensorflow." [Online]. Available: https://www.tensorflow.org/
- [9] 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.
- [10] "Papers with code imagenet benchmark (image classification)." [Online]. Available: https://paperswithcode.com/sota/image-classification-on-imagenet

INCONSISTENT AUDIO DECLIPPING PERFORMANCE ENHANCEMENT BASED ON AUDIO INPAINTING

Ondřej Mokrý and Pavel Záviška

Doctoral Degree Programme (2 and 4), FEEC BUT E-mail: 170583@vutbr.cz, xzavis01@vutbr.cz

> Supervised by: Pavel Rajmic E-mail: rajmic@vutbr.cz

Abstract: Some of the state-of-the-art audio declipping methods are not consistent with the observed signal, meaning that they do not keep the a priori undegraded (reliable) samples. These samples can be directly substituted in the reconstructed signals, but this may create audible artifacts due to the sharp transitions between the declipped and the substituted parts. We propose two methods based on audio inpainting, which deal with these transitions. As a result, we observe a significant improvement of the PEAQ ODG values, but only for some of the declipping algorithms considered.

Keywords: audio, clipping, declipping, inpainting, sparsity, perceptual quality

1 INTRODUCTION

Clipping is a nonlinear distortion that cuts off signal peaks exceeding the allowed dynamic range $[-\theta_c, \theta_c]$, where the value $\theta_c > 0$ is referred to as the clipping threshold. Formally, for the *n*-th sample of the signal $\mathbf{x} \in \mathbb{C}^L$, clipping produces the sample $\theta_c \cdot \operatorname{sgn}(x_n)$ for $|x_n| \ge \theta_c$, and keeps the sample x_n otherwise.

The nonsmooth transitions caused by clipping produce higher harmonics that were not originally present in the signal, which is perceived as an unpleasant distortion. Not only clipping degrades the perceptual quality of audio signals [1], it also worsens further processing conditions, such as automatic speech recognition [2] or voice-based Parkinson's disease detection [3].

The reconstruction of a signal from its clipped observation is generally referred to as *declipping*. For a large overview of the declipping methods, see the audio declipping survey [4].

In declipping, it feels natural to require that the declipped samples should exceed the high (θ_c) and low ($-\theta_c$) clipping thresholds and that the samples that were not altered by clipping (i.e., the *reliable* samples) should remain the same. Signals satisfying these conditions form the *consistent set* and methods that find a solution belonging to this set are naturally referred to as *consistent methods*. On the other hand, *inconsistent methods* only penalize the distance of the solution from the consistent set. See Fig. 1 for the comparison of both approaches. The inconsistent methods include Constrained Orthogonal Matching Pursuit (C-OMP [5]), Plain and Perceptually-motivated Compressed Sensing L1 (CSL1, PCSL1 [6]), Parabola-weighted Compressed Sensing L1 (PWCSL1 [4]), Declipping with Empirical Wiener Shrinkage and Social Sparsity Declipping with Persistent Empirical Wiener (SS EW, SS PEW [7]), or Dictionary Learning (DL [8]). The advantages are usually their low computational complexity, and the possibility to fully exploit any prior about the signal structure, such as the sparsity of its time-frequency representation.

However, sticking to the inconsistent solution implies that the knowledge of the reliable samples from the clipped signal is undervalued. This paper, therefore, addresses the issue of enhancing the quality of inconsistent audio declipping methods by replacing the reliable samples.



Figure 1: Demonstration of consistent and inconsistent declipping solutions on a short piece of audio signal including the application of the straightforward replacement of the reliable samples (RR).

2 PROBLEM FORMULATION

The easiest way to fully exploit the information stored in the reliable samples is simply to substitute them in place of the inconsistent samples. This procedure naturally increases the signal-to-distortion ratio (SDR) and in a large number of cases also improves the perceptual quality according to the perceptual measures, such as PEAQ [9].

On the other hand, the main problem of this technique is the possibility of creating sharp transitions between the clipped and the replaced reliable parts, as illustrated in Fig. 1. Such a nonsmooth phenomenon creates higher harmonic spectral components and thus degrades the perceived audio quality.

To leverage the knowledge of the reliable samples while avoiding the sharp edges on the transitions, we present a novel method based on audio inpainting. The main idea lies in "deleting" a number of samples in the beginning and at the end of each clipped part and then estimating the values of these samples using an arbitrary inpainting method, while the rest of the clipped parts, as well as the (replaced) reliable samples are fixed. The set of signals meeting such conditions will be denoted Γ . Suitable audio inpainting methods, which make use of the just defined set Γ , are described in the following section.

