



ISSN 2545-4889

XV INTERNATIONAL CONFERENCE **ETAI 2021 ETAИ** 23-24 SEPTEMBER, ONLINE CONFERENCE

ЗБОРНИК НА ТРУДОВИ
CONFERENCE PROCEEDINGS

УРЕДНИК: МАРИЈА КАЛЕНДАР ◆ EDITED BY: MARIJA KALENDAR



ЗДРУЖЕНИЕ ЗА
ЕЛЕКТРОНИКА
ТЕЛЕКОМУНИКАЦИИ
АВТОМАТИКА И
ИНФОРМАТИКА
НА РЕПУБЛИКА
МАКЕДОНИЈА



SOCIETY FOR
ELECTRONIC
TELECOMMUNICATIONS
AUTOMATICS AND
INFORMATICS
OF THE REPUBLIC
OF MACEDONIA

ЗБОРНИК НА ТРУДОВИ
ОД XV МЕЃУНАРОДНА КОНФЕРЕНЦИЈА

Уредник: проф. д-р Марија Календар

ЕТАИ 2021 – ЕТАИ 2021

CONFERENCE PROCEEDINGS OF
XV INTERNATIONAL CONFERENCE

Editor: Prof. Dr Marija Kalendar

ISSN 2545-4889, Vol. 2, Issue 1

23 – 24 септември 2021, виртуелна конференција
23 – 24 September 2021, Online Conference



Уредувачки одбор на Зборникот на трудови ЕТАИ 2021

Главен уредник: проф. д-р Марија Календар, ФЕИТ, Скопје

Компјутерска обработка:

Марија Марковска
Бодан Велковски

Editorial board of ETAI 2021 Conference Proceedings

Editor: Prof. Dr. Marija Kalendar

Computer design:

Marija Markovska
Bodan Velkovski

ЕТАИ 2021

<http://www.etai.org.mk>

ПРЕТСЕДАТЕЛ НА КОНФЕРЕНЦИЈАТА

Никола Љушев

ПОТПРЕТСЕДАТЕЛИ НА КОНФЕРЕНЦИЈАТА

Пецо Неделковски и Горан Стојановски

ОРГАНИЗАТОРИ

ЕТАИ – Здружение за Електроника, Телекомуникации, Автоматика и

Информатика на Република Северна Македонија

Факултет за електротехника и информациски технологии на Универзитетот „Св. Кирил и Методиј“ во Скопје, Република Северна Македонија

ПОЧЕСЕН ОДБОР

Никола Јанкуловски

Сашо Коруновски

Нинослав Марина

Imre J. Rudas

Frank Allgoewer

Mogens Blanke

A. Talha Dinibutun

Rolf Isermann

Stephen Kahne

Vladislav Y. Rutkovsky

Kevin Warwick

Janan Zaytoon

Ji-Feng Zhang

Љупчо Коцарев

Леонид Грчев

Љупчо Пановски

Томислав Цеков

Аксенти Грнарлов

Лилјана Гавриловска

Георги Димировски

Миле Станковски

Методија Камиловски

Татјана Колемишевска-

Гугуловска

Ректор на УКИМ Скопје, С. Македонија

Ректор на УКЛО Битола, С. Македонија

Ректор на Универзитет Св. Апостол Павле, Охрид,
С. Македонија

Поранешен ректор на Obuda University Унгарија

Претседател на IFAC, Штутгарт, Германија

Почесен професор на Техничкиот Универзитет во
Лингби, Данска

Поранешен ректор на Догус Универзитетот во
Истанбул, Турција

Почесен професор на Техничкиот Универзитет во
Дармштат, Германија

Поранешен претседател на IFAC, Аеронаут.
Универзитет во Феникс, САД

Руска академија на науките, Москва, Русија

Поранешен заменик проректор на Ковентри
Универзитетот, Обединето Кралство

Поранешен претседател на IFAC, Ремс, Франција

Раководител на Институтот за науки за системи,
CAS, Пекинг, Кина

Македонска академија на науките и уметностите,
Скопје, С.Македонија

Македонска академија на науките и уметностите,
Скопје, С.Македонија

ФЕИТ Скопје, С. Македонија

ФЕИТ С. Македонија

ФЕИТ С. Македонија

ФЕИТ Скопје С. Македонија

ФЕИТ Скопје, DOU Турција

ФЕИТ Скопје С. Македонија

ФЕИТ Скопје С. Македонија

ФЕИТ Скопје С. Македонија

ПРОГРАМСКИ ОДБОР

Антиќ Д. – ЕФ Ниш Србија
Василеска Д. – ASU САД
Chitkushev L. – BU САД
Domenica Di Benedetto M. – ULA
Италија
Karam L. – ASU САД
Leon-Garcia A. – UT Канада
Parageorgiou M. – CANIA Грција
Пепељугоски П. – IBM Research САД
Perunicic (Drazenovic) B. – ASA БиХ
Поповски П. – AAU Данска
Prasad R. – Aarhus Данска
Sasiadek J. – CU Канада
Стефановска А. – LU Обединето
Кралство
Tzanova S. – TUS Бугарија
Џеровски С. – IJS Словенија
Jozsef K. Tar – Универзитет Обуда
Унгарија
Abousleman G. – ASU САД
Акыokus S. – DOU Турција
Гиновска М. – ФЕИТ Скопје РСМ
Давчев Д. – ФИНКИ Скопје РСМ
Десковски С. – ТФ Битола РСМ
Јаневски Т. – ФЕИТ Скопје РСМ
Јоргушески Љ. – ТНО Холандија
Караџинов Љ. – ФЕИТ Скопје РСМ
Кафеџиски В. – ФЕИТ Скопје РСМ
Лазаревска Е. – ФЕИТ Скопје РСМ
Ојлеска Латкоска В. – ФЕИТ Скопје
РСМ
Ставров Д. – ФЕИТ Скопје РСМ
Кухар А. – ФЕИТ Скопје РСМ
Ефнушева Д. – ФЕИТ Скопје РСМ
Денковски Д. – ФЕИТ Скопје РСМ
Ѓорески Х. – ФЕИТ Скопје РСМ
Јакимовски Г. – ФЕИТ Скопје РСМ
Нацински Г. – ФЕИТ Скопје РСМ
Марковска Димитровска М. – ФЕИТ
Скопје РСМ
Марковски Б. – ФЕИТ Скопје РСМ
Пејоски С. – ФЕИТ Скопје РСМ
Раковиќ В. – ФЕИТ Скопје РСМ
Карталов Т. – ФЕИТ Скопје РСМ
Гераров Б. – ФЕИТ Скопје РСМ
Андова В. – ФЕИТ Скопје РСМ

Атанасова С. – ФЕИТ Скопје РСМ
Шуминоски Т. – ФЕИТ Скопје РСМ
Латкоски П. – ФЕИТ Скопје РСМ
Коколански Ж. – ФЕИТ Скопје РСМ
Србиновска М. – ФЕИТ Скопје РСМ
Димчев В. – ФЕИТ Скопје РСМ
Лошковска С. – ФИНКИ Скопје РСМ
Митровски Ц. – ТФ Битола РСМ
Мурговски Н. – ЦУТ Шведска
Нешковик А. – ЕТФ Србија
Pale P. – FER Хрватска
Периќ З. – ЕФ Србија
Поповски Б. – ФЕИТ Скопје РСМ
Susmann P. – NUARI САД
Schaes L. – IBM Research САД
Спасевска Х. – ФЕИТ Скопје РСМ
Станковски Т. – Мед.Ф Скопје РСМ
Стојанов Г. – AUP Франција
Стојановски Т. – Evolve Information
Services Австралија
Тентов А. – ФЕИТ Скопје РСМ
Гавровски Ц. – ФЕИТ Скопје РСМ
Ќосев Ј. – ФЕИТ Скопје РСМ
Хаџијски М. – БАН Бугарија
Хаџи-Велков З. – ФЕИТ Скопје РСМ
Андреевски Ц. – ФТУ Охрид РСМ
Атанасовски В. – ФЕИТ Скопје РСМ
Аџковска Н. – ФИНКИ Скопје РСМ
Богданоски М. – В. Академија Скопје
РСМ
Гаџовски З. – ЕУ Скопје РСМ
Ѓорѓевиќ Д. – ФИНКИ Скопје РСМ
Здравкова К. – ФИНКИ Скопје РСМ
Ивановски З. – ФЕИТ Скопје РСМ
Јолевски И. – ФИКТ Битола РСМ
Калаџиски С. – ФИНКИ Скопје РСМ
Костов М. – ТФ Битола РСМ
Краљевски И. – voiceINTERconnect
GmbH Германија
Кулаков А. – ФИНКИ Скопје РСМ
Ололоска-Гагоска Л. – ФЕИТ Скопје
РСМ
Порјазоски М. – ФЕИТ Скопје РСМ
Ралева К. – ФЕИТ Скопје РСМ
Самак С. – Микросам Прилеп РСМ
Стефановски Ј. – ЈП Стрежево Битола
РСМ

Ташковски Д. – ФЕИТ Скопје РСМ	Jambrošić K. – FER Загреб
Трајанов Д. – ФИНКИ Скопје РСМ	Babic Z. – EТFBL БиХ
Трајковиќ В. – ФИНКИ Скопје РСМ	Cernak M. – Logitech Швајцарија
Утковски З. – УГД Штип РСМ	Ćirić D. – ЕФ Србија
Филиповска С. – ФИНКИ Скопје РСМ	Pombo N. – UBI Португалија
Чорбев И. – ФИНКИ Скопје РСМ	Joshua E. – Univ. Malta Малта
Ристески А. – ФЕИТ Скопје РСМ	Braeken A. – Vrije Universiteit Brussel, INDI Белгија
Јовичиќ С. – ЕТФ Србија	Dobre C. – UPB Романиа
Делиќ В. – ФТН Србија	Spinsante S. – UPM Италија
Христовски Д. – МФ Словенија	Grguric A. – Ericsson Nikola Tesla Хрватска
Villavicencio F. – ОбЕН САД	Velez Lapão L. – Universidade NOVA de Lisboa Португалија
Garcia N. – UBI Португалија	Podobnik V. – Univ. Zagreb Хрватска
Goleva R. – NBU Бугарија	Јакимоски К. – ФОН Универзитет
Димитровски И. – ФИНКИ, Скопје РСМ	Хаџи-Велкова Санева К. – ФЕИТ Скопје РСМ
Маџаров Ѓ. – ФИНКИ, Скопје РСМ	Геговска – Зајкова С. – ФЕИТ Скопје РСМ
Mihaescu C. – УС Романија	Настеска С. – АЕК Скопје РСМ
Nassari M. – ВВС Обединето Кралство	
Cosovic M. – ЕТФ БиХ	
Петровиќ И. – ВИШЕР Србија	
Garner P. – Idiap Швајцарија	
Szaszak G. – ВМЕ Унгарија	

ОРГАНИЗАЦИСКИ ОДБОР

Коколански Живко – ФЕИТ, Скопје С. Македонија (Претседател)
Србиновска Маре – ФЕИТ, Скопје С. Македонија
Марковска Марија – ФЕИТ, Скопје С. Македонија
Кижевска Емилија – ФЕИТ, Скопје С. Македонија
Попоска Марија – ФЕИТ, Скопје С. Македонија
Ставров Душко – ФЕИТ, Скопје С. Македонија
Нацински Горјан – ФЕИТ, Скопје С. Македонија
Велковски Бодан – ФЕИТ, Скопје С. Македонија
Шуминоски Томислав – ФЕИТ, Скопје С. Македонија

ETAI 2021

<http://www.etai.org.mk>

CONFERENCE CHAIRMAN

Nikola Ljusev

CONFERENCE VICE CHAIRMEN

Peco Nedelkovski and Goran Stojanovski

ORGANIZERS

ETAI - Society for Electronics, Telecommunications, Automatics and Informatics of the Republic of North Macedonia

Faculty of Electrical Engineering and Information Technologies at “Ss. Cyril and Methodius University” in Skopje, Republic of North Macedonia

HONORARY COMMITTEE

Nikola Jankulovski	Rector of UKIM Skopje N. Macedonia
Sasho Korunovski	Rector of UKLO Bitola N. Macedonia
Ninoslav Marina	Rector of St. Paul the Apostle University, Ohrid N.Macedonia
Imre J. Rudas	Former Rector of Obuda University, Budapest Hungary
Frank Allgoewer	Acting President of the IFAC, Stuttgart, Germany
Mogens Blanke	Emeritus Professor, Tech. Univ. Lyngby, Denmark
A. Talha Dinibutun	Former Rector of Dogus University, Istanbul, Turkey
Rolf Isermann	Emeritus Professor, Tech. Univ. Darmstadt, Germany
Stephen Kahne	Past President of the IFAC, Aeronaut. Univ. Phoenix, USA
Vladislav Y. Rutkovsky	Russian Academy of Science, Moscow, Russia
Kevin Warwick	Former Deputy Vice-Chancellor, Coventry University, UK
Janan Zaytoon	Former President of the IFAC, Reims, France
Ji-Feng Zhang	Director of Systems Science Inst. CAS, Beijing, China
Ljupco Kocarev	FCSE Skopje N. Macedonia
Leonid Grchev	FEEIT Skopje N. Macedonia
Momchilo Bogdanov	FEEIT Skopje N. Macedonia
Ljupcho Panovski	FEEIT Skopje N. Macedonia
Boris Spasenovski	FEEIT Skopje N. Macedonia
Tomislav Dzekov	EU Skopje N. Macedonia
Aksenti Grnarov	USEEU Tetovo N. Macedonia
Liljana Gavrilovska	FEEIT Skopje N. Macedonia
Georgi Dimirovski	FEEIT Skopje, DOU Turkey
Mile Stankovski	FEIT Skopje N. Macedonia
Metodija Kamilovski	FEEIT Skopje N. Macedonia
Tatjana Kolemishavska-Gugulovska	FEEIT Skopje N. Macedonia

PROGRAM COMMITTEE

Antic D. – EF Serbia
 Vasileska D. – ASU USA
 Chitkushev L. – BU USA
 Domenica Di Benedetto M. – ULA Italy
 Karam L. – ASU USA
 Leon-Garcia A. – UT Canada
 Papageorgiou M. – CANIA Greece
 Pepeljugoski P. – IBM Research USA
 Perunicic (Drazenovici) B. – ASA B&H
 Popovski P. – AAU Denmark
 Prasad R. – Aarhus Denmark
 Sasiadek J. – CU Canada
 Stefanovska A. – LU UK
 Tzanova S. – TUS Bulgaria
 Džeroski S. – IJS Slovenia
 Jozsef K. Tar – Obuda University Hungary
 Abousleman G. – ASU USA
 Akyokus S. – DOU Turkey
 Ginovska M. – FEEIT Skopje RNM
 Davchev D. – FINKI Skopje RNM
 Deskovski S. – TF Bitola RNM
 Janevski T. – FEEIT Skopje RNM
 Jorgusheski Lj. – TNO Netherlands
 Karadzinov Lj. – FEEIT Skopje RNM
 Kafedziski V. – FEEIT Skopje RNM
 Lazarevska E. – FEEIT Skopje RNM
 Ojleska Latkoska V. – FEEIT Skopje RNM
 Stavrov D. – FEEIT Skopje RNM
 Kuhar A. – FEEIT Skopje RNM
 Efniseva D. – FEEIT Skopje RNM
 Denkovski D. – FEEIT Skopje RNM
 Gjoreski H. – FEEIT Skopje RNM
 Jakimovski G. – FEEIT Skopje RNM
 Nadzinski G. – FEEIT Skopje RNM
 Markovska Dimitrovska M. – FEEIT Skopje RNM
 Markovski B. – FEEIT Skopje RNM
 Pejovski S. – FEEIT Skopje RNM
 Rakovikj V. – FEEIT Skopje RNM
 Kartalov T. – FEEIT Skopje RNM
 Gerazov B. – FEEIT Skopje RNM
 Andova V. – FEEIT Skopje RNM
 Atanasova S. – FEEIT Skopje RNM
 Shuminoski T. – FEEIT Skopje RNM
 Latkoski P. – FEEIT Skopje RNM
 Kokolanski Z. – FEEIT Skopje RNM
 Srbinovska M. – FEEIT Skopje RNM
 Dimchev V. – FEEIT Skopje RNM

Loshkovska S. – FINKI Skopje RNM
 Mitrovski C. – TF Bitola RNM
 Murgovski N. – CUT Sweden
 Neshkovic A. – ETF Serbia
 Pale P. – FER Croatia
 Peric Z. – EF Serbia
 Popovski B. – FEEIT Skopje RNM
 Susmann P. – NUARI USA
 Schares L. – IBM Research USA
 Spasevska H. – FEEIT Skopje RNM
 Stankovski T. – FM Skopje RNM
 Stojanov G. – AUP France
 Stojanovski T. – Evolve Information Services Australia
 Tentov A. – FEEIT Skopje RNM
 Gavrovski C. – FEEIT Skopje RNM
 Kjosev J. – FEEIT Skopje RNM
 Hadzijski M. – BAN Bulgaria
 Hadzi-Velkov Z. – FEEIT Skopje RNM
 Andreevski C. – FTU Ohrid RNM
 Atanasovski V. – FEEIT Skopje RNM
 Ackovska N. – FINKI Skopje RNM
 Bogdanovski M. – Military Academy Skopje RNM
 Gacovski Z. – EU Skopje RNM
 Gjorgjevic D. – FINKI Skopje RM
 Zdravkova K. – FINKI Skopje RNM
 Ivanovski Z. – FEEIT Skopje RNM
 Jolevski I. – FIKT Bitola RNM
 Kalajdziski S. – FINKI Skopje RNM
 Kostov M. – TF Bitola RNM
 Kraljevski I. – voiceINTERconnect GmbH Germany
 Kulakov A. – FINKI Skopje RNM
 Ololoska-Gagoska L. – FEEIT Skopje RNM
 Porjazoski M. – FEEIT Skopje RNM
 Raleva K. – FEEIT Skopje RNM
 Samak S. – Mikrosam Prilep RNM
 Stefanovski J. – JP Strezevo Bitola RNM
 Tashkovski D. – FEEIT Skopje RNM
 Trajanov D. – FINKI Skopje RNM
 Trajkovic V. – FINKI Skopje RNM
 Trajkovski I. – FINKI Skopje RNM
 Utkovski Z. – UGD Stip RNM
 Filipovska S. – FINKI Skopje RNM
 Chorbev I. – FINKI Skopje RNM
 Risteski A. – FEEIT Skopje RNM

Kalendar M. – FEEIT Skopje RNM
Jovichic S. – ETF Serbia
Delic V. – FTN Serbia
Hristovski D. – MF Slovenia
Villavicencio F. – ObEN USA
Garcia N. – UBI Portugal
Goleva R. – NBU Bulgaria
Dimitrovski I. – FINKI Skopje RNM
Madzarov Gj. – FINKI Skopje RNM
Mihaescu C. – UC Romania
Naccari M. – BBC UK
Cosovic M. – ETF B&H
Petrovic I. – Viser Serbia
Garner P. – Idiap Switzerland
Szaszak G. – BME Hungary
Jambrošić K. – FER Croatia
Babic Z. – ETFBL B&H
Cernak M. – Logitech Switzerland

Ciric D. – EF Serbia
Pombo N. – Un.da Beira Interior Portugal
Ellul J. – University of Malta Malta
Braeken A. – Vrije Un. Brussel, Belgium
Dobre C. – Un.Pol.of Bucharest Romania
Spinsante S. – Un. Pol.delle Marche Italy
Grguric A. – Ericsson Nikola Tesla
Croatia
Velez Lapão L. – Un. NOVA de Lisboa,
Portugal
Podobnik V. – Un.of Zagreb, Croatia
Jakimoski K. – FON University RNM
Hadzi-Velkova Saneva K. – FEEIT Skopje
RNM
Gegovska – Zajkova S. – FEEIT Skopje
RNM
Nasteska S. – AEC Skopje RNM

ORGANIZING COMMITTEE

Kokolanski Zivko - FEEIT, Skopje, N. Macedonia (Chairman)
Srbinovska Mare - FEEIT, Skopje, N. Macedonia
Markovska Marija - FEEIT, Skopje, N. Macedonia
Kizevska Emilija - FEEIT, Skopje, N. Macedonia
Poposka Marija - FEEIT, Skopje, N. Macedonia
Stavrov Dusko - FEEIT, Skopje, N. Macedonia
Nadzinski Gorjan - FEEIT, Skopje, N. Macedonia
Velkovski Bodan - FEEIT, Skopje, N. Macedonia
Suminovski Tomislav - FEEIT, Skopje, N. Macedonia

ПРЕДГОВОР и ДОБРЕДОЈДЕ

Драги учесници,

Во името на одборите на конференцијата ЕТАИ 2021 и во името на Здружението ЕТАИ, ви посакувам добредојде на 15-тата меѓународна конференција ЕТАИ 2021 која оваа година за прв пат се одржува на нов начин во нашето ново виртуелно секојдневие. Со секоја конференција организирана досега се гордееме што на нашите учесници сме успевале да им ја доловиме уникатната атмосфера на нашиот бисер - Охридското Езеро, познато по својата природна убавина и културното наследство и заштитено од УНЕСКО. За жал оваа година и тоа искуство ќе се обидеме да го доловиме виртуелно, со надеж дека следниот пат традицијата ќе продолжи.

Оваа година одбележуваме 40 години постоење на Здружението ЕТАИ. Целокупното искуство собрано низ годините се пренесува на секоја следна генерација ентузијастички вклучени во работата на Здружението, секогаш со безрезервна поддршка од поистакнатите наши членови. Токму затоа сакам да изразам огромна благодарност на дел од членовите на Здружението кои се тука со нас од почетокот па сè до ден денес и секогаш со несмален ентузијазам успеваат да го задржат и подигнат квалитетот на работата на Здружението и конференциите ЕТАИ низ годините.

Секако, целта на оваа конференција со долгогодишна традиција е да не поврзе преку заедничките интереси и да создаде место каде ќе се поттикне професионална размена на најновите научно-стручни искуства од теориски и апликативен карактер во подрачјата на електрониката, телекомуникациите, автоматиката и информатиката и секако мултидисциплинарните достигнувања кои се денешната наша реалност. Оваа година конференцијата ќе се реализира преку повеќе разновидни и интересни сесии каде ќе се презентираат научни и стручни трудови од актуелни теми. Очекуваме дискусиите на тркалезните маси на тема "Next generation wireless communications - Opportunities and challenges for IoT" и "e-Health and Pervasive Technologies: Challenges and Opportunities", кои се дел од овојгодишната програма на конференцијата да предизвикаат љубопитност кај учесниците и публиката и да отворат широка палета на прашања кои во иднина ќе бидат разгледувани од научната и стручната заедница. Исто така, се надеваме дека конференцијата ќе ги продлабочи старите пријателства и ќе создаде нови, ќе развие соработка и вмрежување на научните работници и деловните луѓе и дека за тоа нема да биде пречка начинот на кој сме принудени да ја изведеме.

Сакам да ја изразам мојата голема благодарност кон нашите пленарни говорници: проф. д-р Пенг Ши од Универзитетот од Аделаида, Австралија, академик проф. д-р Јануш Кацпшик од Институтот за истражување системи при Полската академија на науки и проф. д-р Матјаж Гамс кој доаѓа од Институтот Јожеф Штефан од Словенија. Исто така сакам да го поздравам и истакнатиот млад научник д-р Томе Ефтимов исто така од Институтот Јожеф Штефан од Словенија.

Конференцијата ЕТАИ традиционално е поддржана од повеќе организации и ентузијастички. Најпрво, како нераскинлив дел од ЕТАИ, сакам да му се заблагодарам на Факултетот за електротехника и информациски технологии при УКИМ кој е постојан поддржувач на Здружението и конференциите низ годините. Оваа година конференцијата ќе биде поддржана и од неколку здруженија од македонскиот оддел на меѓународна организација IEEE со кои заеднички се организираат неколку сесии. Исто така, оваа година имаме поддршка и од Меѓународниот проект WideHealth од

делот на Tweening програмата на ЕУ, преку заедничка организација на тркалезната маса "e-Health and Pervasive Technologies: Challenges and Opportunities". На крајот, користам прилика да им заблагодарам на сите кои придонесоа за успешна реализација на овогодишната конференција ЕТАИ 2021 и да им оддадам посебно признание на авторите на трудовите, рецензентите, членовите на Почесниот одбор, Програмскиот одбор и Организацискиот одбор, како и на сите вклучени за придонесот и несебичното залагање во организацијата и успехот на конференцијата.

Сигурна сум дека програмата на овогодишната конференција е доволно актуелна и интересна и за академската и за стручната заедница и дека ќе овозможи услови за унапредување и стекнување со нови пријателства меѓу научните работници, инженерите и студентите. Со надеж дека ќе имаме успешен настан, ви пожелувам пријатна работа.

Марија Календар
Претседател на Здружението
ЕТАИ

FOREWORD and WELCOME

Dear participants,

On behalf of the ETAI 2021 Conference committees and the ETAI Society, I welcome you to the 15th International Conference ETAI 2021 this year, for the first time, hosted in a new way to reflect our new virtual everyday life. With every conference organized so far, we have always been very proud to capture the unique atmosphere of our pearl - Lake Ohrid, known for its natural beauty and cultural heritage and protected by UNESCO as well. Unfortunately, this year, we will have to try and capture that experience virtually, hoping that on our next meeting the tradition will continue.

This year we celebrate the 40th anniversary of the ETAI Society. The entire experience gathered over the years is passed onto every next generation of enthusiasts involved in the work of the Society, always with the unreserved support of our more experienced members. That is why I want to express my gratitude to some of the members of the Society who are here with us from the beginning until today and always with undiminished enthusiasm manage to maintain and raise the quality of the work of the Society and the ETAI conferences over the years.

Of course, the purpose of this conference with a long tradition, is to connect us through common interests and to create a place for professional exchange of the latest scientific and practical experiences of theoretical and applied nature, in the fields of electronics, telecommunications, automation and informatics and of course the multi-disciplinary achievements that are our reality today.

This year the conference will be realized through several diverse and interesting sessions where scientific and professional papers on current topics will be presented. We expect that the roundtable discussions tackling the topics of "Next generation wireless communications - Opportunities and challenges for IoT" and "e-Health and Pervasive Technologies: Challenges and Opportunities" that are part of this year's conference program, will arouse curiosity among participants and audiences with intention to open a wide range of issues that will be addressed in the future by the scientific and professional community. We also hope that the conference will deepen old friendships and create new ones, develop business collaboration and networking of scientists and business people, and that the format we are forced to organize it by, will not present an obstacle to these goals.

I would like to express my great gratitude to our plenary speakers: prof. Dr. Peng Shi from the University of Adelaide, Australia, academician prof. Dr. Janusz Kacprzyk from the Systems Research Institute at the Polish Academy of Sciences and prof. Dr. Matjaz Gams from the Jozef Stefan Institute from Slovenia. I would also like to greet the prominent young scientist Dr. Tome Eftimov also from the Jozef Stefan Institute in Slovenia.

The ETAI conference has traditionally been supported by many organizations and enthusiasts. First, I would like to thank as an inseparable part of ETAI, the Faculty of Electrical Engineering and Information Technologies (FEEIT), that is a constant supporter of the Society and our conferences throughout the years. This year, the conference will be also supported by several Chapters of the Macedonian branch of the international IEEE organization, by joint organization of several sessions. Additionally, we have support from the EU funded WideHealth Project from the Tweening track by

jointly organizing the round table "e-Health and Pervasive Technologies: Challenges and Opportunities".

Finally, I take this opportunity to thank all those who contributed to the successful implementation of this year's ETAI 2021 conference and to pay special tribute to the authors of the papers, reviewers, members of the Honorary Committee, Program Committee, Organizing Committee and volunteers as well as all involved for their contribution. and selfless commitment to the organization and success of the conference.

I am confident that the program of this year's conference is contemporary and interesting enough for both the academic and professional community, and it will provide conditions for promotion and making new friendships between scientists, engineers and students. With the hope that we will have a successful event, I wish you all pleasant work.

Marija Kalendar
President of ETAI Society

СПИСОК НА ТРУДОВИ – ЕТАИ 2021

LIST OF PAPERS – ETAI 2021

ПЛЕНАРНИ ПРЕДАВАЊА / INVITED PLENARY LECTURES	17
FORMATION CONTROL DESIGN FOR MULTI-AGENT SYSTEMS	18
Peng Shi.....	18
IS WEB TRANSFORMING OUR MINDS AND WHERE IS OUR CIVILISATION GOING TO?	19
Matjaž Gams	19
HUMAN-IN-THE-LOOP AI IN DECISION AND CONTROL SYSTEMS: THE ROLE OF LINGUISTIC DATA SUMMARIES	20
Janusz Kacprzyk.....	20
TOWARDS AUTOMATED ALGORITHM PERFORMANCE PREDICTION USING PROBLEM LANDSCAPE DATA: A USE-CASE IN SINGLE-OBJECTIVE OPTIMIZATION	21
Tome Eftimov	21
ЕТАИ СЕСИИ/ ETAI SESSIONS.....	22
ETA1 : CIRCUITS AND SYSTEMS.....	23
A SONAR-BASED OBSTACLE DETECTION SYSTEM FOR THE BLIND AND VISUALLY DISABLED	24
Stefana Hristovska, Kristijan Lazarev and Branislav Gerazov	24
LIGHTING DESIGN, AUTOMATION, EFFICIENCY AND ADVANTAGES MADE WITH LIGHTING LEVEL CONTROL IN INDUSTRIAL FACILITIES	28
Mehmet Gürçan Gür and Yilmaz Uyaroğlu.....	28
DETECTION OF INDIVIDUAL FINGER FLEXIONS USING TWO-CHANNEL ELECTROMYOGRAPHY	34
Blagoj Hristov and Gorjan Nadzinski.....	34
THE SELECTION OF BI-FRACTIONAL ORDER REFERENCE MODEL PARAMETERS FOR MINIMUM SETTLING TIME.....	40
Ertuğrul Keçeci, Erhan Yumuk, Müjde Güzelkaya and İbrahim Eksin.....	40
INTRA-NODAL CACHING ASSISTED UAV BASED DATA ACQUISITION FROM WIRELESS MOBILE AD-HOC SENSOR NETWORKS	45
Umair Chaudhry and Chris Phillips.....	45
ETA1 2: CYBER SECURITY AND MATHEMATICS.....	51
ANALYSIS OF SMART HOME SECURITY BY APPLYING MACHINE LEARNING ALGORITHMS.....	52
Irina Senchuk, Ana Cholakoska and Danijela Efnusheva.....	52
NETWORK SECURITY ANALYSIS BY APPLYING MACHINE LEARNING ALGORITHMS.....	58
Martina Shushlevska, Ana Cholakoska, Danijela Efnusheva.....	58
NUMERICAL SOLUTION OF LAPLACE DIFFERENTIAL EQUATION USING THE FINITE DIFFERENCE METHOD.....	64
Bojana Petrovska, Daniela Janeva, Emilija Tasheva and Andrijana Kuhar.....	64
MODELING POPULATION DYNAMICS AND ECONOMIC GROWTH AS COMPETING SPECIES FOR NORTH MACEDONIA	70
Stefan Boshkovski and Sanja Atanasova.....	70
PERFORMANCE OF GRADIENT ALGORITHMS FOR SOLVING LEAST SQUARES PROBLEM	76
Naum Dimitrieski, Katerina Hadzi-Velkova Saneva and Zoran Hadzi-Velkov.....	76

ETAI 3: CONTROL SYSTEMS AND AUTOMATION	82
MULTI-OBJECTIVE OPTIMIZATION BASED FRACTIONAL ORDER PID CONTROLLER DESIGN	83
Erhan Yumuk, Eda Budak, Müjde Güzelkaya and İbrahim Eksin	83
FRACTIONAL INTEGRATING INTEGER ORDER PI CONTROLLER DESIGN FOR THE FIRST INTEGER ORDER PLUS TIME DELAY SYSTEM	90
Erhan Yumuk, Müjde Güzelkaya and İbrahim Eksin	90
FUZZY LOGIC BASED MAXIMUM POWER POINT TRACKING FOR PHOTOVOLTAIC SYSTEMS	95
Zeynep Bala Duranay and Hanifi Guldemir	95
FUZZY-LOGIC OUTPUT-TRACKING CONTROL FOR UNCERTAIN TIME-DELAY DYNAMICAL PROCESSES: EXPLORING TAKAGI-SUGENO FUZZY MODELS	102
¹ Yuan-Wei Jing, ¹ Xin-Jiang Wei, ² Janusz Kacprzyk, ³ Imre Rudas, and ⁴ Georgi Dimirovski	102
DISCRETE-TIME UNSCENTED KALMAN FILTERS WITH OPERATING OF UNCERTAINTIES: STOCHASTIC STABILITY ANALYSIS	108
¹ Yuanwei Jing, ² Jiahe Xu, ³ Peng Shi and ⁴ Georgi Dimirovski	108
COMPLEX MULTI-NETWORKS WITH FAULTY INTER-NETWORK CONNECTIONS: SYNCHRONIZATION VIA NOVEL PINNING-NODE CONTROL	118
¹ Yuanwei Jing, ² Guanrong Chen, ³ Peng Shi, ⁴ Georgi M. Dimirovski	118
ETAI 4: E-HEALTH	124
INSIEME: A UNIFYING ELECTRONIC AND MOBILE HEALTH PLATFORM	125
Primoz Kocuvan, Erik Dovgan, Tine Kolenik and Matjaž Gams	125
A SYSTEM FOR AUTOMATIC DETECTION OF MAJOR DEPRESSIVE DISORDER BASED ON BRAIN ACTIVITY	129
^{1,2} Daniela Janeva, ² Silvana Markovska-Simoska and ¹ Branislav Gerazov	129
PREDICTING TRENDS AND ANOMALIES IN DAILY ACTIVITIES	134
Vito Janko and Mitja Luštrek	134
FINDING EFFICIENT INTERVENTION PLANS AGAINST COVID-19	139
^{1,2} Nina Reščič, ¹ Vito Janko, ^{1,2} David Susič, ¹ Carlo De Masi, ^{1,2} Aljoša Vodopija, ^{1,3} Matej Marinko, ¹ Tea Tušar, ¹ Erik Dovgan, ¹ Anton Gradišek, ¹ Matej Cigale, ¹ Matjaž Gams and ¹ Mitja Luštrek	139
MACHINE LEARNING BASED ANOMALY DETECTION IN AMBIENT ASSISTED LIVING ENVIRONMENTS	144
¹ Ana Cholakoska, ¹ Valentin Rakovic, ¹ Hristijan Gjoreski, ² Bjarne Pfitzner, ² Bert Arnrich and ¹ Marija Kalendar	144
INVESTIGATING PRESENCE OF ETHNORACIAL BIAS IN CLINICAL DATA USING MACHINE LEARNING	148
¹ Bojana Velichkovska, ¹ Hristijan Gjoreski, ¹ Daniel Denkovski, ¹ Marija Kalendar, ² Leo Anthnoy Celi and ³ Venet Osmani	148
IS WEB TRANSFORMING OUR MINDS AND WHERE IS OUR CIVILISATION GOING TO?	153
Matjaž Gams	153
ETAI 5: COMMUNICATION NETWORKS – 5G	157
EVALUATION OF DISTRIBUTED NFV INFRASTRUCTURES FOR EFFICIENT EDGE COMPUTING IN 5G	158
¹ Gjorgji Ilievski and ² Pero Latkoski	158
PARTICLE SWARM OPTIMIZATION (PSO) BASED RESOURCE ALLOCATION FOR DEVICE TO DEVICE COMMUNICATION FOR 5G NETWORK	164
¹ Wisam Hayder Mahdi and ² Necmi Taspinar	164
INVESTIGATION OF EFFECT OF THE PILOT REUSE FACTOR VIA INTELLIGENT OPTIMIZATIONS ON ENERGY AND SPECTRAL EFFICIENCIES TRADE-OFF IN MASSIVE MIMO SYSTEMS	169
Burak Kürşat Gül and Necmi Taşpınar	169
COMPUTING ON THE EDGE: A SYSTEM AND TECHNOLOGY OVERVIEW	175
Marija Poposka and Zoran Hadzi-Velkov	175
MOBILE EDGE COMPUTING SERVICES WITH QOS SUPPORT FOR BEYOND 5G NETWORKS –USE CASES	180
¹ David Nunev, ² Tomislav Shuminoski, ² Bojana Velichkovska and ² Toni Janevski	180

ETAI 6: INSTRUMENTATION AND MEASUREMENTS.....	186
POSITIONAL VALUE MEASUREMENT FOR A ROOK AND KING VS ROOK CHESS ENDGAME ALGORITHM.....	187
Adrijan Bozinovski and Filemon Jankuloski.....	187
OVERVIEW OF SECURITY AND SAFETY SYSTEMS IN THE AUTOMOTIVE INDUSTRY.....	192
Aleksandra Gjorgjievska, Mare Srbinovska and Martin Gjorgjievski.....	192
DESIGN AND EVALUATION OF COLLABORATIVE LEARNING PLATFORM WITH INTEGRATED REMOTE LABORATORY ENVIRONMENT	200
Zivko Kokolanski, Bodan Velkovski, Tomislav Shuminoski, ³ Dušan Gleich, ³ Andrej Sarjaš, ² Ana B. Kokolanska, ² Anita K. Mijovska, ⁴ Matjaž Šegula, ⁴ Matic Podobnik, ⁵ Zlatko Ruščić and ⁵ Tibor Kratofil.....	200
ERROR EVALUATION IN REACTIVE POWER AND ENERGY MEASUREMENTS ADOPTING DIFFERENT POWER THEORIES	205
Kiril Demerdziev and Vladimir Dimchev.....	205
VIRTUAL REAL TIME POWER QUALITY DISTURBANCE CLASSIFIER BASED ON DISCRETE WAVELET TRANSFORM AND MACHINE LEARNING	212
Petar Vidoevski, Dimitar Taskovski and Zivko Kokolanski.....	212
IMPROVING THE EFFICIENCY OF GROUNDING SYSTEM ANALYSIS USING GPU PARALLELIZATION.....	218
¹ Bodan Velkovski, ¹ Blagoja Markovski, ¹ Vladimir Gjorgievski, ¹ Marija Markovska, ² Leonid Grcev, ³ Stefan Kalabakov and ³ Elena Merdjanovska.....	218
ETAI 7: CLOUD AND IOT TECHNOLOGIES	223
TECHNOLOGICAL, REGULATORY AND BUSINESS ASPECTS OF LPWAN IMPLEMENTATION IN IOT	224
Atanas Godzoski, Toni Janevski and Aleksandar Risteski.....	224
EXTENDED PERFORMANCE EVALUATION OF THE TENDERMINT PROTOCOL	230
Jovan Karamachoski and Liljana Gavrilovska.....	230
ANALYSIS OF SECURITY MECHANISMS OF CONTAINERS IN CLOUD	236
Martina Janakieska and Aleksandar Risteski.....	236
THE APPLICATION OF THE INTERNET OF THINGS IN EVERYDAY EQUIPMENT AFFECTS TO HAVE A MORE EFFICIENT AND QUALITY LIFE	242
Aleksandar Risteski and Avni Rustemi.....	242
USER-TO-CLOUD LATENCY PERFORMANCE CHARACTERISTICS IN AN EUROPEAN CLOUD INFRASTRUCTURE	247
¹ Teodora Kochovska, ¹ Marija Kalendar and ² Simon Bojadzievski	247
ETAI 8: ARTIFICIAL INTELIIGENCE IN AUTOMATION	253
FORECASTING DYNAMIC TOURISM DEMAND USING ARTIFICIAL NEURAL NETWORKS	254
¹ Cvetko Andreeski and ² Biljana Petrevska	254
FORECASTING POWER CONSUMPTION FOR RESIDENTIAL SECTOR	261
Aleksandra Zlatkova, Aneta Buckovska and Dimitar Taskovski.....	261
MODULATION CLASSIFICATION WITH DEEP LEARNING: COMPARISON OF DEEP LEARNING MODELS	267
Selçuk Balsüzen and Mesut Kartal.....	267
MACHINE LEARNING APPROACH FOR AUTONOMOUS CONTROL OF VERTICAL CEMENT ROLLER MILLS	274
^{1,2} Othon Manis, ² Mile Stankovski and ² Gorjan Nadzinski.....	274
SELECTING AN OPTIMISATION ALGORITHM FOR OPTIMAL ENERGY MANAGEMENT IN GRID-CONNECTED HYBRID MICROGRID WITH STOCHASTIC LOAD	280
Natasha Dimishkovska, Atanas Iliev and Borce Postolov.....	280
COMPARATIVE ANALYSIS OF DIFFERENT HELIOSTAT FIELD CONTROL ALGORITHMS	286
Ivan Andonov, Vesna Ojleska Latkoska and Mile Stankovski.....	286
ETAI 9: COMMUNICATION TECHNOLOGIES	295
UNCERTAIN AQM/TCP COMPUTER AND COMMUNICATION NETWORKS: FIXED-TIME CONGESTION TRACKING CONTROL USING GAUSSIAN FUZZY-LOGIC EMULATOR	296
¹ Jindong Shen, ¹ Yuanwei Jing, ² Janusz Kacprzyk, ³ Georgi M. Dimirovski.....	296

WIRELESS POWERED ALOHA NETWORKS WITH FIXED USER RATES AND UAV-MOUNTED BASE STATIONS	305
Slavche Pejoski and Zoran Hadzi-Velkov.....	305
PERFORMANCE INVESTIGATION OF BIDIRECTIONAL OPTICAL IM/DD OFDM WDM-PON USING RSOA AS A COLORLESS TRANSMITTER	311
¹ Mahmoud Alhalabi, ² Necmi Taşpinar and ³ Fady El-Nahal.....	311
DEVELOPMENT AND DEPLOYMENT OF A LORAWAN PERFORMANCE TEST SETUP FOR IOT APPLICATIONS	316
¹ Simeon Trendov, ² Eduard Siemens and ³ Marija Kalendar.....	316
ETAI 10: ARTIFICIAL INTELLIGENCE IN BIOMEDICINE	322
THE REPRESENTATION OF SPOKEN VOWELS IN HIGH GAMMA RANGE OF CORTICAL ACTIVITY.....	323
^{1,2} Daniela Janeva, ² Andrijana Kuhar, ² Lidija Ololoska-Gagoska and ² Branislav Gerazov.....	323
SCORPIANO – A SYSTEM FOR AUTOMATIC MUSIC TRANSCRIPTION FOR MONOPHONIC PIANO MUSIC	328
Bojan Sofronievski and Branislav Gerazov.....	328
FACIAL EMOTION RECOGNITION USING DEEP LEARNING	334
Gjorgji Smilevski and Tomislav Kartalov.....	334
AUTOMATIC COMPOSITION OF TEXT AND MUSIC FOR A SONG IN MACEDONIAN USING DEEP LEARNING.....	339
Angela Najdoska, Emilija Kotevska, Tamara Markachevikj and Hristijan Gjoreski.....	339
MACHINE LEARNING AND DATA SCIENCE AWARENESS AND EXPERIENCE IN VOCATIONAL EDUCATION AND TRAINING FOR HIGH-SCHOOL STUDENTS.....	343
Stefan Zlatinov, Branislav Gerazov, Gorjan Nadzinski and Tomislav Kartalov.....	343
TOWARDS A SYSTEM FOR CONVERTING TEXT TO SIGN LANGUAGE IN MACEDONIAN..	347
Stefan Spasovski ¹ , Branislav Gerazov ¹ , Risto Chavdarov ¹ , Viktorija Smilevska ² , Aneta Crvenkovska, Tomislav Kartalov ¹ , Zoran Ivanovski ¹ and Toni Bachvarovski ³	347
ETAI 2021 TECHNICAL PROGRAMME AGENDA.....	351



ПЛЕНАРНИ ПРЕДАВАЊА INVITED PLENARY LECTURES

FORMATION CONTROL DESIGN FOR MULTI-AGENT SYSTEMS

Peng Shi

*School of Electrical and Electronic Engineering
University of Adelaide, Australia*

Abstract: Multi-agent Systems (MAS) are systems with characteristics of cooperation and decentralization. As the agents often work under complex circumstances, limitations of the hardware that include limited passive sensing and active communication capabilities are likely to be present. As a result of the localization conditions above, the agents need to cooperate in a distributed manner to achieve a common goal. Formation control, which is one of the most popular topics within the realm of multi-agent systems, generally aims to drive multiple agents to achieve a desired scalable formation or time-varying formation changing. In this talk, depending on the agents' sensing and interaction capabilities, the analysis and design of a variety of distributed formation control and some applications are introduced. Under complex circumstances, issues on collision avoidance and system robustness for MAS are also addressed. Simulation and Lab experimental results are given to demonstrate the effectiveness of some design schemes proposed in our group.

IS WEB TRANSFORMING OUR MINDS AND WHERE IS OUR CIVILISATION GOING TO?

Matjaž Gams

Jožef Stefan Institute

Ljubljana, Slovenia

matjaz.gams@ijs.si, <https://dis.ijs.si/mezi/>

Abstract: This paper analyses civilization dangers that Bill Gates and Elon Musk warn about: pandemics, demographic and AI. In addition, the undesired societal changes especially in relation to the mind tampering and the Web are observed. The analysis about longevity of human civilization seems to indicate that we humans on our own are not smart enough to find a solution. Luckily, there is a possible solution: to create and cooperate with superintelligence.

HUMAN-IN-THE-LOOP AI IN DECISION AND CONTROL SYSTEMS: THE ROLE OF LINGUISTIC DATA SUMMARIES

Janusz Kacprzyk

Professor, Researcher
Systems Research Institute, Polish Academy of Sciences
Ul. Newelska 6, 01-447 Warsaw, Poland
Email: kacprzyk@ibspan.waw.pl

Abstract: We consider the problem of how to develop an Artificial Intelligence based – or AI based, for short – system for solving complex decision making and control problems, in both engineering and socioeconomic systems. We consider problems in which, in addition to aspects which are subject to an „objective” evaluation by sensors, there are many relevant aspects which are subject to human judgment, intentions, preferences, etc. which are difficult to quantify, subjective, changeable over time, and subject to many cognitive biases, notably the status quo bias. As the presumably most promising architecture for solving such problems we assume the „human-in-the-loop” paradigm, called also the „human-in-the-loop AI” in which it is postulated and implemented a synergistic cooperation between the human being and the „machine”, that is, approaches and algorithms employed. We use the human-in-the-loop paradigm for decision making and control. We argue that the most promising solution in this context is the use of the human centric systems philosophy, originated at MIT, in which no (additional) interface between the human being and the computer is postulated. Therefore, to attain this we advocate to use in the problem formulation and solution (support) some tools and techniques of natural language which is the only fully natural means of articulation and communication for the humans. Specifically, we use the linguistic data summaries, introduced by Yager and then developed by him and Kacprzyk. They are meant to represent large data sets by short, comprehensible sentences. For instance, if we have a (large) set of data on employees, a static linguistic summary can be „most young employees earn around the mean salary”, and a dynamic summary can be „in most recent years the increase of salaries of experienced employees was slightly growing”. Notice that no matter how big the data set is, such short sentences are fully comprehensible for the human being, and imprecise terms are natural. We present various aspects of such linguistic summaries, such as context dependence, representation of language modalities, etc. We show how they can be used for an effective and efficient human-in-the-loop AI based systems for supporting decision making (mostly static summaries) and control (dynamic summaries). We present an implementation for supporting a day-to-day running of a small computer retailer.

TOWARDS AUTOMATED ALGORITHM PERFORMANCE PREDICTION USING PROBLEM LANDSCAPE DATA: A USE-CASE IN SINGLE-OBJECTIVE OPTIMIZATION

Tome Eftimov

Young Researcher
Department of Intelligence Systems,
Jozef Stefan Institute

Abstract: Many real-world scenarios involve optimization problems, for example, when minimizing risks, minimizing cost, maximizing reliability, and maximizing efficiency. For this reason, evolutionary computation focuses on development of algorithms for global optimization inspired by biological evolution. These algorithms are efficient for finding good solutions to NP-hard problems for which solutions cannot be computed in analytical or semi-analytical form, or by using deterministic algorithms. Additionally, in combination with machine learning algorithms they represent powerful techniques for solving many prediction problems in industry. Benchmarking in evolutionary computation is a crucial task and is used to evaluate the performance of an algorithm against other algorithms. Existing approaches for assessing the performance of algorithms are based on a statistical comparison of the algorithms' results focusing only on the performance data. On the other side, efficient solving of an unseen optimization problem is related to appropriate selection of an optimization algorithm and its hyper-parameters. For this purpose, automated algorithm performance prediction should be performed that in most commonly-applied practices involves training a supervised ML algorithm using a set of problem landscape features. To provide more explainability in the algorithms' behaviour, we are going to present a recently proposed approach known as Deep Statistical Comparison that provides more robust results in benchmarking studies focussing only on performance data, followed by recently proposed approaches for representing the optimization problem landscape data. Finally, we are going to present different ML pipelines that are used for automated algorithm performance prediction that link the problem landscape data to the performance data by exploring the relations between the problem and the performance space. Such kinds of analysis are extremely welcome to understand the algorithm's ensemble and stop treating them as a black-box, which will further allow more easily to transfer the learned academic knowledge into industry. Recent work on vision-correcting displays will also be discussed. Given the measurements of the optical aberrations of a user's eye, a vision correcting display will present a transformed image that when viewed by this individual will appear in sharp focus. This could impact computer monitors, laptops, tablets, and mobile phones. Vision correction could be provided in some cases where spectacles are ineffective. One of the potential applications of possible interest is a heads-up display that would enable a driver or pilot to read the instruments and gauges with his or her lens still focused for the far distance.



ЕТАИ СЕСИИ ETAI SESSIONS



ETAI 1 : CIRCUITS AND SYSTEMS

A Sonar-based Obstacle Detection System for the Blind and Visually Disabled

Stefana Hristovska¹, Kristijan Lazarev², and Branislav Gerazov¹

¹Faculty of Electrical Engineering and Information Technologies, UKIM, Skopje, Macedonia

²Open the windows, Skopje, Macedonia

stefanahristovska@gmail.com, klazarev@openthewindows.org, gerazov@feit.ukim.edu.mk

Abstract—People who are blind and visually impaired face many difficulties in their daily lives. They often suffer while performing even the most basic things that lead them to live at risk. We present an obstacle detection system that is an assistive technology device aimed at helping the blind and visually impaired to move more safely and securely in their surroundings. The system is mounted on the head and detects obstacles that appear in the area from the waist to the head. The device is based on three ultrasonic sensors for obstacle detection. The information that an obstacle is detected is transmitted to the user via vibration. Two sensors are placed horizontally at an angle of 0 deg and these sensors detect up to a maximum distance of 250 cm. The third sensor is placed in the central axis and points downwards at a 14 degree angle with a detection range of 257 cm. A power bank is used for a power supply. Additionally, the user can adjust the vibration intensity via a potentiometer. The device is lightweight at only 220 g. Our initial tests have shown that it provides ample navigation assistance and is comfortable to use.

Keywords—obstacle detection; assistive technology; ultrasound; navigation; sonar

I. INTRODUCTION

Blindness is a condition of lack of visual perception due to physiological or neurological factors, partially or completely. Blindness can be caused by hereditary factors, injuries or disease. If a person, even with glasses, contact lenses, or surgery, does not see well, then he is blind, i.e. he has complete visual impairment. Globally, according to the World Health Organization, at least 2.2 billion people have visual impairments, of which 39 million are completely blind. Most people with poor vision live in developing countries and are over 50 years old. Visual impairments cause significant economic costs, directly due to treatment costs and indirectly due to reduced working capacity. In Macedonia according to the National Union of the Blind in the Republic of Macedonia in October 2020 the number of blind people is 2750. And this number is unfortunately increasing.

The blind and visually impaired face many obstacles in moving and performing daily tasks. In order to be able to move more safely and perform everyday tasks, blind people also use devices – assistive technology, that help them. Ranging from simple walking sticks, Braille books, up to more sophisticated voice enabled devices, such as calculators or wrist watches, and smartphones and computers. Some are also assisted by trained dog guides in their daily movement. Assistive devices

have their advantages and disadvantages, e.g. the white stick can easily break, and does not provide any information on obstacles above the waistline, while more advanced devices can be prohibitively expensive.

There are assistive sonar-based devices that alert users of detected obstacles through vibration [1 – 2]. Some of these systems use sound to alert their users [3 – 5]. However, in most of these systems, the obstacle sensors are mounted on a walking stick, which can be suboptimal for detecting obstacles in front of the user in the region of the head [6 – 8]. Although there are a number of obstacle detection and navigation devices on the market today, they are expensive and not easy to afford. In Macedonia, there are no such devices offered to the blind and visually impaired by public health services.

We present a simple obstacle detection system for the blind and visually impaired that is meant to assist the user in navigating their surrounding. The system targets obstacles located above the waistline that are hard to detect using a walking stick. These obstacles, e.g. tree branches and traffic signs, can bring injury to the head of the user, and thus their timely detection is critical. In that sense, the system is meant to augment the use of a walking stick, and not substitute it. The system is designed to be simple, efficient, lightweight, and comfortable for daily use by the blind and visually impaired. It is mounted on the head, but it does not cover it, adding comfort in warm weather conditions.

The system is based on ultrasonic sensors that use sonar to detect obstacles in the path of the ultrasound waves they emit continuously. Obstacles reflect these ultrasonic waves and their presence is then registered by an embedded microcontroller. The microcontroller then alerts the user via vibration of the presence, position and vicinity of the obstacle. There are three ultrasonic sensors in the system – two are placed horizontally at an angle of 0° and detect obstacles at a maximum distance of 250 cm, and one is placed in the middle, tilted downward at an angle of 14° and detects obstacles up to 257 cm away. The system is powered using a power bank. Additionally, a power switch and a potentiometer for vibration adjustment are integrated, allowing the user to tune the vibration feedback to their liking. The system underwent initial testing and the results show that it is comfortable to use while providing mobility assistance.

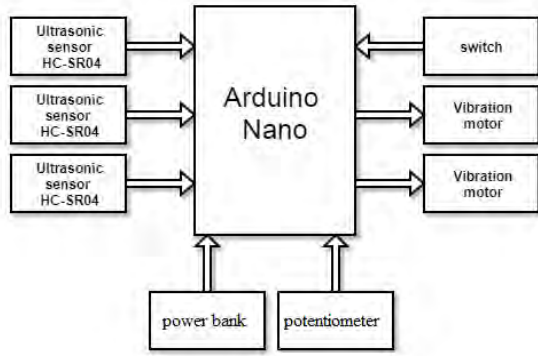


Fig. 1. Block diagram of the obstacle detection system for the blind and visually impaired.

II. SYSTEM DESIGN

The block diagram of the obstacle detection system is shown in Fig. 1. The system is built around a Arduino Nano embedded device, which is the leanest in the Arduino series. The Nano has the needed processing power, and has the advantages of being smaller and lighter, as well as consuming less power. The system uses three HC-SR04 ultrasonic sensors for obstacle detection. Two additional inputs are the optimization switch and the potentiometer for vibration intensity adjustment. As outputs, the system uses two small vibration motors.

The electrical schematics of the system are shown in Fig. 2. The power switch is connected to pin D2, which is grounded with a 10 kΩ resistor. The vibration motors are controlled via Pulse Width Modulation (PWM) and are connected to D5 and D6. The control is realized via transistors BC337 and a flyback diode 1N4001 is included. The ultrasonic sensors are connected to pins D7 – D12. The potentiometer is connected to the analog pin A0.

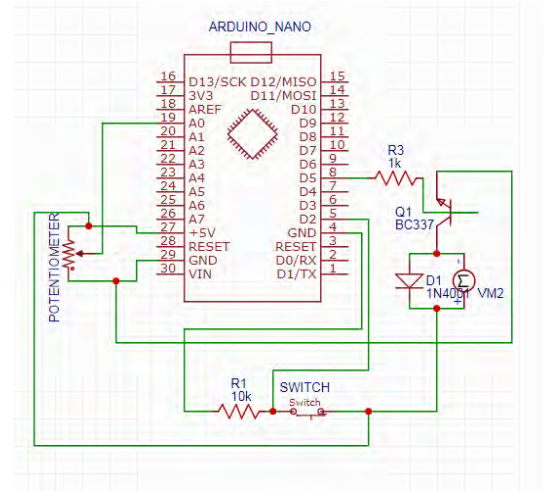
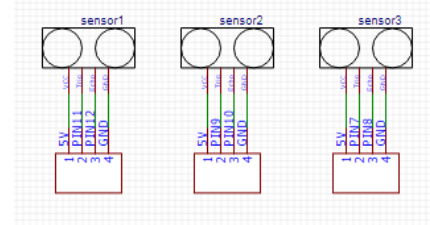
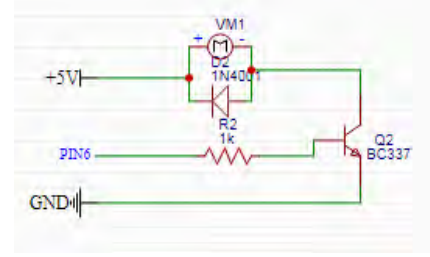


Fig. 3. Electrical schematics for the 5 PCBs for the realization of the obstacle detection system.

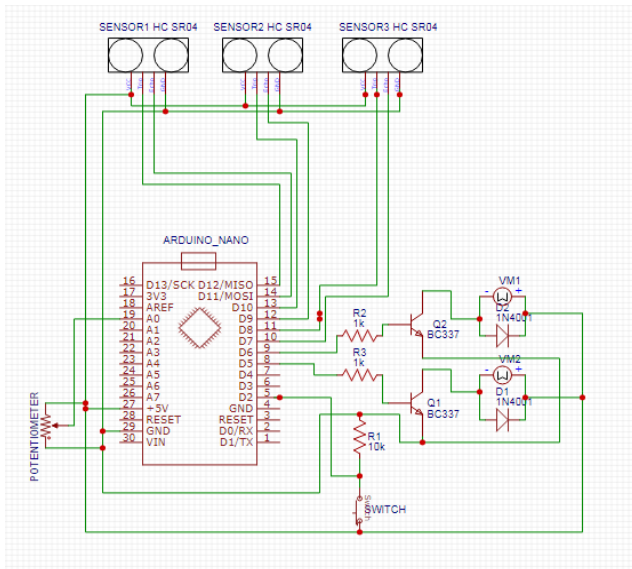


Fig. 2. Electrical schematics of the obstacle detection system.



Fig. 4. The realized PCBs with the assembled components for the obstacle detection system.



Fig. 5. Head mount used for the obstacle detection system.

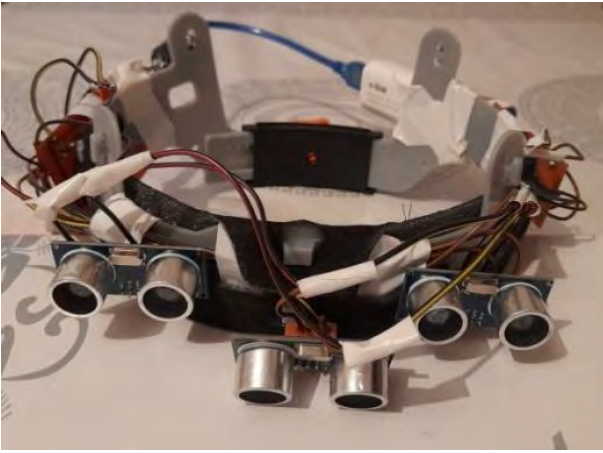


Fig. 6. The assembled prototype of the obstacle detection system.

III. SYSTEM REALIZATION

To make the prototype system, the electrical schematic was split across 5 printed circuit boards (PCBs), as shown in Fig. 3. The first PCB will hold the vibration motor that will be placed on the left side. Three small PCBs will be used for the three ultrasonic sensors, and they will be placed on the front of the system. The fifth and final PCB will house the Arduino, the switch, the potentiometer and the other vibration motor and will be placed on the right side. The PCBs are designed to optimize component placement while reducing overall size. The assembled PCBs are shown in Fig. 4.

The system is designed to be mounted on the head of the user. Mounting it on a cap was not seen as optimal, because it can make wearing the system uncomfortable, especially in warm weather conditions. As a solution the mount mechanism of a construction site helmet was repurposed as the mount for our system.

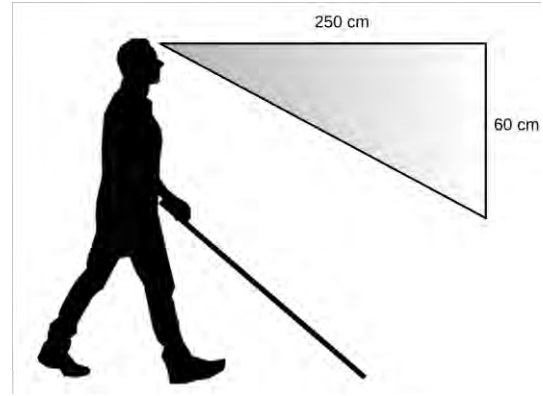


Fig. 7. Target obstacle detection field.

The mount is shown in Fig. 5. It is made of plastic and there is a thin sponge on the front for comfort. On the back it features an adjustment mechanism that can be used to adapt the mount to the head size of the user.

The PCBs are placed on the mount to finalize the prototype of the obstacle detection system, as shown in Fig. 6. Two of the ultrasonic sensors are placed horizontally at a vertical angle of 0° and can detect obstacles up to 250 cm. When one of these two sensors detects an obstacle, the vibration motor mounted on the right will vibrate. The third ultrasound sensor is placed in the middle facing at a 14° angle downwards. The maximum detection distance of 257 cm of this sensor is calculated so that the detection field covered by the sensors is 250 cm horizontally and 60 cm vertically, as shown in Fig. 7.

IV. EVALUATION AND RESULTS

The obstacle detection device for the blind and visually impaired is primarily designed for outdoor use. To evaluate the system, one of our co-authors, who himself is blind, tested the prototype in the premises of the Faculty of Electrical Engineering and Information Technologies. To simulate daily use we assessed the system's performance in several scenarios, including walls and trees and tree branches as obstacles. Fig. 8 shows photos from the evaluation.

The system performed well in the different scenarios. When the user approached a wall the right vibration motor started vibrating at the designed distance of 250 cm. The same was true when the user approached tree branches in the detection field of the system – the system detected the obstacles at a suitable distance and alerted the user via the vibration motors before approaching them too close. Although the system detected most of the tested obstacles, it did not cover obstacles at short distances in front of the shoulders.

The overall impression of using the prototype was positive. The system was very comfortable to use because of the possibility to adjust the mount to the head size, as well as the fact that the device is lightweight.



Fig. 8. Testing the prototype obstacle detection system with different types of obstacles.

Based on the evaluation, several future directions for improvements to the system emerged. One would be to reduce the detection distance to 200 cm or even 150 cm, so that the system would only alert the user of imminent collision danger. Another improvement can be using the motors to communicate the position of the detected obstacles. This can be facilitated by using the right motor to alert about obstacles detected to the right, the left motor for obstacles to the left, and both motors for obstacles directly in front of the user. Finally, the motors could be made to vibrate with different intensity depending on the distance of the detected obstacle.

As additional functionality, a clock and a compass can be added to the system, which would be available at a press of a button. The compass would help for orientation in space, and

the clock for hearing the correct time. These functionalities are currently provided by apps on the smartphone, but it would be better to have them without the need to use a mobile phone.

V. CONCLUSION

Blind and visually impaired pedestrians face a variety of challenges. One of these is the danger of injury caused by obstacle collision in the area above the waistline, and especially the head. These cannot be detected by using the traditional walking stick. We propose an obstacle detection system based on sonar that can be used for assistive navigation for the blind and visually impaired. The system detects obstacles in front of the user using ultrasonic sensors and alerts the user of their presence using vibration. It is meant to assist the user in their mobility as an augmentation to the walking stick.

REFERENCES

- [1] Viswanathan, Kavitha, and Sharmila Sengupta. "Blind navigation proposal using SONAR." 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS). IEEE, 2015.
- [2] Rey, Marina, et al. "Blind Guardian: A sonar-based solution for avoiding collisions with the real world." 2015 XVII Symposium on Virtual and Augmented Reality. IEEE, 2015.
- [3] Bousbia-Salah, Mounir, et al. "An ultrasonic navigation system for blind people." 2007 IEEE International Conference on Signal Processing and Communications. IEEE, 2007.
- [4] Lakde, Chaitali Kishor, and Prakash S. Prasad. "Navigation system for visually impaired people." 2015 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC). IEEE, 2015.
- [5] Harsur, Anushree, and M. Chitra. "Voice based navigation system for blind people using ultrasonic sensor." IJRITCC 3 (2017): 4117-4122.
- [6] Kumar, Krishna, et al. "Development of an ultrasonic cane as a navigation aid for the blind people." 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE, 2014.
- [7] Gupta, Sudeep, et al. "Advanced guide cane for the visually impaired people." 2015 1st International Conference on Next Generation Computing Technologies (NGCT). IEEE, 2015.
- [8] Gbenga, Dada Emmanuel, Arhyel Ibrahim Shani, and Adebimpe Lateef Adekunle. "Smart Walking Stick for visually impaired people using ultrasonic sensors and Arduino." International Journal of Engineering and Technology 9.5 (2017): 3435- 3447.

Lighting Design, Automation, Efficiency and Advantages Made with Lighting Level Control in Industrial Facilities

Mehmet Gürcan GÜR¹

¹Neutec İlaç San.Tic.A.Ş

Technical Operations Directorate, Energy Manager
1.O.S.B. 1.Road 3 Pass Hanli, Arifiye, Sakarya, Turkey
mehmetgur54@gmail.com

Prof.Dr Yılmaz UYAROĞLU²

²Sakarya University Faculty of Engineering
Department of Electrical and Electronics Engineering
Esentepe Campus, Serdivan, Sakarya, Turkey
uyaroglu@sakarya.edu.tr

Abstract— Today, in addition to energy saving and comfort in lighting systems for industry and other areas, the need for control is increasing. It is important to meet this need easily and do so without any technical staff and special training required. Therefore, modern and smart lighting management systems are needed.

With smart lighting systems, control is now provided with a computer, tablet or smart phone via the web interface. No software language is needed. The Internet Protocol IP-based system allows control and monitoring from anywhere. It includes comfort levels and wireless solutions according to European Norm (EN) standards. At the same time, there is the ability to address, periodically check and report emergency lighting and exit signs in accordance with Occupational Health and Safety (OHS) rules.

Lighting methods used to include simple solutions before the use of smart systems, but today, energy efficiency is provided and high savings are achieved thanks to optional control and brightness level adjustment. In addition, it supports environmental practices and efficient maintenance management. In this way, smart lighting systems will be used more in daily life in the near future.

Keywords—Light Level control, Lighting Design, Energy Efficiency, Lighting Automation

I. INTRODUCTION

The lighting system and control are introduced in many areas of the current life and are now done with advanced control methods. This reveals the most suitable lighting in terms of light quality and quantity. Automatic control of lighting also allows for more effective use of consumption and expenditure. Thus, active energy savings are achieved [1]. Parallel to this, the use of lighting automation systems by enterprises is based on some reasons. These include [2];

- Reducing rising energy costs
- The chance to save money on continuously running equipment
- Green buildings have a positive effect on the image of businesses
- Laws and regulations issued by the states and ministries in the form of energy efficiency law
- Effects of environmental rules and procedures

Energy-saving applications are often done with presence and daylight sensors, along with an infrastructure with scenarios and software. It is common for use in industry and commercial buildings. 17% to 60% efficiency rates are reached [3]. Smart lighting systems include conserving equipment as well as human-focused lighting sensors. These are controlled by intelligent algorithms and artificial intelligence applications. It includes building management systems, industry, agriculture and architecture [4]. Applications in this field generally involve increasing the efficiency factor [5], increasing the quality of light, adjusting the circadian rhythm, and increasing the growth rate of plants [6].

Autonomic features such as self-configuration, automatic control and self-healing will be of great benefit in the use of technology. Thus, new types of advanced manufacturing and industrial processes for machine-human cooperation and symbiotic product realization will emerge. As a result of all this, an unprecedented level of operational efficiency will begin to be achieved [7].

At this point, a high-ceilinged drug store is discussed to evaluate the possibilities and applications of energy-efficient

lighting systems and automation. The goal here is good lighting and adequate light for the work done and no glare and glare at the light source or at work, i.e. at the point of operation. Homogeneous lighting is targeted, and proper contrast between work and background is considered essential. It is also important that the light source and the work are in the appropriate colors [8]. Both design and automation were evaluated and analyzed with the advantages they provided and the results were reached. At this point, a high-ceiling drug warehouse is discussed to evaluate the applications mentioned. The goal here is to provide good lighting and sufficient light for the work being done. The other goal is to avoid any glare and flash at the light source and operation point. In addition to a homogeneous lighting, it is imperative to maintain the appropriate contrast between work and background. It is also important to have the light source and the job in the appropriate colors. Analyzes have been conducted and results have been achieved by assessing both design and automation and benefits.

II. HIGH CEILING DRUG WAREHOUSE LIGHTING AUTOMATION APPLICATION AND DESIGN ALGORITHM

The algorithm followed during the lighting application in the warehouse shown in Fig. 1 and the steps of the studies are given under the following topics.

- Selection and features of lighting equipment
- Selection of the angle of light beam
- Selection of the controller
- Adjusting the control element according to the application
- Luminare and sensor placement
- Luminous level calculations
- Software and Communication Protocol Used in Lighting Automation



Fig. 1. High Ceiling Lighting Design in Pharmaceutical Warehouse

A. Selection and Features of Lightning Equipment

When we examine the features of LED luminare in Fig. 2



Fig. 2. High Ceiling Led Luminare Used

- Long operation in aggressive environments
- Having a high quality heat absorbing system, constant light level and lighting power and so able to maintain it for a long time
- High efficiency factor up to (120 lm/W), and able to give the full total light flux, and low light vibration coefficient (less than 1%), which prevents disability glare
- Low light vibration coefficient (less than 1%), which prevents disability glare
- Color rendering index 85 Ra

B. Selection of the Angle of Light

Due to ceiling height 12 meters; for the efficient illuminance level, the light angle was chosen as 30° as in Fig.3.

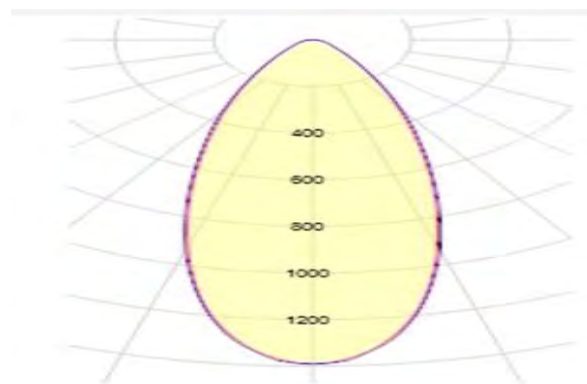


Fig. 3. High Ceiling Led Luminaire Light Angle 30°

C. Selection of the Control Equipment

Warehouse systems contain very large and unused areas. There's only a need for lighting in the event of an operation. The PIR sensors, shown in Fig. 4, are used for the use of the high ceiling and narrow aisle for the control. These sensors and software enable high detection and sensitivity to determine non-operating areas. It is aimed to save energy by keeping the lights on standby.



Fig. 4. Automatically Controlled Presence PIR Detector

These passive infrared sensors can operate optionally automatic, manual and daylight. It is also compatible with communication systems. The operating range according to the ceiling height and sensing distance is shown in Fig. 5.

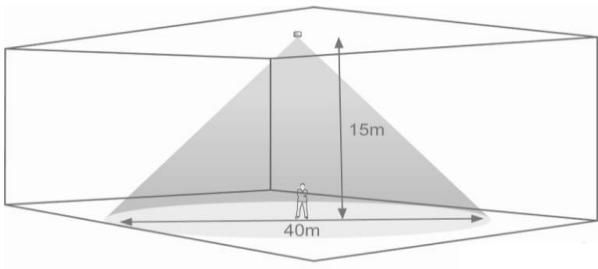


Fig.5. Detection Distances Based on Height

The shelf spacing in the warehouse is 50 meters and the distance from the hanging point of the sensor to the ground is 12 meters.

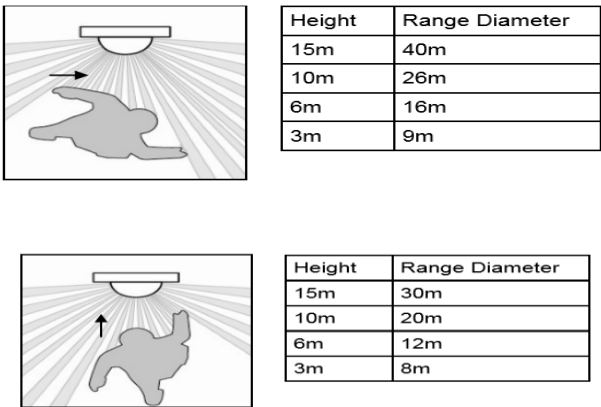


Fig. 6. Sensor Detection Range by Walking Angle

Based on this height and shelf space distance, it was decided to use 2 sensors according to the detection range table given in Fig. 6.

D. Adjusting the Control Element According to the Application

In the design, the masking cover shown in Figure 7 is used to customize the detection in the narrow corridor. The masking cover attached to the sensor is adjusted both transversely and longitudinally, and the 360-degree detection area is adapted only to the relevant shelf area, preventing motion detection from the side shelves

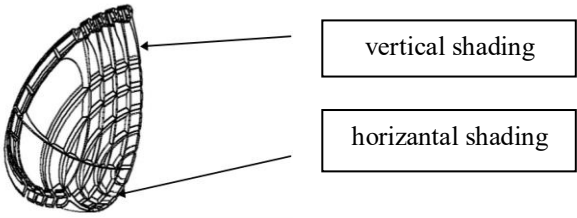


Fig. 7. Sensor Masking Covers

This will allow the final detection angle to be in Fig. 8. At the same time, motion detection areas are as shown in Fig. 9.

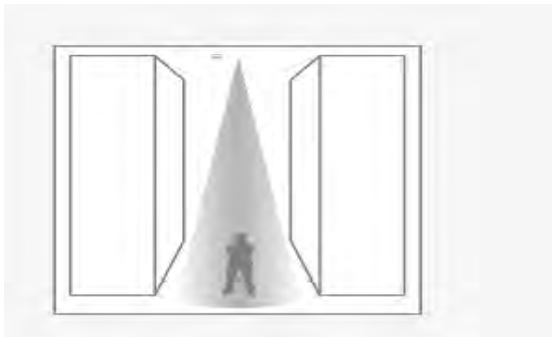


Fig. 8. Narrow Aisle Between Shelf Sensor Detection Angle

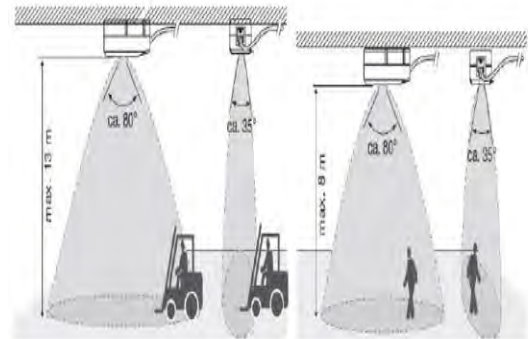


Fig. 9. Sensor Detection Angles According to Forklift and Human Movement

E. Luminaire and Sensor Placement

The luminaire layout, which is created by dialux program by taking the minimum illuminance of 250 lux, is as in Fig. 10.

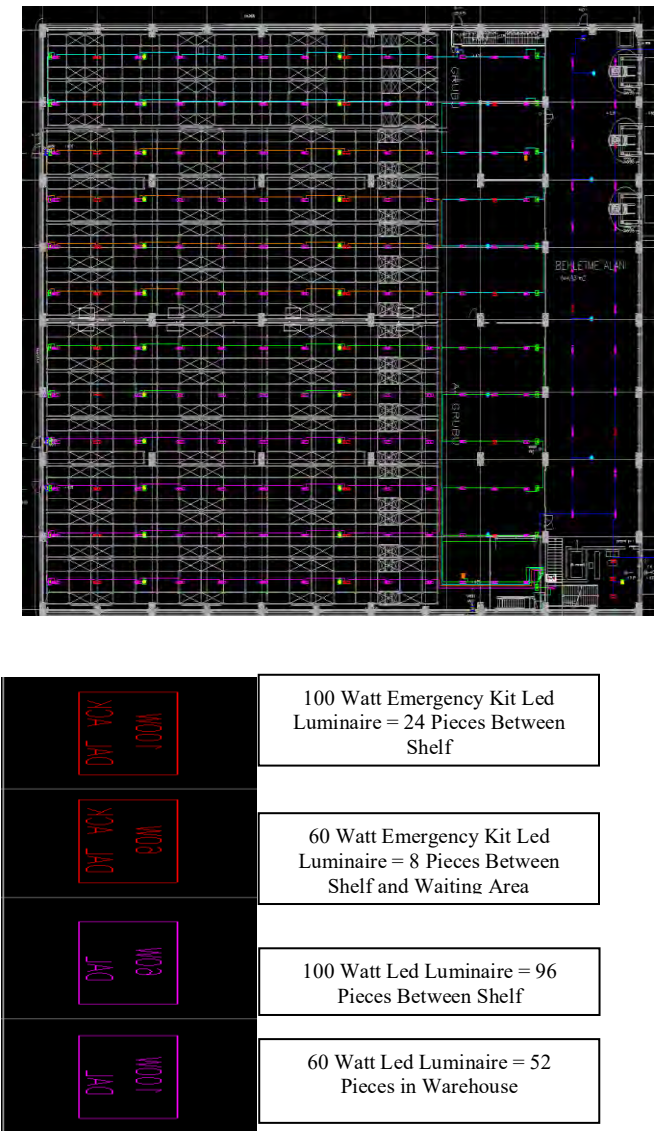


Fig.10. Luminaire Layout Plan

As explained above, the sensors were calculated as 24, 2 in each 12 shelf space. A total of 32 motion sensors are installed in the warehouse, 4 in front of the shelf and 4 in the waiting area.

F. Illuminance Level Calculation

The illuminance level values in Fig. 11-12, were reached by using the dialux program on the related design. Data above 250 lux, which is the minimum luminous intensity, were obtained.

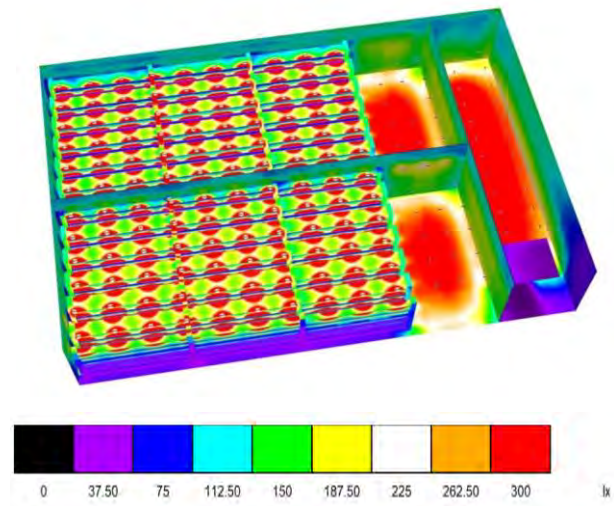


Fig. 11. Luminance Level Outputs

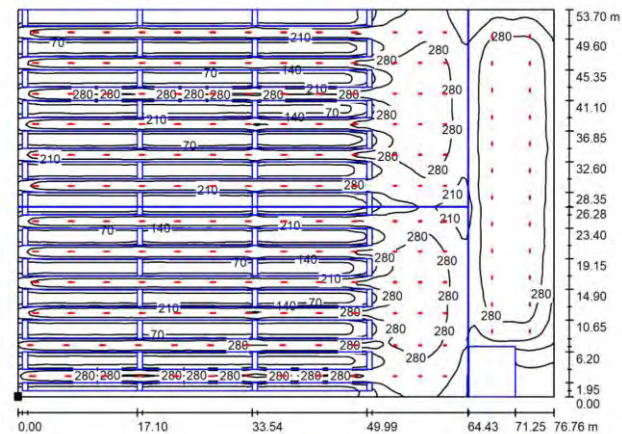


Fig.12. Illuminance Level Values

G. Software and Communication Protocol

WAGO Lighting Management software and DALI communication protocol were used in the warehouse. It is widely used in large and multi-storey structures containing many control areas such as production facilities, warehouses, hospitals. DALI communication protocol complies with the IEC 62386 technical standard. This protocol provides monitoring and control of many parts in the system, such as electronic ballast, emergency lighting kits, sensors and switches. The key features of this software include remote control, intelligent timing based on user-defined scenarios, and both online and real-time monitoring. With advanced reporting techniques, it can show the remaining lamp life and device status. At the same time, it is the display of the locations of all devices on the map, compatibility and interoperability with different brands and models. In Fig. 13, lighting automation

was performed without the need for software language using the web-based lighting control interface.

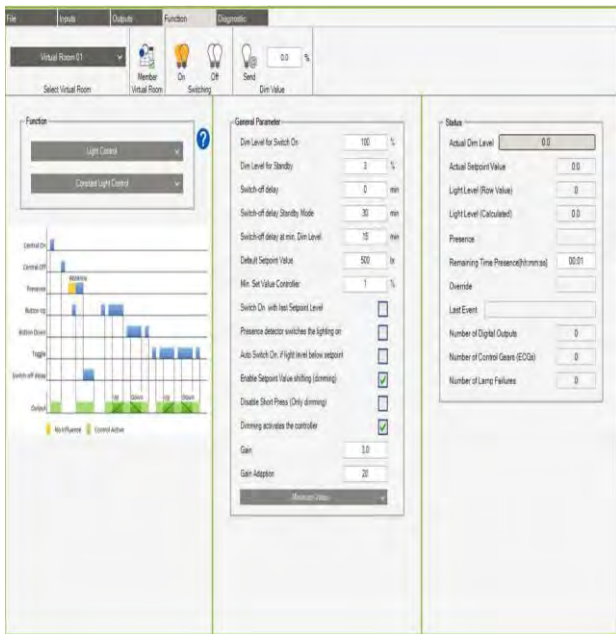


Fig. 13. Web Based Lighting Control Interface

III. ASSESSMENTS

A. Evaluation in Terms of Total Efficient Maintenance Management

In the study, the needs and processes of the business were evaluated. Changing the brightness of the illuminations, operating with presence and motion sensors, benefiting from daylight, constant light function and calendar feature are used. It is provided that the user can enter the time zones automatically on and off during special and public holidays, starting, ending and lunch break. It is provided that the user can enter the time zones automatically on and off during special and public holidays, starting, ending and lunch break. This design and automation provides the following evaluations for maintenance, failure and business management.

- Simple commissioning steps with guided configuration that controlling via the web interface without having to install any program and able to remote access to software and parameters.
- Easy modification of parameters and automatic documentation by factory personnel without any programming effort.
- All equipment's are addressed and identified also software content suitable for planned maintenance

- Detailed diagnostic systems based on maintenance schedules, alarm tables, status indicators and operating hours
- Thanks to the recording of the breakdown and maintenance times, the need for stock and spare parts can be calculated. In this way, it provides the opportunity to save on warehouse and spare parts costs.
- Easy to understand for renovation and new installations,
- Reduced lifecycle costs due to effective light management and the data obtained results within the scope of Quality, ISO, Environmental and Energy Management Systems.

B. Evaluation in Terms of Energy Management Systems

Online energy monitoring module in the warehouse application enables current, voltage and power factor monitoring over 3 phases. In addition, active, reactive, visible power monitoring can be calculated both instantaneous and total amount of energy consumed. This means that the energy savings, depreciation time and efficiency calculation in the warehouse are given in the following topics.

This design consumes 180 LED luminaires in total. The installed power is 15.6 kW. By applying two different methods in this warehouse, the amount of savings and differences were revealed by using both a constant illuminance level (250 lux) and a motion sensor. Measurement criteria are explained. The related measurements were carried out over the same work order, shift schedule and holiday day in 7-day weekly periods. In the case of stand-by on both measurements, the lights are left at 20% safety level. Automation is disabled in reference energy measurement. Lighting control is done through the switch. It is left to the initiative of the personnel to turn off the lighting during breaks, lunch breaks and non-working times.

TABLE I. MEASUREMENT METHOD 1

Measurement Method 1	Consumed Power in 1 hour (kWh)	Total Monthly Consumption (kWh)	Yearly Consumption (kWh)
Automation disabled	11,018	7.932,96	95.195,52
With motion sensor	3,827	2.755,44	33.065,28
Difference (kWh)	7,191	5.177,52	62.130,24
Percent Efficiency Rate (%)	%66		

TABLE II. MEASUREMENT METHOD 2

Measurement Method 2	Consumed Power in 1 hour (kWh)	Total Monthly Consumption (kWh)	Yearly Consumption (kWh)
Automation disabled	11,018	7.932,96	95.195,52
With motion sensor and illuminance level control	2,984	2.148,48	25.781.76
Difference (kWh)	8,034	5.784,48	69.413,76
Percent Efficiency Rate (%)	%73		

The measurements made are given in Table 1. and Table 2. The results in the table are obtained by taking the 1 hour consumption calculation unit time in the measurement interval. With operating method 1, only the motion sensor was activated and a 66% savings was achieved. In Operation Method 2, both the motion sensor and the 250 lux constant illumination level are set. Thanks to the illumination level control, the energy saving rate has increased to 73%, and the comfort and lighting quality have increased. By taking the average energy unit price of 0.45 TL in year 2021, the annual saving amount is 31,236,19 TL over 69,413,76 kWh.

In this study, the initial investment cost of lighting automation and equipment is 10.000 Euros. Efficiency calculation, over the savings rate of 73% and the annual savings of 69,413,76 kWh; If the average energy unit price in the industry for 2021 is 0.45 TL = 0.05625 Euro (1Euro = 8TL)

Automation Cost = 1 Year Amount of Savings x Unit Energy (Euro) x Amortization time than 10.000 Euro = 69.413.75 X 0.05625 Euro x Amortization Time

Amortization Time = 2.56 years.

IV. CONCLUSION

Lighting calculations, design algorithm, equipment selection, software and programming logic are mentioned in the high ceiling drug warehouse discussed. Assessments have been made for maintenance and operating methods. If the energy saving methods, efficiency and depreciation calculations depending on the measurement results are listed

- A more comfortable and efficient lighting was provided by adjusting the fixed brightness level with motion and daylight sensors. While 66% efficiency was achieved with the motion sensor alone, 73% savings were achieved when the constant illumination level (250lx) was controlled together. At the same time, the annual savings with this method were calculated as 69,413,76 kWh.

- According to the savings made with Lighting Automation Systems, the amortization time is 2.56 years only on the investment cost. It is predicted that this period will be shortened if maintenance, operation, control and energy costs increasing every year are added.
- With the DALI communication protocol and lighting management web interface, automation was provided without the need for software language, resulting in high energy savings. The system has been seen to be comfortable, flexible and future oriented. With its commissioning, it is stated that it provides many convenience by the user, operator and caregivers
- The main aim of this study is to show the differences of lighting automation systems, to reveal their easiness of use and concrete data in terms of energy saving. It is aimed to lead the investments to be made related to the energy saving percentages, depreciation calculation and the evaluations made and the results. It is also important in terms of encouraging the planned studies in this area and shedding light on the future scenarios and improvements.

REFERENCES

- [1] Çolak, N., 2003. Lighting Control and Lamps, Best Magazine, Issue 19.
- [2] Alsat, C., (2016). Lighting Automation and Energy Saving Systems, EEC Integrated Building Control Technologies A.Ş.
- [3] B. Von Neida, D. Manicria, A. Tweed, "An analysis of the energy and cost savings potential of occupancy sensors for commercial lighting systems", J. Illum. Eng. Soc. 30 (2) (2001) 111–125.
- [4] J.H. Oh, S.J. Yang, Y.R. Do, "Healthy, natural, efficient and tunable lighting: four-package white leds for optimizing the circadian effect, color quality and vision performance", Light Sci. Appl. 3 (2) (2014) e141.
- [5] R.F. Karliceck, "Smart lighting-beyond simple illumination," 2012 IEEE Photonics Society Summer Topical Meeting Series, IEEE (2012) 147–148.
- [6] G.D. Massa, H.-H. Kim, R.M. Wheeler, C.A. Mitchell, "Plant productivity in response to led lighting", HortScience 43 (7) (2008) 1951–1956.
- [7] Thames, L. and Schaefer, D., 2016. "Software- defined cloud manufacturing for industry 4.0.", Procedia CIRP, vol. 52, pp. 12-17.
- [8] Ilıcak, Ş. 1987. Environmental-Workplace Conditions and Ergonomic Approaches, First National Ergonomics Congress, National Productivity Center, Istanbul.

Detection of Individual Finger Flexions Using Two-channel Electromyography

Blagoj Hristov¹, Gorjan Nadzinski²

Ss. Cyril and Methodius University

Faculty of Electrical Engineering and Information Technologies

Skopje, North Macedonia

e-mail: ¹blhrist@gmail.com, ²gorjan@feit.ukim.edu.mk

Abstract—Due to technological advances in biomedical engineering, electronics, 3D printing and artificial intelligence, there has been a significant increase in the feasibility of producing accurate, fast and fully functional prosthetics. This paper discusses the possibility of designing an artificial forearm prosthetic with the ability of individual control of each finger, which involves using electromyography signals measured by only two sensors placed on the remaining part of the amputated arm. This approach enables crucial improvement in prosthetic control for only a fraction of the cost of prosthetics that are currently available on the market, as expensive hardware components are replaced by inexpensive software solutions. The detection and classification of the type of movement that is being performed is done by using a hybrid convolutional-recurrent neural network. The network provides satisfactory results with the F1 score of the signal predictions being 91.3%. These results are significant because of the fact that using only two-channel electromyography to measure the signals notably reduces the cost of the prosthetic, due to the need for a smaller number of EMG sensors.

Keywords—electric prosthetics, electromyography, signal processing, artificial intelligence, neural networks

I. INTRODUCTION

The number of amputees in the world is rising with each following year, with an increase of 24% in amputations caused by diabetes in the period between 1988 and 2009 [1]. While this statistic might seem disappointing, treatments for amputees have also been constantly improving and becoming more available as medical technology advances. One of these treatments is to replace the lost limb with an artificial prosthetic. In the past, these prosthetics have been purely mechanical and non-functional, meaning they could not replace the functionality of a real arm or leg. Today, more and more prosthetics are being built with functionality in mind, and they have become electrical devices that can be controlled and operated by the user. Unfortunately for amputees, the prosthetics that offer full and fluid functionality are extremely expensive and most regular people could not afford them [2]. The motivation to increase availability for these advanced prosthetics has led to the idea of designing an electric arm prosthetic that achieves decent and robust functionality at a reasonable price.

The main reason high-end prosthetics are so expensive, besides the outer material costs, is the way they achieve such fluid movement and functionality. To be able to mimic a real arm, they must accurately detect what movement the user is trying to make with the prosthetic. This can be done in multiple ways, but one of the most advanced methods is by using electromyography or EMG. This medical technique is used to measure electrical impulses in muscles caused by neuron activation, so the user would be able to control the prosthetic with their brain, just as they would control their real arm.

Electromyography is used in most modern functional prosthetics, although the approach to the classification of the signals as well as the types of movements that can be made vary between them. For example, Open Bionics' "Hero Arm" allows users to toggle between preset grip modes by holding an open hand signal for more than one second, after pressing a button on the back of the prosthetic with their other hand [3]. This, as well as the fact that the wrist also needs to be manually rotated, leads to much room for improvement for future prosthetics. Another example is Ottobock's "Bebionic" hand which is able to perform 14 different grips by pressing a button, some of which can only be accessed by manually rotating the prosthetic's thumb into a specific position [4]. Therefore, the main goal of this paper is to try and improve upon these limitations by attaining adequate to complete functionality of the individual fingers of the prosthetic using just two inexpensive surface EMG sensors placed on the user's forearm.

EMG signals have already been successfully used in multiple cases for decoding specific movements from measurements taken from the forearm muscles. For example, [5] achieves a ~93% accuracy with only two sensors for the detection of four different movements by using neural network classifiers, although instead of the raw signal they use extracted time-domain features. This feature extraction step can take away precious processing time which is needed to be able to make a fast and responsive prosthetic. Another approach that implements neural networks and focuses on individual finger movement detection is proposed in [6] and achieves up to ~98% accuracy. To be able to accomplish this 32 electrodes were used, which would significantly increase the price of the prosthetic as well as make it much less user-friendly.

To be able to use simple dual-channel electromyography without much preprocessing for this complicated task, we use advanced signal processing techniques and implement a hybrid convolutional-recurrent neural network for the detection and classification of the individual finger movements performed by the participant.

This paper is organized as follows:

In Section 2 we describe the use of EMG sensors for signal acquisition. Then, in Section 3 we present the filtering, rectification, and segmentation of the obtained signals in order to extract the meaningful data. Section 4 shows the process of signal classification in order to determine the required movement, before Sections 5 and 6 close the paper by presenting the results and the concluding discussion respectively.

II. SIGNAL ACQUISITION

As mentioned before, we use two EMG sensors to measure the electrical impulses in the forearm of the user. The dataset that is used in this paper is from [7], which measures the EMG signals from eight willing participants. The data is gathered by placing one of the sensors on the flexor digitorum profundus muscle and the other on the extensor digitorum, as shown on Figure 1.

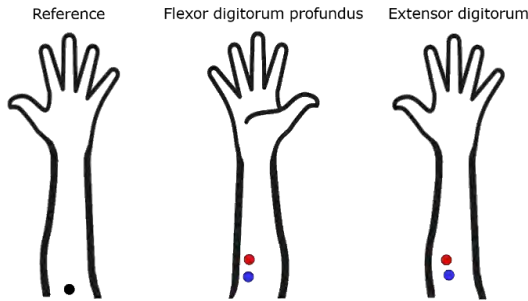


Figure 1. Location of EMG sensors on forearm

The sampling rate at which the data is measured is 4 kHz, and the resolution of the A/D converter of the data acquisition

unit is 12 bits. The types of movements that are included in the dataset are the flexion of each individual finger of the hand, as well as whole hand flexion i.e. forming a fist. By creating a model that can distinguish among these six movements, we will be able to show that two-channel electromyography is accurate enough to simulate adequate or even complete functionality of the prosthetic arm. If the model is successful, in theory we could easily expand its library of possible movements by training it further with new data. An example of what the measured data from one participant in the dataset looks like is given in Figure 2. It is obvious that there are noticeable differences between the signals of each individual movement, as well as between the electrodes, but the amplitude of the signals may vary significantly from person to person. To be able to account for the variability of the measurements we need to create a classifier that is robust and will not be affected by outliers in the data. For this reason, one of the most important steps in this process is the preprocessing of the signals, which we will cover next.

III. SIGNAL PROCESSING

Before we can train a classifier to identify the movements performed in the measured signals, we need to thoroughly analyze and process the data so that we can achieve more accurate results.

A. Signal Filtering

The first method used is the adequate filtering of the measured signals. Most of the important part of the EMG spectrum is located between 20-450 Hz [8], meaning that all frequencies outside of this scope can be safely filtered out using an infinite impulse response band-pass filter. Another noise that is present in the data during the measurement procedure is the so called “mains hum”, which is generated by the electrical power grid located at a frequency of 50 Hz (or 60 Hz for most North and South American countries). To deal with this additional noise we use a notch filter (also called a band-stop filter), which removes the tiny sliver of the signal’s spectrum located at and around the desired frequency.

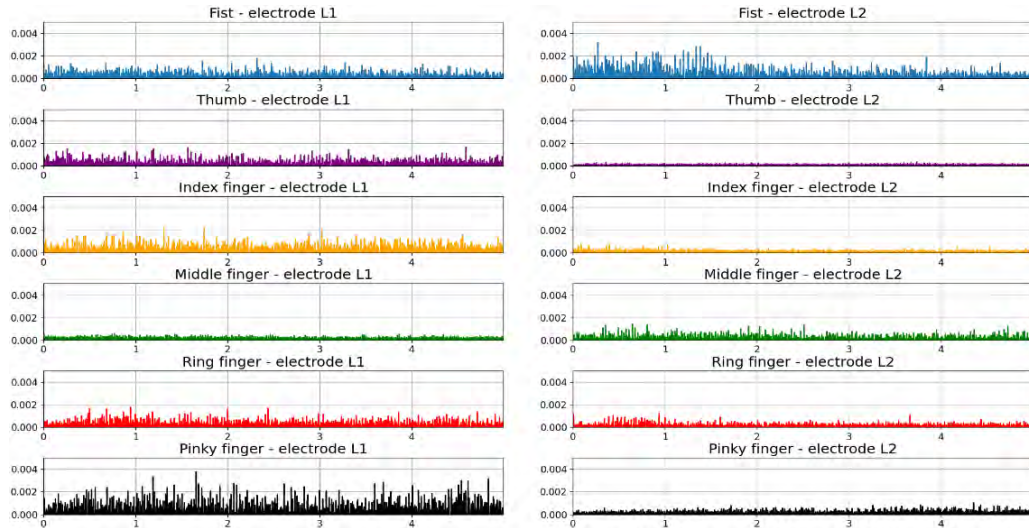


Figure 2. Measured EMG signals from participant 1.

B. DC component and rectification

After the signal has been properly filtered, the next step is to remove the DC bias and rectify it. Because EMG signals are periodic they oscillate between positive and negative numbers, centering on a certain value. This value ideally should be 0 mV, but due to the DC component of the signal (which is a constant level at 0 Hz), it is usually offset by some small amount. The removal of this component is simple, as we can just subtract the mean of the whole signal from each of its values. Unfortunately, this leads to having a new mean of zero which is not an attribute that we want for our data. To amend this without having to remove the negative part of the signal, we calculate its absolute value which preserves all of the information that the signal carries within it.

C. Segmentation

In order for the prosthetic to be used without the users feeling a significant delay in the desired movements, it is necessary for the classification of the signals to take place in a very short period of time. The minimum delay that is unnoticeable to the user is around 250 ms [9] [10]. To be able to achieve this limitation we designate 125 ms for the classification period, so that an additional period of 125 ms is left for the activation of the actuators in the prosthetic and their movement to the required position. In order to ensure that the classification would occur in the given time period, segmentation is performed on the measured signals. Using a sliding window, each signal is segmented into small segments with a length of 250 ms, while the window slides with a step of 125 ms. This means that during the use of the prosthetic, the measurement recorded for each period of 125 ms is then combined with the measurement from the previous 125 ms to form the complete 250 ms segment, which will then be sent to the classifier. The reason for this approach is to maintain the given time limit while also using the information from the previous 125 ms, which will help the classifier to make more accurate predictions. By applying this procedure, the classifier must now learn to recognize individual

segments instead of the whole 5 second signal, which in a real use case situation would not be available. An example of how the segmentation process look like is given in Figure 3.

Another preprocessing step that wasn't mentioned above was the standardization of the whole dataset before training of the classifier, as well as the split into three separate parts. The training dataset contained 67.5% of the whole data, while 22.5% was used for validation and 10% for the test dataset. The data was shuffled while splitting so the continuity between the individual segments was not preserved, as it would only hinder the final results.

IV. CLASSIFICATION

The classifier that is used in this paper is a hybrid convolutional-recurrent neural network. The reasoning behind this decision is because LSTM/GRU networks are very accurate and efficient when classifying temporal data [11] [12], which makes them a perfect choice for this problem. These types of neural networks are an improvement on regular recurrent neural networks because of their ability to learn long-term temporal dependencies, which regular RNNs lack due to the vanishing/exploding gradient problem. The convolutional layers before the recurrent part of the network are used to extract hidden features in the signal which can then be used to improve the overall result of the classification. The architecture of the used network is composed of five layers, including the input and output layers. The input layer is followed by three one-dimensional convolutional layers with a varying number of filters, with each of them being followed by a max-pooling layer. After the convolutional part of the network, the recurrent part is composed of a single Gated Recurrent Unit or GRU layer instead of the more common LSTM, due to its simplicity and faster training time. The output layer of the network has a softmax activation function and six neurons, representing each of the six possible movements. A visual representation of the network's architecture is given in Figure 4.

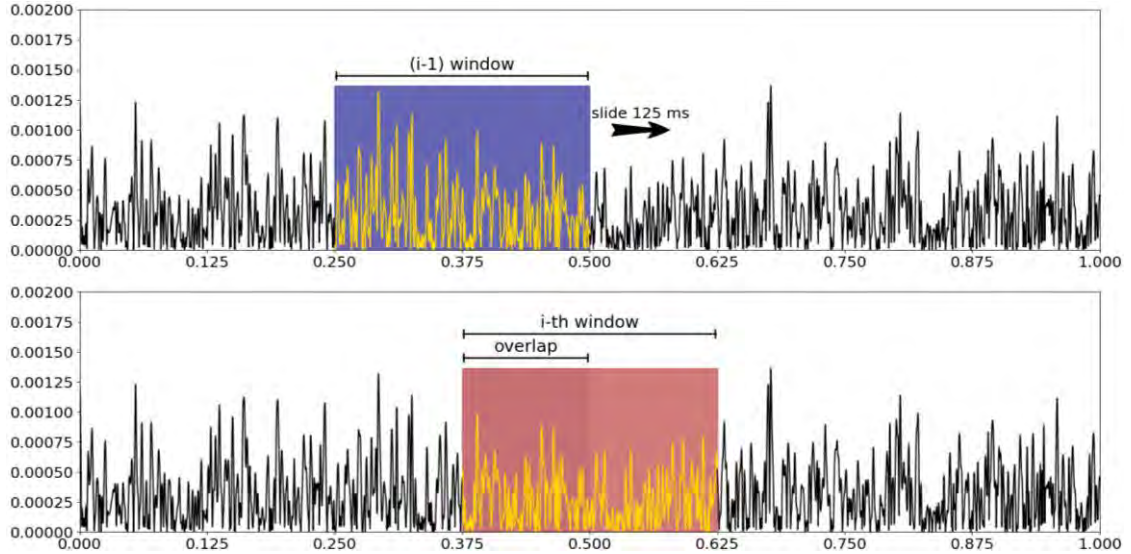


Figure 3. Example of the segmentation process

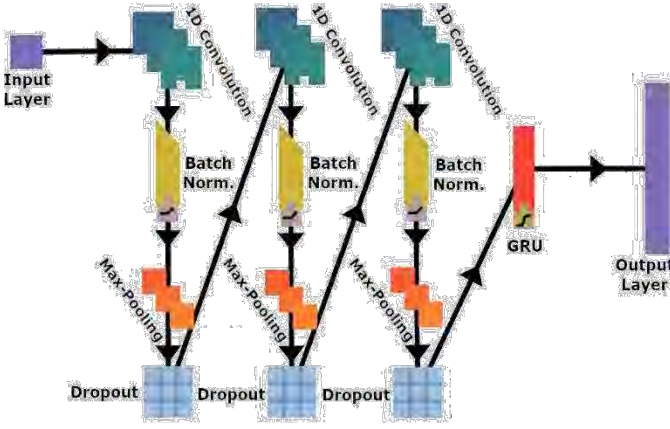


Figure 4. Architecture of the neural network

To avoid overfitting, multiple popular regularization techniques are implemented in the network:

- **Batch normalization** – we use batch normalization at every convolutional layer in the network (before the activation function) to ensure higher stability as well as overall better results.
- **Dropout** – to specifically target the overfitting problem multiple dropout layers with varying rates are used, placed after each of the max-pooling layers. A special case of recurrent dropout is also implemented at the GRU layer, ensuring the removal of unimportant memory links while keeping the temporal dependency of the data.
- **Weight constraints** – placing certain max norm constraints on the trainable weights in the network enforces an upper bound on the magnitude of the weight vector for each neuron.

During training, the *categorical cross entropy* loss function is used:

$$CE = -\log\left(\frac{e^{s_p}}{\sum_j e^{s_j}}\right), \quad (1)$$

which is then optimized by the *AMSGrad* variation of the *Adam* optimizer. A learning rate of $\alpha=10^{-4}$ is used. The network is trained on the segmented data in batches of 64 samples until the early stopping criteria is met, i.e. the validation loss stops decreasing.

V. RESULTS AND DISCUSSION

After the training of the network was complete, the accuracy of the model was evaluated by comparing the predicted classes in the test dataset to the real labels. The model reached a significantly high F1 score of 91.3%. This means that out of 100 movements made with the prosthetic, on average only about 9 of them will be incorrectly predicted which is an acceptable margin of error for the task at hand [13] [14]. By comparing the results achieved in this paper to the results of [7] (which uses the same dataset), we can see that our approach has provided a similar accuracy to their ~90% average between all test subjects. Incorrectly classified movements can be easily identified and

filtered out using an algorithm that will work in real time while the prosthetic is in use. This would not be a problem for the user as they would barely notice this error and it would only appear as a slight delay in the actuation of the prosthetic by a few milliseconds.

True \ Predicted	Predicted					
	Fist	Thumb	Index	Middle	Ring	Pinky
Fist	182	0	3	0	1	2
Thumb	0	161	7	4	0	15
Index	3	6	164	3	3	8
Middle	0	4	4	170	2	8
Ring	4	0	0	2	181	0
Pinky	0	9	4	5	1	168

Figure 5. Confusion matrix

The confusion matrix of the classification results is given in Figure 5. Most of the incorrectly classified labels appear as a slight bias of the model to predict signals as a flexion of the little finger, which is most striking when the true signal is a flexion of the thumb. This error can simply be caused by the anatomical nature of this finger. Unlike the rest of the digits, the little finger is particularly dependent on the flexion of the other fingers and most people cannot flex it without moving the ring finger that is adjacent to it. What makes it interesting in this context is the fact that errors of the type ring-little finger are almost non-existent, while thumb-little finger errors are most prominent, even though these digits are the furthest away from each other. This can also be explained due to the anatomy of the hand, where most people unconsciously stretch out their thumb when flexing the little finger. This subtle movement can then be detected by the electrode that is placed on the extensor muscle, leading to the errors we see in the confusion matrix.

Even though the results obtained from the model are satisfactory, we still need to analyze why there is a deficit of 8.7% where the classification is incorrect. From the visual analysis of the complete signals it is easy to notice that there are significant differences between the measurements taken from two separate participants (especially when it comes to the amplitude of the signals), and there are slight differences even in different measurements from the same participant. This can be seen in Figure 6 where the signal from a flexion of the index finger is compared between participants 1 and 2.

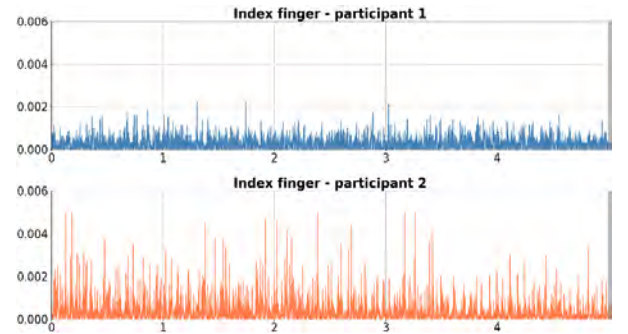


Figure 6. Comparison of measured signals from index finger flexion

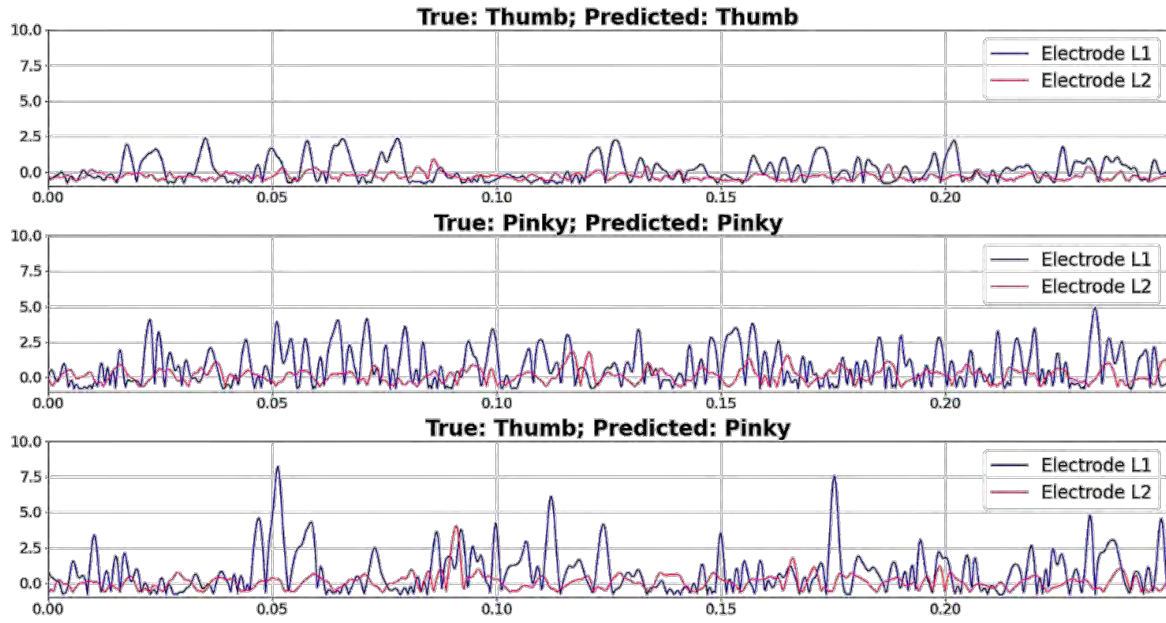


Figure 7. Comparison between correctly and incorrectly classified signals

These differences can be caused by many reasons, but they mostly occur due to the placement of the electrodes themselves, which are very difficult to place on the exact same position on the muscle on different people. In addition to that, differences in skin thickness and fat content in the forearm, as well as the force exerted while executing the flexions can also significantly affect the amplitude of the signals. Seeing as we use the pure measurements to train the neural network (with segmentation and minimal preprocessing), these variations cannot be completely avoided and the only way to suppress them is to increase the generalization capability of the model by increasing the amount of data it uses to train.

To thoroughly examine where the model fails to properly classify the data, we can take a look at a simple example for a misclassified sample. Such comparison of properly and improperly classified signals is given in Figure 7. The graph is divided into three separate plots. The top two plots contain segments of signals in which the true class is a flexion of the thumb and the little finger respectively, which the model has predicted correctly. The bottom plot contains a signal of a flexion of the thumb that the model has predicted as a flexion of the little finger. It can be observed in the correctly predicted samples that the amplitude is higher on average and peaks occur more frequently when flexing the little finger compared to the thumb, yet we see even higher values on the incorrectly predicted sample even though the signal is a flexion of the thumb. This obvious outlier situation could be the cause of the error in this specific case, and it may be present in multiple other incorrectly predicted samples. This confirms the previous hypothesis of varying amplitudes caused by inconsistent flexion force being one of the leading causes of error in the model's predictions.

VI. CONCLUSION

In this paper we showed that by using sophisticated software methods we can apply a cheap, efficient and simple solution to the usability and cost problem that clouds electric prosthetics, which will make this medical device much more accessible to the majority of the population. Electromyography enables fast and efficient prosthetic control that is natural for users, and this will significantly reduce the rate at which unsatisfied users give up on their prosthetics due to poor control and usability. By using an advanced neural network compared to other more traditional classification methods, we created a model that can detect and predict individual finger flexions with an accuracy of 91.3%. This result is more than sufficient and can be further improved by additional training of the network with more data. Furthermore, by applying a software solution to this problem we can significantly reduce the price of the prosthetics due to the need for fewer EMG sensors.

What remains for future implementation is the physical execution of an electric prosthetic based on this classification method, as well as analyzing the model's performance on data from a single participant. Another avenue to explore would be to test different neural network architectures (perhaps simpler ones) in order to simplify the complex nature of the proposed solution, which would allow for an easier, more refined and faster implementation of the model in real-world use cases.

VII. REFERENCES

- [1] "Advanced Amputee Solutions LLC," 2012. [Online]. Available: <https://advancedamputees.com/amputee-statistics-you-ought-know>.
- [2] W. Williams, "Bionic Hand Price List," Bionics for EVERYONE, May 2021. [Online]. Available:

- <https://bionicsforeveryone.com/ottobock-bebionic-hand/>.
- [3] Open Bionics, [Online]. Available: <https://openbionics.com/how-does-a-bionic-arm-work/>.
 - [4] W. Williams, "OttoBock bebionic Hand," Bionics for EVERYONE, May 2021. [Online]. Available: <https://bionicsforeveryone.com/bionic-hand-price-list/>.
 - [5] G. Tsenov, A. H. Zeghib, F. Palis, N. Shoylev and V. Mladenov, "Neural Networks for Online Classification of Hand and Finger Movements Using Surface EMG signals," *2006 8th Seminar on Neural Network Applications in Electrical Engineering*, pp. 167-171, 2006.
 - [6] F. V. G. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings and N. V. Thakor, "Decoding of Individuated Finger Movements Using Surface Electromyography," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1427-1434, 2009.
 - [7] R. N. Khushaba, M. Takruri, S. Kodagoda and G. Dissanayake, "Toward Improved Control of Prosthetic Fingers Using Surface Electromyogram (EMG) Signals," *Expert Systems with Applications*, pp. vol 39, no. 12, pp. 10731–10738, 2012.
 - [8] J. V. Basmajian and C. J. De Luca, "Muscle Interactions," in *Muscles alive*, 1985, pp. 223-245.
 - [9] R. N. Scott, *An Introduction to Myoelectric Prostheses*, vol. UNB Monographs on Myoelectric prostheses, 1984.
 - [10] T. R. Farrell and R. F. Weir, "The optimal controller delay for myoelectric prostheses," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 15, no. 1, pp. 111-118, 2007.
 - [11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
 - [12] M. Simão, P. Neto and O. Gibaru, "EMG-based online classification of gestures with recurrent neural networks," *Pattern Recognition Letters*, vol. 128, pp. 45-51, 2019.
 - [13] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *Journal of Rehabilitation Research & Development*, vol. 48, pp. 643-660, 2011.
 - [14] D. K. Kumar, S. P. Arjunan and V. P. Singh, "Towards identification of finger flexions using single channel surface electromyography – able bodied and amputee subjects," *Journal of NeuroEngineering and Rehabilitation*, vol. 10, p. 50, 2013.

The Selection of Bi-Fractional Order Reference Model Parameters for Minimum Settling Time

Ertuğrul Keçeci, Res. Assist

Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
kececic@itu.edu.tr

Müjde Güzelkaya, Prof. Dr.

Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
guzelkaya@itu.edu.tr

Erhan Yumuk, Res. Assist.

Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
yumuk@itu.edu.tr

İbrahim Eksin, Prof. Dr.

Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
eksin@itu.edu.tr

Abstract—In this study, the effects of bi-fractional order reference model parameters on time domain characteristics are analyzed. A relation between damping ratio and fractional order that exhibits minimum settling time is observed and the damping ratio is expressed in terms of fractional order by fitting a polynomial. Consequently, a bi-fractional order reference model that guarantees minimum settling time is derived. The effectiveness of the newly derived reference model is shown through examples.

Keywords— *bi-fractional order system model; reference model; minimum settling time; time domain characteristics*

I. INTRODUCTION

Fractional calculus has emerged from the correspondence of Leibniz and L'Hospital towards the end of the 17th century [1]. Fractional calculus, which did not show much development for two centuries after this date, started to gain momentum towards the middle of the 19th century. Riemann-Liouville, Grünwald-Letnikov and Caputo proposed definitions to explain fractional integro-differential operator [2].

In the second half part of the 20th century, fractional calculus has attracted the attention of scientists and researchers. The ability to represent the dynamic behavior of some physical systems more accurately has led to various studies [3-4]. Later, with the help of the developing process power of computers, researchers started to model more complex processes in several topics such as thermal systems [5], electric networks [6] and even global dynamics of viruses [7-8] using fractional notion.

However, the applications of fractional calculus on the control field were not restricted only in modeling. In 1991, Oustaloup proposed the first fractional controller definition [9], and three years later Podlubny adapted fractional calculus to classical PID controller [10]. Although the idea of Podlubny brings two more parameters to enhance the performance of classical PID controller, these parameters may cause design and

tuning difficulties. In literature, numerous design methods and real-time applications can be found [11-14]. In addition to the usage of fractional calculus on classic control theory, several researchers expanded this area to modern control theory [15-18]. The transfer function of the closed loop system that satisfies desired control system responses is referred as reference model [19]. Various reference models are used in internal model control (IMC) [20], characteristic ratio assignment (CRA) [21], coefficient diagram method (CDM) [22] and direct synthesis method [23]. Reference model approach simplifies control system design and provides analytical background to the obtained controllers. In this respect, this method requires comprehensive and explanatory analyses on the time and frequency domains.

In this study, we consider bi-fractional order transfer function model [24] with three parameters; namely, natural frequency (ω_n), damping ratio (ζ) and fractional order (γ). The effects of bi-fractional order reference model parameters on time domain characteristics are analyzed. Minimum settling time is selected as the objective and a relationship between damping ratio and fractional order is found. The damping ratio of bi-fractional order transfer function is expressed as a function of fractional order. Consequently, a bi-fractional order reference model that owns only natural frequency and the fractional order parameters is established. The effectiveness of the newly derived reference model is shown through examples.

The paper is organized as follows. Section 2 gives preliminaries to fractional calculus and fractional order transfer functions. Section 3 exhibits the time domain analyses of the bi-fractional order reference model. In Section 4, bi-fractional order reference model is presented for minimum settling time. Finally, conclusion is given in Section 5.

II. FRACTIONAL CALCULUS

Fractional calculus is the superset of classical calculus. The fractional order integro-differential operator can be given as follows:

$${}_r D_t^\gamma f(t) = \begin{cases} \frac{d^\gamma f(t)}{dt^\gamma}, & \gamma > 0 \\ f(t), & \gamma = 0 \\ \int_r^t f(\tau)(d\tau)^{-\gamma}, & \gamma < 0 \end{cases} \quad (1)$$

where the parameters γ , r and t are non-integer order, lower limit and higher limit of operator, respectively. The Grünwald-Letnikov, Riemann-Liouville and Caputo definitions are some of the well-known definitions for fractional order operator [2].

Although various definitions of fractional order integro-differential operator are proposed, it is very complicated and challenging to apply these definitions in real-time applications. To overcome this problem, many filter approximations can be found in the literature. The filter approximation proposed by Oustaloup is probably the most known and precise one [25]. In Oustaloup filter, the approximate integer order transfer function model of fractional order integro-differential operator is calculated as follows:

$$G_f(s) = K \prod_{k=1}^N \frac{s + \omega'_k}{s + \omega_k} \quad (2)$$

where

$$\omega'_k = \omega_b \sqrt[N]{\frac{\omega_h}{\omega_b}^{2k-1-\gamma}} \quad (3)$$

$$\omega_k = \omega_b \sqrt[N]{\frac{\omega_h}{\omega_b}^{2k-1+\gamma}} \quad (4)$$

$$K = \omega_h^\gamma \quad (5)$$

Here, ω_b and ω_h are the frequency values that specify the frequency range of approximation, γ is the fractional order and N is the order of approximation.