3 INPAINTING THE TRANSITIONS

Audio signals are typically sparse in a time-frequency (TF) representation, which is caused by their harmonic and short-time stationary nature. Since the problem of finding a signal in Γ is ill-posed (there are infinitely many solutions meeting the conditions), we may regularize it such that the solution is desired to have a sparse TF representation or a TF representation with minimal ℓ_1 norm [10]. The sparsity-based audio inpainting problem can be formulated as one of the following optimization problems:

$$\underset{\mathbf{c}}{\operatorname{arg\,min}} \|\mathbf{c}\|_{1} + g(A^{*}\mathbf{c}), \tag{1}$$

$$\underset{\mathbf{x}}{\operatorname{arg\,min}} \|A\mathbf{x}\|_{1} + g(\mathbf{x}). \tag{2}$$

The formulations (1) and (2) are referred to as the synthesis and analysis formulations, respectively. The variable $\mathbf{c} \in \mathbb{C}^N$ denotes a TF representation of a signal $\mathbf{x} \in \mathbb{C}^L$. We use a redundant transform, such that the analysis operator $A : \mathbb{C}^L \to \mathbb{C}^N$, N > L, produces TF coefficients of a signal, and the synthesis operator $A^* : \mathbb{C}^N \to \mathbb{C}^L$ reconstructs the signal out of the coefficients. In such a redundant case, the two formulations are not equivalent. Furthermore, we suppose $A^*A = \text{Id}$.
In both formulations, the norm $\|\cdot\|_1$ promotes sparsity of the TF coefficients, and g enforces the signal consistency in the time domain. As indicated above, we may put $g = \iota_{\Gamma}$, i.e., the indicator function, which takes on zero value for elements belonging to Γ and infinity otherwise.

The problem (1) can be solved via the Douglas–Rachford algorithm [11, Sec. IV]. Keeping the notation of [11], let $f_1 = \|\cdot\|_1$ and $f_2 = g \circ A^*$. Then, the proximal mapping $\operatorname{prox}_{\gamma f_1} = \operatorname{prox}_{\gamma \|\cdot\|_1}$ is the soft thresholding with threshold γ and $\operatorname{prox}_{\gamma f_2} = \operatorname{Id} + A \circ (\operatorname{prox}_{\gamma g} - \operatorname{Id}) \circ A^*$ [11, Tab. I, ix]. When $g = \iota_{\Gamma}$, $\operatorname{prox}_{\tau g} = \operatorname{prox}_{\iota_{\Gamma}}$ is the projection onto Γ , i.e., it replaces all the fixed samples and keeps the rest.

The problem (2) can be solved via the Chambolle–Pock algorithm [12, Alg. 1]. Keeping the notation, we put $F = \|\cdot\|_1$ and G = g; the linear operator A is taken care of by the algorithm. The procedure makes use of the proximal mappings $\operatorname{prox}_{\sigma F^*}$ and $\operatorname{prox}_{\tau G}$, where the asterisk in F^* denotes the convex conjugate of the function F. It holds $\operatorname{prox}_{\sigma F^*} = \operatorname{Id} - \sigma \operatorname{prox}_{F/\sigma}(\cdot/\sigma)$ [12, Eq. (6)]. In our settings, $\operatorname{prox}_{\sigma \|\cdot\|_1^*} = \operatorname{Id} - \sigma \operatorname{prox}_{\|\cdot\|_1/\sigma}(\cdot/\sigma)$, i.e., the soft thresholding operator is used with the threshold $1/\sigma$.

4 ADAPTIVE RELIABILITY OF THE DECLIPPED SAMPLES

The model in the previous section can be understood such that we trust the originally reliable samples and also some of the declipped samples, while we do not trust the declipped samples near the transitions *at all*. A natural generalization would be that the reliability of the declipped samples is not only binary. This can be implemented in the function g. If we take the declipped signal and simply substitute the reliable samples, we obtain the signal $\hat{\mathbf{x}}$. Per each declipped sample n, we may put

$$g_n(x_n) = \begin{cases} w_n \cdot |x_n - \hat{x}_n|^2 / 2 & \text{for } \hat{x}_n \text{ declipped,} \\ \iota_{\{\hat{x}_n\}}(x_n) & \text{for } \hat{x}_n \text{ reliable,} \end{cases} \quad g(\mathbf{x}) = \sum_{n=1}^L g_n(x_n), \tag{3}$$

where $w_n \ge 0$ are the *reliability weights*: the greater the w_n , the more we trust the sample \hat{x}_n . To apply the previously described proximal algorithms, we use the proximal mapping of the function g, which is defined elementwise as [13, Thm. 6.6, Ex. 6.65]:

$$x_n \mapsto \frac{w_n}{w_n+1}\hat{x}_n + \frac{1}{w_n+1}x_n. \tag{4}$$

The simple inpainting approach from Sec. 3 fits this model with $w_n = \infty$ for the fixed samples and $w_n = 0$ for the transitions to be filled. For the generalized method, we suggest using a bump function in each declipped segment to assign the weights w_n , since the transitions are less trusted than the declipped samples in the middle of each segment.