The general definition of fractional order transfer function is given as the following,

$$G(s) = \frac{b_1 s^{\mu_1} + b_2 s^{\mu_2} \dots \dots + b_{m-1} s^{\mu_{m-1}} + b_m s^{\mu_m}}{a_1 s^{\gamma_1} + a_2 s^{\gamma_2} + \dots \dots + a_{n-1} s^{\gamma_{n-1}} + a_n s^{\gamma_n}} \quad (6)$$

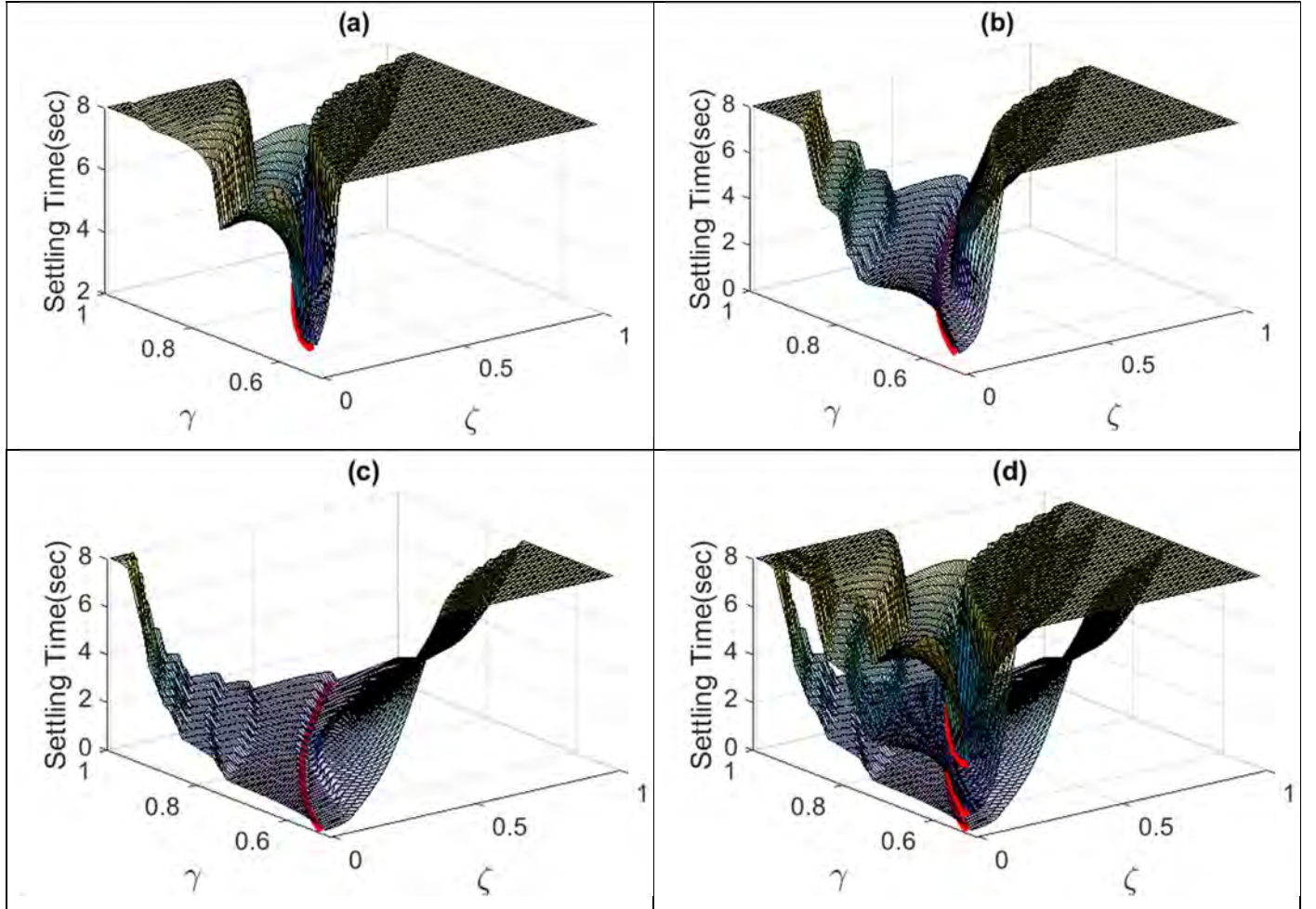


Fig. 1. The effects of model parameters on settling time for a) $\omega_n = 1$, b) $\omega_n = 2$, c) $\omega_n = 5$, d) $\omega_n = 1, \omega_n = 2, \omega_n = 5$

In this equation, a_n and b_m are the coefficients of denominator and numerator respectively, whereas, γ_n represents fractional orders of denominator and μ_m stands for fractional orders of nominator. If the fractional orders in (6) can be represented as integer multiples of their biggest common non-integer divisor γ , the transfer function in (7) is called as commensurate fractional order transfer functions.

$$G(s) = \frac{\sum_{i=0}^m b_m s^{m\gamma}}{\sum_{i=0}^n a_n s^{n\gamma}} \quad (7)$$

III. TIME DOMAIN ANALYSIS OF BI-FRACTIONAL ORDER REFERENCE MODEL

The bi-fractional order reference model [24] is represented as follows:

$$G(s) = \frac{\omega_n^2}{s^{2\gamma} + 2\zeta\omega_n s^\gamma + \omega_n^2} \quad (8)$$

Here, γ refers commensurate non-integer order of the reference model whereas, ζ and ω_n are the damping ratio and natural frequency, respectively.

Fig 1a, 1b and 1c illustrate the effect of $\gamma - \zeta$ model parameters on settling time for $\omega_n = 1$, $\omega_n = 2$ and $\omega_n = 5$, respectively whereas Fig. 1a, 1b and 1c are combined in Fig. 1d. In the figure, the range of the commensurate order γ is set to [0.5 1] so that the model order takes place in the range of [1 2]. On the other hand, ζ value is selected between [0 1] and ω_n of model is chosen as 1, 2 and 5. As it is seen from Fig 1, there is a valley that gives the minimum settling time for a ζ value for any $\omega_n - \gamma$ couples. The red lines on the plots demonstrate this specific pattern and it can be seen from Fig. 1d that the value of ω_n has neglectable effects on the pattern. Moreover, the percentage overshoot of the model response corresponding to the pattern shown by the red line is approximately three.

IV. BI-FRACTIONAL ORDER REFERENCE MODEL FOR MINIMUM SETTLING TIME

ω_n values are varied in the range of [0.25 15] while the range of parameters γ and ζ are in the range of [0.5 1] and [0 1], respectively. Then, the $\gamma - \zeta$ pairs that provide minimum settling time for different ω_n values in the given range are plotted with blue dots in Fig. 2. Since it is seen from the figure that ω_n has neglectable effect, a polynomial can be fitted to present the relationship between γ and ζ parameters for minimum settling time. This polynomial can be derived as in the following form,

$$\zeta = f(\gamma) = 1.4891\gamma^2 - 0.6287\gamma - 0.0883 \quad (9)$$

The red curve in Fig. 2 shows the second order polynomial in (9). Consequently, the bi-fractional order reference model transfer function in (8) can be expressed in terms of two parameters; namely, γ and ω_n as follows:

$$G(s) = \frac{\omega_n^2}{s^{2\gamma} + 2f(\gamma)\omega_n s^\gamma + \omega_n^2} \quad (10)$$

For $\omega_n \in [0.25 15]$ and $\gamma \in [0.5 1]$, the overshoot values of the unit step responses of the reference model in (10) are

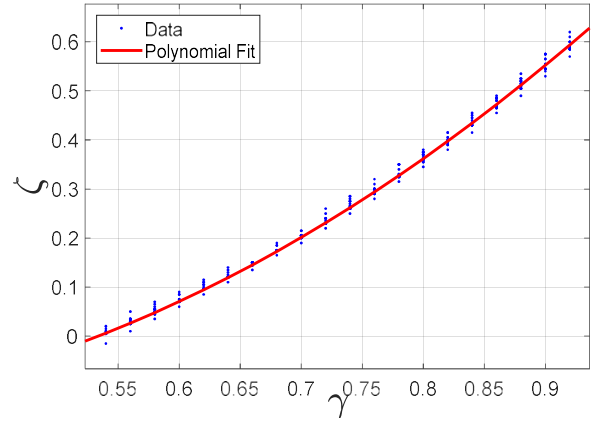


Fig. 2. The fitted polynomial function to data

illustrated in Fig. 3.

It can be observed from Fig. 3 that the unit step responses of the reference model in (10), which corresponds to minimum settling time, possess an overshoot of approximately three percent.

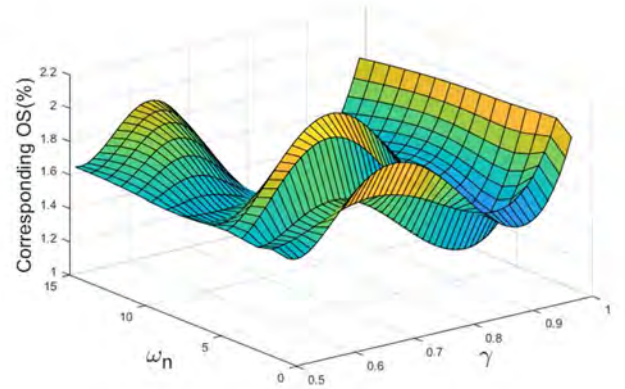


Fig. 3. Overshoot values of model output for different $\omega_n - \gamma$ pairs

In order to demonstrate the effect of the reference model in (10) on the settling time, two different $\omega_n - \gamma$ pairs are selected. The step responses of the reference model in (10) are compared with the step responses of the reference model in (8). In the first case, the ω_n and γ parameters are chosen as 1 and 0.8, respectively. ζ parameters are chosen 0.2 and 0.5 for the reference model in (8) whereas the $\zeta = f(\gamma)$ is calculated as 0.36 for reference model in (10). Their step responses are shown in Fig. 4a. In the second case, the ω_n and γ parameters are chosen as 2 and 0.9, respectively. While the ζ parameters are chosen 0.3 and 0.7 for the reference model in (8), the $\zeta = f(\gamma)$ is calculated as 0.54 for the reference model in (10). Fig. 4b illustrates the step responses of the reference models. Moreover, the dashed lines in Fig. 4a and Fig. 4b represent the $\pm 3\%$ band range of the settling time. As it is observed from Fig. 4, the reference model in (10) owns minimum settling time while its overshoot value is within the $\pm 3\%$ band.

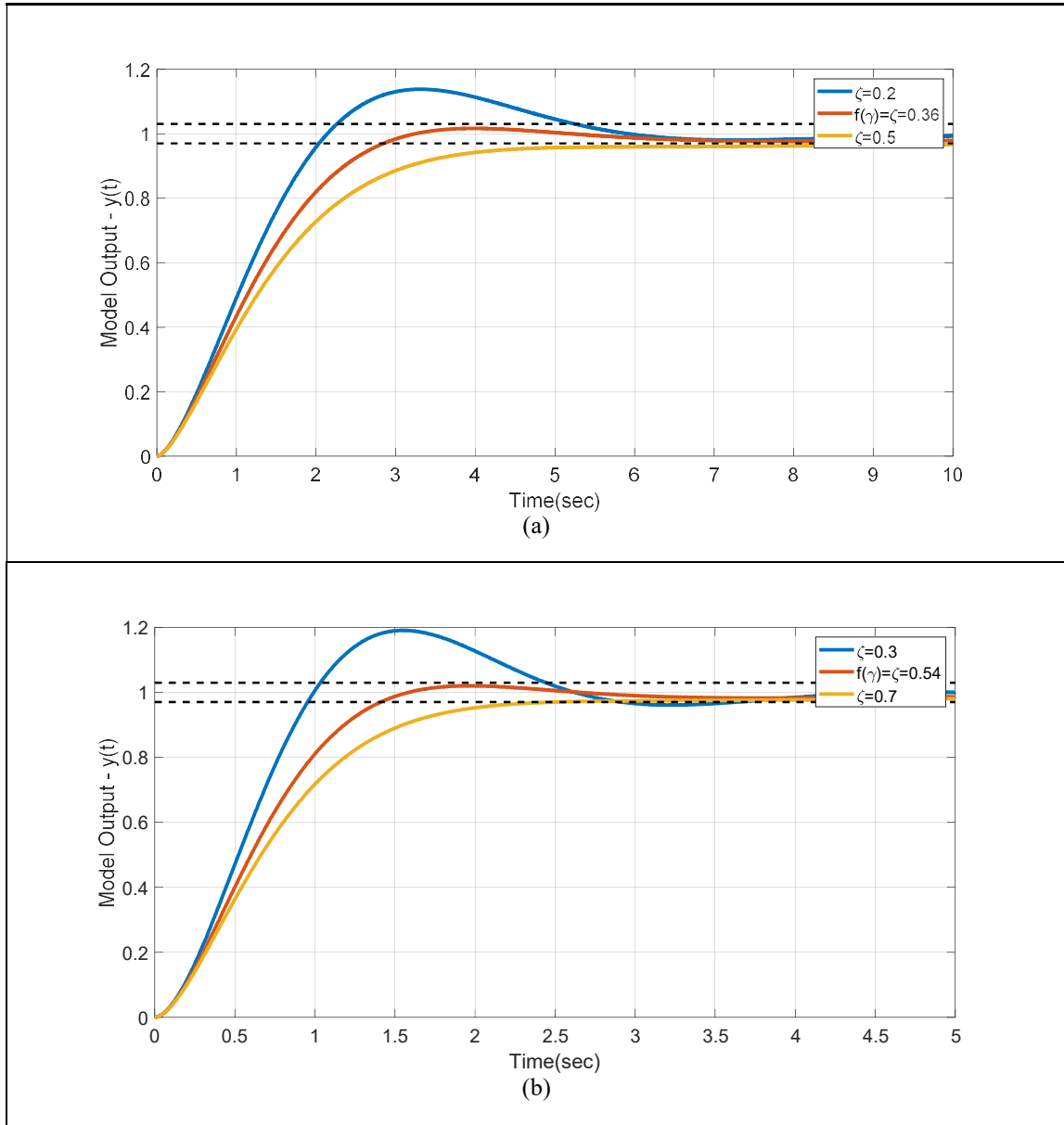


Fig. 4. Step responses of the reference models in (8) and (10)

V. CONCLUSIONS

In this paper, the effects of bi-fractional order reference model parameters on time domain characteristics are examined. Minimum settling time is selected as the objective and it is observed that the natural frequency of the bi-fractional order reference model has neglectable effects on this objective. Therefore, a relationship between the remaining two reference transfer function model parameters is found by using second order polynomial approximation. A bi-fractional order reference model for minimum settling time is derived. It is shown on two examples that the derived bi-fractional order reference model has minimum settling time. Moreover, the overshoot of the step response takes place within the settling time band.

VI. REFERENCES

- [1] Ross, B. (1977). The development of fractional calculus 1695–1900. *Historia Mathematica*, 4(1), 75-89.
- [2] Loverro, A. (2004). Fractional calculus: history, definitions and applications for the engineer. *Rapport technique, Univeristy of Notre Dame: Department of Aerospace and Mechanical Engineering*, 1-28.
- [3] Bagley, R. L., & Torvik, P. J. (1986). On the fractional calculus model of viscoelastic behavior. *Journal of Rheology*, 30(1), 133-155.
- [4] Metzler, R., Glöckle, W. G., & Nonnenmacher, T. F. (1994). Fractional model equation for anomalous diffusion. *Physica A: Statistical Mechanics and its Applications*, 211(1), 13-24.
- [5] Gabano, J. D., & Pointot, T. (2011). Fractional modelling and identification of thermal systems. *Signal Processing*, 91(3), 531-541.
- [6] Galvão, R. K. H., Hadjiloucas, S., Kienitz, K. H., Paiva, H. M., & Afonso, R. J. M. (2012). Fractional order modeling of large three-dimensional RC networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(3), 624-637.

- [7] Naik, P. A., Zu, J., & Owolabi, K. M. (2020). Global dynamics of a fractional order model for the transmission of HIV epidemic with optimal control. *Chaos, Solitons & Fractals*, 138, 109826.
- [8] Jajarmi, A., Yusuf, A., Baleanu, D., & Inc, M. (2020). A new fractional HRSV model and its optimal control: a non-singular operator approach. *Physica A: Statistical Mechanics and its Applications*, 547, 123860.
- [9] Oustaloup, A. (1991). La commande CRONE: commande robuste d'ordre non entier. Hermes.
- [10] Podlubny, I. (1998). Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications. Elsevier.
- [11] Yumuk, E., Güzelkaya, M., & Eksin, İ. (2019). Analytical fractional PID controller design based on Bode's ideal transfer function plus time delay. *ISA transactions*, 91, 196-206.
- [12] Yumuk, E., Guzelkaya, M., & Eksin, I. (2021). Application of fractional order PI controllers on a magnetic levitation system. *Turkish Journal of Electrical Engineering & Computer Sciences*, 29(1), 98-109.
- [13] Zhao, C., Xue, D., & Chen, Y. (2005, July). A fractional order PID tuning algorithm for a class of fractional order plants. In *IEEE International Conference Mechatronics and Automation, 2005* (Vol. 1, pp. 216-221). IEEE.
- [14] Xue, D., Zhao, C., & Chen, Y. (2006, June). Fractional order PID control of a DC-motor with elastic shaft: a case study. In *2006 American control conference* (pp. 6-pp). IEEE.
- [15] Sierociuk, D., & Dzieliński, A. (2006). Fractional Kalman filter algorithm for the states, parameters and order of fractional system estimation.
- [16] Keçeci, E., Yumuk, E., Güzelkaya, M., & Eksin, İ. (2019, November). Comparison of optimal integer and fractional order PID controllers on a stabilized real-time system. In *2019 11th International Conference on Electrical and Electronics Engineering (ELECO)* (pp. 754-758). IEEE.
- [17] Rhouma, A., Bouani, F., Bouzouita, B., & Ksouri, M. (2014). Model predictive control of fractional order systems. *Journal of Computational and Nonlinear Dynamics*, 9(3).
- [18] Yumuk, E., Güzelkaya, M., & Eksin, İ. (2020). Optimal fractional-order controller design using direct synthesis method. *IET Control Theory & Applications*, 14(18), 2960-2967.
- [19] Yumuk, E., Güzelkaya, M., & Eksin, İ. (2019). Analytical fractional PID controller design based on Bode's ideal transfer function plus time delay. *ISA transactions*, 91, 196-206.
- [20] Rivera, D. E., Morari, M., & Skogestad, S. (1986). Internal model control: PID controller design. *Industrial & engineering chemistry process design and development*, 25(1), 252-265.
- [21] Tabatabaei, M., & Haeri, M. (2010). Characteristic ratio assignment in fractional order systems. *ISA transactions*, 49(4), 470-478.
- [22] Maheswari, C., Priyanka, E. B., & Meenakshipriya, B. (2017). Fractional-order PI λ D μ controller tuned by coefficient diagram method and particle swarm optimization algorithms for SO₂ emission control process. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 231(8), 587-599.
- [23] Yumuk, E., Güzelkaya, M., & Eksin, İ. (2020). Optimal fractional-order controller design using direct synthesis method. *IET Control Theory & Applications*, 14(18), 2960-2967.
- [24] Piątek, P., Baranowski, J., Zagórska, M., Bauer, W., & Dziwiński, T. (2015). Bi-fractional filters, part 1: Left half-plane case. In *Advances in modelling and control of non-integer-order systems* (pp. 81-90). Springer, Cham.
- [25] Oustaloup, A., Levron, F., Mathieu, B., & Nanot, F. M. (2000). Frequency-band complex noninteger differentiator: characterization and synthesis. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 47(1), 25-39.

Intra-nodal Caching Assisted UAV Based Data Acquisition from Wireless Mobile Ad-hoc Sensor Networks

Umair B. Chaudhry

School of Electronic Engineering and Computer Science
Queen Mary University of London (QMUL)
London, United Kingdom
U.B.Chaudhry@qmul.ac.uk

Chris I. Phillips

School of Electronic Engineering and Computer Science
Queen Mary University of London (QMUL)
London, United Kingdom
Chris.I.Phillips@qmul.ac.uk

Abstract— Unmanned aerial vehicle assisted data collection is not a new concept and has been used in various mobile ad hoc networks. In this paper, we propose a caching assisted scheme alternative to routing in MANETs for the purpose of wildlife monitoring. Rather than deploying a routing protocol, data is collected and transported to and from a base station using a UAV. Although some literature exists on such an approach, we propose the use of intermediate caching between the mobile nodes and compare it to a baseline scenario where no caching is used. The paper puts forward our communication design where we have simulated the movement of multiple mobile sensor nodes in a field that move according to the Levy walk model imitating wildlife animal foraging and a UAV that makes regular trips across the field to collect data from them. The unmanned aerial vehicle can collect data not only from the current node it is communicating with but also data of other nodes that this node came into contact with. Simulations show that exchanging cached data is highly advantages as the drone can indirectly communicate with many more mobile nodes.

Keywords—UAV; caching; sensors; MANETs; WSN; waypoint

I. INTRODUCTION

The use of wireless sensor networks (WSNs) and mobile ad-hoc networks (MANETs) in various areas such as environmental monitoring, military, vehicular networks and animal tracking has been widely adopted [1]. Applications of such networks vary based on the targeted area. Humidity and seismic sensors, collision avoidance and parking sensors, pulse and temperature sensors are all examples of this. In WSNs, nodes are deployed with the intention of sensing and relaying information to a particular destination for evaluation purposes. In MANETs, nodes are mobile forming temporary networks throughout their runtime. Nodes in these networks are typically small and possess limited resources. They have restricted processing power and run on small batteries hence energy conservation is a serious concern for them. Data is routed from the source to the destination using routing protocols. The convergence and retransmission mechanisms of these protocols impose an additional overhead causing an additional energy

drain. Many efforts have been put into making these protocols as efficient as possible [2]–[4]; however, there is always a trade-off. Conversely, we propose the use of an unmanned aerial vehicle (UAV) to periodically to collect data from caching assisted nodes, hence avoiding routing altogether.

In this paper, we focus on wildlife tracking and monitoring. Traditionally, this is achieved by strapping heavy tracking equipment to animals [5], [6]. Even with current technological trends, wildlife monitoring remains a challenging setting. Typical VHF transmitters are of very restricted range [7] and have a limited battery life and the ones that are longer in range are satellite oriented and hence require even more power, consequently providing a lower lifetime. Table 1 shows some of the existing devices available. Approaches such as [7]–[12] are either too expensive, require dedicated manpower, or the resource constraints of the devices can cause them to fail prematurely. Our aim is to make tracking and monitoring easier and less costly in terms of finance and operation.

TABLE I. TRADITIONAL DEVICES AND THEIR LIMITATIONS [7]

Devices	NANO	MICRO	SMALL	MEDIUM
Weight Range	5g-20g	6g-50g	20g-100g	130g-250g
Example Suitable Animals	Birds, bats, and other tiny mammals	Lizards, tortoises, turtles, frogs, very small mammals	Animals that weight at least 500 grams	Foxes, Tasmanian Devils
Data Recovery Method	UHF Wireless	UHF Wireless	UHF Wireless	Satellite
Drone Data Downloading	Standard	Standard	Standard	N/A
Base Station Battery Life	2 days	5 days	2 days	N/A

This paper highlights a routing-less approach for data collection from mobile sensor nodes for wildlife monitoring

using an unmanned aerial vehicle. In our application scenario, we assume the home range to be a large field or area where the animals roam. Our sensor nodes, according to our use-case scenario, will be strapped on animals and several of these will be dispersed across the home range. The nodes are considered to be mobile with some degree of purpose in their movement but at the same time having some randomness in their behaviour. An unmanned aerial vehicle will make periodic trips across the field. The nodes are equipped with sensing equipment, the type of which is not the focus of our research. The nodes upon encountering the UAV transfer their data to it. In this paper, we test this approach in two scenarios. One is without communication between the nodes and the other with caching enabled among the nodes allowing them to store data from the nodes they come into contact with. Our simulator uses the Levy walk movement model for the nodes.

II. RELATED WORK

Retrieval of data from static and mobile sensor nodes deployed in a large field through an aerial vehicle is a fairly new concept. The authors in [13] describe their method of how a UAV can be used for wildlife monitoring and tracking. In their approach, the field is divided into virtual grids and each grid has a cluster of static sensors deployed. Each cluster has a cluster head which acts as a point of contact for the UAV to collect data. The network model is not a generic one and is highly dependent on data sets obtained from tracking equipment for specific animals. In their case, they used the movement data of zebras from ZEBRANET and the UAV visits the cluster heads of the most active grids for data collection. The authors in [14] propose an automatic tracking system that offers autonomous wildlife monitoring. In their approach, they suggest to equip the wildlife with a system that is a combination of a Global Positioning System (GPS) module and a wireless Subscriber Identity Module (SIM). The GPS coordinates are sent to a central server which is also equipped with a SIM which forwards the information to a SIM-equipped control system from where the received information is fed to a drone in addition to the drone control commands. The drone, using this information and the control commands, navigates to the coordinate location. The information is fed to the drone only at the take-off point. Thus, once the drone reaches the target location, the target might not be there resulting in waste of time, energy and the trip. In addition, the devices used to accomplish this have a limited lifespan which is greatly affected by the presence of two highly energy draining modules on the animals. Aerial assisted data collection using a UAV from limited capacity sensor nodes has also been considered in [15] using a Markov chain to model the movement of the UAV in addition to modelling the irregularities in the movement due to several implicit and explicit factors. The authors in [16] also favour the same concept of acquiring data from sensor nodes deployed in a field using a UAV. They add to the approach by proposing a model that allows the sensor nodes to cooperate with each other to achieve simultaneous transfer of data to the UAV to reduce latency. However, one can argue that cooperation between the nodes is of limited use as multiple nodes can transmit to the UAV simultaneously using different channels. The feasibility of using a UAV for data collection from ground sensors has

been tested in [17] against several parameters including weather, flight height, latency, throughput, jitters and communication channels and authors have recommended a configuration based on their observations. However, it can be argued that the values are highly subjective. Considering the low data transfer rates due to the brief contact duration of a UAV with a sensor node in UAV aided data collection, the authors in [18] propose a modified Media Access Control (MAC) protocol which uses beacon broadcast at the UAV together with Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) at the ground nodes. The nodes have to contend with each other to speak to the UAV in addition to remaining in the listening mode to receive the beacon which is a big drawback considering the limited battery life of the nodes.

Despite the preceding research studies, intermediate caching between the nodes has, to our knowledge, never been considered the way we present it. We believe that intermediate caching between the nodes can be advantageous for several reasons. Firstly, it does not impose additional strain on a particular node (i.e. a cluster-head). Secondly, there are no limitations on the positioning of the nodes and, finally, the UAV has more flexibility in terms of points of contact.

The next section outlines our proposed system. This is followed in Section 4 with a simulation-based evaluation where we employ a Levy movement model for the mobile nodes. Finally we conclude the paper in Section 5.

III. SYSTEM MODEL

We have implemented a discrete time event simulator in Java. Rather than feeding movement traces into the simulator [13], our simulator has the ability to cope with different movement models, allowing for flexibility. We employ the Levy walk movement model as recent evidences have shown that the movement of different animal species are more relatable to the levy walk model [19], [20]. Mobile nodes are dispersed in a field of size $n \times n$. A UAV is sent out from a gateway/base station and the UAV follows a fixed path based on waypoints [21], [22] positioned as shown in Figure 1. The coverage pattern of the UAV can also be understood by the same Figure 1 and the path it will take during its trip across the field.

Our model as shown in Figure 1, considers two versions of this approach. In the first (termed ‘WP’), the UAV only captures the data cached at the designated waypoints (imitating static cluster heads). In the second variation (termed ‘UAV’), if the UAV encounters any node that has data to transmit, the node transmits that data directly to the UAV. WP+C allows the waypoints to not only capture the data of nodes directly in contact with it but also allow the transfer of that node’s cache as well. The same applies for UAV+C where the UAV not only captures the node it comes in contact with directly but also caches information contained in that node’s cache as well. We compare the efficiency by comparing the data capture success by the UAV from nodes in two scenarios. First without intra-node caching and second with intra-node caching.

The caching mechanism can be better understood from Figure 2.

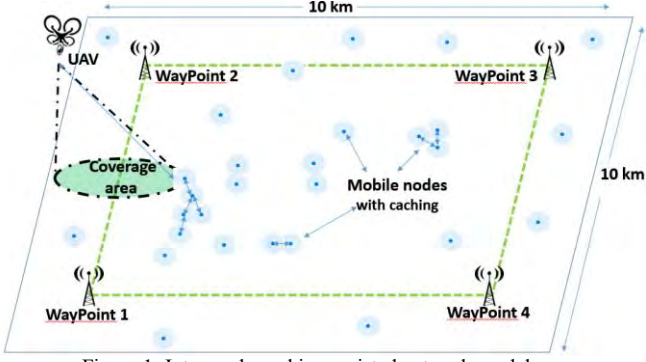


Figure 1: Intra-node caching assisted network model

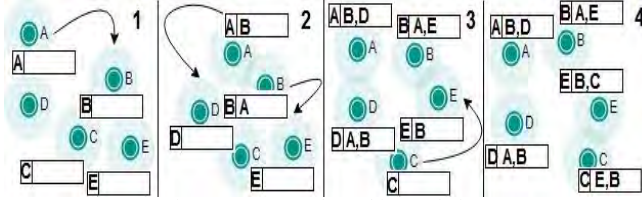


Figure 2: Intra-node caching technique

Snapshot (2) of Figure 2 shows that A and B exchange information upon coming in range of each other. Snapshot (3) shows that when D comes in contact with A, it not only caches A's information but in addition, caches all the information that A has in its cache (information A gathered from B in its previous encounter with B). This means that even if a node has not encountered a certain node directly, it still has the ability to carry this "second-hand" data the node if it came into contact with it indirectly. Hence we refer to this as an 'indirect caching technique'.

Data exchange between the nodes arises using a simple connectionless protocol shown as Algorithm 1.

IV. EXPERIMENTATION AND RESULTS

We have tried to mimic large herbivores such as elephants which according to [23], [24] employ the Levy walk model, normally move with speeds in access of 2mps, have a minimum home range of ten square kilometers [25], and can exist individually or in groups between 8-100 elephants in an area [26]. More details of the simulation parameters are provided in Table II and III. In each simulation, the UAV completes one flight starting and ending at the initial waypoint / take-off point.

TABLE II: NODE ATTRIBUTES

Node attributes	
Node type	Mobile
Node speed	Variable ($\approx 2\text{m/s} - \approx 5\text{m/s}$)
Node coverage	Variable (80m - 720m)
Mobility	Levy model
Cache size	No Constraint
Energy	No Constraint

Algorithm 1: Connection-less data exchange protocol

0 = False, 1 = True

Node as a Sender

```

while DataToSend == 1 do
  Scan for nearby nodes in coverage range;
  if NeighborsFound == 1 then
    transmit message/packet to all neighbors in range;
    DataToSend = 0;
  else
    break;
  end
end

```

Node as a Receiver

```

while DataToReceive == 1 do
  Receive incoming message from the sender;
  if NodeCache == null then
    push message to receiver's NodeCache;
    DataToReceive = 0;
  else
    if message  $\notin$  NodeCache then
      push message to receiver's NodeCache;
    else
      if timestampNewMsg > timestampOldMsg then
        remove old message from receiver's NodeCache;
        push new message to receiver's NodeCache;
      else
        discard received message;
      end
    end
    DataToReceive = 0;
  end
end

```

TABLE III: SIMULATION AND UAV CHARACTERISTICS

Simulation Parameters	
Simulations	10 per scenario
Node distribution area (Nda)	10km x 10km
Simulation duration	UAV's roundtrip ($\approx 45\text{mins}$)
UAV altitude	Constant (100m)
UAV speed	Constant ($\approx 15\text{m/s}$)
UAV coverage radius	Variable (80m - 720m)
UAV mobility	Linear (waypoint-waypoint)
Transfer time (UAV \leftrightarrow node)	Instantaneous
Node Density(KM^{-2})	1 - 7

Node densities are considered using information presented in [27], [28] in addition to the values used for similar work in [29]. For the sake of simplicity, it is also assumed that the data transfer time between the nodes and the UAV is instantaneous. Specific UAV attributes can be seen in Table 3 which have been set considering current domestic and commercial UAV characteristics. It is assumed that the UAV is not resource-constrained compared to the sensor nodes, thus it can house an antenna potentially providing greater coverage than the sensing nodes. The node attributes can be seen in Table II. For the sake

of this research, it is assumed that the nodes are not constrained by data storage or battery life; however, the coverage range has been selected based on nodes available in the market and discussed in [30].

Figure 4 shows the percentage of nodes encountered, i.e. data sent to the UAV on its round trip, when the coverage of the UAV is set to 720m with a node density of 7 per square kilometers while varying the coverage radius of the nodes. We see that the percentage of encountered nodes (all 4 variants) is highly dependent on the coverage range of the nodes. Also, the ability of the UAV to directly interact with the mobile nodes provides considerable benefit. Figure 5 and 6 show the percentage improvement trend in nodes encountered by the UAV when caching is enabled (relative to the non-caching case) between the nodes and the UAV as the coverage range of the sensor nodes is increased. Figure 5 shows the improvement when the UAV is able to capture the messages directly from the nodes whereas Figure 6 is when the UAV only picks up messages from the fixed waypoints it visits on its flight. The characteristics in Figure 5 and 6 show how increasing sensor node coverage range can significantly improve the efficiency of the system.

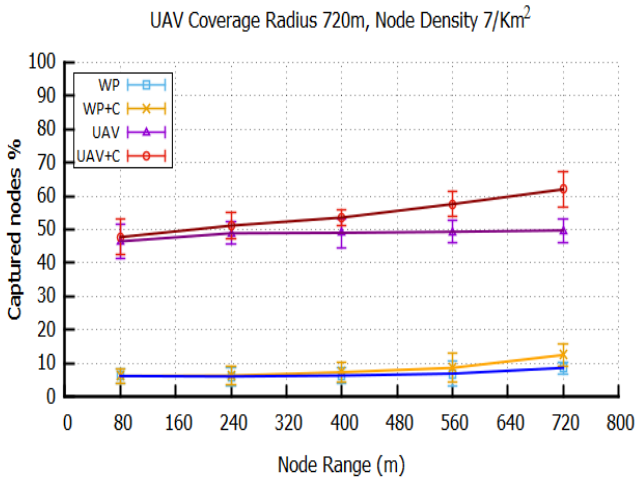


Figure 4: Percentage of encountered nodes with varying node coverage range (inc. 95% CI)

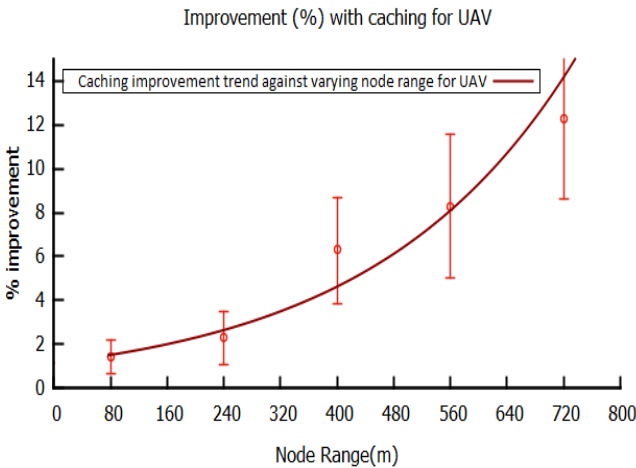


Figure 5: Trend line for Node Range and node encountering % relationship for UAV

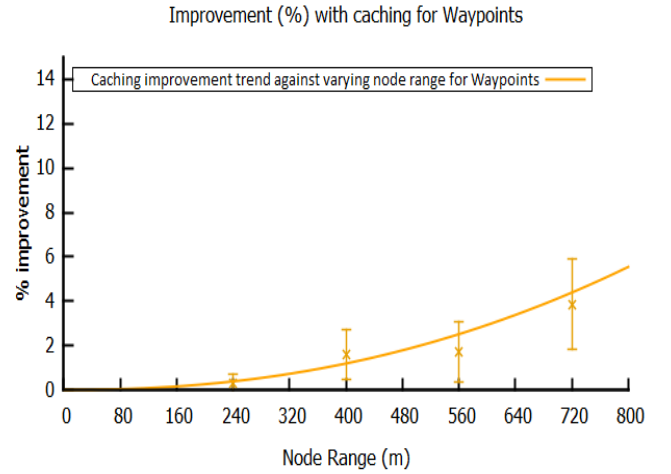


Figure 6: Trend line for Node Range and node encountering % relationship for Waypoints

Figure 7 shows the average number of nodes encountered by the UAV on its trip when the node density is increased in the node distribution area. We can see that with a higher number of nodes deployed in the distribution area, the advantage of using intra-node caching is much more beneficial. Figure 8 and Figure 9 show the trend line plot for gain in performance in terms of the percentage of nodes encountered by the UAV directly from the nodes and only from the waypoints during its trip while increasing the deployed node density. The figures in this scenario suggest that when caching is enable among the nodes, the gain for the UAV picking up data after direct encounter with the nodes follows an exponential trend whereas it follows a logarithmic approach when it obtains data only from the waypoints. Nevertheless, even in this scenario, we see that the advantages of using intra-nodal caching yields a significant advantage when the UAV is allowed to directly collect data from the deployed nodes compared to the scenario when no caching exists. All charts show results with a 95% confidence interval.

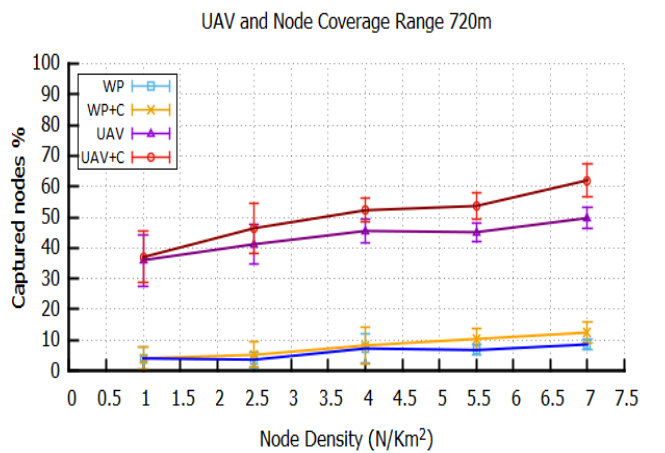


Figure 7: Encountered nodes % with varying deployed node density in node distribution area (inc 95% CI)

V. CONCLUSIONS

It is clear from the results that the use of caching in a mobile ad-hoc sensor network where a UAV is used to relay information from the nodes to the sink and vice versa rather than a routing protocol yields significant benefits compared to the scenario where no caching is used. We can see from the results presented in Section 4 that even for a short duration (45mins), we were able to achieve around 12% improvement on the encountered node percentage when the UAV interacts with the nodes directly and around 4% improvement when the UAV only collects data from the waypoints. Another thing to notice here is that the UAV was allowed only one trip across the field in our experimentations however, we predict that we can further see a significant increase in the nodes seen percentage provided that the UAV makes more than one trips to collect data from the sensor nodes. In addition, the trip timings may also play a reasonable role in the percentage of nodes encountered by the UAV as we did notice in some cases that the number of nodes cached at the sensor nodes and the waypoints were higher than the nodes encountered by the UAV. In our experimentations, we chose the simplest locations for the waypoints, being around the edges of the field however, introducing no-go zones in the area and most visited zones can also have an impact on the results. With the addition of different obstacles in the distribution area, the positioning of the waypoints can greatly affect the performance of the approach. Another interesting area to consider is selective caching with a finite cache size as opposed to our infinite caching model and using selected caching. Deciding what information to cache will open another wide area of research. Additionally, effects with varying node speed and UAV coverage range are also an interesting area to look into and we expect that a UAV with higher coverage range will yield less benefits compared to the scenarios where node ranges are increased.

In addition to what has been discussed, there are several other parameters (node movement speed, UAV coverage area, UAV movement speed, UAV multiple trips, different waypoint locations) that we would like to check including different movement models and the use of multiple UAVs. We believe that our work opens up a very broad area of extensive research that can greatly benefit not only the field of tracking and preservation of wildlife but also, if implemented with custom parameters under different settings, can greatly assist in other areas as well including vehicular ad-hoc networks and disaster area networks as well.

REFERENCES

- [1] T. O. Olasupo, "Wireless Communication Modeling for the Deployment of Tiny IoT Devices in Rocky and Mountainous Environments," *IEEE Sensors Lett.*, vol. 3, no. 7, pp. 1–4, 2019.
- [2] "A detailed survey on Bandwidth efficient cluster based routing schemes in wireless sensor networks," no. Iccsit, pp. 1250–1253, 2019.
- [3] J. D. Á. V, D. L. A. S, R. P. León, and P. Sergio, "Analysis of energetic efficiency in routing protocols and sustaining quality service applied to a wireless sensor network," pp. 27–32, 2019.
- [4] T. M. Behera, S. K. Mohapatra, U. C. Samal, M. S. Khan, M.

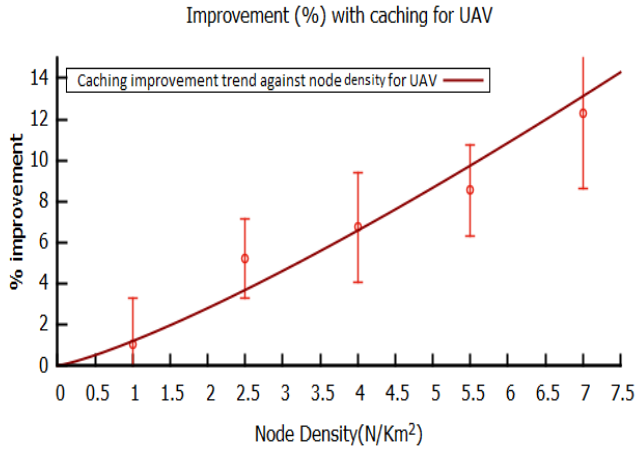


Figure 8: Trend line for Node Density and node encountering % relationship for UAV

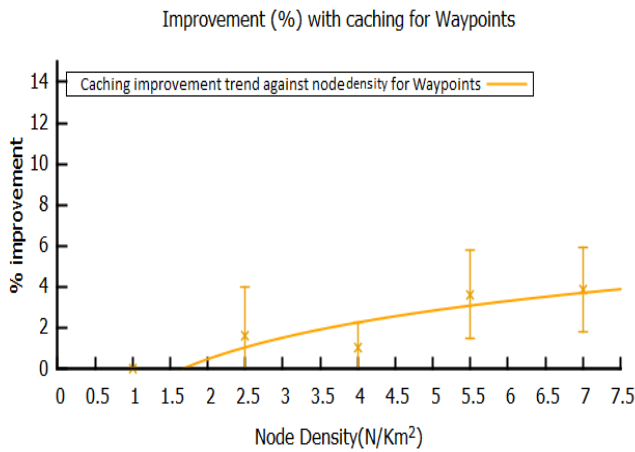


Figure 9: Trend line for Node Density and node encountering % relationship for Waypoints

Figure 10 shows the average number of nodes captured by the UAV on its trip while the node density is increased showing that at densely populated regions, the UAV was able to capture in access of 55% of the nodes deployed in the field.

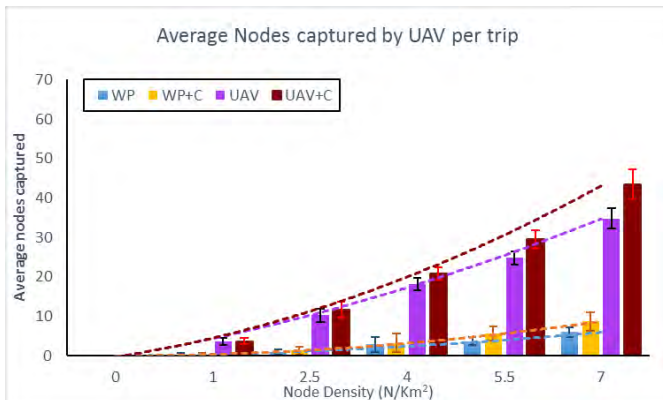


Figure 10: Average number of nodes encountered by the UAV while increasing the deployed node density (inc 95% CI)

- Daneshmand, and A. H. Gandomi, "I-SEP: An Improved Routing Protocol for Heterogeneous WSN for IoT-Based Environmental Monitoring," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 710–717, 2019.
- [5] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with ZebraNet," *Int. Conf. Archit. Support Program. Lang. Oper. Syst. - ASPLOS*, pp. 96–107, 2002.
- [6] N. Adam, C. Tapparello, M. N. Wijesundara, and W. Heinzelman, "JumboNet Elephant Tracking Using Delay-Tolerant Routing with Multiple Sinks," *2018 Int. Conf. Comput. Netw. Commun.*, pp. 689–695, 2018.
- [7] "Wildlife tracking devices - GPS, Sensors, Telemetry." [Online]. Available: <https://www.telemetrysolutions.com/wildlife-tracking-devices/>. [Accessed: 11-Feb-2020].
- [8] "Telematics - Wikipedia." [Online]. Available: <https://en.wikipedia.org/wiki/Telematics>. [Accessed: 11-Feb-2020].
- [9] "Argos system - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Argos_system. [Accessed: 11-Feb-2020].
- [10] "Automatic Packet Reporting System - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Automatic_Packet_Reporting_System. [Accessed: 11-Feb-2020].
- [11] "GPS tracking unit - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/GPS_tracking_unit. [Accessed: 11-Feb-2020].
- [12] "Electronic tagging - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Electronic_tagging. [Accessed: 11-Feb-2020].
- [13] J. Xu, G. Solmaz, R. Rahmatizadeh, D. Turgut, and L. Boloni, "Animal monitoring with unmanned aerial vehicle-Aided wireless sensor networks," *Proc. - Conf. Local Comput. Networks, LCN*, vol. 26-29-Octo, pp. 125–132, 2015.
- [14] S. Tansuriyavong, H. Koja, M. Kyan, and T. Anezaki, "The Development of Wildlife Tracking System Using Mobile Phone Communication Network and Drone," *2018 Int. Conf. Intell. Informatics Biomed. Sci. ICIBMS 2018*, vol. 3, pp. 351–354, 2018.
- [15] A. Arvanitaki and N. Pappas, "Modeling of a UAV-based data collection system," *IEEE Int. Work. Comput. Aided Model. Des. Commun. Links Networks, CAMAD*, vol. 2017-June, 2017.
- [16] S. Say, H. Inata, M. E. Ernawan, Z. Pan, J. Liu, and S. Shimamoto, "Partnership and data forwarding model for data acquisition in UAV-aided sensor networks," *2017 14th IEEE Annu. Consum. Commun. Netw. Conf. CCNC 2017*, pp. 933–938, 2017.
- [17] A. F. Khalifeh, M. AlQudah, R. Tanash, and K. A. Darabkh, "A Simulation Study for UAV- Aided Wireless Sensor Network Utilizing ZigBee Protocol," *Int. Conf. Wirel. Mob. Comput. Netw. Commun.*, vol. 2018-Octob, no. 1, pp. 181–184, 2018.
- [18] X. Ma, R. Kacimi, and R. Dhaou, "Adaptive hybrid MAC protocols for UAV-assisted mobile sensor networks," *CCNC 2018 - 2018 15th IEEE Annu. Consum. Commun. Netw. Conf.*, vol. 2018-Janua, pp. 1–4, 2018.
- [19] A. M. Reynolds, "Current status and future directions of Levy walk research," *Biol. Open*, vol. 7, no. 1, pp. 1–6, 2018.
- [20] M. E. Wosniack, M. C. Santos, E. P. Raposo, G. M. Viswanathan, and M. G. E. da Luz, *The evolutionary origins of Lévy walk foraging*, vol. 13, no. 10. 2017.
- [21] N. Kumbhare, A. Rao, C. Gniady, W. Fink, and J. Rozenblit, "Waypoint-to-waypoint energy-efficient path planning for multi-copters," *IEEE Aerosp. Conf. Proc.*, pp. 1–11, 2017.
- [22] M. T. S. Ibrahim, S. V. Ragavan, and S. G. Ponnambalam, "Way point based deliberative path planner for navigation," *IEEE/ASME Int. Conf. Adv. Intell. Mechatronics, AIM*, pp. 881–886, 2009.
- [23] X. Dai, G. Shannon, R. Slotow, B. Page, and K. J. Duffy, "Short-Duration Daytime Movements of a Cow Herd of African Elephants," *J. Mammal.*, vol. 88, no. 1, pp. 151–157, 2007.
- [24] K. M. Njoki, "Elephant Foraging Behaviour: Application of Levy Flights in Geo-information Science and Remote Sensing," 2009.
- [25] N. Shadrack, M. O. Moses, M. Joseph, M. Shadrack, N. Steve, and I. James, "Home range sizes and space use of African elephants (*Loxodonta africana*) in the Southern Kenya and Northern Tanzania borderland landscape," *Int. J. Biodivers. Conserv.*, vol. 9, no. 1, pp. 9–26, 2017.
- [26] "Basic Facts About Elephants - Global Sanctuary For Elephants." [Online]. Available: <https://globalelephants.org/the-basics/>. [Accessed: 18-Aug-2020].
- [27] K. Kangwana, *Studying Elephants*. 1996.
- [28] S. R. Loarie, R. J. V. Aarde, and S. L. Pimm, "Fences and artificial water affect African savannah elephant movement patterns," *Biol. Conserv.*, vol. 142, no. 12, pp. 3086–3098, 2009.
- [29] M. Dong, K. Ota, M. Lin, Z. Tang, S. Du, and H. Zhu, "UAV-assisted data gathering in wireless sensor networks," *J. Supercomput.*, vol. 70, no. 3, pp. 1142–1155, 2014.
- [30] R. P. Narayanan, T. V. Sarath, and V. V. Vineeth, "Survey on Motes Used in Wireless Sensor Networks: Performance & Parametric Analysis," *Wirel. Sens. Netw.*, vol. 08, no. 04, pp. 51–60, 2016.



ETAI 2: CYBER SECURITY AND MATHEMATICS

Анализа на безбедност во паметен дом со примена на алгоритми од машинско учење

Ирина Сенчук, Ана Чолакоска, Данијела Ефнушева
Институт за компјутески технологии и инженерство
ФЕИТ Скопје, УКИМ
Скопје, Северна Македонија
irinasencuk@yahoo.com, {acholak, danijela}@feit.ukim.edu.mk

Анстракт—Популарноста на паметните домови расте како одговор на експоненцијалниот раст на Интернет на нештата (IoT). Како резултат на тоа, одржувањето на безбедноста на поврзаните уреди станува сè поголем предизвик во мрежите од Интернет на нешта. Во овој труд се проучува детектирање на аномалии и упади во ваквите мрежи, со помош на алгоритми за машинско учење: логистичка регресија и метод на случајна шума (random forest). Двата од споменатите алгоритми се изработени во програмскиот јазик Пајтон, а податоците кои се користат за тренирање и тестирање се од UNSW-NB15 податочното множество. Резултатите кои се добиени укажуваат на Test F1 и AUC (area under the curve) вредности од 78,2% и 81,4% за алгоритмот на логистичка регресија (LR), и 89,3% и 97,7% за методот на случајна шума (CS), соодветно.

Клучни зборови—паметен дом; машинско учење; UNSW-NB15 податочното множество; Интернет на нешта; безбедност; детекција на аномалии;

I. ВОВЕД

Напредокот на технологијата овозможи поврзаност речиси помеѓу било какви уреди преку Интернет [1]. Овој тренд зададе многу безбедносни предизвици и во домот. Популарноста на паметните домови порасна како одговор на експоненцијалниот раст на Интернет на нештата (IoT). Имено, паметен дом може да се дефинира како живеалиште кое вклучува низа сензори, системи и уреди што можат да се пристапат, контролираат и следат од далечина преку комуникациска мрежа. Зголеменото распоредување на уреди поврзани со Интернет во домот, ги изложува корисниците на ризик, бидејќи личните информации стануваат далечински достапни. Напаѓачот може, на пример, да го прислушува безжичниот пренос на сензорите и да ги открие активностите на жителите. Исто така, напаѓачот може од далечина да ја преземе контролата врз домашните уреди користејќи ги за хакирање на домаќинството или како платформа за извршување напади кон други домени, како на пр. за преоптоварување на енергетската мрежа.

Многу истражувања дале добри резултати во откривање на мрежни напади [2-4]; сепак, само ограничен број од нив е насочен кон истражувања поврзани со Интернет на нешта [5-7], а уште помал број истражувања

користат податочни множества создадени од реален Интернет на нешта сообраќај [8]. Најчесто користените податочни множества за дизајнирање на нови системи за детектирање на упади се NSL-KDD множеството [9] и DARPA множеството [10]. Проблемот со овие две податочни множества е што ниту едното ниту другото не опфаќа податоци кои одговараат на мрежи на Интернет на нешта. Исто така, овие две податочни множества се создадени пред повеќе од една деценија и не можат да ги опишат и класифицираат новите типови напади, како што е Мираи ботнет [11]. Во поново време, се повеќе се користат алгоритмите за машинско учење, кои покажуваат ветувачки перформанси во откривање на аномалии во ваквите мрежи [12].

Во ова истражување се користи UNSW-NB15 [12] податочното множество со ботнет напади со податоци за IoT сензори, кое може да се категоризира во нормална активност или малициозен напад. Според тоа, целта на ова истражување е со примена на методи од машинско учење (логистичка регресија и случајна шума) врз избраното податочно множество да се креира модел кој ќе може да детектира напади преку диференцирање на аномалии од нормален податочен проток базиран на однесувањето во мрежата. Една предност на овој пристап е тоа што кога би се случил некој напад, однесувањето на мрежата ќе отстапи од нормалниот шаблон на однесување и аномалијата ќе биде детектирана [13].

Остатокот од трудот е организиран на следниот начин: во глава 2 е даден преглед на тековната состојба. Во глава 3 е направена анализа на податочното множество и неговите карактеристики, при што се дискутира подготовката за машинско учење. Во глава 4 се прикажани резултатите од машинско учење за двата алгоритми: логистичка регресија и метод на случајна шума. Во глава 5 се дадени заклучни согледувања од спроведеното истражување.

II. ТЕКОВНА СОСТОЈБА

Мрежните напади се едни од најголемите безбедносни проблеми во денешниот свет. Порастот на употребата на компјутери, мобилни телефони, сензори, IoT во мрежи, големи податоци, веб апликации, сервери, пресметување во облак и други компјутерски ресурси е голем. Со зголемувањето на мрежниот сообраќај, хакерите и

малициозните корисници постојано смислуваат нови начини на мрежни упади. Развиени се многу техники за откривање на овие упади кои се базираат на методи на податочно рударење и машинско учење.

Машинското учење е област од компјутерските науки која им дава на системите способност за автоматско учење и подобрување од искуство, без експлицитно програмирање. Процесот на учење започнува со набљудувања или податоци, со директно искуство или инструкции, со цел да се пронајдат шаблони во податоците и да се донесат подобри одлуки во иднина врз основа на примерите што биле дадени [14]. Алгоритмите за машинско учење се категоризираат како надгледувани и ненадгледувани (со и без надзор) [15].

Откривањето напади во Интернет на нешта има свои специфични предизвици заради хетерогеноста на уредите, ограничените компјутерски можности и огромниот број на поврзани уреди, што го отежнува вградувањето на системите за откривање на напад. Системите за откривање на напад се користат за заштита на податоците што се разменуваат помеѓу крајни точки и процеси во рамките на информацискиот систем [5-7].

Според Golman [16], ваквиот систем треба да биде дизајниран така што ќе прави разлика помеѓу нормално однесување и абнормално однесување засновано врз ефективен изграден модел на класификација во неговата внатрешност. Конкретно, тој предлага хибриден систем за откривање на напад врз основа на машина со носечки вектори (SVM) и C5.0. Користењето на ваква комбинација на алгоритми би ја подобрила точноста на откривање на напад, во споредба кога би се употребиле одделно.

Најаре [17] предлага различен модел за откривање на упад кој користи MapReduce за обработка на големи структурирани и неструктурирани податоци поставени во парови клучеви/вредности. Начинот на кој MapReduce произведува парови на клучеви/вредности за откривање на напад се потпира на користење на комбинација од Fuzzy CMeans (FCM) и машина со носечки вектори (SVM) за класификација.

Хибриден систем предлага и Tanpure [18]. Системот користи два алгоритми: K-means и Naïve Bayes за групирање и класифицирање на податоците. Моделот е дизајниран за да ги открие следниве типови на напади: оневозможување на услуга(DoS), сонда(Probe), далечинско во локално(R2L) и корисник во корисник со сите привилегии(U2R).

Дизајнирањето на шеми за откривање на аномалии за Интернет на нешта има сличен пристап како и дизајнирањето шеми за откривање аномалии за други уреди. Сепак, треба да се разгледаат уникатните предизвици. На пример, системи за детектирање на упад за кои е потребна значително голема компјутерска моќ или искористеност на меморијата не се соодветни за Интернет на нешта. Понатаму, ваквите уреди се хетерогени и генерираат хетерогени податоци. Тоа значи дека пристапите за откривање на аномалии во ваквите мрежи не треба да бидат ограничени на откривање мрежни напади.

Наместо тоа, тие треба да одат подлабоко и да детектираат аномалии во самите податоци и сообраќајот.

III. ПОДАТОЧНО МНОЖЕСТВО И ПОДГОТОВКА ЗА МАШИНСКО УЧЕЊЕ

UNSW-NB15 е податочно множество за мрежен сообраќај заснован на IoT со различни категории за нормални активности и малициозни напади. Ова податочното множество вклучува конвенционален мрежен сообраќај, како и голем број на мрежни напади изведени со ботнет. Станува збор за ново податочно множество, од 2015 година, каде сообраќајот е категоризиран во девет различни модерни типови на напади и широк спектар на реални нормални активности [12]. Ова податочно множество е генерирано во Австралијански центар за сајбер безбедност, при што содржи 257.673 записи кои се претставени со 49 карактеристики.

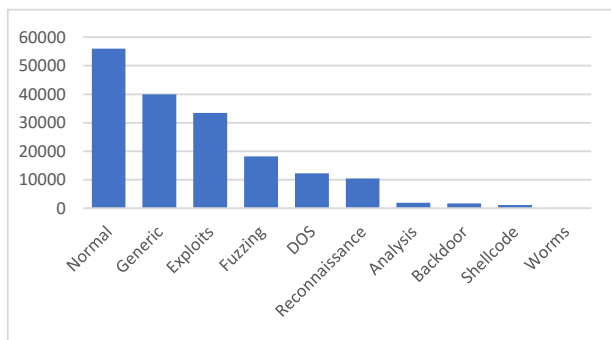
Со примена на алгоритми за машинско учење може да се употребат информациите зададени во карактеристиките со цел да се овозможи идентификување и класифицирање на сообраќајот во нормална или малициозна активност. Всушност, улогата на алгоритмите на машинско учење се состои во развој на мрежен форензички систем заснован на идентификатори на мрежни текови и карактеристики кои можат да следат сомнителни активности во мрежата. Конкретно во ова истражување се применуваат алгоритмите Логистичка регресија и метод на Случајна Шума со цел да се направи детекција на малициозна активност во мрежата.

Како што беше претходно спомнато, UNSW-NB15 податочното множество поддржува девет типови на напади, вклучувајќи:

- Fuzzers: Напад во кој напаѓачот се обидува да открие безбедносни дупки во оперативниот систем, програмата или мрежата и да ги направи овие ресурси суспендирани на одреден временски период, па дури и да ги оневозможи.
- Analysis: Тип на напади кои продираат низ веб-апликациите преку скенирање на порта, злонамерно скриптирање на html фајл и испраќање на несакани пораки итн.
- Backdoor: Техника во која напаѓачот може да ја заобиколи вообичаената автентикација и може да добие неовластен далечински пристап до систем.
- DoS: Напад во кој напаѓачот се обидува да оневозможи пристап до компјутерските ресурси, правејќи ги исклучително зафатени со цел да се спречи овластен пристап до ресурсите.
- Exploit: Напад што ги користат ранливостите, грешките или пропустите во оперативните системи (OC) или софтверот.
- Generic: Овој напад дејствува против криптографскиот систем и се обидува да го открие клучот на безбедносниот систем.

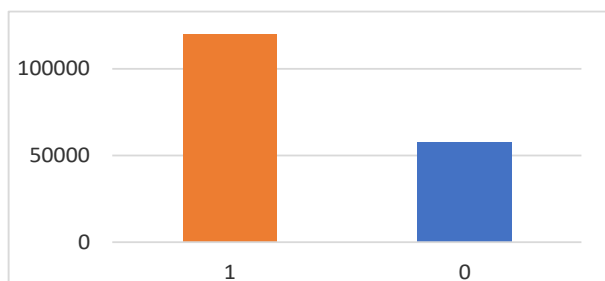
- **Reconnaissance:** Напад што собира информации за целната компјутерска мрежа со цел да се заобиколи нејзината безбедносна контрола.
- **Shellcode:** Напад на малициозен софтвер со кој напаѓачот продира во мало парче код, почнувајќи од школка, со цел да ја контролира компрометираната машина.
- **Worm(црв):** Малициозен софтвер кој се реплицира и се шири на други компјутери со користење на мрежата за ширење на нападот.

Со анализа на податоците откриени се следниве дистрибуции:



Сл. 1. Застапеност на записи за нормален сообраќај и различни типови малициозен сообраќај во UNSW-NB15 под. множество.

Од слика 1 може да се заклучи дека најзастапени напади во UNSW-NB15 податочното множество се Generic и Exploits, со вкупно 40,000 и 33,393 записи, соодветно. Дополнително доколку се направи анализа на бројот на записи од податочното множество кои се малициозни или нормални, се добива распределбата дадена на сл. 2. Тука може да се види дека има поголема застапеност на записи кои се малициозни (68,06%) во однос на застапеноста на записи за нормален сообраќај (31,94%). Во малициозни записи спаѓаат деветте типа на напади објаснети погоре. Вредностите 1 и 0 за малициозни и нормални записи соодветно, се вредностите кои што може да ги има карактеристиката label, со која се одредува за каков податок/запис станува збор.



Сл. 2. Споредба на број на записи за нормален и малициозен сообраќај во UNSW-NB15 податочно множество.

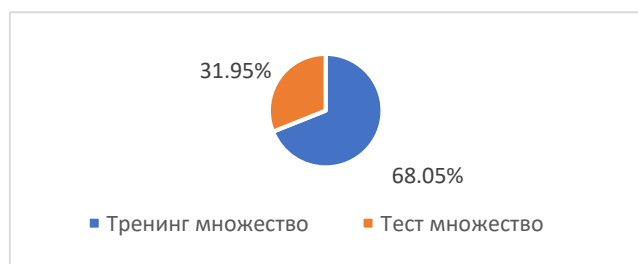
Карактеристиките на UNSW-NB15 податочното множество се класифицирани во шест групи:

- **Карактеристики на проток:** Оваа група ги вклучува атрибутите за идентификација помеѓу домаќините, како што се client-to-server или server-to-client.
- **Основни карактеристики:** Оваа категорија ги вклучува атрибутите што ги претставуваат протоколните врски.
- **Карактеристики на содржина:** Оваа група ги содржи/енкапсулира атрибутите на TCP/IP конекцијата, како и некои атрибути на http услуги.
- **Карактеристики на време:** Оваа група ги содржи атрибутите на времето, на пр. времето на пристигнување помеѓу пакетите, времето на започнување/крај на пакетот и round-trip времето на протоколот TCP.
- **Дополнителни генерирани карактеристики:** Оваа група може да се подели на две групи: карактеристики за општа намена (од 36 до 40) каде секоја од карактеристиките има своја цел да се заштити услугата на протоколите; и карактеристиките за поврзување (од 41 до 47) кои се изградени од проток од 100 запишани конекции засновани на секвенцијалниот редослед на карактеристиката употребена последниот пат.
- **Карактеристики на ознака:** Оваа група ја претставува етикетата (ознаката) на секој запис.

За обработката на UNSW-NB15 податочното множество се користи Пајтон во комбинација со алатката Jupyter Notebook која е open-source веб-апликација што може да се користи за да се креираат и споделуваат документи кои содржат код во живо, равенки, визуелизации и текст. Конкретно, употребени се следниве Пајтон библиотеки [14], при анализата, обработката и креирањето на моделите за класификација:

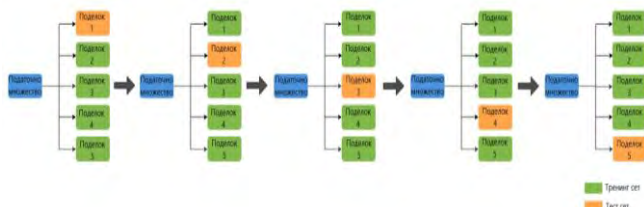
- **Pandas** е најпопуларната Пајтон библиотека која се користи за анализа и манипулација на податоци со високи перформанси, користејќи моќни структури на податоци. Колекцијата на алатки во пакетот Pandas е основен ресурс за подготовка и трансформација на податоци во Пајтон. Библиотеката Pandas се заснова на пакетот NumPy.
- **Numpy** е Пајтон библиотека што се користи за работа со низи и има функции за работа со линеарна алгебра и матрици.
- **matplotlib.pyplot** библиотеката е колекција на командни функции што прават matplotlib да работи како MATLAB.
- **Seaborn** библиотека ги користи matplotlib функциите за цртање на графикони.
- **sklearn - Scikit-learn** е open source библиотека за машинско учење која содржи голем број на алатки за машинско учење и статистичко моделирање вклучувајќи класификација, регресија, кластерирање. Оваа библиотека се користи за градење на модели на машинско учење.

UNSW-NB15 податочното множество е дефинирано од два фајла, множество за учење и множество за тестирање (UNSW_NB15_training-set.csv и UNSW_NB15_testing-set.csv соодветно). Бројот на записи во множеството за тренирање е 175.341, а во множеството за тестирање е 82.332 записи, од различни типови, напади и нормални. Оваа дистрибуција е прикажана на сл. 3.



Сл. 3. Споредба на застапеност на записи во тренинг и тест множество од UNSW-NB15 податочно множество.

При градење на моделите за машинско учење се употребува 5-Fold Cross-Validation, каде множеството на податоци е поделено на 5 дела. Во првата итерација, првиот поделок се користи за валидација на моделот, а остатокот (останатите 4 поделоци) се користат за тренирање на моделот. Во втората итерација, вториот поделок се користи како множество за валидација, додека останатите служат како множества за тренирање. Овој процес се повторува сè додека не се искористи секој поделок од петте како множество за валидација. Оваа постапка илустративно е прикажана на сл. 4.



Сл. 4. 5-Fold Cross-Validation применета за UNSW-NB15 податочно множество.

IV. АНАЛИЗА НА РЕЗУЛТАТИ ОД МАШИНСКО УЧЕЊЕ

Во овој труд е направено истражување со кое се развива систем за детектирање на напади преку диференцирање на аномалии од нормален податочен проток базиран на однесувањето во мрежата. Една предност на овој пристап е тоа што кога би се случил некој напад, однесувањето на мрежата ќе отстапи од нормалниот шаблон на однесување и аномалијата ќе биде детектирана. Во тој контекст, за избраното податочно множество се применуваат два алгоритма од машинско учење со цел прецизно детектирање на аномалии во реално време: логичка регресија (ЛР) и случајна шума (СШ) [15].

Во текот на анализата на моделите добиени се повеќе резултатите за двата алгоритми за машинско учење, кои се прикажани во табела 1. Иако во пракса небалансираноста на класите најчесто се движи во насока на поголем број на податоци кои припаѓаат на негативната класа во ова

истражување може да се забележи дека небалансираноста оди во насока на позитивната класа, бидејќи бројот на напади е поголем. Оваа разлика во бројот на позитивни и негативни класи не е толку екстремна, што може да се забележи и во точноста на моделот (Test Accuracy), при што моделот со подобра ROC (работната карактеристична крива на приемникот) има исто така подобра точност. Сепак, во случај на голема небалансираност, точноста на моделот нема да биде одраз на успешноста на моделот.

ТАБЕЛА 1. СПОРЕДБА НА РЕЗУЛТАТИ ДОБИЕНИ СО ЛОГИСТИЧКА РЕГРЕСИЈА И МЕТОД НА СЛУЧАЈНА ШУМА

	Логистичка регресија	Случајна шума
CV Fit Time	1,861176	32,858635
CV Accuracy mean	0,850936	0,959912
CV Precision mean	0,837826	0,963203
CV Recall mean	0,968452	0,978482
CV F1 mean	0,898415	0,970782
CV AUC mean	0,869224	0,993549
Test Accuracy	0,705983	0,870937
Test Precision	0,659813	0,817716
Test Recall	0,961992	0,985220
Test F1	0,782751	0,893687
Test AUC	0,814547	0,977302

Од горенаведените резултати може да се увиди дека методот на Случајна Шума дава подобри резултати за секоја од горенаведените метрики, при валидација и при тестирање на моделот, во споредба со резултатите добиени за Логистичката Регресија (исклучок е CV Fit Time). Во продолжение е дадена потемелна евалуацијата на резултатите за CV Fit Time, Test F1 и Test AUC параметрите добиени за двата алгоритми.

А. Анализа на CV Fit Time метрика

На сл. 5 е дадена споредба на CV Fit Time за двата алгоритма, каде може да се забележи дека времето за учење и валидација кај Логистичка Регресија е 17,65 пати помало во однос на времето на Случајна Шума.



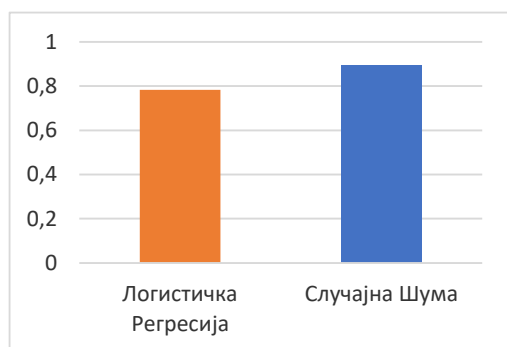
Сл. 5. Споредба на CV Fit Time метрика за Логистичка Регресија и метод на Случајна Шума, применети на UNSW-NB15 податочно множество.

Б. Анализа на Test F1 метрика

Целта на ова истражување е да се истренира модел на машинско учење со цел да ги класифицира типовите на записи и да препознае кога станува збор за напад. Како

резултат на тоа, моделот треба да има релативно висока способност за покривање и висока прецизност. Соодветно на тоа, се избира Test F1 метриката како метрика за проценка. F1 резултатот може да се толкува како просек на прецизноста (precision) и отповикувањето (recall), каде што резултатот F1 ја достигнува својата најдобра вредност кога е 1, а најлош резултат кога е 0.

Според горенаведената табела и сл. 6 може да се увиди дека моделот со Случајна Шума го надминува моделот со Логистичка Регресија бидејќи тој има највисок F1 резултат, но сепак и моделот со Логистичка Регресија е блиску, па и двата се валидни и прифатливи модели. Подесување на хиперпараметрите на овие модели ќе овозможи подобри перформанси во повеќето случаи. Сепак, за ова податочно множество може да се каже дека моделот на Случајна Шума е најдобар основен модел за класификација.

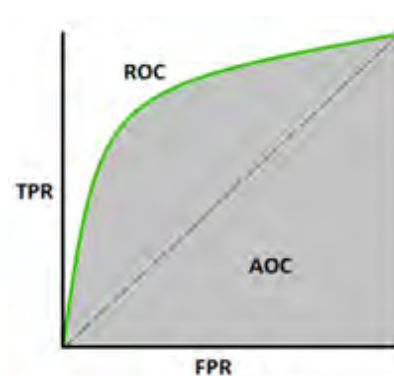


Сл. 6. Споредба на Test F1 метрика за Логистичка Регресија и метод на Случајна Шума, применети на UNSW-NB15 податочното множество.

B. Анализа на ROC и AUC метрики

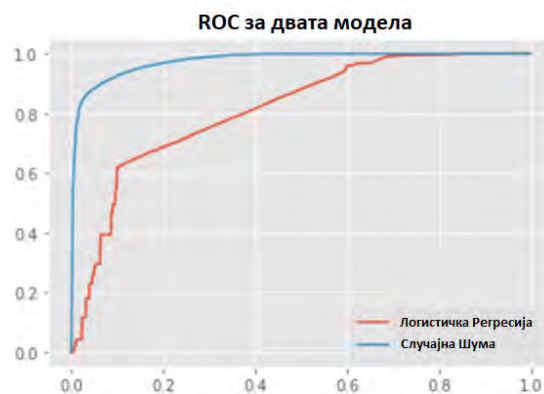
Работната карактеристична крива на приемникот или ROC-кривата (Receiver operating characteristic curve) е метрика за перформансите за проблемите на класификација при различни поставувања на прагот (threshold). ROC е крива на веројатност, а AUC (Area under the ROC curve) претставува степен или мерка на поделливост, односно покажува колку моделот е способен да прави разлика помеѓу класите. Колку е поголема AUC вредноста, толку подобро моделот предвидува 0 како 0 и 1 како 1. Аналогно на тоа, колку е поголема AUC вредноста, толку е подобар моделот во правење на разлика помеѓу напад или нормален повик.

Кривата ROC се создава со цртање на вистинската позитивна стапка или (TPR) наспроти лажната позитивна стапка или (FPR) при различни поставувања на прагот, како што е прикажано на сл. 7. Вистинска позитивна стапка се добива кога претпоставката дека записот е напад е точна. Неточна позитивна стапка се добива кога претпоставката дека записот е напад не е точна (односно тој запис не е напад) [15].

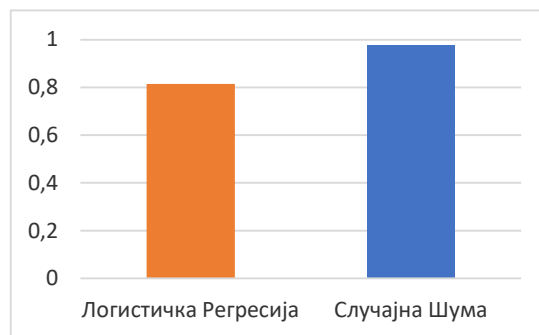


Сл. 7. Споредба на ROC и AUC во зависност од TPR и FPR параметри.

На сл. 8 и сл. 9 се прикажани резултатите за ROC кривите и вредностите на AUC за двата модели во овој практичен дел. Оттука може да се забележи дека ROC кривата како и вредноста на AUC е поголема за моделот со Случајна Шума во споредба со моделот со Логистичка Регресија, што укажува дека моделот со Случајна Шума е подобар во правење разлика помеѓу напад и нормален запис.



Сл. 8. Споредба на ROC криви за Логистичка Регресија и метод на Случајна Шума, применети на UNSW-NB15 податочното множество.



Сл. 9. Споредба на AUC вредности за Логистичка Регресија и метод на Случајна Шума, применети на UNSW-NB15 податочното множество.

V. ЗАКЛУЧОК

Во овој труд се обработени теми поврзани со безбедноста на паметните домови кои вклучуваат безбедносни IoT уреди и користење на машинско учење. Со применетите алгоритми за машинско учење, Логистичка Регресија и метод на Случајна Шума, врз UNSW-NB15 податочното множество извршена е успешна класификација на записите за мрежниот сообраќај, на нормални записи и записи кои се напад.

Доколку се земе во предвид Test F1 (78,2% и 89,3%) и АОС (81,4% и 97,7%) метриците на горенаведените модели (ЈР и СШ), може да се заклучи дека со машинско учење може да се подобри точноста на безбедносните системи кои се неопходни во паметните домови. Од горенаведените анализи на добиените резултати се покажува дека класификацијата со методот на Случајна Шума е поуспешна од онаа со Логистичка Регресија, при работа со UNSW-NB15 податочното множество.

КОРИСТЕНА ЛИТЕРАТУРА

- [1] K. Kimani, V. Oduol, and K. Langat, "Cyber security challenges for IoT-based smart grid networks," *International Journal of Critical Infrastructure Protection*, pp 36-49, 2019.
- [2] D. Satria and H. Ahmadian, "Designing home security monitoring system based Internet of things (IoT) model," *Jurnal Serambi Engineering*, vol. 3, No 1, 2018.
- [3] D. W. F. L. Vilela, A. Lotufo, and C. R. Santos, "Fuzzy ARTM AP neural network IDS Evaluation applied for real IEEE 802.11w data base," in *IEEE International Joint Conference on Neural Networks*, pp. 1-7, 2018.
- [4] Z. Tong and H. Ying, "Application of frequent item set mining algorithm in IDS based on Hadoop framework," in *IEEE Chinese Control and Decision Conference*, pp. 1908-1911, 2018.
- [5] I. Alrashdi, A. Alqazzaz, E. Aloufi, R. Alharthi, M. Zohdy and H. Ming, "AD-IoT: anomaly detection of IoT cyberattacks in smart city using machine learning," in *9th IEEE Annual Computing and Communication Work shop and Conference*, pp. 305-310, 2019.
- [6] A. Mishra and A. Dixit, "Resolving threats in IoT: ID spoofing to DDOS," in *9Th IEEE International Conference on Computing, Communication and Networking Technologies*, pp. 1-7, 2018.
- [7] M. M Shunnan, R. M. Khrais and A. A. Yateem, "IoT denial-of-service attack detection and prevention using hybrid IDS," in *IEEE International Arab Conference on Information Technology*, pp. 252-254, 2019.
- [8] F. A. Bakhtiar, E. S. Pramukantoro and H. Nihri, "A Lightweight IDS based on J48 algorithm for detecting DoS attacks on IoT middleware," in *1st IEEE Global Conference on Life Sciences and Technologies (LifeTech)*, pp. 41-42, 2019.
- [9] M. Tavallae, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed analysis of the KDD CUP 99 data set," in *Second IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009.
- [10] M. A. Ferrag, L. Maglaras, S. Moschogiannis and H. Janicke, "Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study," *Journal of Information Security and Applications*, Vol. 50, 2020.
- [11] J. Fruhlinger, "The Mirai botnet explained: how teen scammers and CCTV cameras almost brought down the internet," *CSO Online*, 2018, Available on: <https://www.csoonline.com/article/3258748/the-mirai-botnet-explained-how-teen-scammers-and-cctv-cameras-almost-brought-down-the-internet.html>
- [12] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," *2015 Military Communications and Information Systems Conference (MilCIS)*, pp. 1-6, 2015.
- [13] V. V. R. P. V. Jyothsna, V. R. Prasad, and K. M. Prasad, "A review of anomaly based intrusion detection systems," *International Journal of Computer Applications*, Vol. 28, Issue 7, pp 26-35, 2011.
- [14] D. James, *Introduction to Machine Learning with Python: a Guide for Beginners in Data Science*, 1st. Ed. USA: CreateSpace Independent Publishing Platform, 2018.
- [15] A. V. Joshi, *Machine Learning and Artificial Intelligence*, Springer, 2020.
- [16] V. Golman, "An Efficient hybrid intrusion detection system based on C5.0 and SVM," in *International Journal of Database Theory and Application*, Vol. 7, No. 2, pp. 59-70, 2014.
- [17] S. A. Hajare, "Detection of network attacks using big data analysis," in *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 4, Issue 5, pp. 86-88, 2016.
- [18] S. S. Tanpure et al., "Intrusion detection system in data mining using hybrid approach," in *International Journal of Computer Applications*, pp. 0975-8887, 2016.

Analysis of Smart Home Security by Applying Machine Learning Algorithms

Irina Senchuk, Ana Cholakoska, Danijela Efnusheva

Institute of Computer Technologies and Engineering
FEEIT, Ss. Cyrill and Methodius University in Skopje
North Macedonia

irinasencuk@yahoo.com, {acholak, danijela}@feit.ukim.edu.mk

Abstract—The popularity of smart homes is growing in response to the exponential growth of the Internet of Things (IoT). Consequently, keeping security of connected devices in IoT networks becomes a huge challenge. This paper studies the detection of anomalies and intrusions in such networks, using machine learning algorithms: random forest and logistic regression. Both of these algorithms are trained and tested using the Python programming language, and the data used for training and testing comes from the UNSW-NB15 dataset. The results obtained show Test F1 and AUC values of 78,2% and 81,4% for logistic regression (LR), and 89,3% and 97,7% for random forest, accordingly.

Keywords—smart home; machine learning; UNSW-NB15 data set; Internet of Things; security; anomaly detection;

Анализа на мрежна безбедност со примена на алгоритми од машинско учење

Мартина Шушлевска, Ана Чолакоска, Данијела Ефнушева

Институт за компјутески технологии и инженерство
ФЕИТ, Универзитет “Св. Кирил и Методиј” во Скопје
Северна Македонија
martinasuslevska@yahoo.com, {acholak, danijela}@feit.ukim.edu.mk

Анстракт—Со порастот на бројот на компјутери и уреди кои се поврзани на Интернет, се зголемија и ризиците од можни нарушувања на нивната безбедност. Како што се зголемува волуменот од собраните податоци за мрежниот сообраќај, така се повеќе се воочува примената на техниките за машинско учење за интелегентна обработка и анализата на овие големи податоци. Во оваа насока е и истражувањето претставено во овој труд, каде со примена на алгоритми за машинско учење се создава модел кој служи за откривање на аномалија, односно детекција на упад во мрежи. За потребите на ова истражување се користи KDD’99 податочното множество, при што развиениот модел се базира на следните алгоритми за класификација: Наивен Баесов (НБ) и машини со носечки вектори (СВМ). Добиените резултатите покажуваат дека НБ успешно ги класифицира нападите со точност од 88%, додека пак СВМ се карактеризира со повисока точност од 99%.

Клучни зборови—машинско учење; KDD’99 податочното множество; детекција на аномалии; безбедност на мрежи; системи за спречување упад;

I. ВОВЕД

Со огромниот раст на компјутерските мрежи и бројот на поврзани уреди, како и дополнителното зголемување на бројот на активирани апликации, се повеќе и повеќе системи се мета на неовластени напади. Всушност, се покажува дека компјутерските системи страдаат од безбедносни пропусти кои се технички тешки и економски скапи за решавање од страна на производителите. Затоа, улогата на системите за откривање на упад (Intrusion Detection System) [1], како уреди за специјална намена за откривање аномалии и напади во мрежата, станува се повеќе значајна при имплементирање на безбедност во компјутерско-комуникациските системи.

Истражувањата во областа на детектирање на упади подолг временски период се базирале на две можни методи: откривање на аномалија односно отстапување од вообичаената активност и сообраќај, и препознавање на обрасци на лошо однесување така што се детектираат

само упади кои содржат познати модели на напад. Вториот пристап вообичаено се фаворизира во комерцијални производи, на што се должи неговата предвидливост и висока точност. Од друга страна, академските истражувања вообичаено се базираат на пристапот на детектирање аномалија како помокен метод поради неговиот теоретски потенцијал за адресирање на напади [2].

Со спроведување на темелна анализа на неодамнешните истражувања за откривање на аномалија, се воочува примена на методи од машинско учење, за кои се вели дека имаат многу висока стапка на детекција од 98% каде стапката на лажно предвидени упади е околу 1% [3]. Техниките на машинско учење се широко распространети и применувани во повеќе истражувања [4-9] каде се покажува дека даваат солидни резултати при детектирање на упади.

Машинското учење го користи однесувањето во минатото за да идентификува обрасци и да гради модели кои помагаат да се предвиди идното однесување и идните настани, без експлицитно програмирање [10]. Процесот на учење започнува со набљудувања или податоци, со директно искуство или инструкции, со цел да се пронајдат шаблони во податоците и да се донесат подобри одлуки во иднина врз основа на примерите што биле дадени [11]. За да се потенцира потребата и придобивките од примената на машинско учење при детектирање на аномалии, во овој труд е направена анализа на алгоритми од машинско учење врз KDD’99 податочното множество [3]. Всушност, креиран е модел базиран на алгоритми за класификација (НБ и СВМ), каде зависно од природата на податоците и студијата на случај истите се подобрени со одреден метод на оптимизација. На крај, моделот е евалуиран, а резултатите се споредуваат преку користење на различни етикети за проверка, со цел утврдување на точноста на моделот.

Остатокот од трудот е организиран на следниот начин: глава 2 дава преглед на тековната состојба; во глава 3 е направена анализа на податочното множество и се дискутира неговата подготовка за машинско учење; во глава 4 се прикажани резултатите од машинско учење за двата применети алгоритми на класификација (НБ и СВМ); и во глава 5 се дадени заклучни согледувања.

II. ТЕКОВНА СОСТОЈБА

Системите за откривање на упади може да бидат софтверски или хардверски базирани, при што нивната намена е да мониторираат мрежни пакети или системи со цел за да откријат некоја штетна активност и да превземат одредени мерки. Постојат неколку типа на системи за откривање на упади, и истите се дискутирани во продолжение.

Мрежен систем за откривање на упади – NIDS (Network intrusion detection system) детектира штетна активност со помош на мониторирање и анализа на мрежниот сообраќај. Овој тип на IDS обично се активира кога пакетите влегуваат во одредена мрежа, при што тој одлучува кои пакети од Интернет ќе ги пропушти во локалната мрежа. Пример за ваков мрежен систем е Snort, [12].

Хост-базиран систем за откривање на упади - HIDS (Host-based intrusion detection system) детектира штетни активности со помош на мониторирање и анализа на системски повици, апликациски логови, листи на контрола на пристап итн. Ваквите системи обично содржат софтверски агент кој треба да биде инсталиран на оперативниот систем. Примери за вакви системи се: Tripwire, OSSEC, [13].

Безжичен систем за откривање на упади - WIDS (Wireless intrusion detection system) мониторира безжични мрежи со цел да открие некое штетно однесување (пр. премногу пакети за деавтентикација, премногу барања за broadcast и сл.). Најчесто ваквите системи работат на точки на пристап - AP (Access Point) и не дозволуваат одредени корисници да се поврзат на нив ако се забележи било каква штетна активност. Примери за вакви системи се Kismet и NetStumbler [14].

Анализа на однесувањето на мрежата - NBA (Network behavior analysis) претставува мониторирање на мрежниот сообраќај на пасивен начин со цел да се детектираат непознати и необични шаблони кои би можеле да предизвикаат одредена штета [15]. Препорачливо е да се користи заедно со огнен ѕид (firewall), како и со други типови на IDS системи.

Системи за спречување на упади - IPS (Intrusion Prevention Systems) се всушност надоградба на системите за откривање на упади. Онаму каде што IDS се користи за откривање и логирање на нападот, таму се користи и IPS – от за да го открие, блокира и логира нападот. IPS системите можат да спречат одредени напади додека истите се случуваат. Постојат повеќе типови на IPS системи, вклучувајќи: NIPS, HIPS, WIPS, NDA, [16].

Со оглед на тоа дека претходно дискутираните системи работат со голема количина на податоци, во последно време се повеќе се актуелизира примената на машинското учење кај IDS системите за детекција на аномалија. Постојат повеќе вакви истражувања кои дале добри резултати [4-9]. На пример, во [4] е претставена имплементација на Fuzzy ARTMAP невронска мрежа како

систем за превенција од упад. Дополнително во [5] е прикажана надоградба на Snort IDS која овозможува решавање на проблемот со кој се соочува Snort кога не може да одлучи каков ќе биде исходот од определен настан. За таа цел се применува посебен алгоритам (frequent itemsets mining) за податочно рударење, кој се базира на MapReduce за обработка на податоците. На сличен начин во [6] се предлага модел за откривање на упад кој користи MapReduce за обработка на големи структурирани и неструктурирани податоци поставени во парови клучеви/вредности. Начинот на кој MapReduce произведува парови на клучеви/ вредности за откривање на напад се потпира на користење на комбинација од Fuzzy CMeans (FCM) и машина за поддршка на вектори (SVM) за класификација. Дополнително, во [7] се предлага хибриден систем за откривање на напад врз основа на машина за поддршка на вектори (SVM) и C5.0. Користењето на ваква комбинација на алгоритми би ја подобрила точноста на откривање на напад, во споредба кога би се употребиле одделно. Хибриден систем е предложен и во [8], каде се применети два алгоритми (K-means и НБ) за групирање и класифицирање на податоците. Уште повеќе, во [9] е дадена компаративна студија околу различните пристапи на длабинско учење за сајбер безбедност при детекција на упад и примена на податочни множества за таа цел. Конкретно, во овој труд се употребува KDD'99 податочното множество со цел да се развие модел базиран на алгоритми за класификација со примена во IDS за откривање на аномалија. Повеќе детали за тоа, се дадени во следната глава.

III. ПОДАТОЧНО МНОЖЕСТВО И ПОДГОТОВКА ЗА МАШИНСКО УЧЕЊЕ

A. Опис на KDD'99 податочно множество

KDD'99 е најкористено податочно множество за проценка на методите за откривање на аномалија, иако датира уште од 1999 година [3]. Ова податочно множество е подготвено и изградено врз основа на податоците кои се добиени во евалуацијата на DARPA'98 IDS податочното множество [17]. DARPA'98 вклучува околу 4 гигабајти сурови (бинарни) tcpdump податоци за 7 неделен мрежен сообраќај, кој може да се обработи во околу 5 милиони конекции, секој со околу 100 бајти. Двете недели од податоците за тестот множеството имаат околу 2 милиони записи за врска. KDD податочното множество се состои од приближно 4.900.000 единечни вектори за поврзување од кои секој содржи 41 карактеристика и е означен или како нормален или како напад, со точно еден специфичен тип на напад. Симулираните напади спаѓаат во една од следниве четири категории:

- Напад на услуга (Denial of Service - DOS): е напад во кој напаѓачот прави ресурсите да се презафатени или преполни, со што се оневозможуваат легитимни барања или им се забранува пристап на легитимните корисници до машината.

- Напад на корен (U2R): е напад во кој напаѓачот започнува со пристап до обична корисничка сметка на системот (можеби стекната со душќање лозинки, напад на речник или социјален инженеринг) и е во состојба да искористи одредена ранливост за да добие root пристап до системот.
- Далечински до локален напад (R2L): се јавува кога напаѓачот кој има можност да испраќа пакети до машина преку мрежа, и при тоа нема сметка на таа машина, искористува одредена ранливост за да добие локален пристап како корисник на машината.
- Напад со пробување (Probe): е обид да се соберат информации за некоја компјутерска мрежа со цел да се заобиколат нејзините безбедносни контроли.

Карактеристиките на KDD'99 податочното множество се класифицирани во три групи. Во првата категорија се вклучени основните карактеристики, прикажани во табела 1. Тука спаѓаат сите атрибути кои можат да се извлечат од TCP/IP конекција. Повеќето од овие карактеристики доведуваат до имплицитно доцнење во откривањето.

ТАБЕЛА 1. ОСНОВНИ КАРАКТЕРИСТИКИ НА TCP КОНЕКЦИЈА

Карактеристика	Опис	Тип
траење	Должина на конекција (во сек)	континуална
тип на протокол	Тип на проткол, пр. Tcp, udp	дискретна
сервис	Мрежен сервис за дестинацијата, пр. http, telnet, итн	дискретна
src_bytes	Број на под. бајти од извор до дест.	континуална
dst_bytes	Број на под. бајти од дест. до извор	континуална
знаме	Нормален или грешка статус на конекција	дискретна
land	1 ако конекцијата е од/до истиот хост/порта; 0 инаку;	дискретна
wrong_fragment	Број на згрешени фрагменти	континуална
итно	Број на итни пакети	континуална

Во втората категорија спаѓаат карактеристиките на сообраќајот. Оваа категорија содржи карактеристики кои се пресметуваат во однос на интервалот на прозорецот и се поделени во две групи: карактеристики на “ист домаќин” (врските од изминатите 2 секунди, кои го имаат истиот дест. хост како тековната конекција), и карактеристики на “ист сервис” (врски кои во изминатите 2 секунди имаат ист сервис, како и тековната врска). Во табела 2 се прикажани само карактеристиките кои се однесуваат на временски прозорец. Покрај нив, во карактеристики на сообраќај спаѓаат уште 10 карактеристики кои се добиени од анализа на серија на конекции (пр. Колку барања се направени до ист хост за x-број на конекции?). Тука спаѓаат: dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_error_rate и dst_host_srv_error_rate.

ТАБЕЛА II. КАРАКТЕРИСТИКИ НА СООБРАЌАЈ ОД ВРЕМЕНСКИ ПРОЗОРЕЦ

Карактеристика	Опис	Тип
број	Број на конекции со ист хост како тековната конекција во последните 2s	континуална
Следните карактеристики се однесуваат на конекциите со “ист домаќин”		
error_rate	% на конекции со “SYN” грешки	континуална
error_rate	% на конекции со “REJ” грешки	континуална
same_srv_rate	% на конекции со ист сервис	континуална
diff_srv_rate	% на конекции со различен сервис	континуална
srv_count	Број на конекции со ист сервис како тековната конекција во последните 2s	континуална
Следните карактеристики се однесуваат на конекциите со “ист сервис”		
Srv_error_rate	% на конекции со “SYN” грешки	континуална
Srv_error_rate	% на конекции со “REJ” грешки	континуална
Srv_diff_host_rate	% на конекции со различен хост	континуална

Во третата категорија спаѓаат карактеристиките на содржина, кои се прикажани во табела 3. Генерално може да се каже дека DOS и probe нападите вклучуваат многу врски до истиот домаќин и за многу краток временски период, па полесно може да се детектираат во споредба со D2L и U2R нападите. Всушност, R2L и U2R нападите вообичаено се вградени во податочниот дел од пакетите, и при тоа вклучуваат само една конекција. За да се откријат ваквите напади, потребни се некои карактеристики со кои ќе може да се следи сомнително однесување во делот за податоци (пр. број на обиди за неуспешно најавање). Тоа се прави преку карактеристиките на содржина.

ТАБЕЛА III. КАРАКТЕРИСТИКИ НА СОДРЖИНА

Карактеристика	Опис	Тип
hot	Број на “hot” индикатори	континуална
num_failed_logins	Број на неуспешни обиди за најава	континуална
logged_in	1 при успешна најава; 0 инаку	дискретна
num_compromised	Број на “compromised” услови	континуална
root_shell	1 ако се пристапи до root shell; 0 инаку	дискретна
su_attempted	1 ако е употребена “su root” наредба; 0 инаку	дискретна
num_root	Број на “root” пристапи	континуална
num_file_ creations	Број на операции на креирање на фајл	континуална
num_shells	Број на shell промптови	континуална
num_access_files	Број на операции до фајлови за контрола на пристап	континуална
num_outbound_cmds	Број на излезни наредби во ftp сесија	континуална
is_hot_login	1 ако најавата припаѓа на “hot” листа; 0 инаку	дискретна
is_guest_login	1 ако најавата е “guest” најава; 0 инаку	дискретна

Б. Обработка на KDD'99 податочно множество

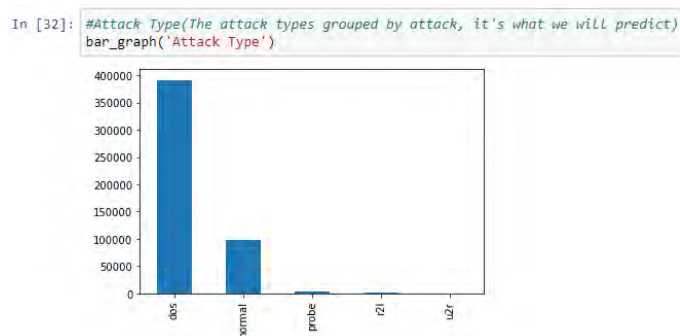
За обработката на ова податочно множество, се користи Пајтон во комбинација со повеќе библиотеки кои се потребни за евалуација на податочното множество. Тука спаѓаат библиотеките: Numpy, Seaborn, Pandas, Matplotlib со чија помош се врши анализа на податоците и нивна графичка репрезентација. Во продолжение се претставени чекорите од обработката на KDD'99 податочното множество

1) Читање на податочно множество

Со читање на KDD'99 податочно множество се добива целосна слика за податочното множество, вклучувајќи ги сите негови карактеристики и типови на напади. Во оваа фаза може да се направи анализа т.е. приказ на типовите на напади (DOS, U2R, R2L и probe) и поднапади, бројот на напади, димензионалноста на самото податочно множество, типот на податоци и сл.

2) Распределба на категориски и нумерички карактеристики

Со користење на методи и функции во Пајтон, се наоѓаат колоните со нумерички и категориски карактеристики од KDD'99 податочното множество. За да се направи визуелизација на овие податоци се користи `bar_graph` функција од Matplotlib библиотеката која како параметар прима одредена карактеристика. Со анализа на графичите, може да се добие слика за карактеристичноста на KDD'99 податочното множество. На пример на сл. 1 е прикажана застапеноста на карактеристиката тип на напад во 10% KDD'99 податочно множество за тренирање.



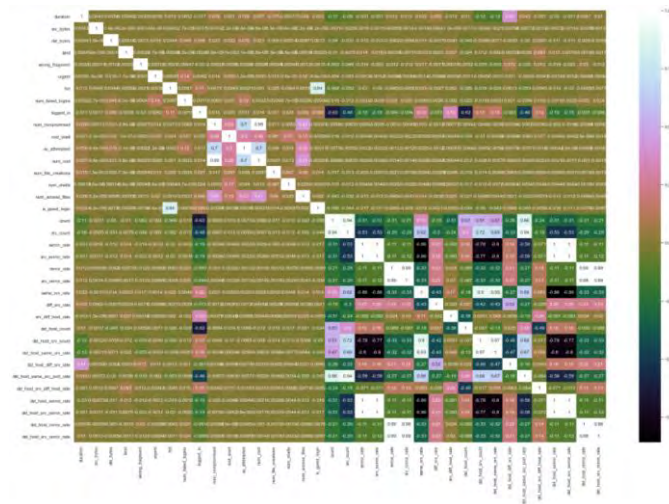
Сл. 1. Застапеност на карактеристиката тип на напад во KDD'99 податочно множество.

Од слика 1 може да се забележи дека 10% KDD'99 множеството за тренирање располага со 391458, 97278, 4107, 1126 и 52 записи кои означуваат DoS, нормален, probe, R2L и U2R тип на сообраќај, соодветно. Дополнително, може да се направи анализа и на малициозниот сообраќај, доколку се прикажат подгрупите на напади за основните типови (DOS, U2R, R2L и probe).

3) Податочна корелација со помош на HEATMAP

Откако со функцијата `drop()` ќе се отстранат колоните кои не се нумерички, следно е да се изврши корелација помеѓу секој пар карактеристики од KDD'99 податочното множество со употреба на функцијата `heatmap()`. `Heatmap()` дава графички приказ како матрица, каде индивидуалните вредности на матрицата се претставени како бои. Оваа функција е многу корисна за визуелизирање на концентрацијата на вредностите помеѓу две димензии на матрицата и на тој начин помага при наоѓање на обрасци и дава перспектива на длабочина.

На сл. 2 е дадена матрица на корелација на карактеристиките на KDD'99 податочното множество. Секој квадрат од матрицата ја покажува корелацијата помеѓу променливите на секоја оска. Корелацијата се движи од -1 до +1. Вредностите поблиску до нула значат дека нема линеарен тренд помеѓу двете променливи. Корелацијата која е близу до 1 ги претставува тие променливи кои се позитивно корелирани. Корелацијата која е поблиску до -1 е негативна корелација, каде наместо двете променливи да се зголемуваат, едната променлива многу опаѓа, а другата расте напоредно. Дијагоналите се сите со вредност 1 и се означени со бела боја затоа што тие квадрати ја корелираат секоја променлива со самата себе (ова е совршена корелација). Всушност, колку е потемна бојата на квадратите од матрицата, толку е поголема корелацијата помеѓу двете променливи.



Сл. 2. Матрица на корелација на карактеристики на KDD'99 податочно множество.

Доколку променливите се независни, тие сами по себе не треба да имаат некоја поврзаност, затоа што тоа може да влијае на точноста при тренирање на самиот алгоритам [18]. За да се отстрани таа зависност, потребно е да се отргне едната променлива, при многу висока корелација.

Како резултат на тоа, следен чекор од ова истражување е да се отстрани една променлива од оние кои се високо корелирани во KDD'99 податочното множество, според матрицата за корелација која е прикажано на сл. 2. Со спроведување на оваа постапка се селектираат 33 од вкупно 41 карактеристики на KDD'99 податочното множество. Истите се прикажани на сл. 3.

```
In [75]: df.columns
Out[75]: Index(['duration', 'protocol_type', 'service', 'flag', 'src_bytes',
'dst_bytes', 'land', 'wrong_fragment', 'urgent', 'hot',
'num_failed_logins', 'logged_in', 'num_compromised', 'root_shell',
'su_attempted', 'num_file_creations', 'num_shells', 'num_access_files',
'is_guest_login', 'count', 'srv_count', 'serror_rate', 'rerror_rate',
'same_srv_rate', 'diff_srv_rate', 'srv_diff_host_rate',
'dst_host_count', 'dst_host_srv_count', 'dst_host_diff_srv_rate',
'dst_host_same_src_port_rate', 'dst_host_srv_diff_host_rate', 'target',
'Attack Type'],
dtype='object')
```

Сл. 3. Селекција на карактеристики од KDD'99 податочното множество.

4) Моделирање

Во оваа фаза се врши вчитување на Пајтон модулите за тренирање и тестирање, при што се прави нормализација на опсегот на карактеристиките, што е прикажано на сл. 4. Тоа се прави со MinMaxScaler() функцијата, така што сите карактеристики се поставуваат во ист опсег, па на тој начин секоја од нив има подеднакво влијание во процесот на учење, бидејќи и грешките ќе бидат во ист опсег. Исто така, многу е важно во оваа фаза на тренирање, MinMaxScaler() да се изврши веднаш по train_test_split функцијата и да се повика за двете множества посебно, бидејќи инаку во тренирањето ќе влијаат и податоците од тест множеството, што не смее да се случи бидејќи во тој случај нема да се добијат реални резултати.

```
In [63]: df = df.drop('target', axis=1)
print(df.shape)

# Target variable and train set
Y = df[['Attack Type']]
X = df.drop(['Attack Type'], axis=1)

# Split test and train data
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.33, random_state=42)
sc = MinMaxScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.fit_transform(X_test)
print(X_train.shape, X_test.shape)
print(Y_train.shape, Y_test.shape)

(494021, 31)
(330994, 30) (163027, 30)
(330994, 1) (163027, 1)
```

Сл. 4. Подготовка на KDD'99 под. множество за машинско учење.

IV. АНАЛИЗА НА РЕЗУЛТАТИ ОД МАШИНСКО УЧЕЊЕ

Множеството за тренирање се состои од 80% од податочното множество, додека пак тест множество е изградено од 20% од податочното множество. За целите на ова истражување се употребуваат: Наивен Баесов алгоритам и машини со носечки вектори. На сл. 5 и 6 е прикажано нагодување на НБ и SVM моделите за тренирање и тестирање, соодветно. Резултатите, кои се сумаризирани во табела 4 покажуваат дека секој од алгоритмите успеал соодветно да научи од тренинг податочното множество и точно да ги класифицира податоците.

```
In [64]: # Gaussian Naive Bayes
from sklearn.naive_bayes import GaussianNB

In [67]: print("Training time: ",end_time-start_time)

Training time: 0.8646891117095947

In [69]: print("Testing time: ",end_time-start_time)

Testing time: 0.5219144821166992

In [71]: print("Train score is:", model1.score(X_train, Y_train))
print("Test score is:",model1.score(X_test,Y_test))

Train score is: 0.8795114110829804
Test score is: 0.882197427419998
```

Сл. 5. Нагодување на Naive Bayes (НБ) моделот. Потребно време за нагодување. Пресметка на точност врз множеството за тестирање.

```
In [72]: from sklearn.svm import SVC

In [79]: print("Training time: ",end_time-start_time)

Training time: 131.92943024635315

In [81]: print("Testing time: ",end_time-start_time)

Testing time: 34.339274644851685

In [82]: print("Train score is:", model4.score(X_train, Y_train))
print("Test score is:", model4.score(X_test,Y_test))

Train score is: 0.9987643280542849
Test score is: 0.998816147018592
```

Сл. 6. Нагодување на SVM моделот. Потребно време за нагодување. Пресметка на точност врз множеството за тестирање.

ТАБЕЛА IV. СПОРЕДБА НА НБ И SVM АЛГОРИТМИ

Алгоритам	Време на тренирање (сек)	Време на тестирање (сек)	Точност на тренирање (%)	Точност на тестирање (%)
НБ	0.8646891	0.5219145	87.95114	88.21974
SVM	131.929430	34.339275	99.8764	99.8816

Во табела 1 може да се забележи дека со НБ алгоритам успешно се решава проблемот на детектирање аномалија, односно тест податочното множество успешно ги класифира нападите со точност од 88%. Од она што е направено во истражувањето може да се каже дека НБ алгоритмот е лесен за примена и брзо ја предвидува класата на податоците за тестот (0.864сек). Исто така, тој се вклопува добро во повеќекласни предвидувања. Кога ќе се одржи претпоставката за независност, НБ класификаторот дава подобри резултати во споредба со други модели како што се: логистичката регресија, и други алгоритми. бидејќи се потребни помалку податоци за тренирање.

Во споредба со НБ класификаторот, табела 1 покажува дека SVM алгоритмот дава успешност/точност на повисоко ниво, односно 99%. Според тоа може да се заклучи дека за KDD'99 податочното множество НБ алгоритмот не е доволен и проблемот на детектирање аномалии во мрежен сообраќај треба да се разгледува во повеќе димензии. SVM алгоритмот дава многу подобра точност затоа што е помокен и поради причината дека НБ алгоритмот дава генерално подобри резултати за мали податочни множества, додека ова множество е огромно.

V. ЗАКЛУЧОК

Секоја година се адресираат нови отворени проблеми и области во обработката и анализата на податоци, каде како можно решение е искористување на алгоритмите за машинско учење. Впрочем, тоа не значи нивно искористување само во големите системи, туку и во разни апликации за секојдневна употреба. Дополнително, постојат најразлични open-source сервиси за анализа и обработка на податоци, кои уште повеќе го прошируваат нивното продажје на примена. Конкретно, во ова истражување се употребуваат НБ и СБМ алгоритми за машинско учење врз KDD'99 податочното множество со цел детектирање на упади во мрежа. Резултатите добиени од истражувањето укажуваат на лесна и сеопфатна примена на алгоритмите за машинско учење при обработка на податоците од KDD'99 податочното множество. Конкретно, со примена на НБ и СБМ алгоритмите се добива висока точност од 88%, односно 99% при класификацијата на нападите. Ова истражување може да се прошири и за други алгоритми и помодерни типови на напади, при што ќе се дозволи употреба и на посложени алгоритми, како и невронски мрежи.

КОРИСТЕНА ЛИТЕРАТУРА

- [1] L. H. Yeo, X. Che and S. Lakkaraju, "Understanding modern intrusion detection systems: a survey," Eastern Michigan University, USA, 2017.
- [2] V. V. R. P. V. Jyothsna, V. R. Prasad, and K. M. Prasad, "A review of anomaly based intrusion detection systems," International Journal of Computer Applications, Vol. 28, Issue 7, pp 26-35, 2011.
- [3] M. Tavallaei, E. Bagheri, W. Lu, and A.A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in Second IEEE International Conference on Computational Intelligence for Security and Defense Applications, pp. 53–58, 2009.
- [4] D. W. F. L. Vilela, A. Lotufo, and C. R. Santos, "Fuzzy ARTM AP neural network IDS Evaluation applied for real IEEE 802.11w data base," in IEEE International Joint Conference on Neural Networks, pp. 1-7, 2018.
- [5] Z. Tong and H. Ying, "Application of frequent item set mining algorithm in IDS based on Hadoop framework," in IEEE Chinese Control and Decision Conference, pp. 1908-1911, 2018.
- [6] S. A. Hajare, "Detection of network attacks using big data analysis," in International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 4, Issue 5, pp. 86-88, 2016.
- [7] V. Golman, "An Efficient hybrid intrusion detection system based on C5.0 and SVM," in International Journal of Database Theory and Application, Vol. 7, No. 2, pp. 59-70, 2014.
- [8] S. S. Tanpure et al., "Intrusion detection system in data mining using hybrid approach," in International Journal of Computer Applications, pp. 0975-8887, 2016.
- [9] M. A. Ferrag et al., "Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study," in Journal of Information Security and Applications, Vol. 50, 2020.
- [10] D. James, Introduction to Machine Learning with Python: a Guide for Beginners in Data Science, 1st. Ed. USA: CreateSpace Independent Publishing Platform, 2018.
- [11] A. V. Joshi, Machine Learning and Artificial Intelligence, Springer, 2020.
- [12] B. Caswell, J. Beale, A. Baker, Snort Intrusion Detection and Prevention Toolkit. Syngress, MA: Burlington, 2007.
- [13] S. B. Ambati, D. Vidyarthi, "A brief study and comparison of open source intrusion detection system and tools," International Journal of Advanced Computational Engineering and Networking, Vol. 1, Issue 10, pp. 26-32, 2013.
- [14] K. Hutchison, Wireless Intrusion Detection Systems, SANS Institute, White Paper, 2005.
- [15] W. Stallings, Network Security Essentials: Applications and Standards, 6th ed, USA: Pearson, 2017.
- [16] D. Stiawan, A. I. Shakhathreh, M. Y. Idris, K. K. A. Bakar and A.H. Abdullah, "Intrusion prevention system: a survey," in Journal of Theoretical and Applied Information Technology, Vol. 40, No. 1, 2012.
- [17] R. P. Lippmann, et al., "Evaluating intrusion detection systems: The 1998 darpa off-line intrusion detection evaluation," in IEEE DARPA Information Survivability Conference and Exposition, Vol. 2, 2000.
- [18] S. Madhavan, Mastering Python for Data Science, UK: Packt Publishing, 2015.

Network Security Analysis by Applying Machine Learning Algorithms

Martina Shushlevska, Ana Cholakoska, Danijela Efnusheva

Institute of Computer Technologies and Engineering
FEEIT, Ss. Cyrill and Methodius University in Skopje
North Macedonia

martinasuslevska@yahoo.com, {acholak, danijela}@feit.ukim.edu.mk

Abstract—With the increasing number of computers and devices connected to the Internet, the risks of possible breaches of their security have also elevated. As the volume of data collected on network traffic increases, so does the application of machine learning techniques for intelligent processing and analysis of this big data. In the research presented in this paper, machine learning algorithms are being used to create a model that serves to detect a network anomaly (network intrusion detection system). For the purposes of this research, the KDD'99 data set is being used, and the developed model is based on the following classification algorithms: Naive Bayes (NB) and Support Vector Machine (SVM). The obtained results show that NB successfully classifies the attacks with an accuracy of 88%, while the SVM algorithm is characterized by a higher accuracy of 99%.

Keywords—machine learning; KDD'99 data set; anomaly detection; network security; intrusion detection systems;

Нумеричко решавање на Лапласовата диференцијална равенка со примена на методот на конечни разлики

Бојана Петровска, Даниела Јанева, Емилија Ташева и Андријана Кухар

Факултет за електротехника и информациски технологии

Универзитет “Св. Кирил и Методиј” во Скопје

kuhar@feit.ukim.edu.mk

Анстракт—Диференцијални равенки кои немаат директно аналитичко решение се многу честа појава како во науката, така и во инженерството. Меѓу нив е и Лапласовата - диференцијална равенка од втор ред со која може да се опишат голем број електрични и физички појави во инженерството, а особено во неговата потесна област биомедицината. Користејќи ги придобивките од развојот на компјутерската технологија развиени се нумерички постапки за решавање на ваквите проблеми. Една од најшироко распространетите нумерички техники е методот на конечни разлики, со кој диференцијалната равенка се претвора во систем од линеарни равенки. Во овој труд е направен придонес кон нумеричкото решавање на Лапласовата равенка со развој на алгоритам за примена на методот на конечни разлики во слободниот програмски јазик Пајтон. Развиениот алгоритам вклучува итеративна техника за решавање на добиените системи од равенки, односно тој е оптимизиран од аспект на зачувување на компјутерските ресурси.

Клучни зборови—Лапласова равенка; нумеричко решавање; метод на конечни разлики; Пајтон; итеративни постапки.

I. ВОВЕД

Лапласовата равенка претставува парцијална диференцијална равенка од втор ред, со која се опишани голем број физички појави од различни области. Во биомедицината решавањето на оваа равенка е од голема важност. Имено, Лапласовата равенка се користи за пресметување на електричното поле кај биомедицинската техника за радиофреквенциска термална аблација на тумори [1]. Уште еден од примерите е пресметувањето на дебелината на мозочниот кортекс [2] со решавање на одредени форми на Лапласовата равенка. Со решавање на Лапласовата равенка се одредува и распределбата на потенцијалот и електричното поле генерирано во биолошките ткива од електродите за електротерапија на тумори [3], итн.

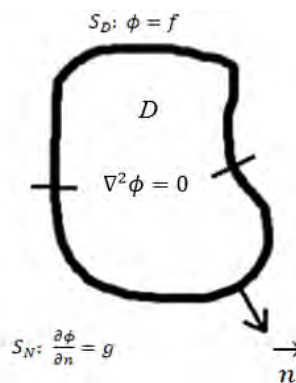
За одредување на еднозначно решение на Лапласовата равенка во даден домен од интерес, потребно е познавање на одредени услови, кои што непознатата функција ги задоволува на границите на доменот [4]. Постојат два главни случаи на дефинирање на граничните услови: граница на Дирихле – кога функцијата е дефинирана на дел

од границата и Нојманова граница - доколку изводот на функцијата е дефиниран на дел од границата. При решавање на дадениот проблем може целата граница да биде граница на Дирихле, или само одреден дел, а остатокот да биде Нојманова граница.

Типичен Лапласов проблем е шематски илустриран на Сл. 1. Во доменот D од Сл. 1 е поставена дво-димензионалната Лапласова равенка по променливата $\phi(x,y)$

$$\nabla^2 \phi = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \quad (1)$$

а на границата важи: $\phi = f$ во S_D и $\frac{\partial \phi}{\partial n} = g$ во S_N , каде што \vec{n} е нормален вектор на границата, S_D е границата на Дирихле, S_N е Нојмановата граница, f е позната вредност на функцијата од интерес, а g е позната вредност на првиот извод од истата.



Сл. 1. Лапласов проблем

Аналитичкото решавање на (1) за проблеми кои се среќаваат во инженерската практика е во најголем број од случаите неизводливо. На пример, во биомедицината пресметките се компликуваат поради сложената геометрија на домените – биолошките ткива и системи, како и нивната нехомогеност. Со развојот на компјутерските системи сè повеќе акцентот е ставен на нумеричките техники за решавање на применети

математички проблеми. Конкретно, Лапласовата равенка може нумерички да се решава со повеќе различни методи, како на пример: методот на конечни разлики (МКР), методот на конечни елементи, методот на гранични елементи итн.

Поради сето тоа, во овој труд е направен придонес кон нумеричкото решавање на Лапласовата равенка со примена на методот на конечни разлики во програмскиот јазик Пајтон. МКР е една од најчесто користените нумерички техники во инженерството. При негова примена Лапласовата диференцијална равенка се дискретизира со замена на парцијалните изводи со нивни апроксимации што се нарекуваат конечни разлики. Математичката постапка за добивање, како и обликот на изразите за конечни разлики, се презентирани во втората секција од овој труд. Во таа секција се опишани и двата главни случаи на дефинирање на граничните услови.

Со примената на МКР парцијалната диференцијална равенка се претвора во множество од линеарни равенки (во матрична форма). Директното решавање на систем од равенки на конечни разлики со помош на вообичаената математичка техника - методот на последователна елиминација, за поголем број на сегменти вклучува извршување на многу голем број математички операции [5]. Еден од прифатените нумерички начини за решавање на многу големи системи од равенки е техниката на итерации. Придобивката од итеративната техника е дека со нејзина примена не е потребно зачувување на големи матрици и на тој начин се оптимизира искористувањето на меморискиот простор.

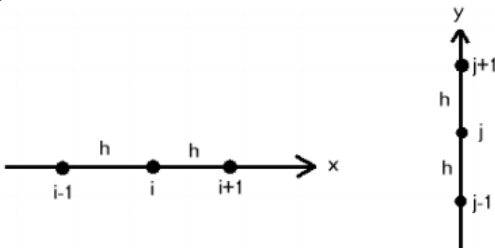
Резултатите од примената на итеративно решение во програмскиот јазик Пајтон за пресметување на електростатичкиот потенцијал ќе бидат изложени во третата секција од овој труд.

Во четвртата секција се сумирани најважните заклучоци и примена на истражувањата од овој труд.

II. МЕТОД НА КОНЕЧНИ РАЗЛИКИ

Методот на конечни разлики е апроксимативен нумерички метод со кој изводите во диференцијалните р-ки се заменуваат со релации на конечни разлики [6], [7] и [8]. Во продолжение ќе бидат изведени споменатите релации за дводимензионален случај.

Нека разгледаме два сегменти дефинирани со трите точки на x - оската поставени на растојание h , како што е прикажано на Сл. 2 - лево. Овие точки се нумерирани како $i-1$, i , $i+1$.



Сл. 2. Точки поставени на x и y оската

Нека вредностите на функцијата $\phi(x,y)$ во тие три точки изнесуваат $\phi_{i-1,j}$, $\phi_{i,j}$ и $\phi_{i+1,j}$. Тајлоровиот развој за $\phi_{i-1,j}$ и $\phi_{i+1,j}$ е од облик

$$\phi_{i-1,j} = \phi_{i,j} - \frac{\partial \phi}{\partial x} |_{i,j} h + \frac{\partial^2 \phi}{\partial x^2} |_{i,j} \frac{h^2}{2!} - \frac{\partial^3 \phi}{\partial x^3} |_{i,j} \frac{h^3}{3!} + \frac{\partial^4 \phi}{\partial x^4} |_{i,j} \frac{h^4}{4!} + O(h^5) \quad (2)$$

и

$$\phi_{i+1,j} = \phi_{i,j} + \frac{\partial \phi}{\partial x} |_{i,j} h + \frac{\partial^2 \phi}{\partial x^2} |_{i,j} \frac{h^2}{2!} + \frac{\partial^3 \phi}{\partial x^3} |_{i,j} \frac{h^3}{3!} + \frac{\partial^4 \phi}{\partial x^4} |_{i,j} \frac{h^4}{4!} + O(h^5) \quad (3)$$

каде што $|_{i,j}$ значи дека изводот е пресметан во точката i,j а h е растојанието помеѓу две точки на x - оската. Доколку ги собереме (2) и (3) се добива

$$\phi_{i-1,j} + \phi_{i+1,j} = 2\phi_{i,j} + \frac{\partial^2 \phi}{\partial x^2} |_{i,j} h^2 + \frac{\partial^4 \phi}{\partial x^4} |_{i,j} \frac{h^4}{12} + O(h^5) \quad (4).$$

Бидејќи станува збор за функција од две променливи, се прави апроксимација од втор ред, и членот $\frac{\partial^4 \phi}{\partial x^4} |_{i,j} \frac{h^4}{12}$ влегува во грешката. Со математичко средување на (4) може да се напише:

$$\frac{\partial^2 \phi}{\partial x^2} |_{i,j} = \frac{\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j}}{h^2} + O(h^2) \quad (5).$$

Десната страна од (5) е апроксимација со конечна разлика на $\frac{\partial^2 \phi}{\partial x^2} |_{i,j}$ со точност од втор ред. Изразот е со точност од втор ред затоа што грешката O е од h^2 ред.