5 EXPERIMENTS AND RESULTS

For the experiments, 10 audio excerpts sampled at 44.1 kHz with their approximate length of 7 seconds were used. These signals were artificially clipped as defined in the Introduction with 7 different threshold based on the input SDR. The restored signals were obtained using the inconsistent methods mentioned in the Introduction and included in the survey [4] (C-OMP, CSL1, PCSL1, PWCSL1, SS EW, SS PEW, and DL). Except for C-OMP and DL, the methods are based on the convex sparsityinspired formulation (1) with g measuring the distance from the consistent set of the declipping problem. C-OMP aims at approximating the non-convex sparse synthesis problem in a greedy way, constraining the solution to lie near the consistent set. Finally, DL enhances the sparsity based approach by searching not only for the signal but also the best operator A^* to allow for sparser solution.

As the TF transform A, we used the discrete Gabor transform, implemented for MATLAB in the toolbox LTFAT [14]. Two settings of the transform were tested: The first system was defined by the Hann window of length 8192 samples, window shift 2048 and 8192 frequency channels. The



Figure 2: Average PEAQ performance of the inconsistent methods and the replacing strategies.

second system used the same window shape, but the numerical parameters were 4096, 1024 and 4096, respectively. Both systems were scaled to satisfy the relation $A^*A = \text{Id}$.

In the experiment, the Douglas–Rachford algorithm was applied with $\gamma = 1$ and $\lambda_n = 1$ for all *n*. It stopped either after 500 iterations or when the relative change of the solution dropped below $5 \cdot 10^{-4}$. Similarly, the Chambolle–Pock algorithm was applied with $\tau = \sigma = 1$ and the same stopping criteria.

The perceptual quality of audio signals was evaluated using PEAQ (Perceptual Evaluation of Audio Quality) [9], specifically the free MATLAB implementation [15]. PEAQ predicts the degradation using a ODG scale ranging from -4 (very annoying) to 0 (imperceptible).

Fig. 2 presents the PEAQ ODG values of the declipped signal, prior to any replacement, and the PEAQ ODG improvement (Δ ODG) for different postprocessing strategies: basic replacement (RR), the simple inpainting method from Section 3 (INP1) and the modified inpainting method from Section 4 (INP2). Only a selection of all the possible settings is plotted, namely INP1 used the synthesis formulation and the TF transform with longer window, INP2 used the analysis formulation and the TF transform with longer segment of length *D*, INP1 filled *d* samples at each of the two transitions, with $d = \max(\lfloor D/3 \rfloor, 10)$. INP2 was applied with the weights according to the function $10\sin(t)^{1/2}$, with *t* representing *D* evenly spaced samples from 0 to π .

The most significant observation here is that the inpainting-based strategies outperform the basic replacement strategy in case of the methods that were inferior prior to any replacement. On the other hand, these strategies fail to enhance the a priori favorable methods, such as SS PEW.

Furthermore, comparing the bottom graphs leads to a clear conclusion that INP2 magnifies both the gains and the losses of INP1.

6 CONCLUSION

We have presented two inpainting-based methods, which serve as a postprocessing procedure for inconsistent audio declipping. The goal was to maximally exploit the reliable samples of the original declipping problem while avoiding sharp transitions in the signal after a simple substitution. The proposed methods enhance the quality of the resulting audio signal based on PEAQ ODG, but only for some of the declipping methods involved. Although the improvement seems to be rather significant in these cases (up to 1 degree on the ODG scale), the inpainting methods do not improve the reconstruction quality, compared to the basic replacement strategy, for the rest of the methods.

ACKNOWLEDGEMENT

The work was supported by the project 20-29009S of the Czech Science Foundation (GAČR).

REFERENCES

- [1] C.-T. Tan, B. C. J. Moore, and N. Zacharov, "The effect of nonlinear distortion on the perceived quality of music and speech signals," *Journal of the Audio Engineering Society*, 2003.
- [2] Y. Tachioka, T. Narita, and J. Ishii, "Speech recognition performance estimation for clipped speech based on objective measures," *Acoustical Science and Technology*, 2014.
- [3] A. H. Poorjam, et al., "Automatic quality control and enhancement for voice-based remote Parkinson's disease detection," *Speech Communication*, 2021.
- [4] P. Záviška, P. Rajmic, A. Ozerov, and L. Rencker, "A survey and an extensive evaluation of popular audio declipping methods," *IEEE Journal of Selected Topics in Signal Processing*, 2021.
- [5] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. Plumbley, "A constrained matching pursuit approach to audio declipping," in *IEEE ICASSP*, 2011.
- [6] B. Defraene, et al., "Declipping of audio signals using perceptual compressed sensing," *IEEE Trans. Audio, Speech, and Language Processing*, 2013.
- [7] K. Siedenburg, et al., "Audio declipping with social sparsity," in IEEE ICASSP, 2014.
- [8] L. Rencker, F. Bach, W. Wang, and M. D. Plumbley, "Consistent dictionary learning for signal declipping," in *Latent Variable Analysis and Signal Separation*. Springer, 2018.
- [9] T. Thiede, et al., "PEAQ The ITU standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, 2000.
- [10] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of The National Academy of Sciences*, 2003.
- [11] P. Combettes and J. Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point* Algorithms for Inverse Problems in Science and Engineering, 2011.
- [12] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, 2011.
- [13] A. Beck, First-Order Methods in Optimization. SIAM, 2017.
- [14] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The Large Time-Frequency Analysis Toolbox 2.0," in *Sound, Music, and Motion*. Springer, 2014.
- [15] P. Kabal, "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality," MMSP Lab Technical Report, McGill University, 2002.