Со одземање на (2) од (3) добиваме:

$$\phi_{i-1,j} - \phi_{i+1,j} = 2 \frac{\partial \phi}{\partial x} |_{i,j} h + \frac{\partial^3 \phi}{\partial x^3} |_{i,j} \frac{h^3}{3} + O(h^5) \quad (6)$$

или:

$$\frac{\partial \phi}{\partial x} |_{i,j} = \frac{\phi_{i+1,j} - \phi_{i-1,j}}{2h} + O(h^2) \quad (7).$$

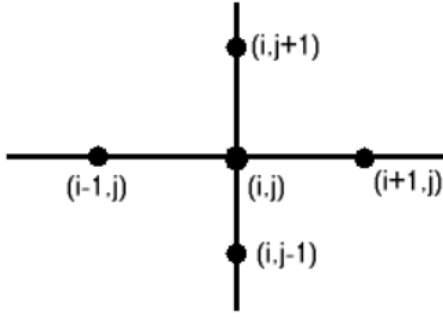
Равенката (7) е апроксимација од втор ред на конечната разлика на $\frac{\partial \phi}{\partial x} |_{i,j}$.

Следен чекор е да се дефинираат два сегменти со три точки $j-1, j$ и $j+1$ по y оската, како што е прикажано на Сл. 2 - десно. За овие точки на сличен начин се добива обликот на конечните разлики:

$$\frac{\partial^2 \phi}{\partial y^2} |_{i,j} = \frac{\phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1}}{h^2} + O(h^2) \quad (8)$$

$$\frac{\partial \phi}{\partial y} |_{i,j} = \frac{\phi_{i,j+1} - \phi_{i,j-1}}{2h} + O(h^2) \quad (9).$$

Со суперпозиција на левата и десната страна од Сл. 2 се добива дводимензионална мрежа со 5 точки кои во МКР се нарекуваат *јазли*. Добиената мрежа е прикажана на Сл. 3.



Сл. 3. Мрежа со 5 јазли добиена со МКР

Со комбинирање на равенките од (5) до (8) можеме да напишеме:

$$\left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2}\right)_{i,j} = \frac{\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j}}{h^2} + \frac{\phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1}}{h^2} \quad (10)$$

Заменувајќи ја (10) во Лапласовата равенка, се добива

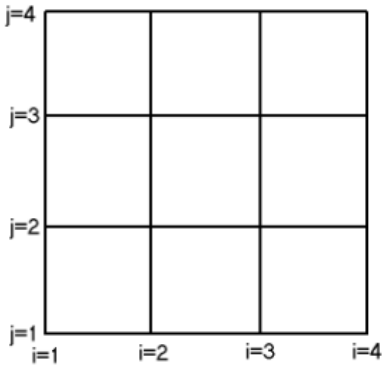
$$\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j} + \phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1} = 0 \quad (11)$$

односно

$$\phi_{i,j} = \frac{1}{4}(\phi_{i+1,j} + \phi_{i-1,j} + \phi_{i,j+1} + \phi_{i,j-1}) \quad (12)$$

A. Принципот на Дирихле

На Сл. 4 е претставен едноставен случај на правоаголен домен со само 4 внатрешни јазли.



Сл. 4. Правоаголен домен со димензии 4x4

Според принципот на Дирихле [9], вредноста на ϕ е позната на сите гранични рабови. Така, $\phi(1,2)$, $\phi(1,3)$, $\phi(2,4)$, $\phi(3,4)$, $\phi(4,3)$, $\phi(4,2)$, $\phi(3,1)$ и $\phi(2,1)$ се познати. Во продолжение ќе биде изложено како се одредуваат вредностите на $\phi(2,2)$, $\phi(3,2)$, $\phi(2,3)$ и $\phi(3,3)$ со примена на итеративната техника.

Го започнуваме итеративниот процес со претпоставката:

$$\phi^{(0)}(2,2) = \phi^{(0)}(3,2) = \phi^{(0)}(2,3) = \phi^{(0)}(3,3) = 0 \quad (13).$$

Во (13) горниот индекс, 0, е бројачот на итерацијата. Почнувајќи од долниот лев агол, се запишуваат вредностите на првата итерација

$$\begin{aligned} \phi^{(1)}(2,2) &= \frac{1}{4}(\phi(1,2) + \phi^{(0)}(3,2) + \phi(2,1) + \phi^{(0)}(2,3)) \\ \phi^{(1)}(3,2) &= \frac{1}{4}(\phi^{(1)}(2,2) + \phi(4,2) + \phi(3,1) + \phi^{(0)}(3,3)) \\ \phi^{(1)}(2,3) &= \frac{1}{4}(\phi(1,3) + \phi^{(0)}(3,3) + \phi^{(1)}(2,2) + \phi(2,4)) \\ \phi^{(1)}(3,3) &= \frac{1}{4}(\phi^{(1)}(2,3) + \phi(4,3) + \phi^{(1)}(3,2) + \phi(3,4)) \end{aligned} \quad (14)$$

Во првата линија на (14) е пресметана првата итеративна вредност на $\phi^{(1)}(2,2)$. Оваа прва итеративна вредност се користи за добивање на првата итеративна вредност на $\phi^{(1)}(3,2)$, во втората линија од (14). Ова е многу лесно остварливо, така што ги чуваме $\phi^{(0)}(2,2)$ и $\phi^{(1)}(2,2)$ на иста мемориска локација, така што го пребришуваме $\phi^{(0)}(2,2)$ со $\phi^{(1)}(2,2)$. Така, за секоја локација (i,j) , $\phi^{(iteration+1)}(i,j)$ го пребришува $\phi^{(iteration)}(i,j)$.

Во текот на итеративната постапка потребно е да се пресметува грешката на секоја итерација. Пред да се пребрише тековната вредност најпрво ја зачувуваме $\phi^{(iteration+1)}(i,j)$ во привремена променлива *temp*. Грешката во (i,j) се пресметува како:

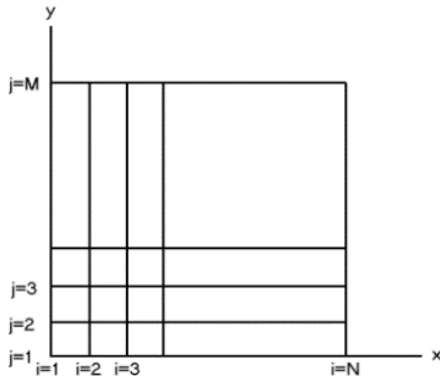
$$\epsilon(i,j) = \left| \frac{temp - \phi^{(iteration)}(i,j)}{temp} \right| \quad (15)$$

Потоа ја пребришуваме $\phi^{(iteration)}(i,j)$ со *temp*. На овој начин ги следиме релативните грешки на сите (i,j) локации. На крајот од итератиската јамка $(iteration + 1)$, ја пресметуваме репрезентативната релативна грешка како максимална вредност од грешките

$$\epsilon_{max} = \max \{\epsilon(i,j)\} \quad (16)$$

Од аспект на зачувување мемориски простор, треба да се следи бројачот на итерацијата. Потребно е да се специфицира „максималната дозволена итерација“ како би се избегнала појава на бесконечни итератиски јамки, поради потенцијална грешка во спецификацијата на податоците [5].

Следно, нека доменот од интерес е сегментиран во мрежа со произволен облик (Сл. 5) дефинирана со $1 \leq i \leq N$ и $1 \leq j \leq M$. Северниот, јужниот, источниот и западниот раб се од типот граници на Дирихле. Граничните услови $\{\phi(1,j); j \in [2, M-1]\}$, $\{\phi(N,j); j \in [2, M-1]\}$, $\{\phi(i,1); i \in [2, N-1]\}$, и $\{\phi(i,M); i \in [2, N-1]\}$ се внесуваат како познати податоци.



Сл. 5. Мрежа со произволни димензии

Секој чекор во постапката за итеративното решавање се врши според формулата:

$$\phi(i, j) = \frac{1}{4}[\phi(i+1, j) + \phi(i-1, j) + \phi(i, j-1) + \phi(i, j+1)] \quad (17)$$

$$\begin{aligned} i &\in [2, N-1], \\ j &\in [2, M-1]. \end{aligned}$$

Може да се забележи дека вредностите на аглиите $\phi(1,1)$, $\phi(N,1)$, $\phi(1,M)$, $\phi(N,M)$ не се јавуваат во итерационата јамка во (17). Овие вредности на аглиите се добиваат од изразите

$$\begin{aligned} \phi(1,1) &= \frac{1}{2}[\phi(1,2) + \phi(2,1)] \\ \phi(N,1) &= \frac{1}{2}[\phi(N-1,1) + \phi(N,2)] \\ \phi(1,M) &= \frac{1}{2}[\phi(1,M-1) + \phi(2,M)] \\ \phi(N,M) &= \frac{1}{2}[\phi(N,M-1) + \phi(N-1,M)] \end{aligned} \quad (18)$$

Како што беше напоменато, мрежа со димензии $M \times N$ ќе произведе систем од MN линеарни равенки.

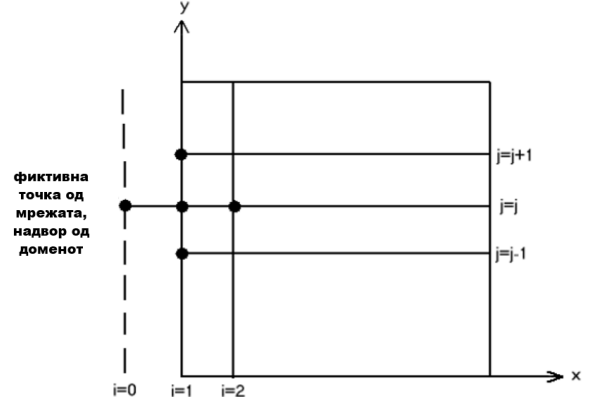
Б. Принципот на Нојман

Да го разгледаме проблемот на мрежата на Сл. 6. Граничните услови на Дирихле се специфицирани на северните, јужните и источните рабови. На западниот раб, не се познати вредности на функцијата, туку на нејзиниот извод. Затоа, се поставуваат Нојмановите гранични услови во облик:

$$\frac{\partial \phi}{\partial n} = -\frac{\partial \phi}{\partial x} = g(y) \quad (19)$$

Како што парцијалните изводи во Лапласовата равенка се апроксимираат со втор ред на конечни разлики во (10), парцијалните изводи во (19) треба, исто така, да бидат апроксимирани со шема од втор ред. За да може да се направи апроксимацијата треба да се избере фиктивна

точка $(0, j)$ од мрежата, надвор од доменот на проблемот, како што е прикажано на Сл. 6.



Сл. 6. Правоаголен домен со надворешна точка

Користејќи ја шемата за втор ред на конечни разлики, за извод од прв ред од (7), релацијата (19) го добива обликот

$$\frac{\partial \phi}{\partial x} \Big|_{(1,j)} = \frac{\phi(2,j) - \phi(0,j)}{2h} = -g(1,j) \quad (20)$$

Ако се запише Лапласовата равенка (12) во точка $(1, j)$ како:

$$\phi(1, j) = \frac{1}{4}[\phi(2, j) + \phi(0, j) + \phi(1, j+1) + \phi(1, j-1)] \quad (21)$$

Ако релацијата (20) се запише во облик:

$$\phi(0, j) = \phi(2, j) + 2hg(1, j) \quad (22)$$

Користејќи ги релациите (21) и (22) добиваме:

$$\phi(1, j) = \frac{1}{4}[2\phi(2, j) + 2hg(1, j) + \phi(1, j+1) + \phi(1, j-1)] \quad (23)$$

Ја користиме (23) за $2 \leq j \leq M-1$, каде $g(1, j)$ е позната функција. Бидејќи условите на Дирихле се специфицирани за северните, јужните и источните рабови, вредностите $[\phi(i, M), 2 \leq i \leq N-1]$, $[\phi(N, j), 2 \leq j \leq M-1]$, $[\phi(i, 1), 2 \leq i \leq N-1]$ се познати.

Оттука, секој чекор во постапката за итеративното решавање се врши според формулите

$$\begin{aligned} \phi(1, j) &= \frac{1}{4}[2\phi(2, j) + 2hg(1, j) + \phi(1, j+1) + \phi(1, j-1)] \\ &\quad (i=1), \quad \text{и} \\ \phi(i, j) &= \frac{1}{4}[\phi(i+1, j) + \phi(i-1, j) + \phi(i, j-1) + \phi(i, j+1)] \quad (24) \\ &\quad (i \neq 1) \\ i &\in [1, N-1] \\ j &\in [1, M-1] \end{aligned}$$

Вредностите на ϕ во јазлите во аглиите се пресметуваат користејќи ја релацијата (18).

III. ПРЕСМЕТУВАЊЕ НА ЕЛЕКТРОСТАТИЧКИ ПОТЕНЦИЈАЛ СО ПРИМЕНА НА МКР ВО ПРОГРАМСКИОТ ЈАЗИК ПАЈТОН

Распределбата на електростатичкиот потенцијал во отсуство на волуменски распределен електричен полнеж ја задоволува Лапласовата равенка (1). За нејзино решавање е потребна само информацијата за граничните услови. Како што беше напоменато во воведниот дел, ретки се проблемите од електростатиката што можат да се решат со употреба на аналитичката форма на Лапласовата равенка. Од тие причини, развиен е алгоритам за нумеричко решавање на споменатата равенка базиран на нумеричкиот метод на конечни разлики во слободниот програмски јазик Пајтон. Во продолжение ќе биде објаснета примената на развиениот алгоритам за пресметување на потенцијалот во дводимензионален домен.

Во доменот од интерес се дефинирани Дирихлеови гранични услови: $\phi = 0$ на сите рабови од доменот, освен на северниот раб каде што потенцијалот изнесува 10V. Доменот од интерес е квадратен и е сегментиран на мрежа од јазли со димензии 6x6. Со замената на релациите на конечни разлики (5) и (8) во секој јазол се добиваат 36 линеарни равенки кои се решаваат со итеративната техника објаснета во секцијата 2.

Решавањето на системот равенки започнува со назначување на иницијална вредност за сите јазли освен граничните - во овој случај таа вредност е 0. Ова може да се забележи во матрицата претставена на Сл. 7 – внатрешните полиња од споменатата матрица (означени со бела боја) се непознатите вредности на потенцијалот. Во неа може да се забележат и зададените гранични вредности кои не се менуваат во текот на итеративните пресметки. Во следните чекори итеративната техника од секцијата 2 се реализира со while циклус, каде контролен параметар е максималната релативна грешка определена со (16).

За илустрација на работата на развиениот итеративен алгоритам, на Сл. 8 се прикажани пресметаните вредности на потенцијалот по првата итерација. На споменатата слика може да се забележи дека во оваа фаза се пресметани само вредностите на потенцијалот во јазлите од првата редица на мрежата. На Сл. 9 се претставени конечните вредности на потенцијалот во сите јазли по завршување на последната итерација. За постигнување на однапред дефинираниот услов за точност кој го поставивме како максимална вредност на релативната грешка од 0.0002%, потребни се 54 итерации. Бројот на итерации зависи во голема мера од поставената иницијална вредност на променливата од интерес. Коректна претпоставка ќе го забрза конвергирањето на нумеричкото решение. Во овој случај, доколку се избере на пример вредноста 5V за иницијална вредност на непознатите, процесот би конвергирал побрзо бидејќи очекуваните вредности се помеѓу 0 и 10 V.

	0	1	2	3	4	5
0	10	10	10	10	10	10
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

Сл. 7. Матрицата со иницијалните (бели полиња) и почетните вредности на потенцијалот.

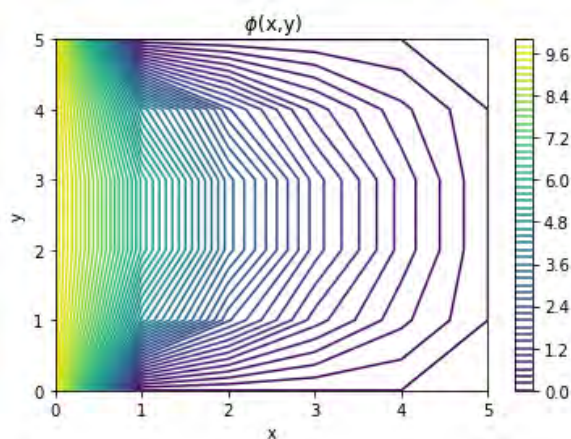
	0	1	2	3	4	5
0	10	10	10	10	10	10
1	0	2.5	2.5	2.5	2.5	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

Сл. 8. Вредноста на потенцијалот при првата итерација

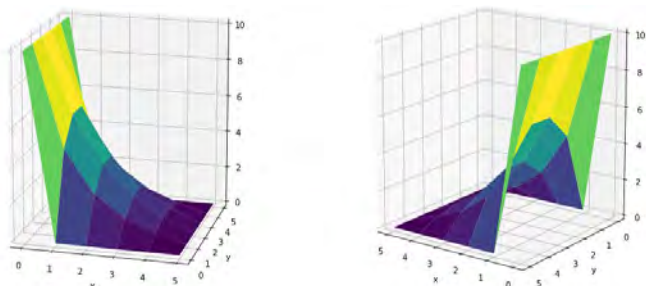
	0	1	2	3	4	5
0	10	10	10	10	10	10
1	0	4.54543	5.94694	5.94694	4.54543	0
2	0	2.23481	3.2954	3.2954	2.23481	0
3	0	1.09845	1.70449	1.70449	1.09845	0
4	0	0.454524	0.719662	0.719662	0.454524	0
5	0	0	0	0	0	0

Сл. 9. Потенцијалот при конечниот број на итерации.

На Сл. 10 е даден графички приказ на распределбата на потенцијалот во дводимензионалниот домен пресметана со развиениот алгоритам. На Сл. 11 се прикажани тридимензионални претстави на распределбата на потенцијалот како функција од координатите x и y , од две различни перспективи.



Сл. 10. Распределбата на потенцијалот во дводимензионалниот домен



Сл. 11. 3D преглед на распределбата на потенцијалот од различен агол

IV. ЗАКЛУЧОК

Лапласовата равенка наоѓа широка примена во опишување на физичките појави во повеќе различни научни гранки. Во биомедицината на пример, со решавање на Лапласовата равенка се одредува распределбата на потенцијал, електрично поле, топлина и притисок, што е од суштинска важност за голем број дијагностички и терапевтски биомедицински процедури. Примената на апроксимативни нумерички методи во биомедицинското инженерство се наметнува како решение на проблемите кои настануваат при решавање на Лапласовата равенка во биолошките ткива и системи кои имаат комплексни геометрии и нехомогени својства.

Од тие причини, во овој труд е изложена нумеричка постапка за решавање на Лапласовата равенка базирана на методот на конечни разлики. Алгоритмот за примена на нумеричката постапка е развиен во слободниот програмски јазик Пajтон. Тој се базира на итеративна техника за

решавање на добиените системи од равенки, а со тоа е оптимизиран од аспект на зачувување на компјутерските ресурси.

Во понатамошната работа алгоритмот би се проширил и на тридимензионални домени со покомлексни геометрии, со што ќе се овозможи пореалистично моделирање на биолошки домени.

КОРИСТЕНА ЛИТЕРАТУРА

- [1] M. Paruch, "Mathematical Modeling of Breast Tumor Destruction Using Fast Heating during Radiofrequency Ablation", <https://doi.org/10.3390/ma13010136>, Special Issue Applied Mathematics and Computer Methods in Materials, Mechanics and Engineering, 2019.
- [2] A. Joshi, C. Bhushan, R. Salloum, J. Wisnowski, David W. Shattuck and Richard M. Leahy, "Using the Anisotropic Laplace Equation to Compute Cortical Thickness", 21st International Conference, Granada, Spain, Proceedings, Part III, September 16-20, 2018.
- [3] B. Pupo et al., "Analytical and numerical solutions of the potential and electric field generated by different electrode arrays in a tumor tissue under electrotherapy", BioMedical Engineering OnLine, 2011.
- [4] A. K. Mitra, Finite Difference Method for the Solution of Laplace Equation, 2016.
- [5] A. Кухар, Предавања од предметот Биомедицинско инженерство, ФЕИТ, Скопје, 2019.
- [6] D. C. & C. Mingham, Introductory finite difference method for PDES, BookBoon, 2010.
- [7] L. Lapidus and G. F. Pinder, Numerical Methods of partial differential equations in science and engineering, New York: Wiley, 1982.
- [8] M. D. & S. Kiwne, "Finite Difference Method for Laplace Equation," Journal of Statistics and Mathematics, vol. 9, no. 1, pp. 11-13, 2014.
- [9] Parag V.Patil, Dr. J.S.V.R. Krishna Prasad, "Numerical Solution for Two Dimensional Laplace Equation with," IOSR Journal of Mathematics, pp. 66-75, 2013.

Numerical Solution of Laplace Differential Equation Using the Finite Difference Method

Bojana Petrovska, Daniela Janeva, Emilija Tasheva and Andrijana Kuhar

Faculty of Electrical Engineering and Information Technologies

"Ss. Cyril and Methodius" University in Skopje

kuhar@feit.ukim.edu.mk

Abstract—Differential equations without a straightforward analytical solution are very common in both science and engineering. Such a differential equation of the second order is the Laplace equation that describes a number of electrical and physical phenomena in engineering and especially in the field of biomedicine. Taking advantage of the benefits of the computer technologies rapid development, numerical procedures have been developed to solve such problems. One of the most widely used numerical technique is the Finite Difference Method, which transforms the differential equation into a system of linear equations. In this paper, a contribution is made to the numerical solution of Laplace's equation by developing an algorithm for applying the Finite Difference Method in the free programming language Python. The developed algorithm includes an iterative technique for solving the obtained systems of equations, i.e. it is optimized for preserving computer resources.

Keywords—Laplace equation; numerical solution; Finite Difference Method; Python; iterative procedures.

Modeling Population Dynamics and Economic Growth as Competing Species for North Macedonia

Stefan Boshkovski

Faculty of Electrical Engineering and Information
Technologies
Ss. Cyril and Methodius University
Skopje, Macedonia
stefan.boskovski.97@gmail.com

Sanja Atanasova

Faculty of Electrical Engineering and Information
Technologies
Ss. Cyril and Methodius University
Skopje, Macedonia
ksanja@feit.ukim.edu.mk

Abstract— *The aim of this paper is to apply nonlinear dynamical systems to models used in data analysis and through which the future flows of the system can be predicted. With this theory we analyze the population dynamics in the Republic of North Macedonia, and its revenues. Depending on the results, an insight is given into what trend we expect for the population and income in the future.*

Keywords—predator, prey, population GDP, Lotka-Volterra

I. INTRODUCTION

Since the beginning of the last century there are important demographic transition in the world which impact economic growth, and the correct understanding of this transition will help to predict important trends. In Republic of N. Macedonia this transition can be analyzed in aspect that population growth is diminishing, but also age structure of the population is changing, the population of young people is decreasing and the elderly proportion is rising. Different countries and regions show different stages of this demographic transition.

Lotka-Volterra (predator-prey) is nonlinear dynamical system well known in the biological, ecological and environmental literature, and has also been applied successfully in other fields. In this paper dynamical systems will be explained, and then used as model for predicting the future states of the economy and the demography of the Republic of North Macedonia. The motivation comes from the work in [4], [5], [6] and [8], and references therein, where dynamical systems are used for such predictions in different countries with different socio-economic growth.

II. NONLINEAR DYNAMICAL SYSTEMS FOR POPULATION MODELING

A. Basic classification of the dynamical systems

There are numerous classification of the dynamical systems, done by many criteria, [1], but the most common classification of the dynamical systems is based on their linear nature as

- Linear dynamical systems;
- Nonlinear dynamical systems.

The mathematical description for the linear dynamical systems has the following form:

$$\frac{dx}{dt} = \dot{x} = \underline{A} \cdot \underline{x}$$

where $\underline{x} \in \mathbb{R}^n$ represents a column vector where $n \in \mathbb{N}$, while the matrix \underline{A} is a square matrix consisting the proper coefficients of the system. It can be seen that from the form of the definition the systems behavior is proportional to the elements of the matrix and it is linear with respect to \underline{x} and $\dot{\underline{x}}$, which results in a linear system overall.

On the other hand, nonlinear systems can be represented in the following form:

$$\dot{x} = P(x, y), \quad \dot{y} = Q(x, y)$$

where P and Q are polynomial functions of x and y , and their exponent is bigger than 2 with respect to x and y . The linear dynamical systems cannot capture the nature of these processes. Nonlinear systems give more freedom in the modeling, but unlike the linear dynamical systems, they cannot be easily solved with common algebraic methods.

B. Predator-prey

Predator-prey is famous examples of a simple nonlinear dynamical systems found in many biological and ecological phenomena describing mostly population dynamics between two species with predator-prey interactions between each other. Usually the use of this model is based on two species that have a very simple diet. With this being said, it should be considered that when choosing these populations, their environment should be as homogenous as possible. This kind of environment will mean that the predator population will be depended on the prey population as a primary source of food, while the prey will be depended on the environment itself, such as the vegetation.

C. The Lotka-Volterra model

Lotka-Volterra model is a natural extension of the Verhulst or logistic model, which has the following form:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right)$$

where P represents the population, r is the proportional growth constant, K is the capacity of the system i.e. the maximum population of prey that the environment can support, and t is the time, so that the population will not grow to infinity, [2].

Assuming two competitive populations, two mathematical functions can be developed based on the growth rate of both populations. Thus, the rate of change in a population is equal to the difference between the total increase or number of newborns and the total decrease, the number of deaths in that population. The first population will be represented with the function $B(t)$ and it will describe the prey species, as for the second population, it will be described by the function $P(t)$, which represent the predator population.

The rate of change of the prey population is $\frac{dB(t)}{dt}$. The primary growth in the prey population is Malthusian, which means that the population grows in proportion to its own population or $a_1(t)B(t)$. It should be considered that the decrease in the prey population is based on the predation of the predator population. Predation is often modeled by assuming random contact between the species in proportion to their populations with a fixed percentage of those contacts resulting in death of the prey species. Mathematically, this is given by a negative term, $-a_2B(t)P(t)$. Clearly, if there are other major predators or if the predator population is low such that starvation due to overcrowding dominates the death rate, then alternative death terms would be more appropriate, [2]. It follows that the growth model for the prey population is:

$$\frac{dB(t)}{dt} = a_1B(t) - a_2B(t)P(t)$$

Likewise the prey population, the predator population rate of change is $\frac{dP(t)}{dt}$. The primary growth for the predator population depends on sufficient food for raising predator's newborns, which implies an adequate source of nutrients from predation on the prey. Thus, the growth of the predator population is similar to the death rate for the prey population with a different constant of proportionality. This implies that the growth of the predator population can be written as $b_2B(t)P(t)$, and the loss of predator can be considered as a type of reverse Malthusian growth. That is, in the absence of prey, the predator population declines in proportion to their own population, which mathematically is given by the negative modeling term, $-b_1P(t)$. The growth model for the predator population gives the following:

$$\frac{dP(t)}{dt} = -b_1P(t) + b_2B(t)P(t).$$

It is obvious that the growth equations of the two types of populations cannot realistically represent their complicated dynamics. This model ignores the climate effects, the interactions with other species, the age factors of the animals themselves as well as the spatial distribution factor of the populations.

D. Equilibrium analysis of the model

An equilibrium point of a dynamical system represents a stationary condition for the dynamics. A dynamical system can have zero, one or more equilibrium points. In order to find the equilibrium points of the dynamical system, the first derivatives should be equal to zero, this means that the solutions will represent non-changeable points through time. Taken the Lotka-Volterra model into account

$$\begin{cases} \frac{dB(t)}{dt} = a_1B(t) - a_2B(t)P(t) \\ \frac{dP(t)}{dt} = -b_1P(t) + b_2B(t)P(t) \end{cases} \quad (1)$$

we need to solve the system of equations [2],

$$\frac{dB(t)}{dt} = 0 \quad \text{and} \quad \frac{dP(t)}{dt} = 0. \quad (2)$$

Taken that B_e and P_e are the solutions for the equilibrium point for the predator and prey population respectively, (2) obtains the following form:

$$a_1B_e - a_2B_eP_e = 0$$

$$-b_1P_e + b_2B_eP_e = 0$$

In general, solving a system of nonlinear equations is not that easy, but in this case, it comes down to solving a system of equations algebraically. These two equations can be factored, which aids in finding the solution. The first equation can be written as $B_e(a_1 - a_2P_e) = 0$, which implies that either $B_e = 0$ or $(-b_1 + b_2P_e) = 0$. Thus, we have possible equilibria at

$$B_e = 0 \quad \text{or} \quad P_e = \frac{a_1}{a_2}.$$

Similarly, the second equation can be factored to give $P_e(-b_1 + b_2B_e) = 0$, which implies that either $P_e = 0$ or $(-b_1 + b_2B_e) = 0$. Thus, there is a possible equilibria at

$$P_e = 0 \quad \text{or} \quad B_e = \frac{b_1}{b_2}.$$

The simultaneous solution of these two equations shows that when $B_e = 0$, then $P_e = 0$, which gives rise to the trivial solution (extinction of both species), $(B_e, P_e) = (0, 0)$.

The other equilibrium is given by

$$(B_e, P_e) = \left(\frac{b_1}{b_2}, \frac{a_1}{a_2}\right). \quad (3)$$

Unfortunately, the equilibria is not enough to help explain the oscillatory behavior reflected in this model. Information about the stability of these equilibria and further numerical and qualitative analysis techniques are necessary. Performing a linear analysis at the equilibria can show how even the equilibrium analysis of this simple model can explain some important ideas in population dynamics of a predator prey system, [2].

E. Linear analysis

Linearization can be used to give important information about how the system behaves in the neighborhood of equilibrium points. This is important because many systems settle into a equilibrium state after some time, so they might tell us about the long-term behavior of the system. The idea is that one can approximate the nonlinear differential equations that govern the behavior of the system by linear differential equations.

A system of differential equations is linearized about its equilibria by finding the Jacobian of the vector function of the nonlinear system of differential equations [2]. For the predator-prey model given with (1), the Jacobian matrix is given by

$$J(B, P) = \begin{pmatrix} a_1 - a_2 P & -a_2 B \\ b_2 P & -b_1 + b_2 B \end{pmatrix}$$

For the equilibrium $(B_e, P_e) = (0, 0)$, the Jacobian matrix is

$$J(0,0) = \begin{pmatrix} a_1 & 0 \\ 0 & -b_1 \end{pmatrix}.$$

The eigenvalues and the associated eigenvectors are,

$$\lambda_1 = a_1, \xi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \lambda_2 = -b_1, \xi_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

which are real number. This shows that the equilibrium $(0, 0)$ is a saddle node with solutions exponentially growing along the B-axis and decaying along the P-axis, [2]. The linear solution of the system then has the form

$$\begin{pmatrix} B_L(t) \\ P_L(t) \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{a_1 t} + c_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-b_1 t}.$$

When the Jacobian is evaluated for the equilibrium $(B_e, P_e) = \left(\frac{b_1}{b_2}, \frac{a_1}{a_2}\right)$, the Jacobian matrix is

$$J\left(\frac{b_1}{b_2}, \frac{a_1}{a_2}\right) = \begin{pmatrix} 0 & \frac{a_2 b_1}{b_2} \\ \frac{a_1 b_2}{a_2} & 0 \end{pmatrix}.$$

This has the purely imaginary eigenvalues,

$$\lambda = \pm i\sqrt{a_1 b_1} \equiv \pm i\omega.$$

which are complex numbers. These eigenvalues indicate that the critical point $\left(\frac{b_1}{b_2}, \frac{a_1}{a_2}\right)$ has a purely oscillatory behavior and is on the verge of stability and it satisfies the assumptions for the oscillatory behavior of predator-prey model. The linearized solution of this equilibrium point can be presented as:

$$\begin{pmatrix} B_L(t) \\ P_L(t) \end{pmatrix} = c_1 \begin{pmatrix} \cos(\omega t) \\ -A \sin(\omega t) \end{pmatrix} + c_2 \begin{pmatrix} \sin(\omega t) \\ A \cos(\omega t) \end{pmatrix},$$

where $A = \frac{b_2}{a_2} \sqrt{\frac{a_1}{b_1}}.$

This produces a structurally unstable model. The model is structurally unstable because small perturbations from the nonlinear terms could result in the solution either spiraling toward or away from the equilibrium, [2].

III. LOTKA-VOLTERRA APPLICATIONS IN PREDICTIVE MODELS

In this section, we use information about the population and the gross domestic product (GDP) for Republic of N. Macedonia, from Trading Economics [3]. Based on the dataset, an attempt will be made to obtain a predictive model, which will be used to predict the upcoming economical and demographical fluctuations in Republic of N. Macedonia. This analysis was inspired by [4], [5] and [6], and references therein.

A. The dataset

We use available information for the population and the GDP from [3], which refers to Republic of N. Macedonia for the period of the last 30 years, more precisely the period from 1990 till 2019, covering the whole period of its independence. The information needed for training and obtaining the model are presented in Table 1.

Table 1 Annual GDP and population of Republic of N. Macedonia

Year	Population in millions	GDP in billions	Year	Population in millions	GDP in billions
1990	1.87	4.7	2005	2.04	6.26
1991	1.89	4.94	2006	2.04	6.86
1992	1.91	2.44	2007	2.04	8.34
1993	2.06	2.68	2008	2.05	9.91
1994	1.94	3.56	2009	2.05	9.49
1995	1.96	4.68	2010	2.05	9.41
1996	1.97	4.65	2011	2.06	10.49
1997	1.99	3.98	2012	2.06	9.75
1998	2.00	3.76	2013	2.06	10.82
1999	2.01	3.86	2014	2.07	11.36
2000	2.02	3.77	2015	2.07	10.06
2001	2.03	3.71	2016	2.07	10.67
2002	2.04	4.02	2017	2.07	11.31
2003	2.02	4.95	2018	2.08	12.63
2004	2.03	5.68	2019	2.08	12.69

B. Creating the model

Just like in [4], [5], [6] and [8] our model has demographical and socioeconomic nature in order to describe and analyze the trends of the acquired data, for the purpose of making future predictions. In this situation the prey will be the GDP, while the predator will be represented by the population.

From Table 1 it can be noted that there is a certain delay in the GDP trend regarding the population, which is characteristic for the Lotka-Volterra models. Further on, there are also some fluctuations in both of the GDP and population trends, which will be discussed and used for modeling later on.

As mentioned, we consider the system (1). Here, the GDP and the population are represented as functions $B(t)$ and $P(t)$, respectively. The GDP equation is consisted of the growth parameter a_1 , which is proportional to the GDP, while the death rate parameter a_2 represents the decrease in GDP, defined

proportionally to the interactions between the two competing species. On the other hand the population equation is similarly defined. Here the growth rate b_2 is proportional to the interactions, while the death rate parameter is proportional to the population and its growth. In eq. (1) the term $a_1B(t)$ gives the positive effect of the GDP on its growth, while the term $-a_2B(t)P(t)$ gives the negative effect that population has on the growth of the GDP.

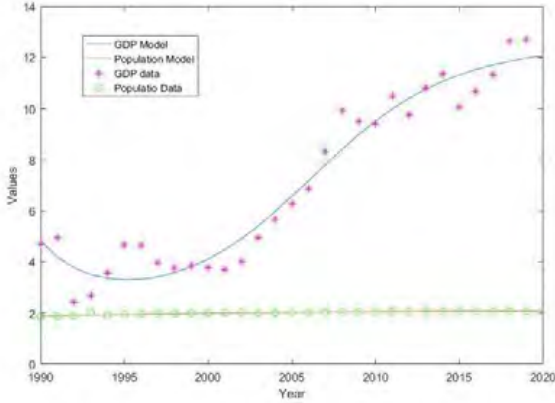


Figure 1 Predator-prey (GDP-Population)

Our analysis and computation of parameters follow the one in [2]. In order to find the initial values of the parameters, we use equation (3). By integrating for one period, or the period between two minima or maxima, we have

$$B_e = \frac{b_1}{b_2} = 3.73 \quad \text{and} \quad P_e = \frac{a_1}{a_2} = 2.02 \quad (4)$$

where the period is 11 years, assumed it represents one cycle. This gives a good approximation for the equilibrium point.

With the equilibrium point acquired the next step is to find the parameters. This will be done by approximating the data as Malthusian growth. The GDP average is calculated in the period from 1992 to 2002. In order to calculate the Malthusian growth the data from 1992 and 1993 will be used, with values 2.44 and 2.68, respectively,

$$B(t) = B_0 e^{a_1 t} \quad \text{so} \quad 2.68 = 2.44 e^{a_1}$$

giving the following estimate

$$a_1 = \ln\left(\frac{2.68}{2.44}\right) = 0.0938.$$

For the population average, the period from 1993 to 2002 was used,

$$P(t) = P_0 e^{-b_1 t}$$

resulting in $b_1 = \ln\left(\frac{2.06}{1.94}\right) = 0.06$.

Analogously, we obtain the other two parameters

$$a_2 = 0.0469, \quad \text{and} \quad b_2 = 0.0245.$$

Next, we use the obtained parameters as initial parameters for training the model, using the Least squares method (LSM) explained in [7], and by using already given function in Matlab. The LSM gives a function that best approximates the given set of points (Table 1), and it does not have to go through the given points.

The results are:

$$B(0) = 3.8186 \quad \text{and} \quad P(0) = 2.0408$$

$$a_1 = 6.2520, \quad a_2 = 3.0753, \quad b_1 = 0.0029 \quad \text{and} \quad b_2 = 0.0004$$

with sum of squared errors $J = 22.6804$. The performances and behavior of the model are shown on Figure 1 and Figure 2.

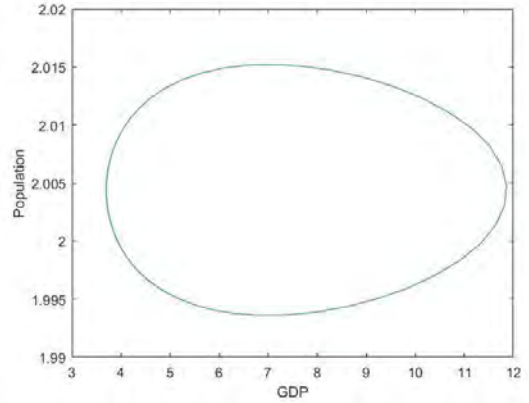


Figure 2 GDP-population phase portrait

C. Predator-prey model with evolution term

In this section, another approach will be used, i.e. an extended predator-prey model with an evolution term, [8]. This model tries to explain the macro-economic events in the society based on the Lotka-Volterra model, and has the following form:

$$\begin{cases} \frac{dB(t)}{dt} = B(t)(a_1 - a_2P(t) - a_3B(t)) \\ \frac{dP(t)}{dt} = P(t)(-b_1 + b_2B(t)) \end{cases}, \quad (5)$$

where $B(t)$ and $P(t)$ are the functions defined in the previous section representing the GDP and the population accordingly, and the parameters: a_1 – GDP growth rate parameter; a_2 – GDP decrease/death rate parameter; a_3 – GDP capacity parameter; b_1 – population death rate parameter, and b_2 – population growth rate parameter, with the following initial values:

$$a_1 = 0.01, \quad a_2 = 0.01, \quad a_3 = 0.0005, \quad b_1 = 0.01 \quad \text{and} \quad b_2 = 0.05$$

The model (5) can be represented also as:

$$\begin{cases} \frac{dB(t)}{dt} = a_1B(t) \left(1 - \frac{a_3B(t)}{a_1}\right) - a_2P(t) \\ \frac{dP(t)}{dt} = P(t)(-b_1 + b_2B(t)) \end{cases}, \quad (6)$$

This gives an opportunity to explicitly define the prey population capacity, in this case the GDP, represented by the a_1/a_3 relationship. If we compare this mode with model (5), it can be noticed that if there is a greater access to food/resources for the prey population (GDP), this will result in smaller value for a_3 with respect to a_1 . This can be also explained by the fact that the term $-a_3B^2(t)$ has a smaller impact on the outcome in reference to the linear term $a_1B(t)$. The scarcity of resources (low carrying capacity), on the other hand implies that the external environment exercises an action of higher intensity on

the system, that is, more intense environmental feedback translates into an increased 'nonlinearity' of the model, [8].

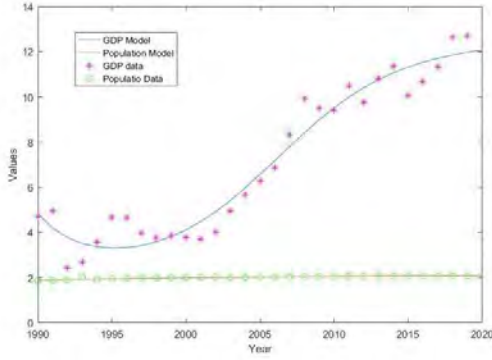


Figure 3 Predator-prey model with evolution term (GDP-population)

The purpose of the second order term $a_3 B^2(t)$ is to explain the evolution of the prey i.e. the GDP, as it can be seen on Figure 3, the result is a stable focus equilibrium point, which is asymptotically stable. The following parameters will be used as initial parameters for training the model

$$B(0) = 4.8310 \text{ and } P(0) = 1.8931$$

$$a_1 = -3.8061, a_2 = -2.0105, a_3 = 0.0339,$$

$$\text{and } b_1 = -0.0079, b_2 = -0.0006,$$

with the following sum of squared errors $J = 16.2390$. The graphical results are shown below.

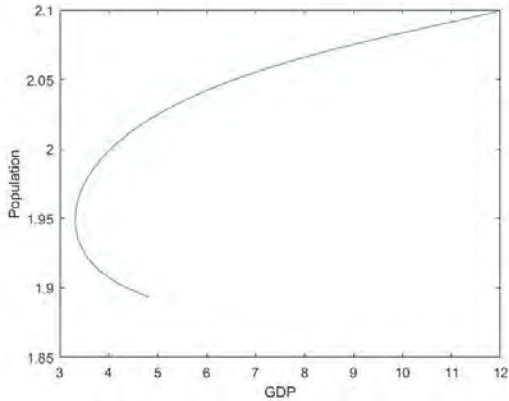


Figure 4 Phase portrait of predator-prey model with evolution term (GDP-population)

D. Model Comparison

With the models trained and the sum of squared errors calculated we can assume that the model with lower J value, has better performances. In this case this will be the model (5). However, further analysis is required based on their other characteristics and dynamical behavior. In this section a comparison will be made between the two trained models in the previous sections. Model 1 will be the model (1), and model 2 is given with model (5), for easier referencing.

We must mention that these two models have a different mathematical definition, and different behavior. With respect to

the GDP, from Figure 1 and Figure 3, we can conclude that there is a proper description of data and both of them capture it well, but on the long term predictions there is a visible difference between their predictions. The results are shown on Figures 5 and 6, below.

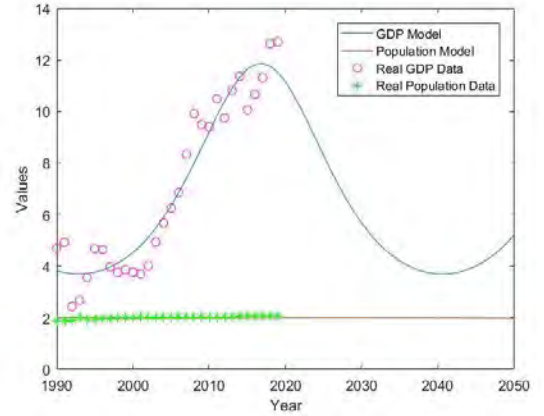


Figure 5 Predator-prey model 1 predictions

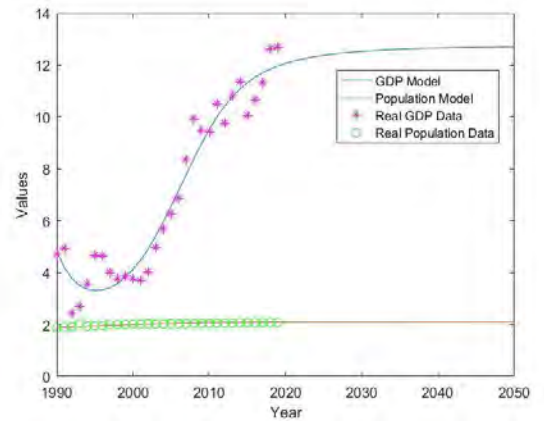


Figure 6 Predator-prey model 2 predictions

From the phase portraits on Figure 2 and Figure 4, it can be assumed that the model 1 has a limit cycle behavior, while model 2 has a stable focus behavior, i.e. the system is asymptotically stable around the equilibrium point.

From this perspective both of the models give a great fit to the real data of the predator population, and their predictions for the future states. Because there is a difference in the sum of squared errors for the cost function J between the two models, i.e., model 1 has a sum of squared errors $J = 22.6804$ and model 2 has a sum of $J = 16.2390$, their performances will be analyzed separately.

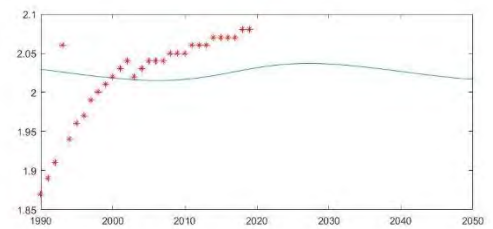


Figure 7 Model 2 population predictions

On Figure 7 it can be seen that the model 1 approximates and predicts the values of population as an oscillation around the average value of 2.02 millions for the real data, while on the Figure 8, the second model gives an almost great fit with the real data. It can be noted that in this case, model 2 has an asymptotically stable behavior, and model 1 has a limit cycle behavior, as analyzed in the previous sections.

From both of the presented models, model 2 gives better results, i.e. it gives better representations of the real values for the populations, unlike model 1. Because some of the parameter values for model 2, are negative, then we conclude that the two species are in direct competition with one another, since they each have a direct negative effect on the other's population.

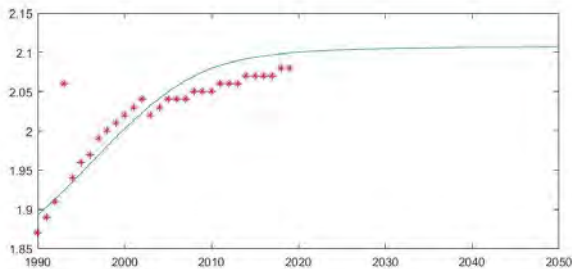


Figure 8 Model 2 populations predictions

IV. CONCLUSION

The purpose of this paper was to examine the dynamical systems, more precisely the nonlinear dynamical systems, and whether they are useful for providing sound assumptions about the future states of some real systems and processes. Although these systems are associated with the chaos and unpredictability they proved to give promising results. At first, the biological and ecological nature of this model was explained, and later it was approached as macroeconomic and demographic systems, with the intention of making a model for the connection between the population and the GDP in R.N. Macedonia.

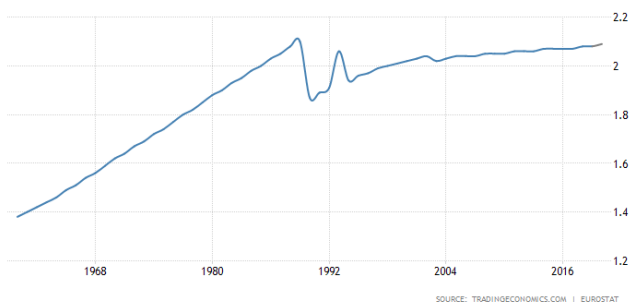


Figure 9 Population of Macedonia 1960-2019 [3]

In this paper, two models were taken for examining such applications of nonlinear systems for modeling population dynamics. The first approach was using the simplest form of the Lotka-Volterra model, whose parameters were fitted to the acquired data using the least square errors method, with the results showing that the system has limit cycle behavior. The second approach aimed to test a model with an implemented quadratic term, and then fitted the parameters to the same

acquired data. Both of the models were analyzed and compared in the previous section, and both of them have satisfactory results.

The results of the model (5) may give an error while predicting the population, but this error can be neglected due to the fact that Macedonian population has long term trend of values, around tens of thousands above 2 million people, and it is most likely it will continue in that manner in the future. This trend comes a consequence from the fact that for almost 20 years R.N. Macedonia has not been able to conduct a state census. If there is a more real data for the population, assumedly there can be more fluctuations in the data, which will do even a better job in the modeling process.

Figure 9 is shown with a purpose to show the existence of larger declines and increases in the population, i.e. existence of a period of recurrence. It can be noticed that for a period of 30 years since the beginning of the 1960's until the end of the 80's, Macedonian population reaches 2.1 million people, the same thing happens again with a drop in the population in the early 90's and reaching again almost 2.1 million population in 2019. This gives another perspective to the whole situation, but this data is not helpful, due to the different economic systems before and after 1991 and the lack of data for the GDP before 1991. All of this indicates that for creating a better macroeconomic predator-prey model, access to data that covers longer period of time can provide a better mathematical model and more realistic representation of the dynamic system.

V. REFERENCES

- [1] Д. М. Георги, Предавања по нелинеарно автоматско управување, Скопје, 1975.
- [2] <https://jmahaffy.sdsu.edu/courses/f09/math636/lectures/lotka/qualde2.html#fitparameter>.
- [3] www.tradingeconomics.com
- [4] M. Oğuz Arslan, and H. Altınok, "A System Dynamics Model of Income Distribution between Labor and Capital for Turkey," Economic Computation and Economic Cybernetics Studies and Research, vol. 52(4), 2018, pp. 241-256.
- [5] N. J. Moura, and M. Byrro Ribeiro, "Testing the Goodwin Growth-cycle Macroeconomic Dynamics in Brazil," Physica A: Statistical Mechanics and its Applications, vol. 392 (9), 2013, pp. 2088-2103.
- [6] S. E. Puliafito, J. L. Puliafito, and M. Conte Grand, 2006. "Modeling population dynamics and economic growth as competing species: An application to CO2 global emissions, " CEMA Working Papers: Serie Documentos de Trabajo. 334, Universidad del CEMA.
- [7] К. Хади-Велкова Санева, Е. Хадиева, С. Геговска-Зайкова, Б. Начевска-Настовска, Нумерички методи, УКИМ, Скопје, 2019.
- [8] C. S. Bertuglia, and F. Vaio, Nonlinearity, Chaos, and Complexity The Dynamics of Natural and Social Systems, Oxford University Press, USA, 2005.

Performance of Gradient Algorithms for Solving Least Squares Problem

Naum Dimitrieski, Katerina Hadzi-Velkova Saneva, and Zoran Hadzi-Velkov

Faculty of Electrical Engineering and Information Technologies

Ss. Cyril and Methodius University, Skopje, Macedonia

E-mail: nonodimitrieski@yahoo.com, saneva@feit.ukim.edu.mk, zoranhv@feit.ukim.edu.mk

Abstract—Stochastic gradient descent is the most important optimization method in machine learning. In this paper, we compare the performances of three stochastic gradient methods to tackle a simple least squares problem: the conventional Stochastic gradient descent (SGD), the Adaptive moment estimation stochastic gradient descent (Adam), and Nesterov-accelerated adaptive moment estimation stochastic gradient descent (Nadam). The input-output data set for the least squares problem is generated by an uncorrelated Gaussian random process with an adjustable variance. We use three performance metrics for benchmarking: the mean squared error between the estimated and the true optimal solutions, and the number of iterations and execution time needed to reach the optimal solution with a certain accuracy. The numerical results show that the conventional SGD outperforms Adam and Nadam. However, Adam and Nadam are much less sensitive to the algorithms' learning rate, which is a crucial advantage that justifies their dominance in most machine learning applications.

Keywords—Least Squares Problem; Optimization; Gradient Descent Algorithms, Machine Learning.

I. INTRODUCTION

Gradient descent is a classic first order iterative method for determining a local minimum of a differentiable function, which is generally attributed to the famous mathematician Cauchy [1]. The method uses repeated steps in opposite direction of the gradient of the function at the current point, which is the direction of the steepest descent [2]. However, the modern engineering design is based on numerical optimization of functions with very high number of variables (typically in thousands). The classic gradient descent cannot tackle such optimization problems (such as those in machine learning), because it requires evaluations of very large sums of gradients that is prohibitive from computational point of view. So, the gradient descent is modified into a stochastic gradient descent (SGD), where the actual gradient is replaced by its estimate. In machine learning, for example, the gradient used in SGD is estimated from a randomly selected subset of the data [3]–[5].

The least-squares problem is an example optimization problem [2], which is the basis for regression analysis and many parameter estimation and data fitting methods, and has many names, e.g., regression analysis or least-squares approximation. Solutions to many classes of optimization problems also reduce to least square problems. So, efficient algorithms for solving various least squares problem are very important for modern numerical analysis. The least squares problem is also found in many engineering disciplines, such as, signal detection in telecommunications, automatic control

systems, machine learning, communications and networks, electronic circuit design.

In this paper we compare the performance of three stochastic gradient methods for solving the classic least squares problem: (1) *conventional SGD*, (2) *Adaptive Moment Estimation Stochastic Gradient Descent* (Adam), and (3) *Nesterov Accelerated Adaptive Moment Estimation Stochastic Gradient Descent* (Nadam), [3]–[7]. For comparing the performance of methods, we use the following metrics: the mean square error (MSE) of the optimal solution, the number of iterations and the algorithm execution time needed to reach the optimal solution with a certain accuracy.

II. OPTIMIZATION PROBLEM

Mathematical optimization and machine learning are heavily based on solving least-squares problem. The least-squares problem (LSP) will be considered in the context of linear regression, where it is aimed at determining the *coefficients* of the hyper plane with dimension N ,

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N)^T = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix}, \quad (1)$$

which fits best the input-output dataset (\mathbf{A}, \mathbf{b}) . The input dataset \mathbf{A} consists of K observations (rows of matrix \mathbf{A}) with N elements ($K \geq N$), $\mathbf{a}_i = (1, a_{i2}, a_{i3}, \dots, a_{iN})^T, i = 1, 2, \dots, K$, i.e.,

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_K^T \end{bmatrix} = \begin{bmatrix} 1 & a_{12} & a_{13} & \dots & a_{1N} \\ 1 & a_{22} & a_{23} & \dots & a_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{K2} & a_{K3} & \dots & a_{KN} \end{bmatrix}. \quad (2)$$

Note, the first element of each observation is unity ($a_{i1} = 1, \forall i$) in order to make sure the hyper plane is offset from the origin with θ_1 as the intercept term. The output dataset \mathbf{b} is given by

$$\mathbf{b} = (b_1, b_2, \dots, b_K)^T = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}. \quad (3)$$

The considered least-squares optimization problem is defined as

$$\text{minimize } J(\boldsymbol{\theta}) = \|\mathbf{A}\boldsymbol{\theta} - \mathbf{b}\|^2 = \sum_{i=1}^K (\mathbf{a}_i^T \boldsymbol{\theta} - b_i)^2, \quad (4)$$

where $\varepsilon_i = \mathbf{a}_i^T \boldsymbol{\theta} - b_i$ is the error of the i th observation, which forms the error vector

$$\boldsymbol{\varepsilon} = \mathbf{A}\boldsymbol{\theta} - \mathbf{b} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_K)^T = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_K \end{bmatrix}. \quad (5)$$

Note, $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{\sum_{i=1}^K x_i^2}$ is the Euclidian norm of the vector $\mathbf{x} = (x_1, x_2, \dots, x_K)^T$.

The solution of (4) determines $\boldsymbol{\theta}$ such that the error is minimized. The analytic solution of (4) is well known and given by $\boldsymbol{\theta}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$, [2].

III. OPTIMIZATION ALGORITHMS

Gradient Descent (GD) is one of the most popular algorithms to perform optimization and by far the most common way to optimize neural networks. Its iterative procedure is given by the following relation between future and current values of the objective variable $\boldsymbol{\theta}$,

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \cdot \nabla J(\boldsymbol{\theta}^{(i)}), \quad (6)$$

where $\boldsymbol{\theta}^{(i)}$ is the i th iteration of the optimization variable, η is the step size (also referred to as the *learning rate*), and $\nabla J(\boldsymbol{\theta}^{(i)})$ is the gradient of the objective function at the i th iteration $\boldsymbol{\theta}^{(i)}$ [2], [5].

For the least-squares problem (4), the gradient of the objective function is given by

$$\nabla J(\boldsymbol{\theta}) = 2\mathbf{A}^T(\mathbf{A}\boldsymbol{\theta} - \mathbf{b}) = \sum_{i=1}^K \nabla J_i(\boldsymbol{\theta}), \quad (7)$$

where $\nabla J_i(\boldsymbol{\theta})$ is the gradient of the i th term of the objective function sum, $(\mathbf{a}_i^T \boldsymbol{\theta} - b_i)^2$, i.e.,

$$\nabla J_i(\boldsymbol{\theta}) = 2 \mathbf{a}_i^T (\mathbf{a}_i^T \boldsymbol{\theta} - b_i). \quad (8)$$

According to (7), the gradient of the objective function, $J(\boldsymbol{\theta})$, is the sum of the gradients at each observation. Therefore, the algorithm (6) used in conjunction with (7) is also called *Batch Gradient Descent*, as it uses the whole batch of training data at every step [4], [5]. This makes GD very slow when the dataset is large. GD also does not allow us to update our model online, i.e. with new examples on-the-fly.

A. Stochastic Gradient Descent Algorithm (SGD)

At the opposite extreme of GD, the *Stochastic Gradient Descent algorithm (SGD)* at each iteration uses only single observation from the dataset and computes the gradient using a single observation (i.e., a single row of matrix \mathbf{A} and vector

\mathbf{b}) [3]-[5]. In the case of the least-squares problem, this gradient is calculated according to (8). Using (8) instead of (7) significantly accelerates the gradient algorithm because it has very little data to manipulate at every iteration. It also makes it possible to train on huge training sets, since only one observation should be kept in the computer memory at each iteration. The iterative procedure of SGD is given by

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \cdot \nabla J_i(\boldsymbol{\theta}^{(i)}). \quad (9)$$

On the other hand, due to its stochastic nature, this algorithm is much less regular than Batch Gradient Descent. Instead of gradually decreasing until it reaches the minimum, the objective function will bounce up and down, decreasing only on average. Over time it will end up very close to the minimum, but once it gets there it will continue to bounce around, never settling down. So once the algorithm stops, the final iteration is probably close to the optimal value [4], [5].

B. Adaptive Movement Estimation SGD Algorithm (Adam)

The *Adaptive Moment Estimation SGD (Adam)* is a more complex algorithm which adjusts weights for each individual iteration [6]. Similarly to SGD, Adam requires only the value of the gradient at the i th observation, $\nabla J_i(\boldsymbol{\theta}^{(i)})$, but necessitates the usage of auxiliary vectors. The auxiliary vectors \mathbf{m} and \mathbf{v} are adjusted with respect to the new value of the gradient and the previous one. They are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradient respectively, hence the name of the method. The i th iteration of the vectors \mathbf{m} and \mathbf{v} are given by

$$\mathbf{m}^{(i)} = \beta_1 \mathbf{m}^{(i-1)} + (1 - \beta_1) \cdot \nabla J_i(\boldsymbol{\theta}^{(i)}), \quad (10)$$

$$\mathbf{v}^{(i)} = \beta_2 \mathbf{v}^{(i-1)} + (1 - \beta_2) \cdot [\nabla J_i(\boldsymbol{\theta}^{(i)})]^2, \quad (11)$$

where β_1 and β_2 are constants whose values do not change throughout the optimization process. The authors of [6] propose default values of 0.9 and 0.999 for β_1 and β_2 , respectively.