IMPACT OF LOSS FUNCTION ON MULTI-FRAME SUPER-RESOLUTION

Anzhelika Mezina

Doctoral Degree Programme 1st year, FEEC BUT E-mail: xmezin00@vutbr.cz

> Supervised by: Radim Burget E-mail: burgetrm@feec.vutbr.cz

Abstract: Nowadays, one of the most popular topics in image processing is super-resolution. This problem is getting more actual even in security, since monitoring cameras are everywhere and in the case of an incident, it is necessary to recognize a person from records. A lot of approaches exist, which are able to reconstruct image, and the most of them are based on deep learning. The main focus of this work is to analyze, which loss function for neural networks is more effective for real-world image reconstruction. For this experiment chosen architecture and dataset are used for multi-frame super-resolution for $\times 8$ scaling.

Keywords: super-resolution, image processing, loss function, deep learning

1 INTRODUCTION

Last time, the image processing, especially, improvement of image quality, takes a big place in research area. In spite of the fact, that there are a lot of techniques for image processing, and it is not difficult to get cameras with good optics, there is still a need to reconstruct images from low quality. Especially, this problem is critical in the security cameras. These cameras usually have low quality of records and in the case of incidents the police is not able to recognize a person in such videos.

Recent years, many approaches have been introduced and a lot of them are based on neural networks, which require huge datasets for training. Currently, the most works use the datasets, which were created from Youtube videos or films, and these images are really nice, because of light and color correction. The security cameras do not have such good quality and recognition of a person from such records is getting more complicated. For that reason, it is necessary to have algorithms and methods, which would be able to improve quality of image.

The main focus in research area is on the methods for single frame super-resolution, however, there are not so many methods, which perform reconstruction from a couple of images. Additionally, there are not so many works, which pay attention on loss functions for deep learning methods. For these reasons, this paper provides the comparison study of loss functions which can be applied for multi-frame super-resolution task. In this work the experiment is performed with dataset of images from real-world for 8 scale factor. The results introduced in this paper can be useful for future study of problem of image reconstruction from low quality images.

2 RELATED WORK

The often used methods in the problem of super-resolution are based on convolutional neural networks (CNN) and generative adversarial networks (GAN). The well known methods are based on single frame super-resolution, for example, SRCNN [1], which is based on CNN and has only three layers. The authors applied Mean Squared Error (MSE) as a loss function. It achieved better results than

methods without application of deep learning. Another method introduced in 2018 is ESRGAN [2]. It is based on GAN and has been developed as enhanced version of SRGAN [3]. The authors used perceptual loss function, which is better for reconstruction tasks.

Multi-frame super-resolution is not so popular area and the most methods are based on mathematical approaches, however, there are some methods, which utilize the deep leaning. One of the latest works [4] introduced the method based on CNN to perform motion-based multi-frame super-resolution. Also, some approaches use residual learning, for example, in the work [5] authors combined sub-pixel registration method for mapping into the high-resolution grid and deep residual learning approach for restoring features without noise. Both mentioned methods utilized MSE as a loss function.

3 METHODOLOGY

This work is focused on impact of loss functions used in neural network on quality of a reconstructed image from a sequence of images. There are some kind of loss functions, which are frequently used in image reconstruction, such as Mean Square Error (MSE), Feature reconstruction loss, Charbonnier loss and, relatively new one, is Learned Perceptual Image Patch Similarity (LPIPS). In this section, the experiment with these loss functions is described.

3.1 DATASET

Nowadays, there are not so many datasets for multi-frame super-resolution. In this work the new dataset Multi-Frame Labeled Faces Database (MLFDB) [6] is used for experiments. This dataset contains sequences of 7 images, extracted from real-life videos. Moreover, this dataset allows to do experiments with different scale factors: 2, 4, 8. For this experiment the chosen scale factor is 8 and training, validation, testing sets have 2,504, 837 and 753 images respectively. The example of input and output (label) images are shown in Fig. 1. The input size of images is 32×32 px and output is 256×256 px.



Figure 1: Example from dataset.

3.2 Loss functions

For this work used such loss functions, as MSE, Feature reconstruction loss, Charbonnier loss and LPIPS.