Since \mathbf{m} and \mathbf{v} are initialized at 0, they will be biased toward 0 during the first several iterations, so the additional auxiliary vectors $\hat{\mathbf{m}}$ and $\hat{\mathbf{v}}$ will help boost \mathbf{m} and \mathbf{v} at the beginning of training,

$$\hat{\mathbf{m}}^{(i)} = \frac{\mathbf{m}^{(i)}}{1 - (\beta_1)^i}, \quad (12)$$

$$\hat{\mathbf{v}}^{(i)} = \frac{\mathbf{v}^{(i)}}{1 - (\beta_2)^i}. \quad (13)$$

Based on (10) - (13), the Adam iteration algorithm is given by

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \cdot \frac{\hat{\mathbf{m}}^{(i)}}{\sqrt{\hat{\mathbf{v}}^{(i)} + \varphi}}, \quad (14)$$

where φ is a constant whose numerical value is in interval $[10^{-6}, 10^{-8}]$ and prevents division by 0. Since Adam is an adaptive learning rate algorithm, its convergence behavior is more robust with respect to the selection of the learning rate η , [5].

C. Nesterov-Accelerated Adaptive Movement Estimation SGD Algorithm (Nadam)

The *Nesterov-Accelerated Adaptive Moment Estimation SGD (Nadam)* is a modification of the Adam algorithm [7]. It is derived from Adam, by implementing the Nesterov Accelerated Gradient Principle in the last step of the algorithm. Based on (10) - (13), the Nadam iteration algorithm is given by

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \cdot \frac{1}{\sqrt{\hat{\mathbf{v}}^{(i)} + \varphi}} \cdot \left(\beta_1 \hat{\mathbf{m}}^{(i)} + \frac{1-\beta_1}{1-(\beta_1)^i} \nabla J_i(\boldsymbol{\theta}^{(i)}) \right) \quad (15)$$

Note, compared to Adam, the Nadam algorithm is a bit slower as it used more complex rendering operations, but is slightly more resilient to oscillations, [4], [5], [7]. The recommended values for β_1 and β_2 correspond to those of Adam.

IV. PERFORMANCE BENCHMARKING

The performance comparison of the three considered algorithms is realized by using a method similar to that in [8]. Computer simulations in MATLAB are used to obtain $M = 50$ input-output datasets (\mathbf{A}, \mathbf{b}) . Similar to [9], the optimal solution of the least-squares optimization problem (4) is preset by the following vector with dimension $N = 9$:

$$\boldsymbol{\theta}^* = [0, 3, 1.5, 0, 0, 2, 0, 0, 0]^T. \quad (16)$$

In the m th simulation run ($1 \leq m \leq 50$), the input dataset $\mathbf{A}(m)$ defined by (2) and error terms $\boldsymbol{\varepsilon}(m)$ defined by (5), are randomly generated such that their elements follow the normal distribution, i.e.

$$a_{ij} \sim \mathcal{N}(0, 1), \quad 1 \leq i \leq K, \quad 2 \leq j \leq 9, \quad (17)$$

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2), \quad 1 \leq i \leq K, \quad (18)$$

where σ^2 is the error variance. As presented in the following sections, the error variance is the main input parameter to our analysis. In our simulations, the number of observations is set to $K = 10^4$. The output dataset in the m th simulation run, $\mathbf{b}(m)$ is calculated as

$$\mathbf{b}(m) = \mathbf{A}(m)\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}(m). \quad (19)$$

Based on the input-output dataset $(\mathbf{A}(m), \mathbf{b}(m))$, the corresponding gradient method is applied to obtain the optimal solution $\boldsymbol{\theta}(m)$. After $M = 50$ simulation runs, we determine the *mean square error* (MSE) between the estimated solution after the final (K th) iteration, $\boldsymbol{\theta}^{(K)}$, and the true optimal solution, $\boldsymbol{\theta}^*$, given by (16), as follows:

$$MSE = \frac{1}{50} \sum_{m=1}^{50} \|\boldsymbol{\theta}^{(K)}(m) - \boldsymbol{\theta}^*\|^2. \quad (20)$$

We also calculate the number of iterations, n^* , needed to achieve a predefined accuracy, δ_0 , of the estimated solution relative to the true solution $\boldsymbol{\theta}^*$:

$$\delta = \frac{\|\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta}^*\|}{\|\boldsymbol{\theta}^*\|} \leq \delta_0. \quad (21)$$

The value of n^* determined from (21) is called the *number of algorithm steps* (NAS). We also estimate the needed time to execute n iterations from (21), which is denoted as the *algorithm execution time* (AET). Again, $M = 50$ input-output datasets (\mathbf{A}, \mathbf{b}) are used, identically as in the first metric, with randomly generated elements whose probability distribution function is identical to (17) and (18). The optimal solution is the same as (16) and $\mathbf{b}(m)$ is calculated exactly as (19). For the purpose of the experiment δ_0 is set to 5%.

Next, we present the results from comparing three stochastic gradient methods: SGD, Adam and Nadam. We set the following parameters for the Adam and Nadam algorithms, $\beta_1 = 0.9$ and $\beta_2 = 0.999$, as they are the recommended values [6].

A. Algorithm Convergence

The convergence of the three algorithms is illustrated on Figs. 1 and 2, which respectively depict the evolution of the optimization variables θ_2 and θ_3 with each iteration towards their optimal values $\theta_2^* = 3$ and $\theta_3^* = 1.5$. To plot the figures, we assume a learning rate $\eta = 0.005$ for SGD, and $\eta = 0.01$ for Adam and Nadam. The error variance of the dataset is set to $\sigma^2 = 0.1$.

For the selected learning rates, the three algorithms need much less than 10000 observations to arrive at the optimal solution, i.e., SGD takes about 400 iteration steps, whereas Adam and Nadam take about 1200 iteration steps. We also note that SGD converges faster than Adam and Nadam.

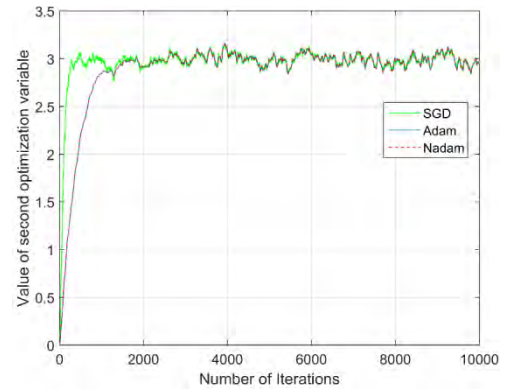


Fig. 1. Convergence of optimization variable θ_2 to the optimal solution (16)

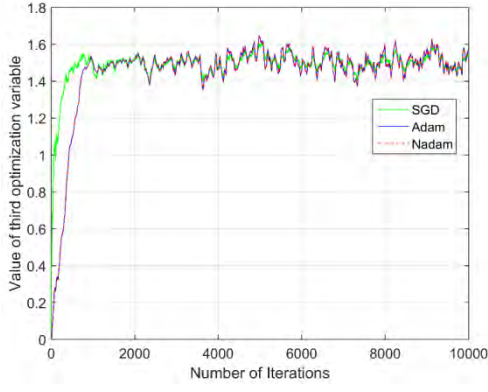


Fig. 2. Convergence of optimization variable θ_3 to the optimal solution (16)

B. The impact of the learning rate

Next, we study the dependence of MSE in function of the learning rate η . The error variance σ^2 is set to $\sigma^2 = 0.1$. Fig. 3 shows that the MSE metric first decreases to some minimum, and then increases in η . Clearly, for each algorithm, there is an *optimal learning rate* that minimizes the MSE: $\eta^* = 0.0003$ for SGD, and $\eta^* = 0.001$ for Adam and Nadam.

As η increases near to 0.1, SGD diverges but Adam and Nadam still converge to the optimal solution. Unlike SGD, Adam and Nadam converge steadily even for very high learning rates because of the normalization operations employed at each iteration step, according to (12) and (13).

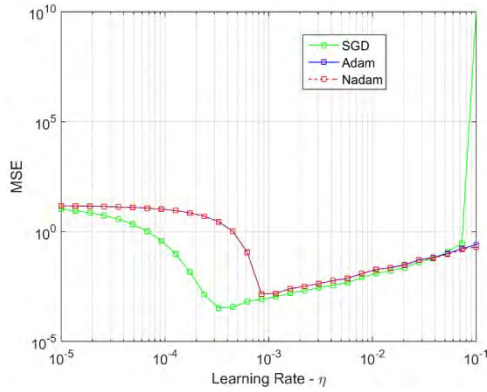


Fig. 3 MSE in function of the learning rate

C. The impact of noise in the dataset

In this section, we study the impact of the noise variance σ^2 on the MSE of the three algorithms. Figs. 4, 5 and 6 depict the MSE vs. σ^2 for SGD, Adam and Nadam, respectively, with the learning rate η appearing as a parameter on those figures. Note, MSE and σ^2 are plotted on logarithmic scales, which allows to observe the algorithm behavior over a wide range of noise variances.

Generally, there is a linear dependence of MSE in σ^2 . When η is very small, MSE is independent of σ^2 . The independence of MSE from σ^2 is desirable, but comes at a cost of increased execution time of the algorithms. Fig. 4 shows that MSE of SGD algorithm is minimal for $\eta = 0.005$, while $\eta = 0.01$ yields the best results for Adam and Nadam (Fig. 5 and 6, respectively).

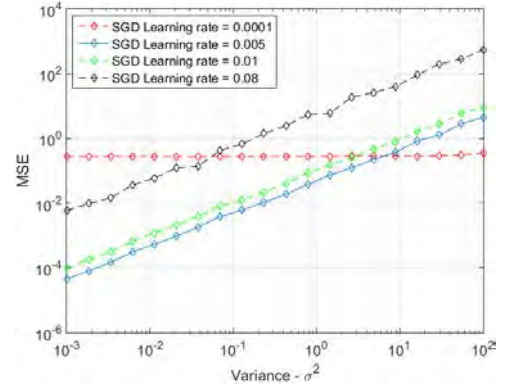


Fig. 4. MSE performance of SGD algorithm for different learning rates

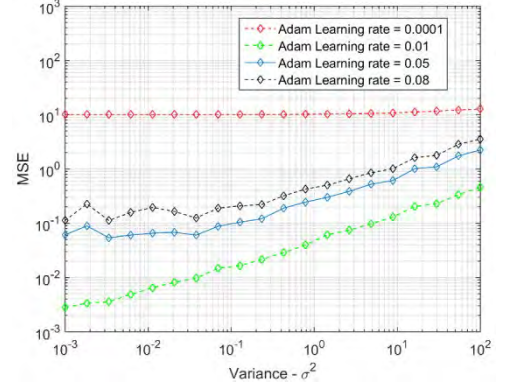


Fig. 5. MSE performance of Adam for different learning rates

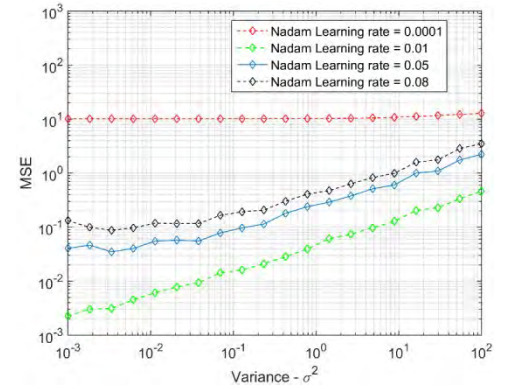


Fig. 6. MSE performance of Nadam for different learning rates

Figs. 7 and 8 compare the MSE performances of the three algorithms in function of the error variance σ^2 . In both figures, the learning rate of SGD is set to $\eta = 0.005$. The learning rates of Adam and Nadam are set to $\eta = 0.01$ (Fig. 6), and $\eta = 0.05$ (Fig. 7).

We conclude that Adam and Nadam perform similarly for a given learning rate, but differently than SGD. Depending on the noise error variance, their MSEs can be lower or higher than the MSE of the SGD. When the noise error variance is below some threshold, the MSE of SGD is lower than the MSEs of Adam and Nadam. When the noise error variance exceeds this threshold, MSE of SGD is higher than the MSEs of Adam and Nadam.

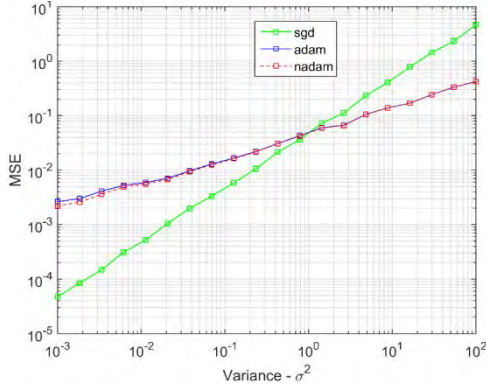


Fig. 7. MSE performance for optimal learning rates

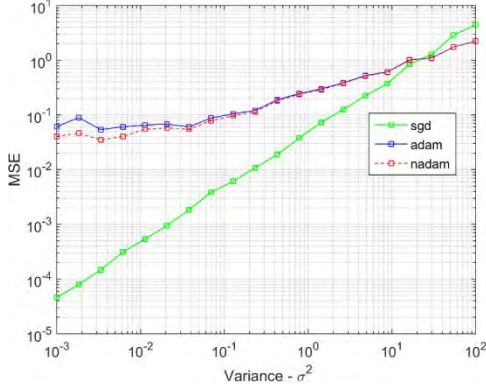


Fig. 8. MSE performance for sub-optimal learning rates

D. The needed number of algorithm steps

Next, we consider NAS, n^* vs. the noise variance σ^2 , for a threshold accuracy of $\delta_0 = 5\%$, c.f. (21). The obtained results are given at Fig. 9. It shows the required number of iterations in function of the error variance. As in the first metric, we use the following values for the learning rate η : for Adam and Nadam, $\eta = 0.01$, and $\eta = 0.005$ for SGD. It can be seen that greater error variances require more iterations. Indeed, for smaller variance values, SGD yields better results, however as the variance increases its results deteriorate to a value of variance at which both Adam and Nadam start to yield better results. In consequence, we conclude that Adam and Nadam algorithms converge faster for greater error variance, whilst SGD for smaller variances.

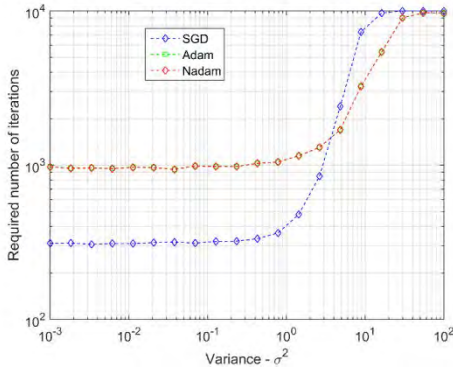


Fig. 9. The needed number of algorithm steps for threshold accuracy $\delta_0 = 5\%$ in function of error variance

On the other hand, Fig. 10 shows the required number of iterations for the algorithms to reach the 5% threshold in function of the learning rate, for a set error variance value of 0.1. Note, for extremely small values of η , all three algorithms require the maximally allowed number of iterations $K = 10^4$, indicating extremely poor convergence. Furthermore, the SGD algorithm performs best for $\eta \in [0.005, 0.04]$, while for higher values of η it diverges. Adam and Nadam show their best results for $\eta > 0.03$. It is empirically proven that for such an experiment, even when the learning rate is 0.25, both Adam and Nadam still show good convergence with limited oscillations around the optimal solution [4]-[7].

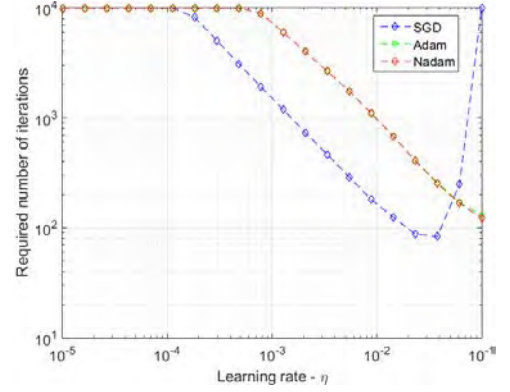


Fig. 10. The needed number of algorithm steps for threshold accuracy $\delta_0 = 5\%$ in function of learning rate

E. Algorithm Execution Time

In Fig. 11, we compare the algorithm execution times for the three algorithms. Again, the algorithm stops execution when it converges to the optimal values under the criterion $\delta_0 = 5\%$. Note, Nadam requires slightly more time in order to converge to the given threshold than Adam, as a result of the greater number of complex instructions used in the algorithm. As a result of its simplicity, SGD requires significantly less time to execute successfully than Adam or Nadam. Although the number of necessary iterations needed by SGD to converge to the target threshold for large variance values is greater than that of both Adam and Nadam (Fig. 9), SGD still manages to execute faster than either of them.

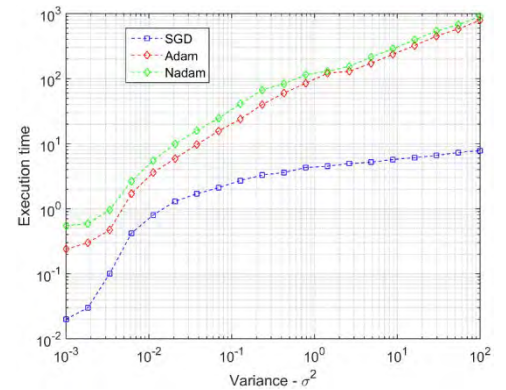


Fig. 11. Execution time in function of error variance

V. CONCLUSION

In this paper, we have compared the performances of three famous gradient descent algorithms for solving a simple least squares optimization problem. The conventional SGD algorithm is superior to both Adam and Nadam algorithms when the variance of Gaussian noise in the input-output dataset is below some variance threshold. Beyond this threshold, the convergence rate of SGD gets lower or sometimes even diverges away from the optimal solution. Although slower, both Adam and Nadam prove to be robust to the choice of the learning rate, which is very important for machine learning.

It is also worth noting that there is very little difference in performance between Adam and Nadam, which is expected, since the latter is a modification of the former. The usage of Nadam increases the execution times by as much as 25%, while its MSE performance is very close to that of Adam. Compared to Nadam, Adam is used much more frequently in practice.

Overall, the choice between Adam and conventional SGD in practice depends mostly on the dataset. The conventional SGD is preferable when the dataset values are predominantly close to zero. Examples include the linear regression and support vector regression, because the scaling of the dataset affects the accuracy of the model. On the other hand, the random forest algorithm, for example, does not require scaling,

and so, the usage of Adam is preferable over SGD as it guarantees convergence and the learning rate need not be adjusted.

REFERENCES

- [1] A. L. Cauchy, "Méthode générale pour la résolution des systèmes d'équations simultanées," *C. R. Acad. Sci. Paris*, 25, pp. 536–538, 1847.
- [2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [3] L. Bottou, "Stochastic gradient descent tricks," *Neural networks: Tricks of the trade*, pp. 421–436, Springer, 2012.
- [4] S. Ruder, "An overview of gradient descent optimization algorithms," pp. 1 – 13, 2017, arXiv:1609.04747.
- [5] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow, Concepts, Tools, and Techniques to Build Intelligent*, O'Reilly Media, 2019.
- [6] D. P. Kingma and J. Lei Ba, "Adam: a method for stochastic optimization," *International Conference on Learning Representations*, pp. 1–13, 2015.
- [7] T. Dozat, "Incorporating Nesterov momentum into Adam," *ICLR Workshop*, (1): 2013–2016, 2016.
- [8] M. Dimovski and I. Stojkovska, "Regularized least-square optimization method for variable selection in regression models," *Mat. Bilten*, 40 (LXVI), no. 4, pp. 80 – 100, 2016.
- [9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1996.



ETAI 3: CONTROL SYSTEMS AND AUTOMATION

Multi-Objective Optimization Based Fractional Order PID Controller Design

Erhan Yumuk, Res. Assist
Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
yumuk@itu.edu.tr

Müjde Güzelkaya, Prof. Dr.
Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
guzelkaya@itu.edu.tr

Eda Budak, System Analyst
CicekSepeti.com / Lolaflora.com
Eindhoven, Netherlands

İbrahim Eksin, Prof. Dr.
Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
eksin@itu.edu.tr

Abstract—In this study, multi-objective optimization based integer and fractional order PID controllers are designed for first order plus dead time system models using two different cases. For the first case, the objective function is selected as the integral square error (ISE) and for the second case it is chosen as the integral time square error (ITSE). In both cases, multi-objective optimization is carried out for both set-point following and load disturbance rejection responses together. All optimization problems are solved by genetic algorithm to obtain the coefficients of the integer and fractional order PID controllers. After a set of Pareto optimal solutions is obtained, the Nash bargaining point is specified in order to make comparisons. The simulation results show that fractional order PID is superior to integer order PID when the dead time is small. However, fractional and integer order PID controllers show similar performances as the amount of dead time increases.

Keywords— *fractional order PID controllers; multi-objective optimization; Nash solution; first order plus dead time models*

I. INTRODUCTION

The notion of fractional derivative and integral, which is accepted as a generalization of derivative and integral operations of integer numbers, is first developed by Leibniz and Newton in 17th century [1]. At the beginning, fractional calculus was a field only studied by a few mathematicians and theoretical physicists, but today it has managed to become popular in various application fields due to the rapid developments in computer technology and studies [2]. It can be suggested that a significant part of these applications is of modeling complex dynamic systems and designing fractional order controllers.

As it is known, dynamic systems are modeled under certain assumptions. Usually, the elements of the system are considered to be ideal, and differential equations involving integer order integrals and derivatives are used system modeling. However, it is also known that integer order differential equations are insufficient in modeling many dynamic systems. For this reason, fractional order differential

equations have been used for modeling dynamical systems in various studies [3-5].

In terms of controller design, integer order PID (IOPID) controllers provide only the coefficients of the controller to be selected freely. On the other hand, when a fractional order PID controller (FOPID) is used, the integral and derivative orders can also be selected besides the coefficients of the controller. This shows that FOPID controllers provide more flexibility in the design process, increase the performance of the system and are less sensitive to parameter changes in the system [6-10]. Aside from these advantages, the complexity of obtaining the responses of fractional order systems in the time domain causes problems when analyzing and designing controllers. Due to this problem, most of the proposed design methods in literature are developed either in the frequency domain [6-8] or by using a numerical search algorithm [9-10].

Many tuning rules for the coefficients of IOPID controller is devised based only one requirement such as fast set point changes or fast load disturbance response. However, it is important to take into account conflicting requirements simultaneously such as load disturbance rejection and set-point following performances or settling time and control effort. Since an intermediate tuning between the conflicting requirements is required, Multi-objective Optimization (MO) tools are employed to find the most advantageous PID controller. A tuning rule for IOPID controllers for first integer order plus dead time (FOPDT) model is proposed in [11] by creating a Multi-objective Optimization (MO) problem while considering a trade-off between set-point following and load disturbance rejection responses while constraining the maximum sensitivity value. After the generation of the Pareto front, which is optimal solution set for a MO problem, the Nash bargaining solution is selected in order to determine the coefficients of IOPID controllers. The values of the objective functions, which are integral square error (ISE) values of set-point following and load disturbance rejection responses are calculated and the system responses are compared.

In this study, multi-objective optimization based IOPID and FOPID controllers are designed for the first order plus dead time system models using two different cases. For the first case, the objective function is selected as the integral square error (ISE) and for the second case it is chosen as the integral time square error (ITSE). In both cases set-point following and the load disturbance rejection responses are considered together. The multi-objective optimization problems are solved by using genetic algorithm to obtain the coefficients of the IOPID and FOPID controllers. After a set of Pareto optimal solutions is obtained, the Nash bargaining point is specified in order to make comparisons. The simulation results show that FOPID controllers have satisfactory performance compared to integer order counterparts

The organization of this study is as follows: In Section II, a brief information about fractional calculus and multi-objective optimization is given. Section III, the simulation studies for design of IOPID and FOPID controllers are carried out on various first order plus time delay system models. Lastly, in Section IV, the results of these simulations are analyzed and some suggestions for future studies are made.

II. THEORETICAL BACKGROUND

A. Fractional Calculus

The operator d^n/dt^n represents derivation for positive values of n and integration for negative values of n . Fractional order calculus investigates the behavior of this operator when n is not an integer number. The integro-differential operator ${}_rD_t^\alpha$ is used in fractional calculus and its expression is shown as follows:

$${}_rD_t^\alpha = \begin{cases} d^\alpha/dt^\alpha & \text{Re}(\alpha) > 0 \\ 1 & \text{Re}(\alpha) = 0 \\ \int_r^t (d\tau)^\alpha & \text{Re}(\alpha) < 0 \end{cases} \quad (1)$$

where α is a fractional number. For zero initial condition, the Laplace transform of the fractional operator is given as:

$$\mathcal{L}\{ {}_0D_t^\alpha g(t) \} = s^\alpha G(s) \quad (2)$$

Some numerical approximation techniques can be used in order to simulate the fractional operator, since there is no exact implementation. Among these approximation techniques, the Oustaloup approximation is the most known one [12]. Oustaloup approximation can be given as:

$$s^\alpha \approx G_f(s) = K \prod_{k=1}^N \frac{s + \omega_{z,k}}{s + \omega_{p,k}}, \quad 0 < \alpha < 1 \quad (3)$$

where

$$K = \omega_h^\alpha, \quad \omega_{z,k} = \omega_l \left(\frac{\omega_h}{\omega_l} \right)^{\frac{2k-1-\alpha}{2N}}, \quad \omega_{p,k} = \omega_l \left(\frac{\omega_h}{\omega_l} \right)^{\frac{2k-1+\alpha}{2N}}$$

Here, ω_h and ω_l represent the upper and lower boundaries of the frequency, respectively and N is the number of poles/zeros in the approximation. As the value of N increases, which means that as the order of the approximation increases, the obtained integer order transfer function represents the fractional operator better. Ideally, the closest representation of the fractional operator would be obtained if the order of the approximation were infinitely large.

B. Multi Objective Optimization

A multi-objective optimization (MO) problem can be defined as

$$\min F(x) = [F_1(x) \quad F_2(x) \quad \dots \quad F_k(x)]^T \quad (4)$$

subject to

$$\begin{aligned} g_j(x) &\leq 0, j = 1, 2, \dots, m \\ h_l(x) &= 0, l = 1, 2, \dots, e \\ x_L &\leq x_i \leq x_U, i = 1, 2, \dots, n \end{aligned}$$

where k , m , and e represent the numbers of objective functions, inequality constraints and equality constraints, respectively. $x = [x_1 \quad x_2 \quad \dots \quad x_n]^T$ is the vector of n decision variables and x_L and x_U denote lower and upper bounds for each of the decision variables x_i , respectively.

Unlike single objective optimization problem, the outcome of MO problem in (4) is an optimal solution set rather than just a single optimal solution point. Two types of optimal solution points are used to find the optimal solution set: (i) dominated and (ii) non-dominated ones, as it can be seen in the Figure 1. A point is called dominated if there is at least one other solution point that generates better results regarding both of the objective functions and a solution point is referred as non-dominated if there is no other solution point that gives better results for both of the objective functions. The non-dominated solution points create a curve called the Pareto front. A Pareto front example for two conflicting objective functions is shown in Fig 2. None of the solutions on the Pareto front is better than the others however they differ in the degree of performance between the objective functions. Moving along this curve causes improvement in one objective while worsening another objective. In other words, it is impossible to improve a Pareto optimal solution with respect to all of the objective functions.

Even though all of the solution points at the Pareto front are adequate solutions and do not dominate each other, a decision has to be made in order to complete the multi-objective optimization decision process. One procedure to select a fair point is to use bargaining games. For this respect, one of the commonly used method is the Nash solution [13]. The Nash bargaining solution is subtracted from the worst possible solutions for each of the objective functions. Then, the product of these differences is maximized in order to shift the points away from the worst cases possible.

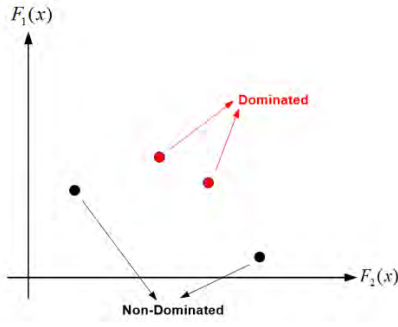


Fig. 1. Dominated and non-dominated solutions in Pareto set.

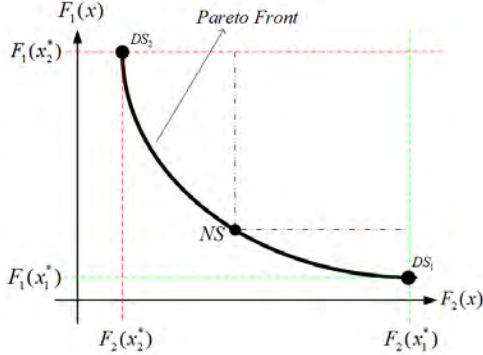


Fig. 2. A pareto front example for two objective functions.

The Nash solution can be found by solving the following maximization problem.

$$\max (F_1(x_2^*) - F_1^{NS})(F_2(x_1^*) - F_2^{NS}) \quad (5)$$

subject to

$$\begin{aligned} F_1^{NS} &\leq F_1(x_2^*) \\ F_2^{NS} &\leq F_2(x_1^*) \end{aligned}$$

where F_1^{NS} and F_2^{NS} denote the cost function values at Nash point x^{NS} . DS_1 is the point which minimizes $F_1(x)$ at x_1^* , then corresponding $F_2(x_1^*)$ value is found. On the other hand, DS_2 is the point which minimizes $F_2(x)$ at x_2^* then corresponding $F_1(x_2^*)$ value is found. These points shown in Fig. 2 are called dictatorial solutions for each of objective functions.

III. FRACTIONAL AND INTEGER ORDER PID DESIGN

In this study, a unity feedback control system whose block diagram is shown in Fig. 3 is considered

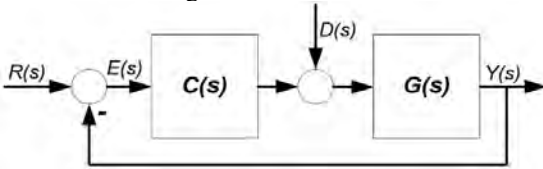


Fig. 3. Unity feedback control system block diagram.

In the control system, $R(s)$, $D(s)$, $E(s)$ and $Y(s)$ denote the Laplace transforms of the reference $r(t)$, disturbance (t) , error

$e(t)$ and output $y(t)$ signals, respectively. Moreover, the output signal

$$y(t) = y_r(t) + y_d(t) \quad (6)$$

where $y_r(t)$ and $y_d(t)$ are set-point following response and load disturbance response, respectively. A trade-off between performances of these two responses exists. In other words, a good set-point following performance results in a bad disturbance rejection and vice versa.

As system model type, FOPDT system model will be examined for different values of L and T :

$$G(s) = \frac{K}{Ts+1} e^{-Ls} \quad (7)$$

We accept that $K = 1$. Four different value sets for $\{T, L\}$ are chosen as $\{1, 0.25\}, \{1, 0.5\}, \{1, 0.75\}, \{2, 0.75\}$. As controller type, the following IOPID and FOPID controllers are used :

$$C_{IOPID}(s) = K_p + \frac{K_i}{s} + K_d s \quad (8)$$

$$C_{FOPID}(s) = K_p + \frac{K_i}{s^\lambda} + K_d s^\mu \quad (9)$$

As performance measures, ISE and ITSE are selected. The resultant multi objective optimization problem is solved by using genetic algorithm. As a result, a set of Pareto optimal solutions is obtained and among these solutions the Nash bargaining solution is selected in order to make comparisons.

A. ISE Based FOPID Controller Design

The problem formulation can be shown as

$$\min J(\theta) = [J_{sp}(\theta) \quad J_{ld}(\theta)] \quad (10)$$

where

$$J_{sp}(\theta) = ISE_{sp} = \int_0^\infty e_r(t) dt, \quad d(t) = 0 \quad (11)$$

$$J_{ld}(\theta) = ISE_{ld} = \int_0^\infty e_d(t) dt, \quad r(t) = 0 \quad (12)$$

Here, $J_{sp}(\theta)$ and $J_{ld}(\theta)$ denote integral square error for set point step response and disturbance rejection, respectively. Moreover, θ denotes $\{K_p, K_i, K_d\}$ for IOPID design, $\{K_p, K_i, K_d, \lambda, \mu\}$ for FOPID design.

As a result of this optimization, the Pareto fronts for four different FOPDT processes are obtained using IOPID and FOPID controllers. These Pareto fronts are shown in Fig. 4. The points on the upper right corner of the boxes in Fig.4 represent the Nash bargaining solutions that are obtained by using IOPID controllers. The set of solutions that are located inside these boxes are generated by using FOPID controllers and are better solutions compared to the Nash bargaining solutions generated by IOPID controllers. However, for fair comparison, the values related to the Nash bargaining solutions for both IOPID and

FOPID controllers are given in Table I. As it can be seen from the table, FOPID controllers' Nash bargaining solutions manage to generate lower values for the ISE for both characteristics with lower dead times. As the dead time increases, the results obtained from both IOPID and FOPID controllers become non-dominant among each other. In this case, it can be said that both solutions are equally acceptable since it cannot be clearly stated that one of them is better than the other one.

The system response of the first system ($L = 1, T = 0.25$) for the Nash bargaining solutions regarding set point following and disturbance rejection is given in Fig. 5. From the graph and Table I, it can be seen that the FOPID controller decreases the rise time value while increasing the settling time and the overshoot values. This is not an unusual outcome due to the fact that improving the time domain responses of these systems were not expected from the FOPID. The aim of this study is to observe whether or not the FOPID controller decreases the ISE

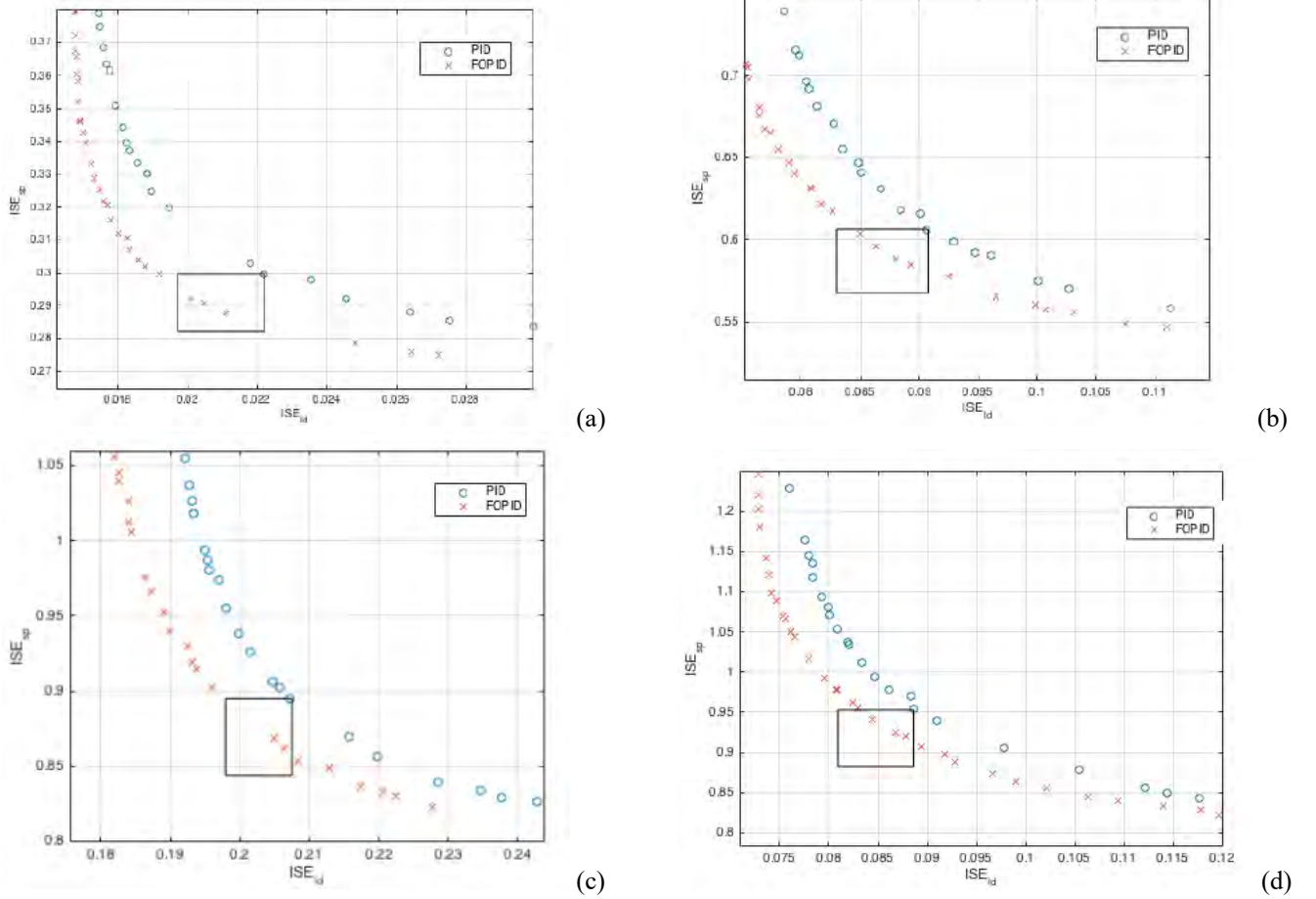
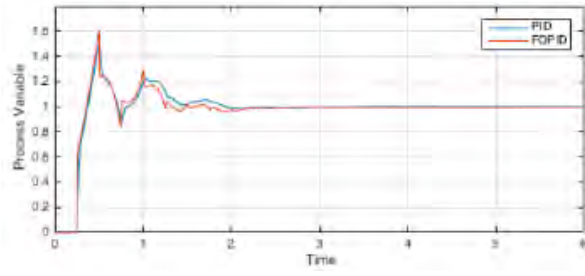


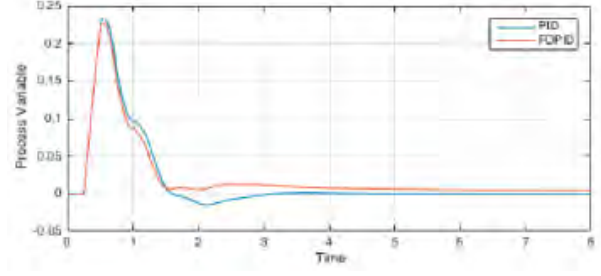
Fig. 4. Pareto front of control system using IOPID and FOPID for (a) $T=1, L=0.25$ (b) $T=1, L=0.5$ (c) $T=1, L=0.75$ and (d) $T=2, L=0.75$ (ISE case)

TABLE I. CONTROLLER PARAMETERS AND THEIR PERFORMANCE VALUES

System	Cont.	ISE_{sp}	ISE_{ld}	K_p	K_i	K_d	λ	μ	t_r	t_s	OS(%)
T=1 L=0.25	IOPID	0.299	0.022	3.834	7.490	0.568	1	1	0.093	1.900	50.252
	FOPID	0.292	0.020	3.621	7.913	0.627	0.859	0.996	0.080	2.126	60.626
T=1 L=0.5	IOPID	0.605	0.095	2.354	3.270	0.631	1	1	0.143	4.253	64.051
	FOPID	0.588	0.088	1.934	3.738	0.699	0.868	0.995	0.132	4.511	68.228
T=1 L=0.75	IOPID	0.895	0.207	1.742	1.961	0.664	1	1	0.168	6.251	65.412
	FOPID	0.902	0.195	1.379	2.479	0.775	0.833	0.990	0.150	9.773	76.840
T=2 L=0.75	IOPID	0.954	0.088	3.185	2.653	1.267	1	1	0.197	6.458	74.398
	FOPID	0.907	0.089	2.397	3.257	1.387	0.797	0.998	0.187	6.113	75.068



(a)



(b)

Fig. 5. (a) The set point following and (b) load disturbance responses for the system ($L=1$, $T=0.25$) using IOPID and FOPID (ISE case)

values for both set-point following and load disturbance rejection responses, which FOPID managed to accomplish successfully.

B. ITSE Based FOPID Controlller Design

The problem formulation can be shown as

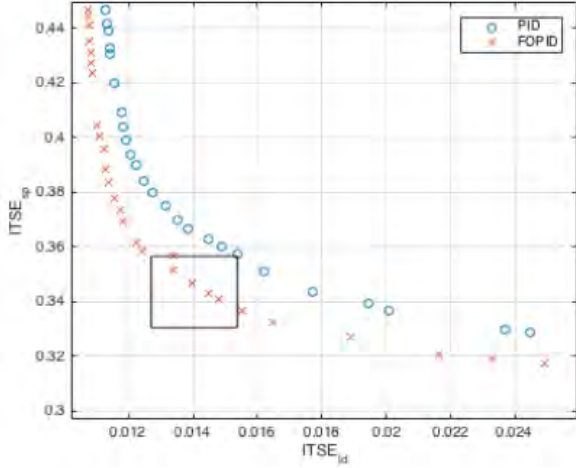
$$\min J(\theta) = [J_{sp}(\theta) \quad J_{ld}(\theta)] \quad (12)$$

where

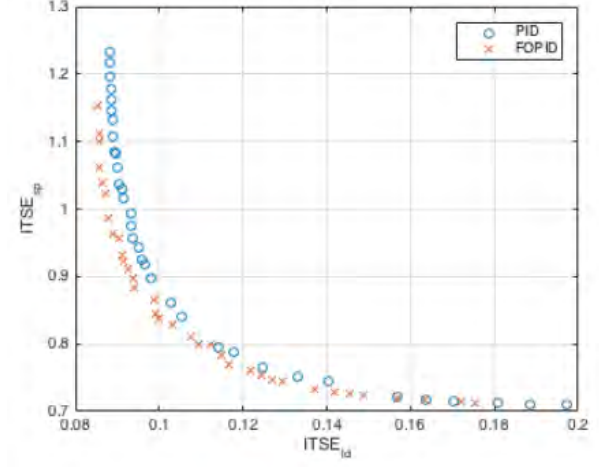
$$J_{sp}(\theta) = ITSE_{sp} = \int_0^\infty te_r(t)dt, \quad d(t) = 0 \quad (13)$$

$$J_{ld}(\theta) = ITSE_{ld} = \int_0^\infty te_d(t)dt, \quad r(t) = 0 \quad (14)$$

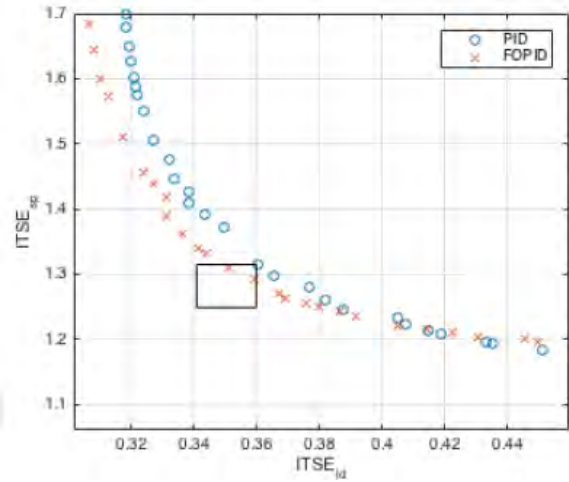
Similar to ISE case, $J_{sp}(\theta)$ and $J_{ld}(\theta)$ denote integral time square error for set point step response and disturbance rejection, respectively. Moreover, θ denotes $\{K_p, K_i, K_d\}$ for IOPID design, $\{K_p, K_i, K_d, \lambda, \mu\}$ for FOPID design.



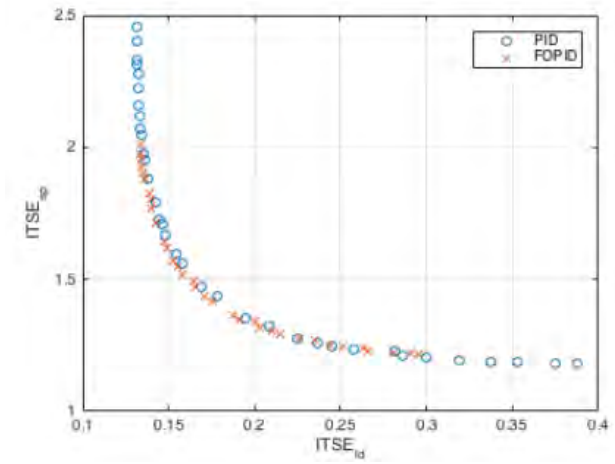
(a)



(b)



(c)

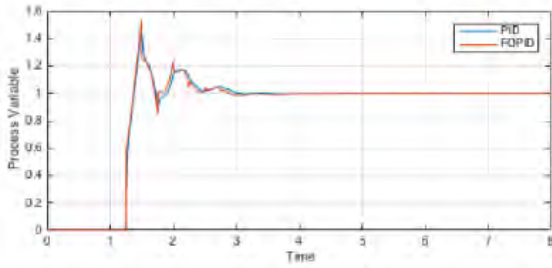


(d)

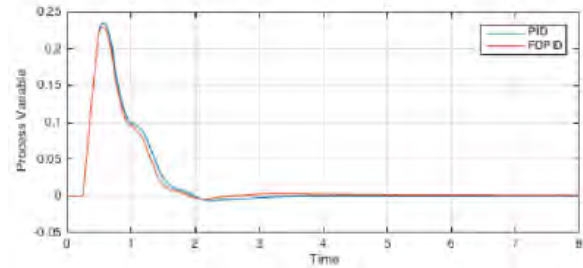
Fig. 6. Pareto front of control system using IOPID and FOPID for (a) $T=1$, $L=0.25$ (b) $T=1$, $L=0.5$ (c) $T=1$, $L=0.75$ and (d) $T=2$, $L=0.75$ (ITSE case)

TABLE II. CONTROLLER PARAMETERS AND THEIR PERFORMANCE VALUES

System	Cont.	$ITSE_{sp}$	$ITSE_{ld}$	K_p	K_i	K_d	λ	μ	t_r	t_s	OS(%)
T=1 L=0.25	IOPID	0.351	0.016	3.939	6.650	0.528	1	1	0.106	2.920	44.859
	FOPID	0.342	0.014	3.958	7.243	0.571	0.967	0.998	0.091	2.813	54.446
T=1 L=0.5	IOPID	0.794	0.114	2.286	2.945	0.568	1	1	0.177	4.463	55.863
	FOPID	0.837	0.100	2.150	3.765	0.657	0.930	0.989	0.144	5.103	69.778
T=1 L=0.75	IOPID	1.315	0.360	1.706	1.750	0.592	1	1	0.258	6.179	55.437
	FOPID	1.340	0.341	1.368	2.180	0.739	0.868	0.976	0.194	7.013	65.718
T=2 L=0.75	IOPID	1.474	0.170	3.165	2.325	1.147	1	1	0.241	6.527	64.909
	FOPID	1.438	0.171	2.381	3.017	1.381	0.839	0.955	0.217	7.242	69.890



(a)



(b)

Fig. 7. (a) The set point following and (b) load disturbance responses for the system ($L=1$, $T=0.25$) using IOPID and FOPID (ITSE case)

The Pareto fronts obtained after optimization procedure are shown in Fig. 6. Similar to the first case, the points on the upper right corner of the boxes in Fig 6 represent the Nash bargaining solutions that are obtained by using IOPID controllers. The set of solutions that are located inside these boxes are generated by using FOPID controllers and are better solutions compared to the Nash bargaining solutions generated by IOPID controllers. As it can be seen from the Pareto fronts of various cases, it is safe to say that the optimal solutions sets that are obtained by using FOPID controllers are better than the ones obtained by using PID controllers. However, for fair comparison, the values related to the Nash bargaining solutions for both PID and FOPID controllers are given in the Table II. .

The system response of the first system ($L = 1, T = 0.25$) for the Nash bargaining solutions regarding set point following and disturbance rejection is given in Fig. 7. From the graph and the Table II, it can be seen that the FOPID controller decreases the rise time value while increasing the settling time and the overshoot values. This is not an unusual outcome due to the fact that improving the time domain responses of these systems were not expected from the FOPID.

IV. CONCLUSION

In this study, multi-objective optimization based integral and FOPID controllers is designed using four different first order plus dead time system model. In order to accomplish this goal, a multi-objective optimization problem is defined that addressed a trade-off between the set-point following and load disturbance performance. The objective functions of the MO problem is selected as ISE and ITSE values for both set point following and the disturbance rejection tasks, which conflict

with each other. Then, genetic algorithm is used in order to obtain the optimal solutions set for each situation. After the Pareto fronts for both integer order PID controllers and FOPID controllers is obtained, it is observed that the FOPID controllers' results are better than the PID ones. FOPID controller manages to generate lower ISE and ITSE values for both operation modes. This situation is an expected result due to the fact that FOPID controllers have two more parameters compared to the integer order PID controllers

For fair comparison, the coefficients of both the controllers are specified using Nash solutions on the obtained Pareto fronts. When comparing the Nash solutions, FOPID controllers give better results in systems with small dead time values than IOPID controllers. As the amount of dead time increases, it is observed that FOPID and IOPID controllers are not superior to each other. In this case, the performance values of both controllers are regarded as equally acceptable. When the system responses are examined, it is seen that the FOPID controllers speed up the systems by reducing their rise times while increasing their settling times and overshoot values.

REFERENCES

- [1] M. Dalir, M. Bashour, "Application of fractional calculus," Applied Mathematical Sciences, 2010, vol 4, no 21, pp. 1021-1032.
- [2] C.A: Monje, Y.Q. Chen, B.M. Vinagre, D. Xue, V. Feliu, Fractional order Systems and Controls: Fundamentals and Applications. Springer London, 2010.
- [3] E. Yumuk, M. Güzelkaya, İ. Eksin, "Design of an integer order PI/PID controller based on model parameters of a certain class of fractional order systems," Proc of the Ins of Mech Eng, Part I: J of Sys and Con Eng, 2019, vol. 233, no 3, pp. 320-334

- [4] E. Yumuk, M. Güzelkaya, İ. Eksin, "Optimal fractional-order controller design using direct synthesis method," *IET Cont Theory&App*, 2020, vol. 14, no. 18, pp. 2960-2967.
- [5] R. Azarmi, M. Tavakoli-Kakhi, A. K. Sedigh, A. Fatehi, "Analytical design of fractional order PID controllers based on the fractional set-point weighted structure: Case study in twin rotor helicopter," *Mechatronics*, 2015, vol 31, pp. 222-233.
- [6] M. Bettayeb, R. Mansouri, "Fractional IMC-PID-filter controllers design for non integer order systems," *Journal of Process Control*, 2014, vol. 24, no 4, pp. 261-271.
- [7] E. Yumuk, M. Güzelkaya, İ. Eksin, "Analytical fractional PID controller design based on Bode's ideal transfer function plus time delay," *ISA transactions*, 2019, vol. 91, pp. 196-206.
- [8] E. Yumuk, M. Güzelkaya, İ. Eksin, "Application of fractional order PI controllers on a magnetic levitation system," *Turk J Elec Eng & Comp Sci*, 2021, vol. 29, pp. 98-109
- [9] C.A. Monje, B.M. Vinagre, V. Feliu, Y.Q. Chen, "Tuning and auto-tuning of fractional order controllers for industry applications," *Cont Eng Prac*, 2008, vol. 16, no 7, pp. 798-812.
- [10] Y. Luo, Y.Q. Chen, C.Y. Wang, Y.G. Pi, "Tuning fractional order proportional integral controllers for fractional order systems," *J Process Cont*, 2010, vol.20, pp. 823-831.
- [11] H.S. Sanchez, A. Visioli, R. Vilanova, "Optimal Nash tuning rules for robust PID controllers," *Journal of the Franklin Institute*, 2017, vol. 354, no 10, pp. 3945-3970.
- [12] D. Valerio, J.S. da Costa, "Introduction to single input, single output fractional control," *IET Cont Theory&App*, 2011, vol. 5, no 8, pp. 1033-1057.
- [13] J.F. Nash, "The bargaining problem," *Econometrica*, 1950 vol 18, no 2, pp155-162.

Fractional Integrating Integer Order PI Controller Design for the First Integer Order Plus Time Delay System

Erhan Yumuk, Res. Assist
Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
yumuk@itu.edu.tr

Müjde Güzelkaya, Prof. Dr.
Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
guzelkaya@itu.edu.tr

İbrahim Eksin, Prof. Dr.
Dept. of Control and Automation Engineering
Istanbul Technical University (ITU)
Istanbul, Turkey
eksin@itu.edu.tr

Abstract—In this study, a robust fractional integrating integer order PI controller design is proposed for the first integer order plus time delay systems. Phase and gain margins are selected as the robustness criteria. Moreover, the delayed Bode's ideal transfer function is employed as a reference model to design an analytical controller. The phase and gain margin specifications are exactly determined by virtue of this transfer function. In simulation studies, the proposed fractional integrating integer order PI controller is compared with integer order PI controllers using various phase and gain margin values. The simulation results show that the proposed controller is superior to integer order PI controller in both frequency and time domain criteria aspects.

Keywords— *fractional integrating PI controllers; phase and gain margins; delayed Bode's ideal transfer function; first order plus time delay system models*

I. INTRODUCTION

Fractional calculus is a mathematical tool that has been employed in control and automation engineering area for approximately half a century although it emerged three centuries ago. Three more configurations have been appeared using fractional calculus in the area: (i) Fractional order controller design for integer order model [1, 2], (ii) Integer order controller design for fractional order model [3] and (iii) Fractional order controller design for fractional order model [4, 5]. This study takes place in Configuration (i) since a fractional integrating integer order PI controller is designed for the first order plus time delay model.

The first order plus time delay (FOPTD) model is one of the most commonly used model in control engineering area. This model provides a satisfactory approximation to represent the dynamics of the real-time systems with S-shaped step response encountered in the areas such as chemical, thermal, mechanics

etc [6]. Moreover, it is sufficient in describing higher order systems and even nonlinear behavior of a real-time system. Various identification methods have been introduced to find the model parameters; namely gain K , time constant T and time delay L . [7-9] The most prominent one is based on graphical method using step response. In this identification method, the model gain K is specified via the ratio of process input and output steady-state values. The time delay L is found using the intercept of tangent with the largest slope on system response and horizontal time axis. Finally, the time constant T is equal to difference between L and the time when step response reaches 0.63 times of its final value.

It is well known that most controller design methods are model based and the performance of control system can be affected negatively in case of model mismatch. To overcome this problem, it is crucial to design a robust controller. Two commonly used robustness specifications are phase and gain margins (GPM). Various researches regarding both integer order (IO) and fractional order (FO) controller designs have been carried out to satisfy these specifications [10-12]. While designing IOPI controller using GPM specifications, the problem is transformed into solving four nonlinear equations with four unknowns (phase and gain cross-over frequencies, ω_g , ω_p and the coefficients of IOPI controller, K_c , τ_i) [10]. For integer order PID controller design, the problem is to solve four nonlinear equations with five unknowns (phase and gain cross-over frequencies, ω_g , ω_p and the coefficients of IOPID controller, K_c , τ_i , τ_d). One additional parameters is determined using extra specification, e.g. the minimization of integral square error [10]. On the other hand, four nonlinear equations with five and seven unknowns must be solved to design FOPI and FOPID controllers using the same specifications, respectively. It is a challenging task to solve these nonlinear equations. Internal Model Control (IMC) strategy can be

employed to obtain an analytical solution [13, 14]. However, GPM specifications cannot exactly be satisfied using this strategy since it has only one tuning parameter.

In this study, a robust fractional integrating integer order PI controller design is proposed for the first integer order plus time delay systems. Phase and gain margins are selected as the robustness criteria. Moreover, the delayed Bode's ideal transfer function [15, 16] is employed as a reference model to design an analytical controller. The delayed Bode's ideal transfer function provides the exact determination of the phase and gain margin specifications. Simulations are performed to compare the proposed controller with integer order PI controllers using various phase and gain margin values. The simulation results show that the proposed controller is superior to integer order PI controller in the aspects of both frequency and time domain criteria.

The organization of this study is as follows: In Section II, a brief information about fractional calculus and delayed Bode's ideal transfer function is given. Section III provides a robust fractional integer order PI controller design. In Section IV, the simulation studies are carried out. Lastly, study concludes with Section V.

II. THEORETICAL BACKGROUND

A. Fractional Calculus

The fractional operator ${}_r D_t^\alpha$ is employed in fractional calculus, the expression of which is described as follows:

$${}_r D_t^\alpha = \begin{cases} d^\alpha/dt^\alpha & \text{Re}(\alpha) > 0 \\ 1 & \text{Re}(\alpha) = 0 \\ \int_r^t (d\tau)^\alpha & \text{Re}(\alpha) < 0 \end{cases} \quad (1)$$

where α is a non-integer number. This operator defines fractional derivative and fractional integration for $\text{Re}(\alpha) > 0$ and $\text{Re}(\alpha) < 0$. The Laplace transformation of the fractional operator using zero initial condition is found as follows:

$$\mathcal{L}\{ {}_0 D_t^\alpha g(t) \} = s^\alpha G(s) \quad (2)$$

Various approximations are used in either time or frequency domain to simulate the fractional operator. The most popular approximations in time and frequency domains are Grünwald-Letnikov and Oustaloup, respectively [17, 18]. Grünwald-Letnikov approximation is given as

$${}_r D_t^\alpha g(t) \approx \frac{1}{h^\alpha} \sum_{i=0}^{\lfloor \frac{t-r}{h} \rfloor} (-1)^i \binom{\alpha}{i} g(t - ih) \quad (3)$$

where h is the step size, $\lfloor \frac{t-r}{h} \rfloor$ denotes the integer part of the number and $\binom{\alpha}{i}$ is binomial coefficients. On the other hand, Oustaloup approximation can be given as:

$$s^\alpha \approx G_f(s) = K \prod_{k=1}^N \frac{s + \omega_{z,k}}{s + \omega_{p,k}}, \quad 0 < \alpha < 1 \quad (4)$$

where

$$K = \omega_h^\alpha, \quad \omega_{z,k} = \omega_l \left(\frac{\omega_h}{\omega_l} \right)^{\frac{2k-1-\alpha}{2N}}, \quad \omega_{p,k} = \omega_l \left(\frac{\omega_h}{\omega_l} \right)^{\frac{2k-1+\alpha}{2N}}$$

Here, the upper and lower boundaries of the frequency are denoted by ω_h and ω_l , respectively and N is the approximation order. Grünwald-Letnikov approximation in Eq. (3) would represent the fractional operator exactly if the step size were taken as infinitely small. On the other hand, Oustaloup approximation in Eq. (4) would describe the fractional operator precisely if its order were taken as infinitely large.

B. Delayed Bode's Ideal Transfer Function

Bode [19] proposed the following open loop transfer function:

$$L(s) = \frac{K_b}{s^\gamma} \quad (5a)$$

where $K_b, \gamma \in \mathbb{R}$ denote system gain and fractional order of Bode's ideal transfer function, respectively. Bode's ideal transfer function for $\gamma > 0$ and $\gamma < 0$ describes fractional integrator and fractional derivative, respectively. The slope of the amplitude curve of the transfer function is constant value $-20\gamma \text{ dB/dec}$ on log-log scale. Similarly, its phase curve has a constant value $-\pi\gamma/2 \text{ rad}$ for all frequencies. On the other hand, the delayed Bode's ideal transfer function [15] is given as

$$L(s) = \frac{K_b}{s^\gamma} e^{-\theta s}, \quad 1 < \gamma < 2 \quad (5b)$$

The only difference between Bode's and delayed Bode's ideal transfer functions is the delay term of θ . The slope of the amplitude curve of the transfer function is the same as in Bode's ideal transfer function. Its phase curve has a value of $-\pi\gamma/2 - \omega\theta \text{ rad}$. Unlike the Bode's ideal transfer function case, the phase curve depends on frequency ω .

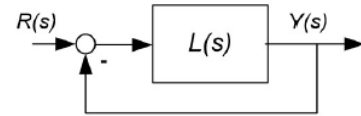


Fig. 1. Delayed Bode's ideal transfer function block diagram.

Fig. 1 illustrates the unity feedback reference system where delayed Bode's ideal loop transfer function is replaced in the forward path. Here, $R(s)$ and $Y(s)$ denote the Laplace transformations of the reference $r(t)$ and output $y(t)$ signals, respectively. In the next section, this configuration is used as a reference system so as to design a controller based on frequency domain specifications, i.e. phase and gain margins.

III. FRACTIONAL INTEGRATING INTEGER ORDER PI CONTROLLER DESIGN

In this study, a unity feedback control system whose block diagram is shown in Fig. 2 is considered.

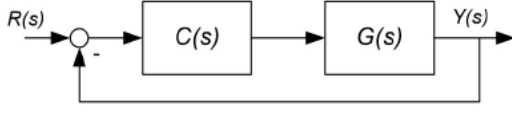


Fig. 2. Unity feedback control system block diagram.

The following FOPDT model is used as the system model $G(s)$ in Fig. 2:

$$G(s) = \frac{K}{Ts+1} e^{-Ls} \quad (6)$$

where T, K and L are time constant, gain and time delay of the system model, respectively. The specifications to be satisfied by the controller $C(s)$ in Fig 2 are as follows:

- Phase margin (ϕ_m)

$$\angle C(j\omega_g)G(j\omega_g) = \phi_m - \pi \quad (7a)$$

- Gain margin (A_m):

$$|C(j\omega_p)G(j\omega_p)| = \frac{1}{A_m} \quad (7b)$$

where ω_g and ω_p are the gain and phase crossover frequencies. In this design method, we use delayed Bode's ideal transfer function as the reference transfer function given in Fig 1. Moreover, its delay term θ is accepted to be equal to the delay term L of the system model. Then, the following fractional order controller could be designed as

$$C(s) = \frac{L(s)}{G(s)} = \frac{K_b(Ts+1)}{Ks^\gamma} \quad (8)$$

This controller can be rewritten as

$$C(s) = \underbrace{\frac{1}{s^{\gamma-1}}}_{\text{fractional integrator}} \underbrace{\frac{K_b T}{K} \left(1 + \frac{1}{Ts}\right)}_{\text{integer order PI}} \quad (9)$$

and can be named as Fractional Integrating Integer Order PI controllers (FI_IOPI).

In order to calculate the parameters, γ and K_b , the following four non-linear equations must be solved:

$$|L(j\omega_g)| = 1 \Rightarrow K_b = \omega_g^\gamma \quad (10a)$$

$$\angle L(j\omega_g) = \phi_m - \pi \Rightarrow \frac{\gamma\pi}{2} = -\pi + \phi_m + \omega_g\theta \quad (10b)$$

$$\angle L(j\omega_p) = -\pi \Rightarrow \frac{\gamma\pi}{2} = -\pi + \omega_p\theta \quad (10c)$$

$$|L(j\omega_p)| = \frac{1}{A_m} \Rightarrow K_b = \frac{\omega_p^\gamma}{A_m} \quad (10d)$$

Here, the first two nonlinear equations, (10a) and (10b) are used to satisfy the desired phase margin and the last two, (10c) and

(10d) are employed to meet the desired gain margin. The fixed point iteration method [20] is used to solve these four nonlinear equations.

IV. SIMULATION STUDIES

In this section, the following FOPTD model is considered, which is given in [11]:

$$G(s) = \frac{1}{s+1} e^{-0.1s} \quad (11)$$

In [11], four different gain and phase margin (GPM) specifications are selected as $\{(3, 45^\circ), (5, 45^\circ), (3, 60^\circ), (5, 60^\circ)\}$. Using each of the specifications, IOPI controllers are designed as in the following form:

$$C(s) = K_c \left(1 + \frac{1}{T_i s}\right) \quad (12)$$

The coefficients of IOPI controllers are given in Table I. On the other hand, the coefficients of the delayed Bode's ideal transfer function in (5) can be found by solving the four nonlinear equations in (10a)-(10d) for each of the given specifications. After that, FI_IOPI controller is designed using Eq. (9) as follows:

$$C(s) = \frac{1}{s^{\gamma-1}} K_d \left(1 + \frac{1}{s}\right) \quad (13)$$

The coefficients of FI_IOPI controller are also found in Table I.

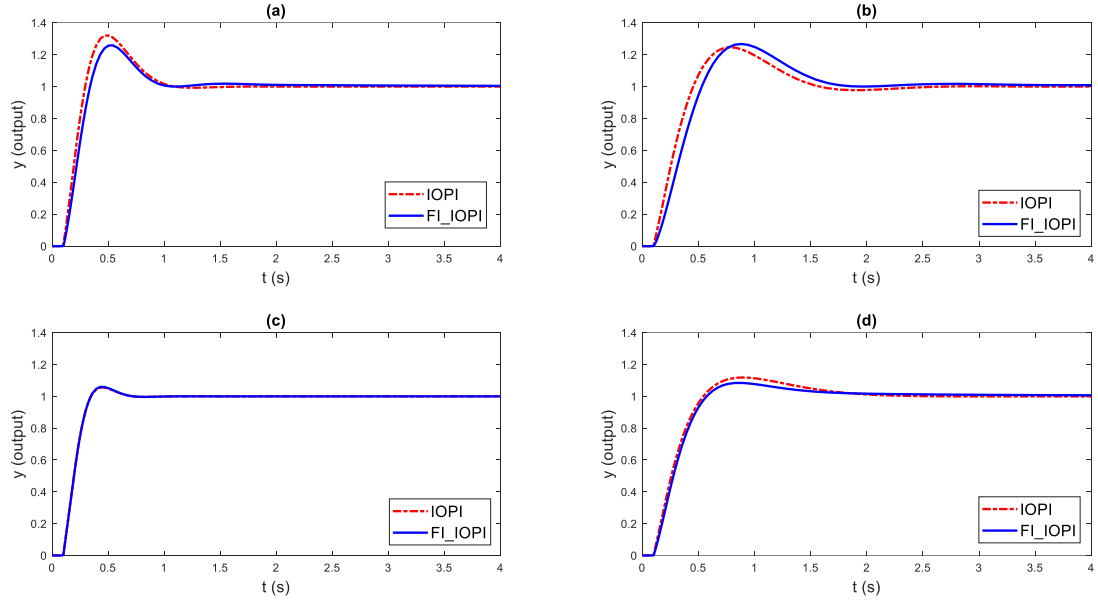
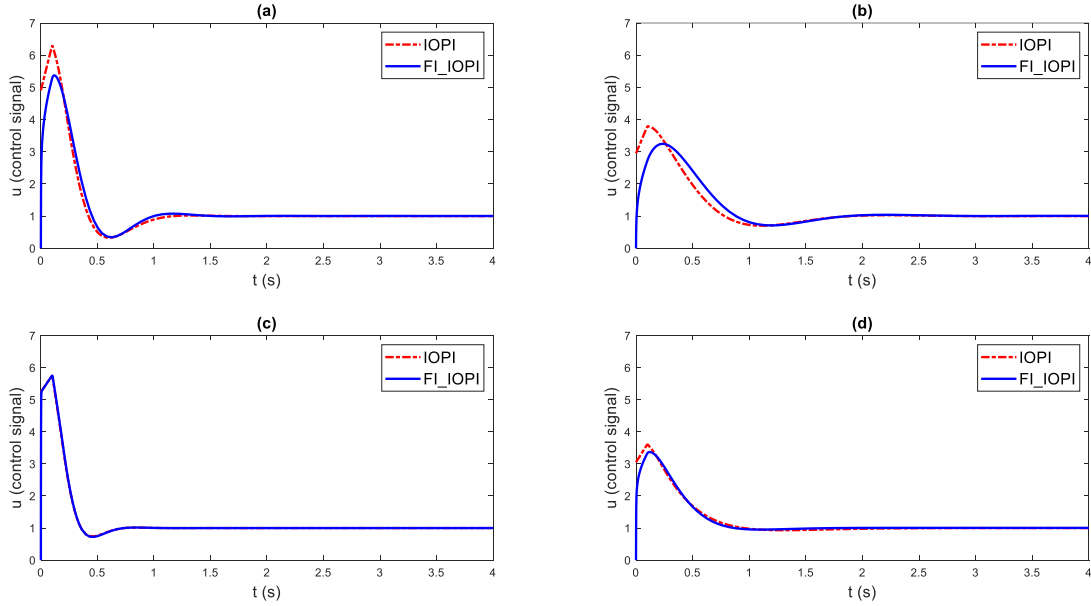
TABLE I. PARAMETERS OF IOPI AND FI_IOPI CONTROLLERS

Design Specifications		IOPI controller parameters		FI_IOPI controller parameters	
A_m	ϕ_m	K_c	T_i	K_d	γ
3	45°	4.91	0.35	6.77	1.17
5	45°	2.95	0.35	4.50	1.29
3	60°	5.24	1.00	5.24	1.00
5	60°	3.05	0.54	3.81	1.12

The step responses of the control systems using IOPI and FI_IOPI are illustrated in Fig. 3 for four specifications. Moreover, the time and frequency domain characteristics of the control systems are summarized in Table II. It can easily be seen from the table that the proposed controller satisfies GPM specifications exactly. Examining the time domain characteristics in detail, it is seen that the proposed controller has less overshoot and less settling time while integer order PI is superior to proposed controller in terms of rise time. On the other hand, Fig. 4 depicts the control signals of the control systems for four specifications. Even though the control signals are close to each other, the proposed controller is superior to IOPI in terms of the maximum amplitude of the control signals.

TABLE II. FREQUENCY AND TIME DOMAIN CHARACTERISTICS OF CONTROL SYSTEMS USING IOPI AND FI_IOPI CONTROLLERS

Design Specifications		Frequency domain characteristics				Time domain characteristics					
A_m	ϕ_m	A_m		ϕ_m		M_p		t_r		t_s	
		IOPI	FI_IOPI	IOPI	FI_IOPI	IOPI	FI_IOPI	IOPI	FI_IOPI	IOPI	FI_IOPI
3	45°	2.91	3	41.6°	45°	32.0116	25.6603	0.1515	0.1797	0.9880	0.9362
5	45°	4.83	5	46.6°	45°	24.5897	26.3566	0.2691	0.3183	2.0509	1.6479
3	60°	3	3	60°	60°	5.5769	5.5769	0.1774	0.1774	0.6054	0.6054
5	60°	3.05	5	58.5°	60°	11.8474	8.1899	0.3212	0.3362	1.8424	1.7327


 Fig. 3. Step responses using IOPI and FI_IOPI for (a) $A_m = 3, \phi_m = 45^\circ$ (b) $A_m = 5, \phi_m = 45^\circ$ (c) $A_m = 3, \phi_m = 60^\circ$ and (d) $A_m = 3, \phi_m = 60^\circ$.

 Fig. 4. Control signals using IOPI and FI_IOPI for (a) $A_m = 3, \phi_m = 45^\circ$ (b) $A_m = 5, \phi_m = 45^\circ$ (c) $A_m = 3, \phi_m = 60^\circ$ and (d) $A_m = 3, \phi_m = 60^\circ$.

V. CONCLUSION

In this study, a robust fractional integrating integer order PI controller design is proposed for the first integer order plus time delay systems. The robustness criteria are selected as phase and gain margins. Moreover, the delayed Bode's ideal transfer function is employed as a reference model to design an analytical controller. In simulation studies, the proposed fractional integrating integer order PI controllers are compared with integer order PI controllers using four different phase and gain margin values. The control system using the proposed controller satisfies GPM specifications exactly because of the delay term in the reference model. Furthermore, the proposed controller has less overshoot and less settling time while integer order PI is superior to the proposed controller in terms of rise time. On the other hand, the proposed controller is superior to IOPI in terms of the maximum amplitude of control signals.

REFERENCES

- [1] P. Chen, Y. Luo, Y. Peng, Y.Q. Chen, "Optimal robust fractional order PI²D controller synthesis for first order plus time delay systems," ISA transactions, 2021, Available online
- [2] B. Maamar, M. Rachid, "IMC-PID-fractional-order-filter controllers design for integer order systems," ISA transactions, 2014, vol. 53, no. 5, pp. 1620-1628.
- [3] E. Yumuk, M. Güzelkaya, İ. Eksin, "Design of an integer order proportional–integral/proportional–integral–derivative controller based on model parameters of a certain class of fractional order systems." Proc of the Ins of Mech Eng, Part I: J Sys and Cont Eng, 2019, vol. 233, no.3, pp. 320-334.
- [4] E. Yumuk, M. Güzelkaya, İ. Eksin, "Optimal fractional-order controller design using direct synthesis method," IET Cont Theory&App, 2020, vol. 14, no. 18, pp. 2960-2967.
- [5] R. Azarmi, M. Tavakoli-Kakhi, A. K. Sedigh, A. Fatehi, "Analytical design of fractional order PID controllers based on the fractional set-point weighted structure: Case study in twin rotor helicopter," Mechatronics, 2015, vol 31, pp. 222-233.
- [6] Q. Bi, W.J. Cai, E.L.Lee, Q.G. Wang, C.C. Hang, "Robust identification of first-order plus dead-time model from step response," Control Engineering Practice, 1999, vol. 7, no.1, pp. 71-77..
- [7] G. Fedele, "A new method to estimate a first-order plus time delay model from step response," Journal of the Franklin Institute, 2009, vol. 346, no.1, pp. 1-9
- [8] Q. Wang, Z. Yong, "Robust identification of continuous systems with dead-time from step responses," Automatica, 2001, vol. 37, no 3, pp.377-390.
- [9] K.J. Astrom, T. Hagglund, PID Controllers: Theory, Design and Tuning, Instrument Society of America, NC, USA, 1995.
- [10] W.K. Ho, K W Lim, W. Xu., "Optimal gain and phase margin tuning for PID controllers." Automatica, 1998, vol. 34, no. 8, pp. 1009-1014.
- [11] W.K Ho, C.C. Hang, L.S. Cao., "Tuning of PID controllers based on gain and phase margin specifications," Automatica, 1995, vol. 31, no. 3, pp 497-502.
- [12] K. Li., "PID tuning for optimal closed-loop performance with specified gain and phase margins.", IEEE transactions on control systems technology, 2012, vol. 21, no. 3, pp 1024-1030.
- [13] P.P. Arya, S. Chakrabarty, "Robust internal model controller with increased closed-loop bandwidth for process control systems," IET Control Theory & Applications, 2020, vol. 14, no. 15, pp.2134-2146.
- [14] İ. Kaya, "Tuning PI controllers for stable processes with specifications on gain and phase margins," ISA transactions, 2004, vol. 43, no. 2, pp.297-304.
- [15] E. Yumuk, M. Güzelkaya, İ. Eksin, "Analytical fractional PID controller design based on Bode's ideal transfer function plus time delay," ISA transactions, 2019, vol. 91, pp. 196-206.
- [16] E. Yumuk, M. Güzelkaya, İ. Eksin, "Application of fractional order PI controllers on a magnetic levitation system," Turk J Elec Eng & Comp Sci, 2021, vol. 29, pp. 98-109
- [17] D. Valerio, J.S. da Costa, "Introduction to single input, single output fractional control," IET Cont Theory&App, 2011, vol. 5, no 8, pp. 1033-1057.
- [18] C.A: Monje, Y.Q. Chen, B.M. Vinagre, D. Xue, V. Feliu, Fractional order Systems and Controls: Fundamentals and Applications. Springer London, 2010.
- [19] H.W. Bode, Network Analysis and Feedback Amplifier Design, 1945.
- [20] J.H. Mathews, K.D. Fink, Numerical Methods Using MATLAB, Pearson Prentice Hall, New Jersey, 2004.

Fuzzy Logic Based Maximum Power Point Tracking for Photovoltaic Systems

Zeynep Bala Duranay*

Electrical-Electronics Engineering Department
University of Firat, Faculty of Technology
Elazig, Turkey
zbduranay@firat.edu.tr

Hanifi Guldemir

Electrical-Electronics Engineering Department
University of Firat, Faculty of Technology
Elazig, Turkey
hguldemir@firat.edu.tr

Abstract—The power demand has been continuously increasing due to the increasing population and increasing technological requirements. There is a need to meet this power demand from alternate energy sources. Solar energy is one of the most important energy sources of this kind. The PV systems are used to obtain electrical power from solar energy. The power produced by PV system changes with the change in weather conditions. Therefore, it is important to draw maximum power in all weather conditions. For this purpose, maximum power point tracking algorithms are developed for use with PV based energy producing systems. Fuzzy logic based maximum power point tracking is studied in this paper. The developed PV fed boost converter with fuzzy logic maximum power point tracking is simulated with various weather conditions. The system is implemented in Matlab/Simulink. The simulations have been done with various solar irradiation and temperature values. The results are checked with the PV producers' technical data for extracting maximum power. Results show that the PV system with fuzzy logic based maximum power point tracking ensures drawing maximum power in all weather conditions.

Keywords—boost converter, fuzzy logic control, maximum power point tracking, photovoltaic system

I. INTRODUCTION

Renewable energy is of great importance for consumers which lack of fossil energy resources. It also has the merits of being clean, cheap and sustainable. These natural energy sources can be listed as sun, wind and wave energy which can be thought as continuous energy sources alternative to fossil fuels which are running out in every day. Due to its availability and effectiveness, the energy received from the sun which is the solar energy, is the most important energy among the other renewable energy resources. Photovoltaic panels which produce electric power are used to generate electrical energy from solar energy. Increasing energy demand increases the wide use photovoltaic (PV) systems which resulting new studies on PV systems which can be grid connected or off-grid system. The surrounding ambient conditions such as cloud, rain, snow, dust, humidity which affect the solar radiation and temperature and these are not constant over time, strongly affect the power produced by PV panels [1-2]. Thus, the most of the studies are on the areas of obtaining reliable, regular and efficient output power from the PV system. The developed new techniques aim to raise the efficiency of the PV system

together with decreasing the cost of the produced energy by increasing the maximum power drawn from the PV system. One of the solution to overcome the drawbacks researches are made to improve the semiconductors used to construct the PV panels by testing various semiconductors. The cost is the limiting factor to test new semiconductors to improve PV panels. The other solution is focused on improving the performance of PV panels by tracing the maximum power point (MPP).

Many works have been presented to increase the performance of PV systems by using various algorithms to operate at MPP which are named as maximum power point tracking (MPPT) techniques. It has been shown that a system with MPPT is more efficient than a non MPPT system [3].

Some popular MPPT techniques such as perturbation and observation technique (P&O) [4] and hill climbing technique (HC) [5] adjust the PV voltage to track the maximum point of the voltage. These techniques have the problem of encountering tracking errors in case of rapidly changing operation points hence lack of accurate MPP tracking.

The incremental conductance (INC) algorithm is one of the most used methods for MPP tracking due to accurate tracking at steady state and rapidly changing operation point [6-7]. INC method uses the slope of the PV panel power characteristic to reach MPP at which the slope of the power curve is zero.

Classical MPPT methods have low convergence speed, high oscillation around MPP, and slow dynamic response in case of sudden environmental changes [8].

Besides these conventional methods, artificial intelligence methods such as neural network (NN) [9-10] and fuzzy logic control (FLC) [11] techniques are also used for tracking MPP applications which are getting more popular with increasing computing power. These algorithms are used to improve the shortcomings of the classical methods such as tracking speed and oscillation around MPP.

NN takes the PV parameters such as short circuit current, open circuit voltage or environmental information such as solar irradiation and PV temperature. The most important part of the NN is the training phase. In order to get improved results, the training should be made for changing weather conditions and for each specific PV panel.

FLC technique is used to get the PV to operate around MPP. It has the capability of handling nonlinearity. Fuzzy based MPPT technique does not require a correct mathematical model. It incorporates human thinking and decisions into the system to produce a control action.

II. DC-DC BOOST CONVERTER

A dc-dc boost converter circuit as presented in Fig. 1, includes a DC supply V_i , inductor L , capacitor C , a mosfet switch, a diode, and a load resistance.

Boost converter provides a voltage higher than its supply voltage by repeatedly making the switching element on and off. The magnitude of the output voltage is changed by changing the duty cycle (D) of the switching signal which is known as pulse width modulation.

Due to existence of switch the converter has two modes of operation. The input and output relationship can be obtained by the switch ON and OFF states assuming ideal components and continuous conduction mode. The two modes of the operation are the switch ON, diode OFF and the switch OFF, diode ON modes.

Equivalent circuits of the boost representing the ON and OFF states of the switch are given in Fig. 2 (a) and (b).

If the switch is closed, the inductor accumulates energy and capacitor feeds the load by releasing its stored energy. If the switch is off, the inductor delivers energy and capacitor accumulates energy.

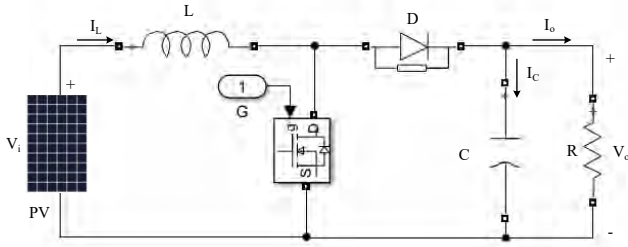


Fig. 1. Boost converter circuit.

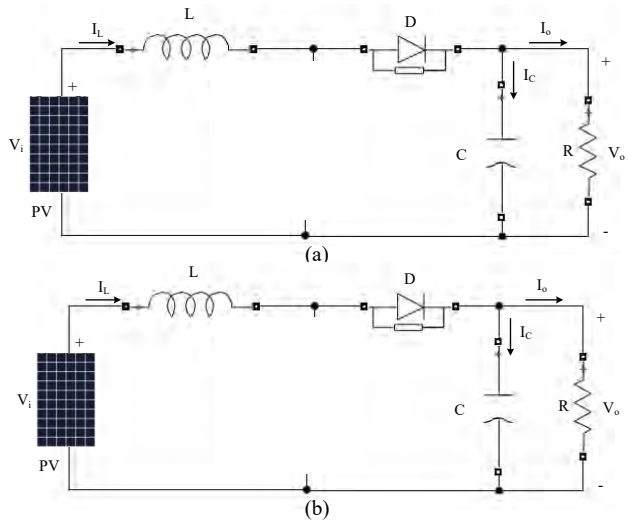


Fig. 2. Boost Circuit when (a) switch is closed (b) switch is open.

The input-output relation is obtained by inductor volt-second and capacitor charge balance. The energy balance requires the input energy to be equal to the output energy that is:

$$P_i = P_o \Rightarrow I_i V_i = I_o V_o \quad (1)$$

The input current provides charge to the output for the time $(1-D)T$ that the switch is open. The expression for the charge balance is

$$Q_i = Q_o \Rightarrow I_i (1-D)T = I_o T \quad (2)$$

Combining (1) and (2), the following expression representing the input-output relationship of a boost converter is obtained.

$$V_o = 1 / (1-D) V_i \quad (3)$$

where D is a positive number less than 1. Thus (3) shows that the output voltage is higher than the input supply voltage.

III. MAXIMUM POWER POINT TRACKING

As the environmental parameters are time varying parameters, the power produced by a PV panel is also varying. To draw a higher power from PV panel it need to be operating at MPP. To operate the system at its MPP an MPPT algorithm needs to be used. Different approaches such as HC techniques [12-13], fractional methods (open circuit voltage and short circuit current [14-15], fuzzy logic based [11, 16], and NN based methods [17-18] are developed for MPPT. Here in this study, perturb and observe (PO) which is one of the HC techniques, based FLC is used for MPPT. The voltage and current are the inputs for PO algorithm as summarized in Fig. 3.

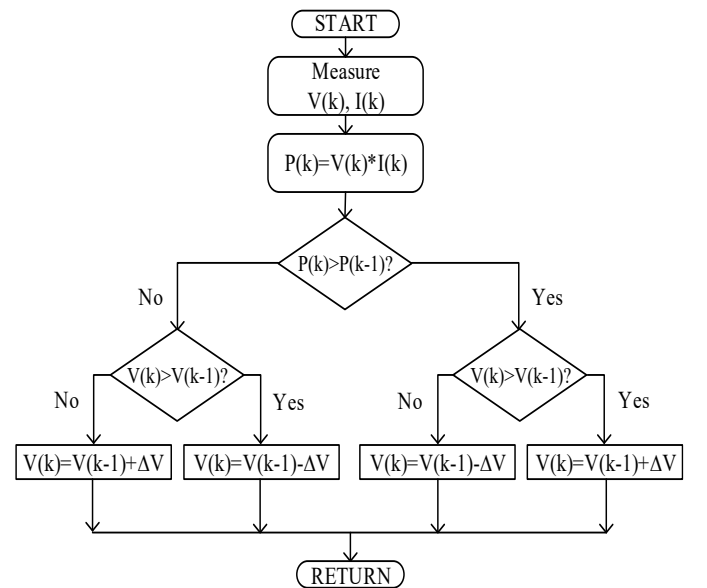


Fig. 3. PO algorithm flowchart.

IV. FUZZY LOGIC CONTROLLER TECHNIQUE FOR PHOTOVOLTAIC MPPT

The problems from real world are complex. This complexity makes it difficult to express the problem mathematically. Fuzzy logic has the advantage of incorporating human deductions into the system and hence, it does not need an accurate mathematical model. Thus fuzzy logic is used in applications where the system models are complex and cannot be easily obtained.

The main parts of the FLC are as follows:

- **Fuzzification:** Receives the real data from the system and maps them into a fuzzy set. The fuzzy set uses linguistic variables and membership functions.
- **Rule Base:** Contains fuzzy IF-THEN rules. The action of the controller is determined by these rules.
- **Inference:** It is the stage where the knowledge of the rules are interpreted to obtain the control action.
- **Defuzzification:** In this step, fuzzy inputs are converted into real output.

The aim of using FLC is to force the PV to operate near the MPP. The FLC has two input, the error (E) and change of error (ΔE) and a single output which is change of duty cycle (ΔD) signal is obtained as

$$E(k) = (P(k) - P(k-1)) / (V(k) - V(k-1)) \quad (4)$$

$$\Delta E(k) = E(k) - E(k-1) \quad (5)$$

Where P, V and I are the power, voltage and current of the PV panel. Fig. 4 shows the block producing the error and change of error signals. ΔE represents the direction of moving.

The voltage and power of the PV module are used to define the E and ΔE . Five fuzzy levels are used for input and output variables. These are expressed using linguistic terms PB (Positive Big), PS (Positive Small), ZE (Zero), NS (Negative Small), NB (Negative Big) each of which are described by a membership function. Table 1 shows fuzzy rules linking change of input variables to the output value. This table produces 25 fuzzy rules to define the control action as follows:

R12: If E is ZE and ΔE is NS then ΔD is NS

R23: If E is PB and ΔE is ZE then ΔD is ZE

...

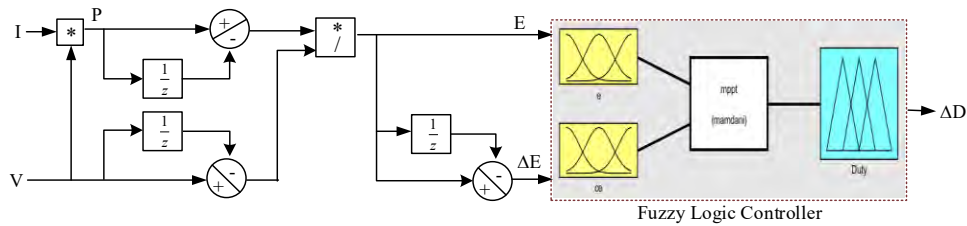


Fig. 4. Block for producing error and change of error signals.

R25: If E is PB and ΔE is PB then ΔD is PS.

The input and output membership functions are presented in Fig 5. E and ΔE are calculated and converted into linguistic terms using these membership functions.

TABLE I. FUZZY RULE TABLE

E \ ΔE	NB	NS	ZE	PS	PB
NB	PS	PB	NB	NB	NS
NS	PS	PS	NS	NS	NS
ZE	ZE	ZE	ZE	ZE	ZE
PS	NS	NS	PS	PS	PS
PB	NS	NB	PB	PB	PS

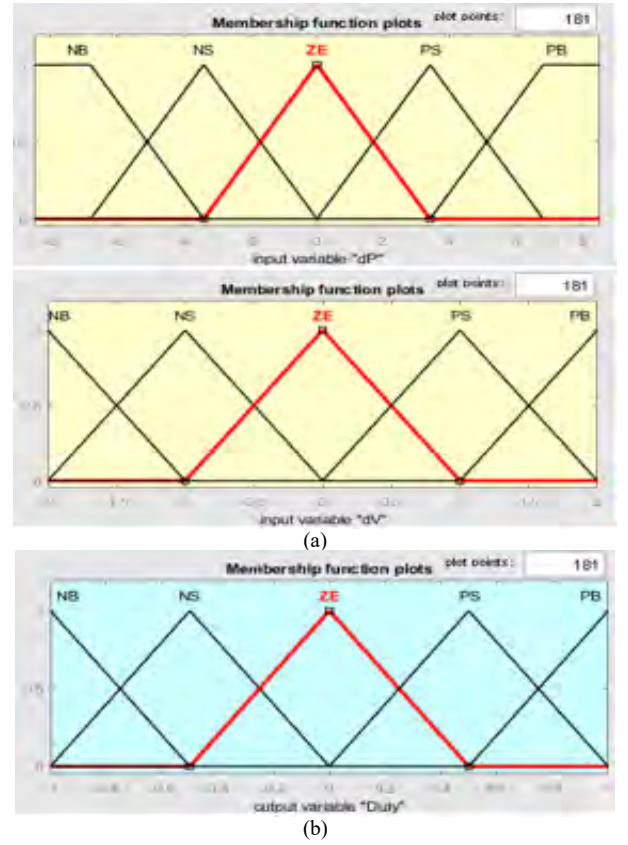


Fig. 5. Boost Circuit when (a) the switch ON (b) the switch OFF.

Mamdani's method with max-min and the center of gravity for the inference and defuzzification is employed to obtain the output. In this step, FLC output is converted from a linguistic variable to numerical values using membership functions. The FLC output is the change in the D. D is calculated as

$$D(k) = D(k-1) + \Delta D \quad (6)$$

The system flowchart is presented in Fig. 6.

The fuzzy rules are defined to follow the MPP of the PV system with varying climatic conditions. Using the P-V characteristics of PV panel given in Fig 7, the following criteria are used to drive the rules:

- If the error determined from (4) is big that is the power is far from the MPP, then D should be big to take the power to the MPP quickly.
- If the power is near the MPP, a small change in D is required.
- If the MPP is achieved, there should be no change in D.

As an example, if the working point is far in the left of MPP, that is E is PB, and CE is ZE, then D should be largely increased i.e., D must be PB to get the MPP. Using P-V characteristic, the algorithm to reach the MPP is summarized in Table 2.

V. PV PANEL

The PV panel is a Kyocera Solar KG200GT module. The P-V characteristics of this panel with various irradiation conditions with constant 25°C temperature are given in Fig 8. The same characteristics with different temperatures and 1000W/m² irradiation are given in Fig. 9. Table 3. lists the electrical characteristics of this PV panel.

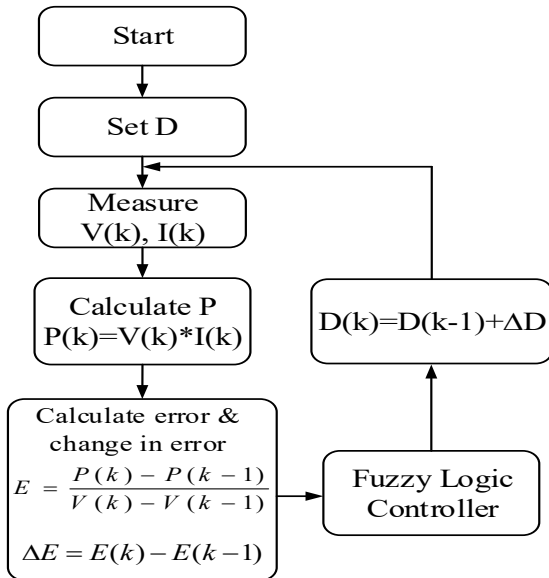


Fig. 5. MPPT flowchart.

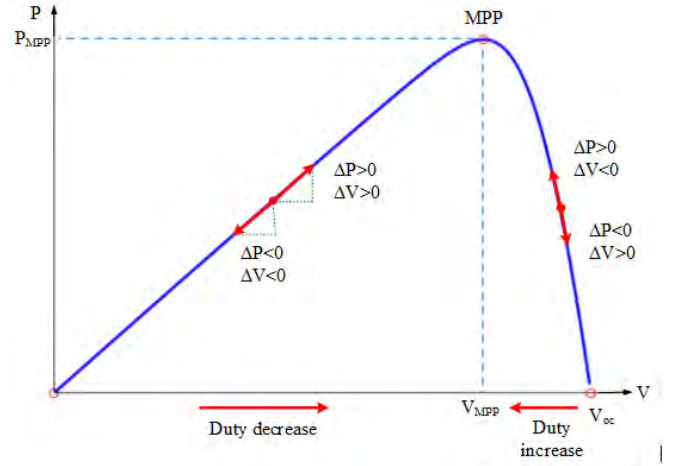


Fig. 6. P-V characteristic curve of a solar panel.

TABLE II. ACTIONS TAKEN TO REACH MPP [19]

$\Delta P(P2-P1)$	$\Delta V(V2-V1)$	Action
Positive	Positive	Increase Vr
Positive	Negative	Decrease Vr
Negative	Positive	Decrease Vr
Negative	Negative	Increase Vr
Zero	-	Constant

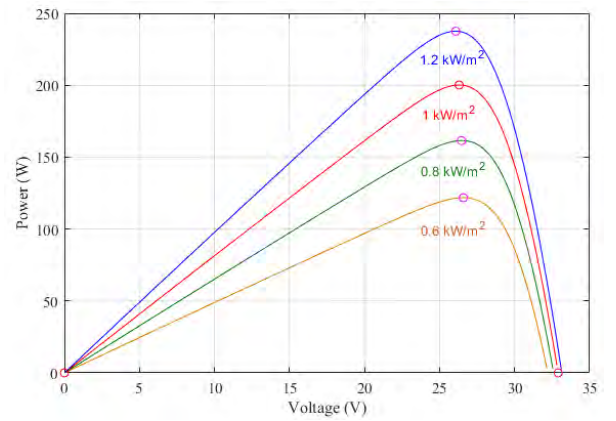


Fig. 7. P-V characteristic of the PV panel under 25°C.

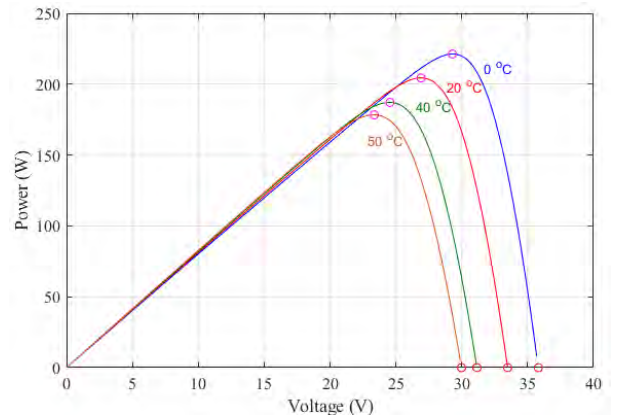


Fig. 8. P-V characteristic curve of the PV panel under 1000W/m² irradiation.

TABLE III. THE ELECTRICAL CHARACTERISTIC OF PV PANEL

Parameter	Value
Maximum Power	200.143 W
Open Circuit Voltage	32.9 V
Voltage at MPP	26.3 V
Short Circuit Current	8.21 A
Current at MPP	7.61

The power produced by the solar panel is dependent on climatic conditions. The increase in irradiation and decrease in temperature results increase in power. The system here is implemented to draw maximum power from the PV panel for all climatic conditions.

VI. SIMULATIONS AND RESULTS

The simulink block of PV system with FLC based MPPT is represented in Fig. 10. This system is used to obtain maximum power from PV module under different environmental conditions.

A boost converter with the parameters given in Table 4 is designed and controlled for PV system.

For the control action, D is used for tracing of the MPP by comparing with the triangular carrier to produce a PWM signal for the boost converter as in Fig 11.

The PV fed boost converter with FLC based MPPT is simulated with different solar irradiance values. During these irradiation variations the temperature is maintained at 25°C. The Fig. 12 represents the solar irradiance variation affecting the PV module. The solar irradiance values are changed as 600 W/m², 800 W/m², 1000 W/m², and 1200 W/m² for the times corresponding to 0s, 0.1s, 0.2s, and 0.3s respectively.

The voltage, the current, and the power obtained at the boost converter output is given in Fig. 13.

TABLE IV. BOOST CONVERTER PARAMETERS

Parameter	Value
L	10mH
C	50 μ F
f_{sw}	10KHz
R	50 Ω

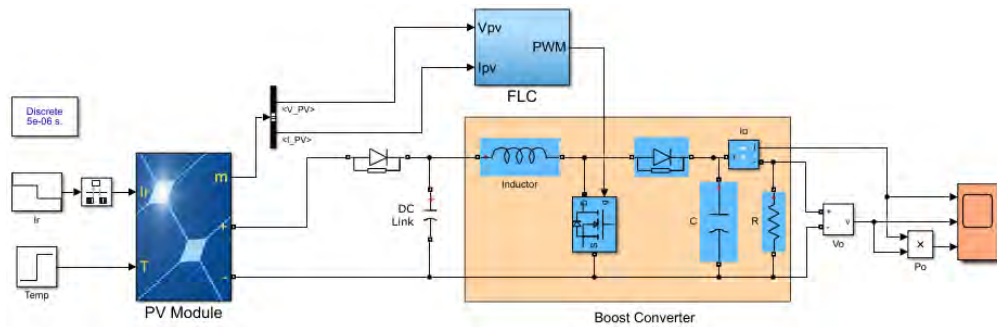


Fig. 9. Simulink block of the FLC based MPPT boost converter.

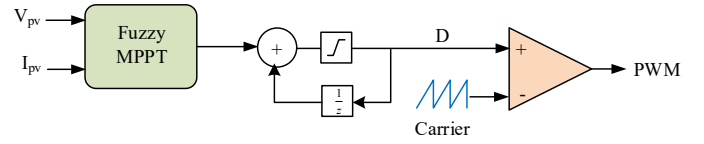


Fig. 11. PWM generation of fuzzy logic based MPPT control.

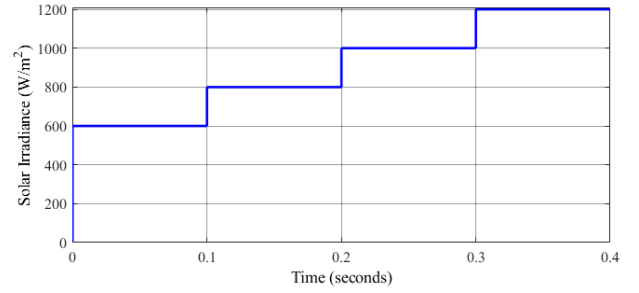


Fig. 12. The solar irradiance variation of the PV module.

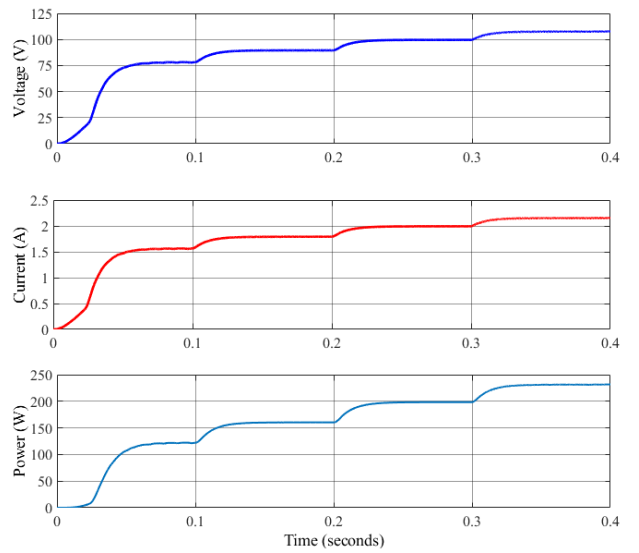


Fig. 13. The voltage, current and the power waveforms of PV system with various irradiances.

As Fig. 8 in section 5 and Fig. 13 are compared, it is seen that, the power at the converter output is the same as the maximum power that the PV module can produce with the given solar radiation values.

As a second case, the PV fed boost converter with FLC based MPPT is simulated with different temperature values while the solar radiation is kept constant at 1000W/m^2 . Fig. 14 represents the temperature variation influencing the PV module. The temperature values are changed as 0°C , 20°C , 40°C , and 50°C for the times corresponding to 0s, 0.1s, 0.2s, and 0.3s respectively.

The voltage, the current and the power waveforms of the PV fed boost converter with FLC based MPPT under varying temperature condition is given in Fig. 15.

As Fig. 9 in section 5 and Fig. 15 are compared, it is clear that, the power at the converter output is the same as the maximum power that the solar module can produce with the given temperature values.

In practice, the temperature and irradiation values change smoothly. But, here in this study, sudden changes are applied to see the impact of these changes on the output values.

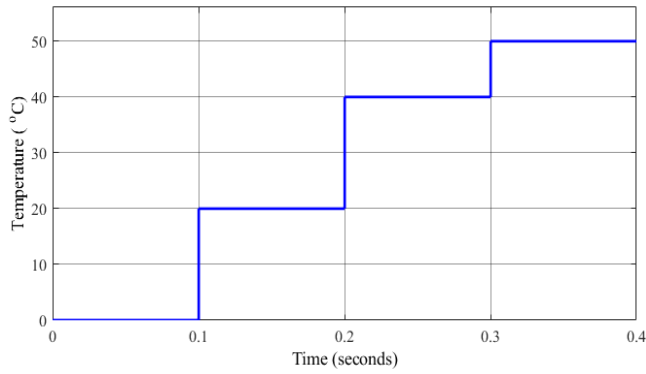


Fig. 14. Temperature variation of the PV module.

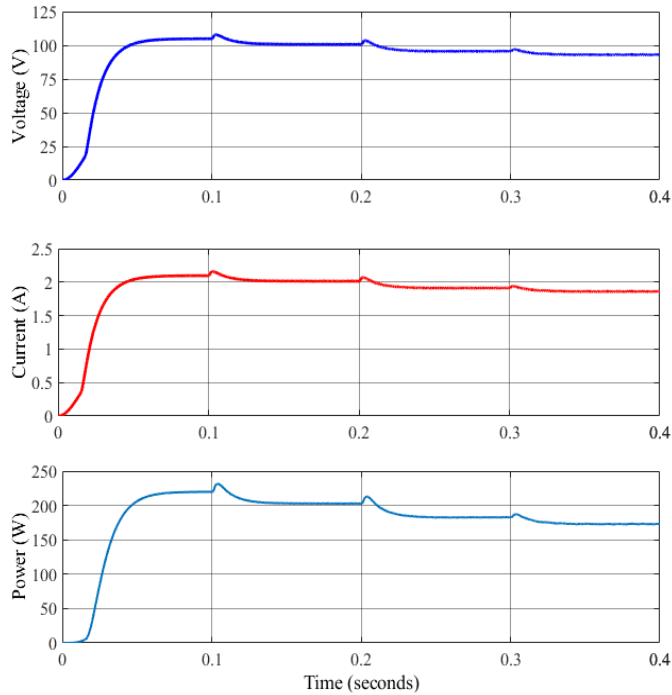


Fig. 15. The voltage, current and the power waveforms of PV system with various temperatures.

CONCLUSION

As the generation of electrical power from PV systems dependent on solar radiation and temperature, it is crucial to obtain high power from the PV system. For extracting maximum power from the PV system in all environmental conditions MPPT algorithms are developed. Besides traditional algorithms, artificial intelligent based MPPT algorithms are also developed for this purpose. Research efforts continue to develop more efficient and cheap MPPT techniques together with converter and control schemes.

In this study, extraction of maximum power from a PV fed boost converter is implemented using fuzzy logic based MPPT technique.

The simulation based performance analysis of PV fed boost converter with FLC based MPPT has been presented. The simulation results are compared with the module technical data. The MPPT technique has been provided to extract maximum power in the cases of environmental changes. The simulation results showed the effectiveness of FLC based MPPT technique in terms of maximum power extraction.

REFERENCES

- [1] M. Berrera, A. Dolara, and S. Leva, "Experimental test of seven widely adopted MPPT algorithms," IEEE Bucharest Power Tech Conference, Bucharest, Romania, pp. 1–8, 2009.
- [2] S. Sreekanth, and I.J. Raglend, "A comparative and analytical study of various incremental algorithms applied in solar cell," International Conference on Computing, Electronics and Electrical Technologies (ICCEET), Nagercoil, India, pp. 452–456, 2012.
- [3] D. K. Chy, and M. Khaliluzzaman, "Experimental assessment of PV arrays connected to buck-boost converter using MPPT and Non-MPPT technique by implementing in real time hardware," International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, pp. 306–309, 2015.
- [4] Y. Liu, M. Li, X. Ji, X. Luo, M. Wang, and Y. Zhang, "A comparative study of the maximum power point tracking methods for PV systems," Energy Conversion and Management, vol. 85, pp. 809–816, 2014.
- [5] M. S. Bouakkaz, A. Boukadoum, O. Boudebouz, I. Attoui, N. Boutasseta and A. Bouraiou, "Fuzzy logic based adaptive step hill climbing MPPT algorithm for PV energy generation systems," International Conference on Computing and Information Technology (ICCIT-1441), pp. 1–5, 2020.
- [6] F. Liu, S. Duan, F. Liu, B. Liu, and Y. Kang, "A variable step size INC MPPT method for PV systems," IEEE Transactions on Industrial Electronics, vol. 55(7), pp. 2622–2628, 2008.
- [7] A. Safari, and S. Mekhilef, "Simulation and hardware implementation of incremental conductance MPPT with direct control method using cuk converter," IEEE Transactions on Industrial Electronics, vol. 58(4), pp. 1154–1161, 2010.
- [8] S. Dadfar, K. Wakil, M. Khaksar, A. Rezvani, M. R. Miveh, and M. Gandomkar, "Enhanced control strategies for a hybrid battery/photovoltaic system using FGS-PID in grid-connected mode," International Journal of Hydrogen Energy, vol. 44(29), pp. 14642–14660, 2019.
- [9] P. Joshi, and S. Arora, "Maximum power point tracking methodologies for solar PV systems—A review," Renewable and Sustainable Energy Reviews, vol. 70, pp. 1154–1177, 2017.
- [10] M. Seyedmahmoudian, B. Horan, T. K. Soon, R. Rahmani, A. M. T. Oo, S. Mekhilef, and A. Stojcevski, "State of the art artificial intelligence-based MPPT techniques for mitigating partial shading effects on PV systems—A review," Renewable and Sustainable Energy Reviews, vol. 64, pp. 435–455, 2016.

- [11] M. Fathi, and J. A. Parian, "Intelligent MPPT for photovoltaic panels using a novel fuzzy logic and artificial neural networks based on evolutionary algorithms," *Energy Reports*, vol. 7, pp. 1338–1348, 2021.
- [12] O. Waszynczuk, "Dynamic behavior of a class of photovoltaic power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. 9, pp. 3031–3037, 1983.
- [13] Y.T. Hsiao, and C. H. Chen, "Maximum power tracking for photovoltaic power system," *IEEE Industry Applications Conference*, vol. 2, pp. 1035–1040, 2002.
- [14] K. Kobayashi, H. Matsuo, and Y. Sekine, "A novel optimum operating point tracker of the solar cell power supply system," *IEEE Power Electronics Specialists Conference*, vol. 2143, pp. 2147–2151, 2004.
- [15] G. W. Hart, H. M. Branz, and C. H. Cox Iii, "Experimental tests of open-loop maximum-power-point tracking techniques for photovoltaic arrays," *Solar Cells*, vol. 13(2), pp. 185–195, 1984.
- [16] L. Zhanghong, Zhangxiaonan, and Xiayilan, "MPPT control strategy for photovoltaic cells based on fuzzy control," *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 450–454, 2016.
- [17] R. Divyasharon, R. Narmatha Banu, and D. Devaraj, "Artificial neural network based MPPT with cuk converter topology for PV systems under varying climatic conditions," *IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, pp. 1–6, 2019.
- [18] H. Elaissoui, M. Zerouali, A. E. Ougli, and B. Tidhaf, "MPPT algorithm based on fuzzy logic and artificial neural network (ANN) for a hybrid solar/wind power generation system," *International Conference On Intelligent Computing in Data Sciences (ICDS)*, pp. 1-6, 2020.
- [19] Z. B. Duranay, and H. Guldemir, "Modelling and simulation of a single phase standalone PV system," *International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1-6, 2019.

Fuzzy-logic Output-tracking Control for Uncertain Time-delay Dynamical Processes: Exploring Takagi-Sugeno Fuzzy Models*

Yuan-Wei Jing, Xin-Jiang Wei, Janusz Kacprzyk, Imre Rudas, and Georgi Dimirovski

Abstract—A novel solution to the fuzzy-logic output tracking control for uncertain time-delay dynamic processes (with uncertain perturbations) is derived by employing the synergy of Takagi-Sugeno fuzzy-logic representation models and theory of variable structure sliding-mode control. Firstly, the sliding mode is chosen by applying the variable structure control theory. According to the reaching condition, the variable control method was proposed for two condition cases when the time delay is available known precisely and when it is unavailable unknown, respectively. The proposed design synthesis does guarantee the trajectory of controlled system to arrive at the sliding surface in finite time interval and be kept on it thereafter. Secondly, the sufficient condition for globally bounded plant state in the closed loop is derived by using the ISS theory and LMI method. An example and its simulation results are explored to illustrate the validity and effectiveness of the proposed design, which apparently outperforms many known previous solutions in the literature.

Keywords—Fuzzy Takagi-Sugeno models; nonlinear time-delay dynamical processes; output tracking control; sliding-mode control.

I. INTRODUCTION

The tasks of stabilization and tracking are the fundamental as well as typical control problems. In general, tracking problems are more difficult than stabilization problems especially for nonlinear or non-amenable to mathematical modelling dynamic plants [1]. In due time Takagi and Sugeno [2] have invented the class of Takagi-Sugeno (T-S) fuzzy systems and also T-S model based designs of fuzzy controllers have been successfully applied to the stabilization control design nonlinear and time-delays systems, e.g. see [3–11]. In most of these applications, the T–S fuzzy model has been proved to be a good representation for a certain class of nonlinear dynamic systems. In the studies, a nonlinear plant was represented by a set of linear models interpolated by membership functions (T-S) fuzzy model and then a model-based fuzzy controller was developed to stabilize the T-S fuzzy model. On the other hand, tracking control designs are also important issues for practical applications, such as in

process set-point tracking, robot trajectory tracking, missile tracking and attitude tracking control of flying objects.

However there are studies focused on the periodic reference tracking control design based on the T-S fuzzy model, especially for continuous-time systems [3, 4, 5, 6, 7, 9]. Tseng et al. [6] proposed an interesting fuzzy tracking control design for nonlinear dynamic systems via T-S fuzzy model similar to that in [5]. However, their method decomposing LMI is greatly conservative and not applicable to operate. In addition, their work did only concern on the tracking control of nominal T-S fuzzy model without time-delay and uncertainty. The robustness of the whole control tracking system thus cannot be guaranteed. In this paper, following works [6, 7] but based on a new synergy of fuzzy T-S model [9, 20] and variable structure control theory [15, 17, 18, 19], the output tracking control problem for fuzzy time-delay systems in the existence of parameter perturbations were developed. Firstly, the sliding mode was selected by variable structure control theory [17, 18]. According to reaching condition, the variable structure control method was proposed for both cases when the time-delay was precisely available and when not available, respectively. The method guaranteed the trajectory of the system to arrive at the slide surface in finite time interval and be kept on sliding surface thereafter. Secondly, the sufficient condition for globally bound of the state was proposed by using theories of Sontag [17] on Input-to-State Stability (ISS) and S-procedure of Yakubovich [12] which yielded the unique Linear Matrix Inequalities (LMI) [13]. Thus these novel developments are computable using MathWorks LMI toolbox [14, 15].

Recently these authors proposed a modified T-S fuzzy model that takes into account most of possible dynamic phenomena in real-world processes exhibiting time-delays and uncertainties [11]. The fuzzy controller was successfully designed by using the theory of sliding-mode variable-structure control, the theory of ISS stability and S-procedure via the LMI technique. The formulation of the investigated problem in the setting of adopted fuzzy-logic and fuzzy-system theory is presented in the next section. In the subsequent section the proposed novel design synthesis is developed. Then the benchmark illustrative example of a continuously steered chemical process plane and the essential simulation results are presented. The concluding remarks and references follow thereafter.

II. PROBLEM FORMULATION IN FUZZY REPRESENTATION OF TIME-DELAY UNCERTAINTY PLANT PROCESS

The celebrated T-S dynamic fuzzy model had been initially proposed by Takagi and Sugeno [2] so as it is

*This research has been supported by Chinese National Natural Science Foundation (grants, 61473073; 61104074) and Program for Liaoning Excellent Talents in University (grant LJQ2014028).

Yuan-Wei Jing (jingyuanwei@ise.neu.edu.cn) and Xin-Jiang Wei (weixinjiang@eyou.com) are with the College of Information Science & Engineering, Northeastern University, Shenyang, Liaoning 110004, P.R. China.

Janusz Kacprzyk (kacprzyk@ibspan.waw.pl) is with the Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland.

Imre Rudas (rudas@uni-obuda.hu) is with Faculty of Engineering, Obuda University, Budapest, HU 1034, Hungary.

Georgi Dimirovski (dimir@jfeit.ukim.edu.mk) is with the School FEIT, SS Cyril & Methodius University, MK 1000 Skopje, R.N. Macedonia.

feasible to represent arbitrary dynamical processes possessing various phenomena, which are either impossible or much too difficult to describe analytically [1]. The most successful T-S fuzzy model, however, is a piecewise interpolation of several linear models through their membership functions and respective grades of membership [21, 22]. The fuzzy model is described by fuzzy If-Then rules and it has been successfully used in wide variety of applications [6-9, 22]. Therefore it will be employed here to deal with the control problem for the uncertain time-delay dynamical processes of industrial plants [23] and to explore the effects that can be achieved via applying to Takagi-Sugeno fuzzy models.

The i -th rule of the fuzzy model for the nonlinear time-delay system is of the following form:

Plant Rule i :

If $\theta_1(t)$ is μ_{i1} and...and $\theta_p(t)$ is μ_{ip} , Then

$$\begin{aligned}\dot{x}(t) &= (A_{1i} + \Delta A_{1i})x(t) + \\ &+ (A_{2i} + \Delta A_{2i})x(t - \tau) + Bu(t) \quad (1) \\ y(t) &= Cx(t), \quad (i = 1, 2, \dots, r). \quad (2)\end{aligned}$$

In this model, there μ_{ij} denotes fuzzy set; $x(t) \in \mathbb{R}^n$ denotes state vector; $u(t) \in \mathbb{R}^m$ denotes the control input; A_{1i} , A_{2i} denotes some constant matrices of compatible dimensions; τ denotes the time-delay affecting the state, which is assumed bounded $0 \leq \tau \leq d$ by a real-valued constant d . Quantities ΔA_{1i} , ΔA_{2i} denote the uncertainty perturbations in the local linear dynamics, while $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{m \times n}$ are the respective input and output matrices. Naturally, the premise variables are independent of the input control variables $u(t)$ which are generated by the controller to be synthesized. It is well known [21, 22], the overall fuzzy system model is obtained by fuzzy blending of each individual rule which yields:

$$\begin{aligned}\dot{x}(t) &= \sum_{i=1}^r h_i(\theta(t))[(A_{1i} + \Delta A_{1i})x(t) + \\ &+ (A_{2i} + \Delta A_{2i})x(t - \tau) + Bu(t)] \quad (3) \\ y(t) &= Cx(t). \quad (4)\end{aligned}$$

In equation (3), the $\theta_i(t)$ ($i = 1, 2, \dots, p$) are the premise variables and $\nu_{ij}(\theta_j(t))$ denotes the grade of membership of $\theta_j(t)$, while $\theta(t)$ is the vector of premise variables $[\theta_1(t), \theta_2(t), \dots, \theta_p(t)]^T$. It should be noted furthermore, according to [22, 23] also in (3) there are:

$$w_i(\theta(t)) = \prod_{j=1}^r \nu_{ij}(\theta_j(t)), \quad h_i(\theta(t)) = \frac{w_i(\theta(t))}{\sum_{j=1}^r w_j(\theta(t))}.$$

In addition, the next standard assumptions are needed.

Assumption 1: The matrix CB is nonsingular.

Assumption 2: All the perturbations ΔA_{1i} , ΔA_{2i} satisfy the following condition: there exist $\|\Delta \bar{A}_{1i}\| \leq M_{\Delta A_1}$, $\|\Delta \bar{A}_{2i}\| \leq M_{\Delta A_2}$, such that $\Delta A_{1i}(x) = B\Delta \bar{A}_{1i}$, $\Delta A_{2i}(x) = B\Delta \bar{A}_{2i}$, where $M_{\Delta A_1}$, $M_{\Delta A_2}$ are known real-valued numbers.

Assumption 3 [2]: There exists known real-valued number $q > 1$, such that $\|x(t - \tau)\| \leq q\|x(t)\|$ for $\tau \in [0, d]$ with d is a constant upper bound.

The objective in this paper is to derive a design synthesis for a variable structure controller such that the output $y(t)$ of (3) will track a desired given reference trajectory $y_d(t)$.

III. AN INNOVATED SETTING OF OUTPUT TRACKING CONTROL PROBLEM

In what follows, firstly the definition of the Input-to-State Stability (ISS) and a necessary and sufficient condition for the ISS of nonlinear dynamic systems is introduced.

Consider the general nonlinear system

$$\dot{x} = f(x, u), \quad (5)$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ is a continuous function satisfying $f(0, 0) = 0$, which means system (5) possesses a resting equilibrium state. According to [4, 5], there are needed a couple of lemmas.

Lemma 1 [6, 15]: System (5) is ISS if and only if there is a smooth function $V: \mathbb{R}^n \rightarrow \mathbb{R}_+$ such that there exist K_∞ function ν_1, ν_2 and K function ν_3, ν_4 , such that

$$\begin{aligned}\nu_1(\|\xi\|) &\leq V(\xi) \leq \nu_2(\|\xi\|) \quad \forall \xi \in \mathbb{R}^n \\ \dot{V}(\xi) &\leq -\nu_3(\|\xi\|), \text{ so that } V(\xi) \geq \nu_4(\|\xi\|)\end{aligned}$$

A function $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is called K function if it is continuous, strictly increasing and $\gamma(0) = 0$; if γ further satisfies $\lim_{t \rightarrow \infty} \gamma(t) = \infty$, it is called K_∞ function.