MSE loss function is the most frequently used in image processing. This loss function calculates squares between original image and predicted one. It can be formulated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i)^2,$$
(1)

where X_i – original high-resolution image, Y_i – predicted super-resolution image, n – number of training samples.

Charbonnier loss was also applied for super-resolution task [7]. This loss function might better handle outliers and improve the reconstruction accuracy.

$$L = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(Y_i - X_i)^2 + \varepsilon^2},$$
(2)

where X_i – original high-resolution image, Y_i – predicted super-resolution image, ε – constant, 0.001, n – number of training samples.

In spite of high Peak Signal-to-Noise Ratio (PSNR) value of the image, which is reconstructed with MSE loss, there is still a problem, that quality of image do not correlate well with human perception [7]. For that reason, the perceptual loss function is very popular in image super-resolution task. Feature Reconstruction Loss [8], which is a kind of perceptual loss, is used in this work. In contrast to MSE loss function, it encourages the similar of feature representation, instead of matching of pixels. This loss function can be defined as:

$$\ell_{feat}^{\theta,j}(\hat{I},I) = \frac{1}{C_j H_j W_j} ||\phi_j(\hat{I}) - \phi_j(I)||^2,$$
(3)

where *I* – original high-resolution image, \hat{I} – predicted super-resolution image, $\phi_j(\hat{I})$ – activations of *j*th layer of the VGG19 network ϕ for processing the image *I*, $C_i H_j W_j$ – the shape of the feature map.

LPIPS was applied for extreme super-resolution with 16 scale factor [9]. The authors state, that for extreme super-resolution the VGG-based loss function is not the best choice, since VGG network is used for classification task. But the authors proposed to apply the AlexNet network as a perceptual loss function. This loss function can be described as:

$$\ell_{lpips} = \tau(\phi_j(\hat{I}) - \phi_j(I)), \tag{4}$$

where I – original high-resolution image, \hat{I} – predicted super-resolution image, ϕ – feature extractor, τ – transformation of deep embedding to scalar LPIPS score.

3.3 ARCHITECTURE

For this experiment used neural network architecture, which was proposed and applied for multiframe super-resolution task, is U-Net based network with GEU blocks [6]. This model processes the sequence of images. Originally, in the paper it was used for 2 scale factor, however, for this experiment it was adapted for 8 scale factor: some layers with subpixel convolution operation for upsampling were added. The scheme of architecture is shown in Fig. 2. The model was trained with hyperparameters: optimizer is Adam with learning rate 0.0001, the number of epochs is 300 with 500 steps for epochs, batch size is 4.



Figure 2: Adapted for 8 scale U-Net+GEU 3 from [6].

RESULTS 4

The U-Net+GEU 3 model was trained on mentioned dataset with different loss functions which were described in this work. The results of evaluation of methods are shown in Table 1 and Fig. 3, where the last column, FFR, is the number of failed face recognition. According to the metrics the best objective results - SSIM, MSE, PSNR, achieved model with Charbonnier loss function. It achieved SSIM - 0.7251, MSE - 302.5284, PSNR - 24.1890 dB. However, it can be seen from Table 1, that the number of failed face recognition is the least for LPIPS loss function -435, that can be a useful point in case of person identification. On the other hand, from subjective side it seems that image reconstructed with feature reconstruction loss looks better and more details are seen. Moreover, the sharpness difference between original and reconstructed images is the least for feature reconstruction loss.

Table 1	Results for	' 8 scale factor.

T C II	COD (LOD	DOME ID	G1 11 66	TED
Loss function	SSIM	MSE	PSNR, dB	Sharp. difference	FFR
MSE	0.7083	352.6861	23.4785	0.2071	473
Charbonnier	0.7251	302.5284	24.1890	0.2073	466
Feature reconst.loss	0.6490	616.5493	20.8621	0.0270	450
LPIPS	0.5240	435.4415	22.3790	-0.4884	435





(a) Ground truth

(e) LPIPS

Figure 3: Results of evaluation

5 CONCLUSION

The aim of this work was to investigate the impact of loss functions on image reconstruction from a sequence of frames. For experiment the new dataset and the method for multi-frame super-resolution were used, and applied loss functions were Charbonnier, MSE, feature reconstruction loss and LPIPS. According to results, the best metrics such as SSIM, MSE, PSNR, achieved images reconstructed with Charbonnier loss. However, the resulting image seems very smooth and for human it is difficult to recognize a person there. The number of failed face recognition is the least for LPIPS loss. Subjectively, more details are seen in the image reconstructed with feature reconstruction loss, but the face has some deformations. In this way, it is obviously that loss function takes a big role in image reconstruction.

As the conclusion, it is necessary to prepare more experiments with architectures and loss functions. There is still a lack in research field for image super-resolution for real-world application. Nowadays, there are a lot of approaches with nice reconstructed images, which were tested on datasets with good quality images. However, these methods can fail on images from monitoring cameras. For real-world application, it is better to use loss functions in neural networks, which are focused on improvement quality for human perception. If the reconstructed image is used by police for person identification, it does not matter, which objective metrics it is achieved, it is more important to recognize a culprit.

REFERENCES

- [1] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." IEEE transactions on pattern analysis and machine intelligence 38.2 (2015): 295-307.
- [2] Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018.
- [3] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [4] Elwarfalli, Hamed, and Russell C. Hardie. "FIFNET: A convolutional neural network for motionbased multiframe super-resolution using fusion of interpolated frames." Computer Vision and Image Understanding 202 (2021): 103097.
- [5] Noor, Dewan Fahim, et al. "Multi-frame super resolution with deep residual learning on flow registered non-integer pixel images." 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019.
- [6] Rajnoha, Martin, Anzhelika Mezina, and Radim Burget. "Multi-frame labeled faces database: Towards face super-resolution from realistic video sequences." Applied Sciences 10.20 (2020): 7213.
- [7] Lai, Wei-Sheng, et al. "Deep laplacian pyramid networks for fast and accurate super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [8] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." European conference on computer vision. Springer, Cham, 2016.
- [9] Jo, Younghyun, Sejong Yang, and Seon Joo Kim. "Investigating loss functions for extreme superresolution." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.

PSEUDO-DIFFERENTIAL FRACTIONAL-ORDER FREQUENCY FILTER

Ondrej Sladok, Alkanan Mohamad

Doctoral Degree Programme (4th year), FEEC BUT, Bachelor Degree Programme (3rd year), FEEC BUT E-mail: sladok@phd.feec.vutbr.cz, xalkan00@stud.feec.vutbr.cz

Supervised by: Jaroslav Koton

E-mail: koton@feec.vutbr.cz

Abstract: This paper describes a pseudo-differential fractional $(1+\alpha)$ -order frequency filter. The structure operating in voltage-mode and and works as a low-pass filter. The proposed structure is characterized by high input impedance, high output impedance and the minimum number of passive elements when all are grounded. The properties and functionality of the $(1+\alpha)$ -filter are verified by simulations.

Keywords: Pseudo-differential filter, fractional-order filter, frequency filter, $(1+\alpha)$ -order, voltagemode, current conveyors, foster, FOE.

1 INTRODUCTION

In analog technology, there is still interest in frequency filters, which can be found in a large number of electronics. A new type of frequency filters are pseudo-differential frequency filters [1]. These filters combine the properties of fully differential filters and single-ended filters. This type offers other advantageous possibilities in the functionality of the construction. They allow very promising results in the suppression of the common-mode signal, as well as the reduced harmonic distortion of the processed signal as fully differential filters. At the same time, their construction is much simpler, their complexity resembles single-ended filters.

In addition, this article describes another interesting industry in frequency filters. These are fractional-order filters [3]. The presence of fractional-order α , usually in the range $0 < \alpha < 1$, allows to create structures of non-integer order. This type allows a new view and further use of such structures. In the case of analog fractional-order frequency filter design, the presence of the fractionalorder element (FOE) is assumed. So far, unfortunately, FOE cannot be purchased as a single component such as today's resistors or other components. The first approach approximates directly the term s^{α} being present in the transfer function by integer-order polynomial function. The second possible proposal of FOE is very often used technology – a network of RC ladders, such as Foster I, Foster II, Cauer I and Cauer II.

This article combines the two research topics discussed above. The design of a frequency filter of pseudo-differential fractional $(1+\alpha)$ -order is described here. The proposed filter operating in voltage-mode provides a low-pass response, advantageously employing current conveyors as active elements, both input and output impedance of the filter is infinitely high in the ideal case.

2 THEORY OF PSEUDO-DIFFERENTIAL FREQUENCY FILTERS

The pseudo-differential fractional-order frequency filters have many advantages [1], the biggest advantage being undoubtedly the simpler complexity of the resulting structure, further have a high common-mode rejection ratio, low a total harmonic distortion, where even signal components are suppressed. In addition, they can become part of new electronic components or open up new possibilities for many applications. The main theoretical feature is the use of only differential input and

output signals. At the same time, however the internal structure is single-ended (nondifferential). This phenomenon causes the retention of properties with both fully differential frequency filters and single-ended filters and is just less complicated. The following mathematical description applies to the differential input and output signal in voltage mode [2].

$$v_{1d} = v_{1+} - v_{1-}, \quad v_{2d} = v_{2+} - v_{2-}, \quad A_d = \frac{v_{2d}}{v_{1d}} = \frac{v_{2+} - v_{2-}}{v_{1+} - v_{1-}},$$
 (1)

where signal v_{1d} is the difference between the two input signals v_{1+} and v_{1-} . Signal v_{2d} is the difference between the two output signals v_{2+} and v_{2-} . Differential voltage gain A_d is the ratio of the differential output signal to the differential input signal. v_{1d} , v_{2d} and A_d denote differential input voltage, the differential output voltage and differential voltage gain, respectively.

3 PROPOSAL OF FRACTIONAL-ORDER AND PROPERTIES OF ACTIVE ELEMENTS

The fractional-order was designed according to Freeborn as stated in [3]. Freeborn uses for the lowpass filter a traditional approximation according to Butterworth. The fractional-order transfer function can then be expressed as follows:

$$A_{\rm d}^{1+\alpha}(s) = \frac{k_1}{s^{1+\alpha} + k_2 s + k_3},\tag{2}$$

where $0 < \alpha < 1$ and at the same time the capacitor C_2 is replaced by the pseudo-capacitance C_{α} . Capacitor C_1 remains traditional. For coefficients k_1 and k_2 were selected to approximate the flat passband response of the Butterworth filters using a numerical search comparing the fractional transfer function and first-order Butterworth passbands over the frequency range $\omega = 0.01-1$ rad / s. The coefficients yielding the lowest error during the search, which was limited to $0 < k_2 < 2$ and $0 < k_3 < 1$ when $k_1 = 1$.

$$k_2 = 1.0683\alpha^2 + 0.161\alpha + 0.3324$$

$$k_3 = 0.2937\alpha + 0.7122$$
(3)

Active elements from the field of current conveyors were used to construct the pseudo-differential fractional $(1+\alpha)$ -order filter. One Differential Difference Current Conveyor (DDCC) and one Differential Voltage Current Conveyor (DVCC) were used for the design. The schematic symbols are shown in Fig 1.

$$\begin{array}{c|c} \underbrace{I_{Y1}}_{0} & \underbrace{DVCC}_{1} & \underbrace{I_{Z+}}_{1} & \underbrace{I_{Y2}}_{1} & \underbrace{V_{1} & Z_{+}}_{1} & \underbrace{I_{Z+}}_{1} & \underbrace{I_{Z+}}_{1} & \underbrace{I_{Z+}}_{1} & \underbrace{I_{Z-}}_{1} &$$

Fig. 1. Schematic symbols of a) DVCC and b) DDCC

For all types it holds that the Y terminal currents are zero. The $Z_{+/}$ terminal currents are defined as:

$$I_{Z+} = I_X, \ I_{Z-} = -I_X. \tag{4}$$

The only significant difference between the assumed current conveyors are the formulas specifying the voltage at X terminal, which for DVCC and DDCC it holds respectively:

$$V_{\rm X} = V_{\rm Y1} - V_{\rm Y2}, V_{\rm X} = V_{\rm Y1} - V_{\rm Y2} + V_{\rm Y3}.$$
(5)

4 PSEUDO-DIFFERENTIAL FRACTIONAL-ORDER FREQUENCY FILTER

For the design of the filter became the lit. [4], using the transformation of the single-ended filter into a pseudo-differential filter. Furthermore, the filter was modified to fractional-order according to the literature [3].

The new proposed pseudo-differential $(1+\alpha)$ -order filter operating in voltage-mode is shown in Fig. 2. The structure can realize the low-pass response. The filter uses two active elements of DDCC and DVCC. The input signal is applied to DDCC 1 to terminals Y₁ and Y₂ it ensures high input impedance of the filter. Furthermore, the filter uses four passive elements, of which two resistors R_1 and R_2 , one capacitor C_1 and the pseudo-capacitance C_{α} , where they all are grounded. Pseudo-capacitance C_{α} was constructed using a capacitive fractional-order element (FOE).



Fig. 2. Pseudo-differential fractional $(1+\alpha)$ -order filter operating in voltage-mode

Assuming the general transfer functions (1) and (2), the mathematical description of the new pseudo-differential fractional $(1+\alpha)$ -order low-pass filter is displayed as follows:

$$A_{\rm d}^{(1+\alpha)}(s) = \frac{1}{s^{1+\alpha}C_{\alpha}C_{1}R_{1}R_{2} + sC_{1}R_{1} + 1}.$$
(6)

5 SIMULATIONS

The OrCAD and Snap programs were used for the performed simulations. In OrCAD, simulations were performed for ideal elements (UCC_1L) and for real elements (UCC_3L). The MATLAB program was then used for mathematical calculations, where fractional-order was designed. The filter was simulated and verified for fractional-order values $\alpha = 0.4$, $\alpha = 0.6$ and $\alpha = 0.8$. The structure was designed at the pole frequency $f_0 = 100$ kHz, next the capacitor $C_1 = 1$ nF was chosen. Resistors R_1 and R_2 were calculated according to (7). All values of the proposed pseudo-differential (1+ α)-order filter can be found in Tab. 1. Next, to approximate the capacitive FOE, the Foster-II 7th-order RC network was used as shown in Fig. 3.

Mathematical description of resistors R_1 and R_2 for selected α values.

$$R_{1} = \frac{1}{\omega_{0}C_{1}} \cdot \frac{k_{2}}{k_{3}}, R_{2} = \frac{1}{\omega_{0}^{1+\alpha}C_{\alpha}k_{2}}$$
(7)

Using [5] and the MATLAB program of the values capacitors and resistors of Foster II network were obtained. Furthermore, the central frequency $f_c = 100$ kHz and the approximation range from 1 kHz to 1 MHz were chosen for the FOE. The values of individual resistors and capacitors of the FOE are described in Tab. 2.



Fig. 3. The 7th-order structures approximating the FOE – type Foster II

Tab. 1	. The	values	for	three	values	α

α	0.4	0.6	0.8
Z _C	1592	1592	1592
k_2	0.56768	0.81348	1.14472
<i>k</i> ₃	0.82968	0.88842	0.94716
$R_{1}\left(\Omega ight)$	1088	1457	1923
$R_{2}\left(\Omega ight)$	2805	1957	1391
C_{α} (F/s ^{1-α})	3.01E-06	2.08E-07	1.44E-08

values α 0,4 0,6 0,8 α $R_0(\Omega)$ $R_1(\Omega)$ $R_2(\Omega)$ $R_3(\Omega)$ $R_4(\Omega)$ $R_5(\Omega)$ $R_6(\Omega)$ $R_7(\Omega)$ C_1 (pF) C_2 (pF) C_3 (pF) C_4 (pF) C_5 (pF) $C_6 (pF)$ $C_7 (pF)$

The magnitude responses and phase responses of the three approximated FOEs ($\alpha = 0.4, 0.6, 0.8$) obtained by the simulat11ions are shown in Fig. 4. and Fig. 5. It is clear from the phase responses characteristics for all values α . Resulting responses confirm the correct approximation in the specific frequency range. From Fig.4 the values of the modulus at the mean frequency $f_c = 100$ kHz 1645 Ω (for $\alpha = 0.4$), 1592 Ω (for $\alpha = 0.6$) and 1462 Ω (for $\alpha = 0.8$) in compared to Tab.1 show minimum deviations and confirm the correct behavior of the FOE. The magnitude and phase responses of the proposed pseudo-differential fractional (1+ α)-order frequency filter obtained by simulations are displayed in Fig. 6. and Fig. 7.





Tab. 2. The values of the Foster II for three



From the characteristics in Fig. 6. and Fig. 7. it is obvious that the resulting responses confirm the correct functionality and properties of the pseudo-differential fractional $(1+\alpha)$ -order filter. The simulations show very close, almost the same behavior as the theoretical assumptions.

6 CONCLUSION

In this paper, a new pseudo-differential fractional $(1+\alpha)$ -order frequency filter operating in voltage mode is presented. The filter connects two very frequently discussed topics in the field of frequency filters, which are fractional-order filters and pseudo-differential filters. The proposed structure uses two active elements (DDCC and DVCC) and four passive elements (two resistors, traditional capacitor and pseudo-capacitance), all are grounded. The FOE was built using the RC network type Foster II. The filter is characterized by high input impedance. Furthermore, high common-mode rejection ratio, low sensitivity and low harmonic distortion. The performance of the structure was verified by simulations for three values α , which confirm the functionality and properties of the proposed $(1+\alpha)$ -order filter.

REFERENCES

- [1] O. Sladok, J. Koton, N. Herencsar. Universal Pseudo-Differential Filter Using DDCC and DVCCs. Elektronika Ir Elektrotechnika, (2017), vol. 23, no. 6, p. 46-52. ISSN: 1392-1215. http://dx.doi.org/10.5755/j01.eie.23.6.19694
- [2] U. Tietze, Ch. Schenk, E. Gamm. Electronic circuits: handbook for design and application. Berlin: Springer; 2008. p. 1543.
- [3] T. Freeborn. Comparison of $(1+\alpha)$ fractional-order transfer functions to approximate lowpass butterworth magnitude responses. Circuits Systems, and Signal Processing. (2015). 35. 10.1007/s00034-015-0226-y.
- [4] J-W. Chiu, W-Y. Horng, High-input and low-output impedance voltage-mode universal biquadratic filter using DDCCs. Circuits and Systems II: Express Briefs, IEEE Transactions on. 54. 649-652. (2007). 10.1109/TCSII.2007.899460.
- [5] G. Tsirimokou, "A systematic procedure for deriving RC networks of fractional-order elements emulators using MATLAB", AEU-International Journal of Electronics and Communications, vol. 78, pp. 7–14, (2017). DOI: 10.1016/j.aeue.2017.05.003.



BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION TECHNICKÁ 3058/10, 616 00 BRNO, CZECH REPUBLIC



BRNO FACULTY OF ELECTRICAL UNIVERSITY ENGINEERING OF TECHNOLOGY AND COMMUNICATION



www.fekt.vut.cz