When time-delay τ is precisely known and available, the sliding mode was selected for system (3)-(4) by using variable structure control theory [17, 19] as follows:

$$s(t) = (CB)^{-1}(y(t) - y_r(t))$$

By defining the output tracking error

$$e(t) = y(t) - y_r(t),$$

the following variable structure controller is obtained:

$$\begin{aligned} u(t) = & -\sum_{i=1}^r h_i(\theta(t))(CB)^{-1}C[A_{1i}x(t) + A_{2i}x(t-\tau)] \\ & -\alpha_1 s - \alpha_2 \operatorname{sgn} s - [M_{\Delta_1} \|x(t)\| + \\ & + M_{\Delta_2} \|x(t-\tau)\|] \operatorname{sgn} s + (CB)^{-1} \dot{y}_r(t) \end{aligned} \quad (6)$$

where α_1, α_2 are two positive real-valued numbers.

Substituting the controller (6) into (4), yields

$$\begin{aligned} \dot{x}(t) = & \sum_{i=1}^{i=r} h_i(\theta(t))[(A_{1i} - B(CB)^{-1}CA_{1i})x(t) + \\ & h_i(\theta(t))[(A_{1i} - B(CB)^{-1}CA_{1i})x(t) + \\ & + (A_{2i} - B(CB)^{-1}CA_{2i})x(t-\tau) + \\ & + B(\Delta \bar{A}_{1i}x(t) + \Delta \bar{A}_{2i}x(t-\tau))] - \\ & - B[\alpha_1 s + \alpha_2 \operatorname{sgn} s + (M_{\Delta_1} \|x(t)\| + \\ & + M_{\Delta_2} \|x(t-\tau)\|) \operatorname{sgn} s - (CB)^{-1} \dot{y}_r(t)] \\ y = & Cx(t) \end{aligned} \quad (7)$$

Since matrix C is of full row rank, a nonsingular matrix T_1 can be found such that

$$CT_1 = [0 \quad \bar{C}_2] = \bar{C},$$

where \bar{C}_2 is nonsingular. Let $T_1^{-1}B = [\bar{B}_1 \quad \bar{B}_2]^T = \bar{B}$,

where $\bar{B}_1 \in R^{(n-m) \times m}$, $\bar{B}_2 \in R^{m \times m}$, since

$CB = \bar{C}\bar{B} = \bar{C}_2\bar{B}_2$ it follows that matrix \bar{B}_2 is nonsingular. Next, let define

$$T_2 = \begin{bmatrix} I & -\bar{B}_1\bar{B}_2^{-1} \\ 0 & \bar{C}_2 \end{bmatrix}, T_0 = T_2T_1^{-1} = \begin{bmatrix} T_{01} \\ T_{02} \end{bmatrix}$$

where $T_{01} = [I \quad -\bar{B}_1\bar{B}_2^{-1}]T_1^{-1}$, $T_{02} = [0 \quad \bar{C}_2]T_1^{-1}$. It can

be shown that $T_{02} = C$, $T_{01}B = 0$. Let $T_0^{-1} = [T_{0inv1} \quad T_{0inv2}]$,

where $T_{0inv1} \in R^{(n-m) \times m}$ and $T_{0inv2} \in R^{n \times m}$, for closed-loop system (7), the following theorem can be derived.

Theorem 1: Consider system (3) which satisfies the Assumptions 1-3. Suppose there exist positive-definite matrices P and R , such that the following inequality

$$\begin{pmatrix} PN_{1i} + N_{1i}^T P + R & PT_{01}A_{2i}T_{0inv1} \\ T_{0inv1}^T A_{2i}^T T_{01}^T P & -R \end{pmatrix} < 0$$

holds, where $N_{1i} = T_{01}A_{1i}T_{0inv1}$. The variable structure controller (6) will enforce output (8) of the closed-loop system (7) to track the desired given reference signal $y_r(t)$.

Proof: The proof is divided into two parts. In part a), it is shown the output of (3) can follow exactly the desired signal $y_r(t)$ after a finite time interval. In part b), it is shown the state of (3) is bounded globally.

a) The derivative of $s(t)$ along the trajectory of closed-loop system (7) is

$$\begin{aligned} \dot{s}(t) = & (CB)^{-1}C\dot{x}(t) - (CB)^{-1}\dot{y}_r(t) \\ = & \sum_{i=1}^r h_i(\theta(t))(CB)^{-1}C[A_{1i}x(t) + A_{2i}x(t-\tau)] + \\ & + u(t) + \sum_{i=1}^r h_i(\theta(t))[\Delta \bar{A}_{1i}x(t) + \Delta \bar{A}_{2i}x(t-\tau)] - (CB)^{-1}\dot{y}_r(t) \\ = & -\alpha_1 s - \alpha_2 \operatorname{sgn} s - [M_{\Delta_1} \|x(t)\| + M_{\Delta_2} \|x(t-\tau)\|] \operatorname{sgn} s \\ & + \sum_{i=1}^r h_i(\theta(t))[\Delta \bar{A}_{1i}x(t) + \Delta \bar{A}_{2i}x(t-\tau)] \end{aligned}$$

when $s_j > 0$, thus it appears

$$\begin{aligned} \dot{s}_j = & -\alpha_1 s_j - \alpha_2 - M_{\Delta_1} \|x(t)\| - M_{\Delta_2} \|x(t-\tau)\| + \\ & + (\sum_{i=1}^r h_i(\theta(t))[\Delta \bar{A}_{1i}x(t) + \Delta \bar{A}_{2i}x(t-\tau)])_j \leq \\ \leq & -\alpha_1 s_j - \alpha_2 - M_{\Delta_1} \|x(t)\| - M_{\Delta_2} \|x(t-\tau)\| + \\ & + \sum_{i=1}^r h_i(\theta(t))(M_{\Delta_1} \|x(t)\| + M_{\Delta_2} \|x(t-\tau)\|) \\ = & -\alpha_1 s_j - \alpha_2 \end{aligned}$$

Similarly, we can show that when $s_j < 0$, the following inequality holds $\dot{s}_j \geq -\alpha_1 s_j + \alpha_2$. From above, it can be seen that all s_j will arrive at zero in finite time interval and kept here thereafter.

b) The state of system (3) is bounded globally.

Under the following state transformation

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = T_0 x(t) = \begin{bmatrix} T_{01}x \\ T_{02}x \end{bmatrix}$$

it can be derived

$$\begin{aligned} \dot{z}_1(t) = & \sum_{i=1}^r h_i(\theta(t))[T_{01}A_{1i}T_{0inv1}z_1(t) + \\ & + T_{01}A_{1i}T_{0inv2}z_2(t) + T_{01}A_{2i}T_{0inv1}z_1(t-\tau) + \\ & + T_{01}A_{2i}T_{0inv2}z_2(t-\tau)] \end{aligned} \quad (9)$$

$$\begin{aligned} \dot{z}_2(t) = & \sum_{i=1}^r h_i(\theta(t))[CB(\Delta \bar{A}_{1i}x(t) + \Delta \bar{A}_{2i}x(t-\tau))] - \\ & - CB[\alpha_1 s + \alpha_2 \operatorname{sgn} s + \\ & + (M_{\Delta_1} \|x(t)\| + M_{\Delta_2} \|x(t-\tau)\|) \operatorname{sgn} s - (CB)^{-1}\dot{y}_r(t)] \end{aligned} \quad (10)$$

Since $z_2 = T_{02}x = Cx = y$, notice that in the previous argument, when $t \geq t_r$, $y(t) = y_r(t)$. So our attention can be concentrated in (9).

Now let view $z_2(t)$ as the input of (9) and choose an appropriate ISS-Lyapunov function candidate as follows:

$$V(z_1(t)) = z_1^T P z_1 + \int_{t-\tau}^t z_1^T(s) R z_1(s) ds$$

The derivative of $V(z_1(t))$ along the trajectory of system (9) is

$$\begin{aligned} \dot{V}(z_1) &= \sum_{i=1}^r h_i(\theta(t)) z_1^T(t) (P N_{li} + N_{li}^T P) z_1(t) + \\ &\quad + 2 \sum_{i=1}^r h_i(\theta(t)) z_1^T(t) T_{01} A_{li} T_{0inv2} z_2(t) + \\ &\quad + 2 \sum_{i=1}^r h_i(\theta(t)) z_1^T(t) P T_{01} A_{2i} T_{0inv1} z_1(t - \tau) + \\ &\quad + 2 \sum_{i=1}^r h_i(\theta(t)) z_1^T(t) P T_{01} A_{2i} T_{0inv2} z_2(t - \tau) + \\ &\quad + z_1^T(t) R z_1(t) - z_1^T(t - \tau) R z_1(t - \tau) \\ &= \sum_{i=1}^r h_i(\theta(t)) \begin{bmatrix} z_1^T(t) & z_1^T(t - \tau) \end{bmatrix} \times \\ &\quad \times \begin{pmatrix} P N_{li} + N_{li}^T P + R & P T_{01} A_{2i} T_{0inv1} \\ T_{0inv1}^T A_{2i}^T T_{01}^T P & -R \end{pmatrix} \begin{bmatrix} z_1(t) \\ z_1(t - \tau) \end{bmatrix} + \\ &\quad + 2 \sum_{i=1}^r h_i(\theta(t)) z_1^T(t) P T_{01} A_{li} T_{0inv2} z_2(t) + \\ &\quad + 2 \sum_{i=1}^r h_i(\theta(t)) z_1^T(t) P T_{01} A_{2i} T_{0inv2} z_2(t - \tau) \end{aligned}$$

Next, let define matrix

$$W_i = \begin{pmatrix} P N_{li} + N_{li}^T P + R & P T_{01} A_{2i} T_{0inv1} \\ T_{0inv1}^T A_{2i}^T T_{01}^T P & -R \end{pmatrix}$$

Let $\lambda = \min\{\lambda_{\min}(-W_i), 1, 2, \dots, r\}$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of the concerned matrix. In addition, we have

$$\begin{aligned} \dot{V}(z_1) &\leq -\lambda(\|z_1\|^2 + \|z_1(t - \tau)\|^2) + \\ &\quad + \gamma_1 \|z_1\| \|z_2\| + \gamma_2 \|z_1\| \|z_2(t - \tau)\| \end{aligned}$$

where γ_1, γ_2 are two positive real numbers, from assumption 3, There exists known number $\beta_1 > 1$, such as $\|x(t - \tau)\| \leq q \|x(t)\|$, so we obtain

$$\begin{aligned} \dot{V}(z_1) &\leq -\lambda \|z_1\|^2 + \gamma \|z_1\| \|z_2\| \\ &\leq -\frac{1}{2} \lambda \|z_1\|^2 + (-\frac{1}{2} \lambda \|z_1\|^2 + \gamma \|z_1\| \|z_2\|) \end{aligned}$$

where $\gamma = \gamma_1 + \gamma_2 \beta$, when $\|z_1\| \geq \frac{2\gamma}{\lambda} \|z_2\|$, such that

$$\dot{V}(z_1) \leq -\frac{1}{2} \lambda \|z_1\|^2$$

According to Lemma 1, when $t \geq t_r$, the system (9) is ISS.

This completes the proof. \square

In the presentation above, the design of the controller is given when the time-delay τ is known and available. Similarly to the above conclusive derivation, when time-delay τ is not available, still it follows

$$\begin{aligned} u(t) &= -\sum_{i=1}^r h_i(\theta(t)) (CB)^{-1} C A_{li} x(t) - \\ &\quad - \sum_{i=1}^r h_i(\theta(t)) H_i \|x(t)\| \operatorname{sgn} s - \\ &\quad - \alpha_1 s - \alpha_2 \operatorname{sgn} s + (CB)^{-1} \dot{y}_d(t) \end{aligned} \quad (11)$$

where $H_i = q \|(CB)^{-1} C A_{2i}\| + M_{\Delta A_1} + q M_{\Delta A_2}$. Notice that the selection of the sliding mode is the same as above. Thus, Theorem 2 can be derived in the same way and proved Theorem 1 was derived.

Theorem 2: Consider system (3) which satisfies the Assumptions 1-3. Suppose there exist positive-definite matrices P and R , such that the following inequality

$$\begin{pmatrix} P N_{li} + N_{li}^T P + R & P T_{01} A_{2i} T_{0inv1} \\ T_{0inv1}^T A_{2i}^T T_{01}^T P & -R \end{pmatrix} < 0$$

holds, where $N_{li} = T_{01} A_{li} T_{0inv1}$. Then the variable structure controller (11) will enforce the output of the closed-loop system (3) along with (11) to track the desired given reference signal $y_r(t)$.

IV. THE CSTR PLANT PROCESS AND SIMULATION RESULTS

Consider a continuous stirred tank reactor (CSTR) as in article [5]; however, complete theoretical treatment of this chemical industrial plant is found in [23]. The respective analytical representation model is described by means of the following equations:

$$\begin{aligned} V \frac{dA}{dt} &= \lambda q A_0 + q(1 - \lambda) A(t - \alpha) - \\ &\quad - q A(t) - V K_0 \exp\left(\frac{-E}{RT(t)}\right) A(t) \end{aligned}$$

$$\begin{aligned} V C_\rho \frac{dT}{dt} &= q C_\rho [\lambda T_0 + (1 - \lambda) T(t - \alpha) - T(t)] - \\ &\quad - V(-\Delta H) K_0 \exp\left(\frac{-E}{RT(t)}\right) A(t) - U(T(t) - T_\omega) \end{aligned}$$

By means of an appropriate transformation, the following nonlinear time-delay model it can be found:

$$\dot{x}_1(t) = f_1(x) + \left(\frac{1}{\lambda} - 1\right) x_1(t - \tau)$$

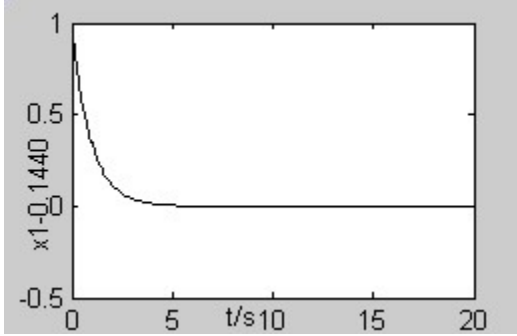
$$\dot{x}_2(t) = f_2(x) + \left(\frac{1}{\lambda} - 1\right) x_2(t - \tau) + \beta u(t)$$

As in source article [5] and illustrative examples in [23], a steady-state at $x_e(t) = [0.1440; 0.8862]$ is considered.

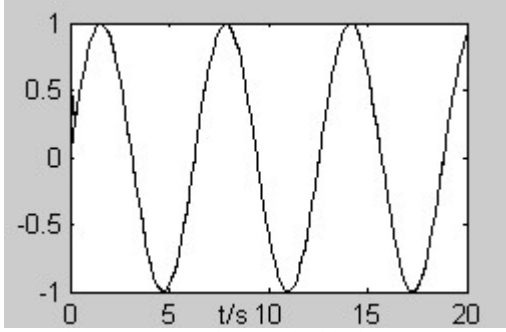
By using variations new state variables can be defined as

$$\delta x_1 = x_1 - x_e(1), \delta x_2 = x_2 - x_e(2)$$

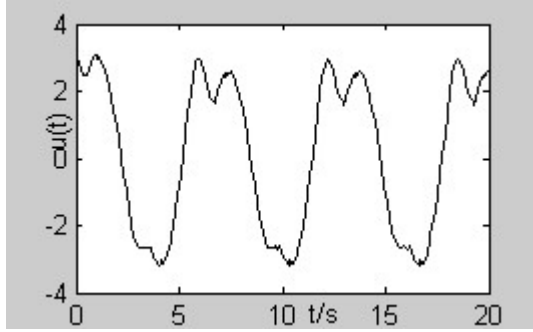
in the neighbourhood of the important equilibrium steady-state.



(a) Trajectory of internal state δx_1



(b) Output tracking trajectory exhibit negligible discrepancy.



(c) Tracking control effort follows approximately the required periodic reference.

Fig. 1 Close-loop system responses when time-delay is precisely known and available to the controller.

Then, as in work [4], based on crucial importance of state variable x_2 , δx_2 respectively, the following fuzzy system model can be observed and investigated in the same manner:

Rule 1: If δx_2 is *small* (e.g., δx_2 is 0.886), Then

$$\begin{aligned} \delta \dot{x}(t) = & (A_{11} + \Delta A_{11})\delta x(t) + \\ & + (A_{21} + \Delta A_{21})\delta x(t - \tau) + B\delta u(t) \end{aligned}$$

Rule 2: If δx_2 is *medium* (e.g., δx_2 is 2.7520), Then

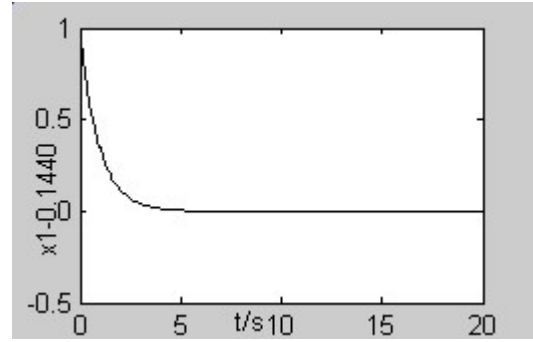
$$\begin{aligned} \delta \dot{x}(t) = & (A_{12} + \Delta A_{12})\delta x(t) + \\ & + (A_{22} + \Delta A_{22})\delta x(t - \tau) + B\delta u(t) \end{aligned}$$

Rule 3: If δx_2 is *large* (e.g., δx_2 is 4.7052), Then

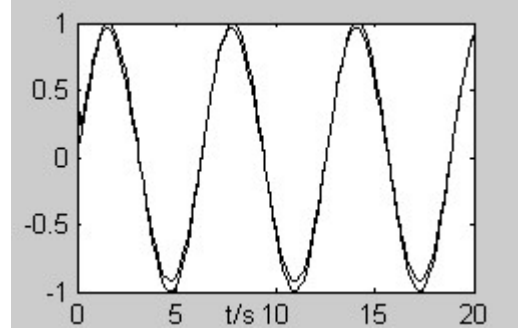
$$\begin{aligned} \delta \dot{x}(t) = & (A_{13} + \Delta A_{13})\delta x(t) + \\ & + (A_{23} + \Delta A_{23})\delta x(t - \tau) + B\delta u(t) \end{aligned}$$

For computer simulations, the chosen periodic reference $y_r(t) = \sin(t)$ while initial condition is $\delta x(0) = [1 \ 1]^T$.

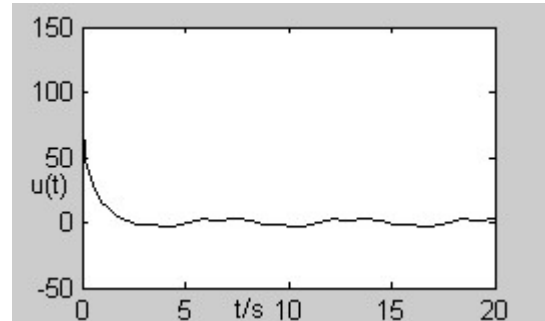
In turn, the time-domain system responses in closed loop were obtained as shown in Figures 1 and 2, respectively.



(a) Trajectory of internal state δx_1



(b) Output tracking trajectory exhibit negligible discrepancy..



(c) Tracking control effort still follows approximately the required periodic reference; however, at the initial times its controlling magnitude has an increase of ten times approximately.

Fig. 2 Closed-loop system responses when time-delay is unknown to fuzzy-logic based hybrid controller.

From the simulation results in Figures 1 and 2, it can be inferred the proposed fuzzy-control design does guarantee simultaneously: the internal states remain globally bounded, while at the same time the plant output is enforced to track the periodic reference almost ideally. This demonstrates the close-loop system does possess not only stability but also rather good tracking performance. Moreover, the chattering phenomenon is considerably suppressed by this fuzzy-logic based variant of sliding-mode control.

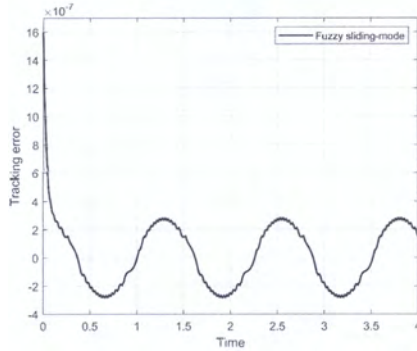


Fig.3 The initial time tracking control error in closed-loop system operation.

It should be noted that in the case of unknown time delays in the plan process the control effort at the initial time is being increase considerably. The controlling magnitude at the initial time is increased ten times approximately in order to enforce the plant on the periodic reference to be tracked. It is therefore worth have a closer look into the tracking error discrepancy during this short segment of initial times. This simulation result is depicted in Figure 3. Apparently, the tracking error is rather small except during the very short segment following the initial instant of reference command excitation of the closed-loop system.

V. CONCLUDING REMARKS AND FUTURE RESEARCH

Based on the synergy of modeling technique the class of T-S fuzzy systems and of sliding-mode variable structure systems control theory, the output tracking control problem for fuzzy time-delay systems in presence of parameter uncertainty perturbations was developed. The proposed design-synthesis of fuzzy-logic based variant of sliding-mode control has capacity to overcome inherited potential deficiencies of both the H_∞ control theory and the LMI computational technique. The proposed fuzzy control techniques can considerably weaken the inherent chattering phenomenon [17] in the sliding mode control. The envisaged future research evolves towards extending this design so as to employ fuzzy-neural synergy [24] and also into nonlinear fuzzy-time delay T-S system models [10, 22].

REFERENCES

- [1] G. M. Dimirovski, N. E. Gough, S. Barnett, "Categories in systems and control." *International Journal of Systems Science*, vol. 8, no. 9, pp. 1081-1090, 1977.
- [2] T. Takagi, M. Sugeno, M., "Fuzzy identification of systems and its application to modeling and control." *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no.1, pp. 116 – 132, 1985.
- [3] B. S. Chen, H. J. Uang, C. S. Tseng, "Robust tracking enhancement of robot systems including motor dynamics: A fuzzy-based dynamic game approach." *IEEE Transactions on Fuzzy Systems*, vol. 6, no.2, pp. 538–552, 1998.
- [4] W. J. Wang, H. R. Lin, "Fuzzy control design for the trajectory tracking on uncertain nonlinear systems." *IEEE Transactions on Fuzzy Systems*, vol. 7, no.1, pp. 53–62, 1999.
- [5] C. H. Chou, C. C. Cheng, "Design of adaptive variable structure controllers for perturbed time-varying state delay systems." *Journal of the Franklin Institute*, vol. 338, pp. 35-46, 2001.
- [6] C. S. Tseng, B. S. Chen, H. J. Uang, "Fuzzy tracking control design for nonlinear dynamic systems via T-S fuzzy model." *IEEE Transactions on Fuzzy Systems*, vol. 9, no.3, pp. 381–392, 2001.
- [7] S. C. Tong, T. Wang and H. X. Li, "Fuzzy robust tracking control design for uncertain nonlinear dynamic systems," *International Journal of Approximate Reasoning*, vol. 30, no. 2, pp. 73-90, 2002.
- [8] J. C. Lo and M. L. Lin, "Robust H_∞ nonlinear modelling and control via uncertain fuzzy systems," *Fuzzy Sets and Systems*, vol. 143, no.2, pp. 189-209, 2004.
- [9] H. O. Wang, K. Tanaka, and M. Griffin, "An approach to fuzzy control of nonlinear systems: Stability and design issues." *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 1, pp. 14-23, 2006.
- [10] D. Zhang, Y. Jing, Q. Zhang, G. M. Dimirovski, "Stabilization of singular T-S fuzzy Markovian jump system with mode-dependent derivative-term coefficient via sliding mode control." *Applied Mathematics & Computations*, vol. 364, art. 124643, 2020.
- [11] V. A. Yakubovich, "The S-procedure in non-linear control theory" (in Russian). *Vestnik Leningradskogo Universsiteta, Matematika i Mekhanika*, vol. 4, pp. 73-93, 1971
- [12] S. P. Boyd, L. El Ghaoui, Feron, E. and V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control Theory*, SIAM Studies in Applied Mathematics, vol.15. Philadelphia, PA: The SIAM, 1994.
- [13] P. Gahinet, A. Nemirovskii, A. J. Laub, M. Chilali, *The LMI Tool Box*. Natick: NJ: The MathWorks, Inc., 1995.
- [14] Mathworks, MATLAB, LIMULINK, Natick, MA: The Mathworks, Inc, 1995.
- [15] H. K. Khalil, *Nonlinear Systems* (3rd ed.). Upper Saddle River, NJ: Prentice Hall, 2002, Sections 4.4 and 4.5.
- [16] E. D. Sontag, Y. Wang, "On the characterization of input-to-state stability property." *Systems and Control Letters*, vol. 34, pp. 351-359, 1995.
- [17] V. L. Utkin, *Sliding Modes and their Application in Variable Structure Systems*. Moscow, RF, USSR: MIR Publishers, 1978.
- [18] S. H. Zak, *Systems and Control*. New York and Oxford: Oxford University Press, 2003, Sections 6.6 and 6.7.
- [19] D. S. Yoo, M. J. Chung, "Variable structure control with simple adaptation laws for upper bounds on the norm of the uncertainties." *IEEE Transactions on Automatic Control*, vol. 37, no. 6, pp. 860-864, 1992.
- [20] K. Tanaka and H. O. Wang, *Fuzzy Control System Design and Analysis: A Linear Matrix Inequality Approach*. New York, NY: J. Wiley & Sons, 2001.
- [21] L. A. Zadeh, "Inference in fuzzy logic." *The IEEE Proceedings*, vol. 68, pp. 124-131, 1980.
- [22] L. A. Zadeh, "Is there a need for fuzzy logic?" *Information Sciences*, vol. 178, pp. 2751-2779, 2008.
- [23] D. E. Seborg, T. F. Edgar, D. A. Mellichamo, F. J. Doyle III, *Process Dynamics and Control*. Hoboken, NJ: J. Wiley & Sons, Inc., 2011
- [24] C.-T. Lin, C. S. G. Lee, *Neural Fuzzy Systems*. Upper Saddle River, NJ: Prentice Hall, 1996.

Discrete-Time Unscented Kalman Filters with Operating of Uncertainties: Stochastic Stability Analysis

Yuanwei Jing¹, Jiahe Xu², Peng Shi³, Georgi Dimirovski^{*4}

Abstract – The performance of the Unscented Kalman Filter (UKF) for a class of general nonlinear stochastic discrete-time systems in presence of uncertainties is investigated in this paper. It is proved that the estimation error of the UKF remains bounded provided certain conditions are satisfied. It is further shown that the estimation error remains bounded provided the system satisfies the nonlinear observability rank condition. Furthermore, it is shown that the design of noise covariance matrix plays an important role in improving the stability of the UKF algorithm. These results are verified by simulations for a given illustrative example of an inherently nonlinear plant.

Keywords – Estimation, Kalman filtering, stochastic nonlinear systems, stochastic stability, unscented Kalman Filter.

I. INTRODUCTION

State estimation and Kalman filtering never ceased to attract considerable research efforts worldwide since the famous seminal papers [1, 2] by Rudolf E. Kalman based on new insights into dynamic systems and following his new approach to general theory of control systems [2, 3]. These new insights involved not only concept of systems state but also system mechanism properties such as observability and controllability in addition to mechanisms of stability and instability. A great many researchers have extended these ideas and insights, and some of recent ones are found in [5-8]. In a rather condensed summary these may be envisaged as in Figure 1.

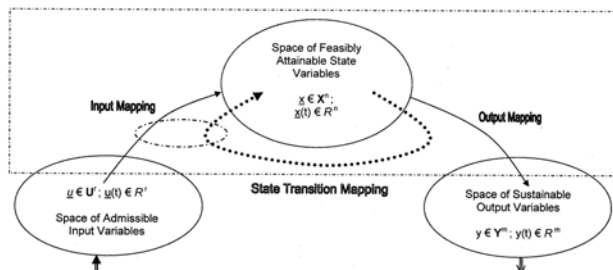


Figure 1. The ure of mapping dynamic system processors as interpreted in engineering terms and consistent with the semi-group algebraic setting.

At the same time, however, it has to be immediately emphasised that Kalman filtering has transcended into a unique discipline in systems and signal processing sciences, which yielded a number of successful technologies for various applications [9]. Although initially introduced and developed for estimation problems in linear systems, in due time, it has been extended to nonlinear systems as well within the context of Extended Kalman Filters (EKF) [9-14]. Moreover, articles by Unbehauen and co-authors [15-17] provided the first analytical results on stochastic stability for both discrete-time

and continuous-time EKF. This way they contributed sound background foundations for the EKF system theory for nonlinear estimation applications. Nonetheless, the issues of state estimation and Kalman filtering for nonlinear systems remain open to further research because of the uniqueness of phenomena in nonlinear dynamic systems.

In the early 1990-ties Julier and co-authors proposed the Unscented Kalman Filter (UKF) in [18], [20] precisely for the purpose of essential extension of Kalman filtering to nonlinear estimation problems. Then they have shown the UKF was a considerable improvement in comparison with the EKF [18], [20], [21], [24]. The UKF is based on employing the special transformation technique that is a mechanism for propagating mean and covariance through a nonlinear transformation [19], [25] and called unscented transformation (UT). Thus it is no longer necessary to use a linearization technique and compute the system Jacobian and Hessian matrices for the UKF [20-24]. This invention has enabled essential avoidance of the error produced by the interruption of higher-order terms and the precision can reach the second-order even higher, e.g. as precise as third-order to the Gauss noise [20]. By nowadays, the UKF is widely used in various applications, ranging from target tracking [18] to position determination [23], multi-sensor fusion [26], estimation in flight control under wind shear [27], etc. Author-inventor Julier himself compared the performances of both the UKF and the EKF for a inherently nonlinear system and showed that the UKF outperforms considerably the EKF [21]*; the same example is examined in this paper. A similar comparison has been established in [26].* These findings are further supported with the published results in [28-32] as well as in work [36].

It should be noted, nonetheless, superior performance of the UKF and its practical usefulness is accompanied with a certain heuristics in its original theoretical derivation [18-20], which makes difficult mathematical rigorous derivations of unscented Kalman filters in various applications. Furthermore, the properties of stability and convergence for the UKF are considerably hard to analyze hence have been developed only for special applications where the considered nonlinear

¹ Yuanwei Jing is with Northeastern University (NEU) of Shenyang, NEU School of Information Science & Engineering, Liaoning 110004, P.R. of China. E-mail: ywjing@mail.neu.edu.cn.

² Jiahe Xu is with Research Institute of Wood Industry, Chinese Academy of Sciences, Beijing, 100091, P.R. of China. E-mail: ellipsis@qq.com.

³ Peng Shi is with the University of Adelaide, School of Electrical & Electronic Engineering, SA 5005 Adelaide, Australia, peng.shi@adelaide.edu.au

⁴ Georgi Dimirovski is with SS Cyril and Methodius University at the FEEIT Doctoral School, ASE Institute, Karpos 2 BB, MK-1000 Skopje, R. of N. Macedonia, E-mail: dimir@feit.ukim.edu.mk; *Correspondence Author.

systems are assumed along with linear measurement equation [28].

In addition to these results, one of the two recent research direction is towards studying extensions of the UKF to operation circumstances with intermittent observations [34], [35], [37] or in the presence of packet dropouts for the case of discrete-time systems [36]. The other direction is in studying the UKF for a more general nonlinear case in a stochastic framework [11], [38], which remains of primary interest given the variety of potential applications. Moreover, some interesting relationships between the observability of nonlinear systems [39], [40] and detectability of time-varying linear systems [41], [34] in conjunction with the existence of positive definite solutions for the UKF have also emerged as an important research area.

The here reported research was motivated by the encouraging results on stability analysis both for the standard and extended Kalman filtering [9], [13] as well as on stochastic stability analysis for more general nonlinear estimation problems [12], [15-17]. Additional motivation was the recently solved continuous-time UKF case in [33]. Thus, via studying the dynamics of the discrete-time UKF, in this paper the relevant result on stochastic stability and its underlying relation to nonlinear system observability are derived. The main contribution of this paper is the proof that, under certain conditions, the estimation error of the UKF remains bounded in the sense of mean square. In particular, in order to improve the stability, slight modifications of the standard UKF were performed by introducing an additive, positive definite matrix into the noise covariance matrix. It is shown that if this extra matrix is properly selected the performance of the UKF used for general nonlinear systems may be improved significantly even in the presence of big initial estimation error. Furthermore, the role of nonlinear observability in this context is also established. This way modified UKF is used as an estimator for an example system to illustrate and test the applicability of theoretical results as well as to seek verification via simulations.

II. THE UNSCENTED KALMAN FILTER

The study technique employed in this paper was inspired by the research work of Reif and co-authors on extended Kalman filters [16, 17] for discrete-time case [16] in conjunction with advance studies by Julier and co-authors in [23, 24]. In addition, the considered class of fairly general nonlinear discrete-time systems is assumed to be represented by

$$\begin{aligned} x_k &= f(x_{k-1}) + G_k w_{k-1}, \\ y_k &= h(x_k) + D_k v_k \end{aligned} \quad (1)$$

where, $k \in N$ discrete time, N denotes the set of natural numbers including zero. In (1), $x_k \in R^r$ represents the state and $y_k \in R^m$ the measurement output of stochastic systems. Nonlinear functions $f(\bullet)$ and $h(\bullet)$, possessing uncertainties, are assumed to be continuously differentiable with respect to

x_k , w_k and v_k . The latter two represent uncorrelated, zero-mean white noise, R^k and R^i vector-valued stochastic processes with identity covariance. It is assumed x_0 are uncorrelated with w_k and v_k , and $E(x_0) = \hat{x}_0$, $\text{cov}(x_0) = P_0$. The variances of w_k and v_k satisfy the following expressions

$$E[w_i w_j^T] = \begin{cases} Q_k & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}, \quad E[v_i v_j^T] = \begin{cases} R_k & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

where, Q_k is the system's noise sequence covariance matrix, and it is a symmetrical non-negative definite matrix. Matrix R_k is measurement noise sequence covariance matrix and it is a symmetrical positive definite matrix.

The procedure for implementing the UKF, on the grounds of source article [25], can be summarized as follows.

The n -dimensional random variable x_k with mean \hat{x}_k and covariance \hat{P}_k can be approximated by sigma points $\chi_{i,k}$ selected from the columns of $\hat{x}_{k-1} \pm \left(a \sqrt{n \hat{P}_{k-1}} \right)_{i=1:L}$, $i = 0, \dots, 2n$.

The opposite weights are $\omega_0 = 1 - (1/a^2)$, $\omega_i = 1/(2na^2)$, $i = 1, 2, \dots, 2n$.

Each point is instantiated through the process model to yield a set of transformed samples; the predicted mean and covariance are computed as

$$\chi_{i,k|k-1} = f(\chi_{i,k-1}) + G_{k-1} \chi_{i,k-1}, \quad \hat{x}_{k|k-1} = \sum_{i=0}^{2n} \omega_i \chi_{i,k|k-1} \quad (2)$$

$$\hat{P}_{k|k-1} = \sum_{i=0}^{2n} \omega_i (\chi_{i,k|k-1} - \hat{x}_{k|k-1})(\chi_{i,k|k-1} - \hat{x}_{k|k-1})^T + G_k Q_k G_k^T + \Delta Q_k \quad (3)$$

where, ΔQ_k is an extra positive definite matrix introduced in the calculated covariance matrix as a slight modification of the UKF so that the stability will be improved.

Then the measurement update can be performed with the equations as follows.

$$y_{i,k|k-1} = h(\chi_{i,k|k-1}), \quad \hat{y}_k = \sum_{i=0}^{2n} \omega_i y_{i,k|k-1} \quad (4)$$

$$\hat{P}_{yy} = \sum_{i=0}^{2n} \omega_i (y_{i,k|k-1} - \hat{y}_k)(y_{i,k|k-1} - \hat{y}_k)^T + D_k R_k D_k^T \quad (5)$$

$$\hat{P}_{xy} = \sum_{i=0}^{2n} \omega_i (\chi_{i,k|k-1} - \hat{x}_{k|k-1})(y_{i,k|k-1} - \hat{y}_k)^T \quad (6)$$

$$W_k = \hat{P}_{xy} \hat{P}_{yy}^{-1} \quad (7)$$

$$\hat{x}_k = \hat{x}_{k|k-1} + W_k (y_k - \hat{y}_k) \quad (8)$$

$$\hat{P}_k = \hat{P}_{k|k-1} + W_k \hat{P}_{yy} W_k^T \quad (9)$$

Clearly, the implementation of the UKF is extremely convenient because it does not need to evaluate the Jacobian matrices, which is necessary in the case of the EKF.

III. STABILITY ANALYSIS OF THE UKF

In this section, a simple approach to represent the error dynamics of the UKF for general nonlinear systems is given.

A. Instrumental Diagonal Matrix and Extra Positive Definite Matrix

Firstly the instrumental time-varying matrices are introduced in order to give a formulation for the UT technique of Julier and his co-authors [18-21]. Define the estimation error and prediction error by

$$\tilde{x}_k = x_k - \hat{x}_k, \quad (10)$$

$$\tilde{x}_{k+1|k} = x_{k+1} - \hat{x}_{k+1|k}. \quad (11)$$

Expanding x_k in (1) by means of a Taylor series about x_k gives,

$$x_k = f(\hat{x}_{k-1}) + \nabla f(\hat{x}_{k-1})\hat{x}_{k-1} + \frac{1}{2}\nabla^2 f(\hat{x}_{k-1})\hat{x}_{k-1}^2 + \dots + G_k w_{k-1} \quad (12)$$

where $\nabla^i f(\hat{x})\hat{x}^i = \left(\sum_{j=1}^L \tilde{x}_j \frac{\partial}{\partial x_j} \right)^i f(x) \Big|_{x=\hat{x}_{k-1}}$, x_j denotes the j -th component of x . Expanding $\hat{x}_{k|k-1}$ given in (2) by a Taylor series yields,

$$\begin{aligned} \hat{x}_{k|k-1} &= \left(1 - \frac{1}{a^2}\right)f(\hat{x}_{k-1}) + \frac{1}{2La^2} \sum_{j=1}^L f \left[\hat{x}_{k-1} + \left(a\sqrt{L\hat{P}_{k-1}}\right)_j \right] \\ &\quad + \frac{1}{2La^2} \sum_{i=L+1}^{2L} f \left[\hat{x}_{k-1} - \left(a\sqrt{L\hat{P}_{k-1}}\right)_i \right] \quad (13) \\ &= f(\hat{x}_{k-1}) + \frac{1}{2}\nabla^2 f(\hat{x}_{k-1})P_{k-1} + \dots \end{aligned}$$

Substituting (12) and (13) into (11) gives an approximate equality

$$\tilde{x}_{k|k-1} \approx F_k x_{k-1} + G_k w_{k-1}, \quad (14)$$

where $F_k = \left(\frac{\partial f(x)}{\partial x} \Big|_{x=\hat{x}_k} \right)$

In (14), it is evident that there always exist residuals of state error prediction $\tilde{x}_{k+1|k}$. In order to take these residuals into account and obtain a more exact equality, an unknown instrumental diagonal matrix $\Lambda_k = \text{diag}(\lambda_{1,k}, \lambda_{2,k}, \dots, \lambda_{M,k})$ is introduced, so that

$$\tilde{x}_{k+1|k} = \Lambda_k F_k x_k + G_k w_{k-1} \quad (15)$$

The residual of the measurement can also be defined by

$$\tilde{y}_k = y_k - \hat{y}_k = \Gamma_k H_k \tilde{x}_{k|k-1} + D_k v_k \quad (16)$$

where $H_k = \left(\frac{\partial h(x)}{\partial x} \Big|_{x=\hat{x}_k} \right)$ and $\Gamma_k = \text{diag}(\gamma_{1,k}, \gamma_{2,k}, \dots, \gamma_{M,k})$ is also an unknown instrumental diagonal matrix as Λ_k .

In contrast, the real prediction error covariance matrix is

$$\begin{aligned} P_{k|k-1} &= E[\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^T] = E[(\Lambda_k F_k \tilde{x}_k + G_k w_k)(\Lambda_k F_k \tilde{x}_k + G_k w_k)^T] \quad (17) \\ &= \Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \Delta P_{k|k-1} + G_k Q_k G_k^T \end{aligned}$$

where $\Delta P_{k|k-1}$ is the difference between $\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k$ and $E[\Lambda_k F_k \tilde{x}_{k-1} \tilde{x}_{k-1}^T F_k^T \Lambda_k]$. Suppose $\delta P_{k|k-1}$ denotes the difference between the real covariance matrix $P_{k|k-1}$ and the sampled one $\hat{P}_{k|k-1} = \sum_{i=0}^{2n} \omega_i (\chi_{i,k|k-1} - \hat{x}_{k|k-1})(\chi_{i,k|k-1} - \hat{x}_{k|k-1})^T + G_k Q_k G_k^T + \Delta Q_k$; then the calculated covariance matrix shown in (3) becomes

$$\hat{P}_{k|k-1} = P_{k|k-1} + \delta P_{k|k-1} + \Delta Q_k = \Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k \quad (18)$$

where

$$\hat{Q}_k = \Delta P_{k|k-1} + G_k Q_k G_k^T + \delta P_{k|k-1} + \Delta Q_k \quad (19)$$

The real measurement error covariance matrix is

$$\begin{aligned} P_{yy} &= E[\tilde{y}_{k|k-1}\tilde{y}_{k|k-1}^T] = E[(\Gamma_k H_k \tilde{x}_{k|k-1} + D_k v_k)(\Gamma_k H_k \tilde{x}_{k|k-1} + D_k v_k)^T] \quad (20) \\ &= \Gamma_k H_k \hat{P}_{k|k-1} H_k^T \Gamma_k + \Delta P_{yy} + D_k R_k D_k^T \end{aligned}$$

where ΔP_{yy} is the difference between $\Gamma_k H_k \hat{P}_{k|k-1} H_k^T \Gamma_k$ and $E[\Gamma_k H_k \tilde{x}_{k|k-1} \tilde{x}_{k|k-1}^T H_k^T \Gamma_k]$. Now suppose δP_{yy} denotes the difference between the real covariance matrix P_{yy} and the sampled one $\hat{P}_{yy} = \sum_{i=0}^{2n} \omega_i (y_{i,k|k-1} - \hat{y}_k)(y_{i,k|k-1} - \hat{y}_k)^T + D_k R_k D_k^T$; then the calculated covariance matrix shown in (5) becomes

$$\hat{P}_{yy} = P_{yy} + \delta P_{yy} = \Gamma_k H_k \hat{P}_{k|k-1} H_k^T \Gamma_k + \hat{R}_k \quad (21)$$

where

$$\hat{R}_k = \Delta P_{yy} + D_k R_k D_k^T + \delta P_{yy} \quad (22)$$

And, the real error covariance matrix P_{xy} is

$$\begin{aligned} P_{xy} &= E[\tilde{x}_{k|k-1}\tilde{y}_{k|k-1}^T] = E[\tilde{x}_{k|k-1}(\Gamma_k H_k \tilde{x}_{k|k-1} + D_k v_k)^T] \quad (23) \\ &= E[\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^T H_k^T \Gamma_k + \tilde{x}_{k|k-1} v_k^T D_k^T] \\ &= \hat{P}_{k|k-1} H_k^T \Gamma_k = \hat{P}_{xy} \end{aligned}$$

The difference between the real covariance matrix P_{xy} and the sampled one $\hat{P}_{xy} = \sum_{i=0}^{2n} \omega_i (\chi_{i,k|k-1} - \hat{x}_k)(v_{i,k|k-1} - \hat{y}_k)^T$ is less than the residuals accounted in $\hat{P}_{k|k-1}$. Hence it can be neglected here.

B. Stochastic Boundedness of Estimation Error

For analysis of the error dynamics some standard results about the boundedness of stochastic processes from [11], [38] are recalled.

Lemma 3.1: Assume that ζ_k is the stochastic process and there is a stochastic process $V(\zeta_k)$ as well as real numbers $\nu_{\min}, \nu_{\max} > 0$, $\mu > 0$, and $0 < \alpha \leq 1$ such that for any k

$$\nu_{\min} \|\zeta_k\|^2 \leq V(\zeta_k) \leq \nu_{\max} \|\zeta_k\|^2 \quad (24)$$

$$E[V(\zeta_k) | \zeta_{k-1}] - V(\zeta_{k-1}) \leq \mu - \beta V(\zeta_{k-1}) \quad (25)$$

are fulfilled. Then the stochastic process is bounded in the mean square,

$$E\{\|\zeta_k\|^2\} < \frac{\nu_{\max}}{\nu_{\min}} E\{\|\zeta_0\|^2\} (1-\alpha)^k + \frac{\mu}{\nu_{\min}} \sum_{i=1}^{k-1} (1-\alpha)^i. \quad (26)$$

For the purpose of establishing the sufficient conditions that ensure stability of the UKF another two lemmas, given below, are needed.

Lemma 3.2: Assume that matrices $A \in R^{m \times n}$, $B \in R^{m \times n}$ and $C \in R^{n \times n}$, if $A > 0$ and $C > 0$, then

$$A^{-1} > B(B^T AB + C)^{-1} B^T \quad (27)$$

Lemma 3.3: Assume that matrices $A \in R^{n \times n}$, $C \in R^{n \times n}$, if $A > 0$ and $C > 0$, then

$$A^{-1} > (A + C)^{-1} \quad (28)$$

With Lemmas 3.1 – 3.3 and the formulations shown in (15), (16), (18) and (21), it becomes possible to state the first main result of this paper.

Theorem 3.1: Consider general nonlinear stochastic systems as represented by (1) and the UKF described by (2)-(9). Further, suppose the following assumptions hold true:

(1) There are real numbers $f_{\min}, h_{\min}, \lambda_{\min}, \gamma_{\min} \neq 0$, and $f_{\max}, h_{\max}, \lambda_{\max}, \gamma_{\max} \neq 0$, such that the following bounds on various matrices are fulfilled for every $k \geq 0$:

$$f_{\min}^2 I \leq F_k F_k^T \leq f_{\max}^2 I, \quad (29)$$

$$h_{\min}^2 I \leq H_k H_k^T \leq h_{\max}^2 I, \quad (30)$$

$$\lambda_{\min}^2 I \leq \Lambda_k \Lambda_k^T \leq \lambda_{\max}^2 I, \quad (31)$$

$$\gamma_{\min}^2 I \leq \Gamma_k \Gamma_k^T \leq \gamma_{\max}^2 I. \quad (32)$$

(2) There are real numbers $q_{\min}, q_{\max}, \hat{q}_{\min}, \hat{q}_{\max}, r_{\max}, \hat{r}_{\max}, p_{\max}, p_{\min} > 0$, such that the following bounds are fulfilled:

$$p_{\min} I \leq \hat{P}_k \leq p_{\max} I, \quad (33)$$

$$q_{\min} I \leq Q_k \leq q_{\max} I, \quad (34)$$

$$\hat{q}_{\min} I \leq \hat{Q}_k \leq \hat{q}_{\max} I, \quad (35)$$

$$\hat{R}_k \leq \hat{r}_{\max} I. \quad (36)$$

Then the estimation error \tilde{x}_k given by (10) is exponentially bounded in the mean square.

Proof: Due to Lemma 3.1, let choose

$$V_k(\tilde{x}_k) = \tilde{x}_k^T \hat{P}_k^{-1} \tilde{x}_k \quad (37)$$

From (33) it follows

$$\frac{1}{p_{\max}} \|\tilde{x}_k\|^2 \leq V(\tilde{x}_k) \leq \frac{1}{p_{\min}} \|\tilde{x}_k\|^2 \quad (38)$$

In order to satisfy the requirement for applying Lemma 1, it is necessary to have an upper bound on $E[V(\zeta_k) | \zeta_{k-1}] - V(\zeta_{k-1})$. Substituting (21) and (23) into (9) yields

$$\hat{P}_k = \hat{P}_{k|k-1} - \hat{P}_{xy} \hat{P}_{yy}^{-1} \hat{P}_{xy}^T = (I - W_k \Gamma_k H_k) \hat{P}_{k|k-1} \quad (39)$$

where

$$W_k = \hat{P}_{k|k-1} H_k^T \Gamma_k (\Gamma_k H_k \hat{P}_{k|k-1} H_k^T \Gamma_k + \hat{R}_k)^{-1} \quad (40)$$

By making use of (8), (10), (16) as well as (40) it follows

$$\begin{aligned} \tilde{x}_k &= x_k - \left(\hat{x}_{k|k-1} + \left(\hat{P}_{k+1|k} H_{k+1}^T (\Gamma_k H_{k+1} \hat{P}_{k+1|k} H_{k+1}^T \Gamma_k + \hat{R}_{k+1})^{-1} \right) \tilde{y}_k \right) \\ &= \tilde{x}_{k|k-1} - W_k \tilde{y}_k \end{aligned} \quad (41)$$

The from (37) and (41) it can be found

$$\begin{aligned} V_k(\tilde{x}_k) &= (\tilde{x}_{k|k-1} - W_k \tilde{y}_k)^T \hat{P}_k^{-1} (\tilde{x}_{k|k-1} - W_k \tilde{y}_k) \\ &= \tilde{x}_{k|k-1}^T \hat{P}_k^{-1} \tilde{x}_{k|k-1} - \left(\Gamma_k H_k \tilde{x}_{k|k-1} + D_k v_k \right)^T W_k^T \hat{P}_k^{-1} \tilde{x}_{k|k-1} \\ &\quad - \tilde{x}_{k|k-1}^T \hat{P}_k^{-1} W_k \left(\Gamma_k H_k \tilde{x}_{k|k-1} + D_k v_k \right) \\ &\quad + \left(\Gamma_k H_k \tilde{x}_{k|k-1} + D_k v_k \right)^T W_k^T \hat{P}_k^{-1} W_k \left(\Gamma_k H_k \tilde{x}_{k|k-1} + D_k v_k \right) \end{aligned} \quad (42)$$

Rearranging (40) yields

$$W_k = (I_n - \Gamma_k H_k W_k) \hat{P}_{k|k-1} H_k^T \Gamma_k \hat{R}_k^{-1} = \hat{P}_k H_k^T \Gamma_k \hat{R}_k^{-1} \quad (43)$$

On the other hand, applying the well known matrix inversion lemma on (39) gives

$$\hat{P}_k^{-1} = \hat{P}_{k|k-1}^{-1} H_k^T \Gamma_k \hat{R}_k^{-1} \Gamma_k H_k \quad (44)$$

Inserting (43), (44) and (15) into (42), and taking the conditional expectation yields:

$$\begin{aligned} (45) \quad E[V_k(\tilde{x}(k)) | x(k-1)] &= \\ &= E\{(\Lambda_k F_k \tilde{x}_k + G_k w_k)^T \hat{P}_{k|k-1}^{-1} (\Lambda_k F_k \tilde{x}_k + G_k w_k) - \\ &\quad - [\Gamma_k H_k (\Lambda_k F_k \tilde{x}_k + G_k w_k)]^T \times (\hat{R}_k^{-1} - \hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k \hat{R}_k^{-1}) \times \\ &\quad \times [\Gamma_k H_k (\Lambda_k F_k \tilde{x}_k + G_k w_k)] + v_k^T D_k^T \hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k \hat{R}_k^{-1} D_k v_k | \tilde{x}_{k-1}\}. \end{aligned}$$

Now let examine the term $(\hat{R}_k^{-1} - \hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k \hat{R}_k^{-1})$ on the right side of (45). By using (43) and (40) it can be verified

$$\begin{aligned}
& (\hat{R}_k^{-1} - \hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k \hat{R}_k^{-1}) \\
& = \hat{R}_k^{-1} [1 - \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k (\Gamma_k H_k \hat{P}_k H_k^T \Gamma_k \hat{R}_k^{-1})^{-1}] \quad (46) \\
& = (\Gamma_k H_k \hat{P}_k H_k^T \Gamma_k + \hat{R}_k)^{-1} > 0
\end{aligned}$$

Substituting (46) and (18) into (45), and the applying Lemma 3.3, the expectation (45) becomes:

$$\begin{aligned}
& E[V_k(\tilde{x}_k) | \tilde{x}_{k-1}] \\
& \leq E\{(\Lambda_k F_k \tilde{x}_{k-1})^T (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k)^{-1} (\Lambda_k F_k \tilde{x}_{k-1}) + w_k^T G_k^T (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k \\
& \quad + \hat{Q}_k)^{-1} G_k w_k - (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1})^T [\Gamma_k H_k (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} H_k^T \Gamma_k \\
& \quad + \hat{R}_k]^{-1} (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1}) - (\Gamma_k H_k G_k w_k)^T [\Gamma_k H_k (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k \\
& \quad + \hat{Q}_k)^{-1} H_k^T \Gamma_k + \hat{R}_k]^{-1} (\Gamma_k H_k G_k w_k) + v_k^T D_k^T R_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k R_k^{-1} D_k v_k | \tilde{x}_{k-1}\}. \quad (47)
\end{aligned}$$

Notice that inequalities (29) and (31) imply that $(\Lambda_k F_k)^{-1}$ exists. It is therefore that it may be established

$$\begin{aligned}
& E\{[(\Lambda_k F_k \tilde{x}_{k-1})^T (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k)^{-1} (\Lambda_k F_k \tilde{x}_{k-1}) | \tilde{x}(k-1)]\} \\
& = \tilde{x}_{k-1}^T \hat{P}_{k-1}^{-1} \tilde{x}_{k-1} = V_{k-1}(\tilde{x}_{k-1})
\end{aligned} \quad (48)$$

Subtracting (48) from both sides of (47) then gives

$$\begin{aligned}
& E[V_k(\tilde{x}_k) | \tilde{x}_{k-1}] \\
& \leq E\{w_k^T G_k^T (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} G_k w_k \\
& \quad - (\Gamma_k H_k G_k w_k)^T [\Gamma_k H_k (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} H_k^T \Gamma_k + \hat{R}_k]^{-1} (\Gamma_k H_k G_k w_k) \\
& \quad + v_k^T D_k^T \hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k R_k^{-1} D_k v_k | \tilde{x}_{k-1}\} - (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1})^T \\
& \quad \times [\Gamma_k H_k (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} H_k^T \Gamma_k + \hat{R}_k]^{-1} (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1})
\end{aligned} \quad (49)$$

Now focus on the last term in (49). According to Lemma 3.2 in may be inferred

$$\begin{aligned}
& \hat{P}_{k-1}^{-1} > (\Gamma_k H_k \Lambda_k F_k)^T [(\Gamma_k H_k \Lambda_k F_k) \hat{P}_{k-1} (\Gamma_k H_k \Lambda_k F_k)^T \\
& \quad + \Gamma_k H_k \hat{Q}_k H_k^T \Gamma_k + \hat{R}_k]^{-1} (\Gamma_k H_k \Lambda_k F_k)
\end{aligned} \quad (50)$$

By means of pre- and post-multiplying both sides of (50) by \tilde{x}_{k-1}^T and \tilde{x}_{k-1} , respectively, for $\forall \tilde{x}_{k-1} \neq 0$ it holds

$$\begin{aligned}
& \tilde{x}_{k-1}^T \hat{P}_{k-1}^{-1} \tilde{x}_{k-1} > (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1})^T [(\Gamma_k H_k \Lambda_k F_k) \hat{P}_{k-1} (\Gamma_k H_k \Lambda_k F_k)^T \\
& \quad + \Gamma_k H_k \hat{Q}_k H_k^T \Gamma_k + \hat{R}_k]^{-1} (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1})
\end{aligned} \quad (51)$$

Upon denoting

$$\begin{aligned}
\alpha_k & = (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1})^T [(\Gamma_k H_k \Lambda_k F_k) \hat{P}_{k-1} (\Gamma_k H_k \Lambda_k F_k)^T \\
& \quad + \Gamma_k H_k \hat{Q}_k H_k^T \Gamma_k + \hat{R}_k]^{-1} (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1}) / \tilde{x}_{k-1}^T \hat{P}_{k-1}^{-1} \tilde{x}_{k-1},
\end{aligned} \quad (52)$$

it follows from (51) that $\alpha_k < 1$. Under the assumed inequalities (29)-(36), on the other hand, it follows

$$\begin{aligned}
\alpha_k & \geq p_{\min} (\gamma_{\min} h_{\min} \lambda_{\min} f_{\min})^2 [p_{\max} (\gamma_{\max} h_{\max} \lambda_{\max} f_{\max})^2 + \hat{q}_{\max} h_{\max}^2 + \hat{r}_{\max}]^{-1} \\
& \stackrel{\Delta}{=} \alpha_{\min} > 0
\end{aligned} \quad (53)$$

Thus, by using (51) and (53) it is readily shown that

$$\begin{aligned}
& -(\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1})^T [(\Gamma_k H_k \Lambda_k F_k) \hat{P}_{k-1} (\Gamma_k H_k \Lambda_k F_k)^T \\
& \quad + \Gamma_k H_k \hat{Q}_k H_k^T \Gamma_k + \hat{R}_k]^{-1} (\Gamma_k H_k \Lambda_k F_k \tilde{x}_{k-1}) \leq -\alpha_{\min} V_{k-1}(\tilde{x}_{k-1})
\end{aligned} \quad (54)$$

Next, the other terms in (49) are considered. By means of appropriate analysis of the existing matrix quantities it may well be established:

$$\begin{aligned}
\mu_k & = E\{w_k^T G_k^T (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} G_k w_k \\
& \quad - (\Gamma_k H_k G_k w_k)^T [\Gamma_k H_k (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} H_k^T \Gamma_k \\
& \quad + \hat{R}_k]^{-1} (\Gamma_k H_k G_k w_k) + v_k^T D_k^T \hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k \hat{R}_k^{-1} D_k v_k | \tilde{x}_{k-1}\}
\end{aligned} \quad (55)$$

Since both sides of (55) are scalars, taking the trace will not change its value. Application of $\text{tr}(AB) = \text{tr}(BA)$ gives

$$\begin{aligned}
\mu_k & = E\{\text{tr}[(\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} - \Gamma_k H_k^T [\Gamma_k H_k (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k \\
& \quad + \hat{Q}_k) H_k^T \Gamma_k + \hat{R}_k]^{-1} H_k \Gamma_k (G_k w_k w_k^T G_k^T)] \\
& \quad + \text{tr}[\hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k \hat{R}_k^{-1} D_k v_k v_k^T D_k^T] | \tilde{x}_{k-1}\} \\
& = E\{\text{tr}[(\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} - \Gamma_k H_k^T [\Gamma_k H_k (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k \\
& \quad + \hat{Q}_k) H_k^T \Gamma_k + \hat{R}_k]^{-1} H_k \Gamma_k Q_k] + \text{tr}[\hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k]
\end{aligned} \quad (56)$$

By virtue of Lemma 3.2 it follows

$$\begin{aligned}
& (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} - \Gamma_k H_k^T [\Gamma_k H_k (\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k \\
& \quad + \hat{Q}_k) H_k^T \Gamma_k + \hat{R}_k]^{-1} H_k \Gamma_k > 0
\end{aligned} \quad (57)$$

Thus it is found $\mu_k > 0$ ($\mu(k) > 0$). From (28) and (29)-(36)

$$\begin{aligned}
\mu_k & \leq \text{tr}[(\Lambda_k F_k \hat{P}_{k-1} F_k^T \Lambda_k + \hat{Q}_k)^{-1} Q_k] + \text{tr}[\hat{R}_k^{-1} \Gamma_k H_k \hat{P}_k H_k^T \Gamma_k] \\
& \leq \frac{q_{\max}}{\hat{q}_{\min}} \bullet L + \frac{\gamma_{\max}^2 h_{\max}^2 p_{\max}}{\hat{r}_{\min}} \bullet M = \mu_{\max}^{\Delta}
\end{aligned} \quad (58)$$

Therefore, it is possible to apply Lemma 3.1 along with (49), (54) and (58). Consequently, the inequality

$$E[V_k(\tilde{x}_k) | \tilde{x}_{k-1}] - V_{k-1}(\tilde{x}_{k-1}) \leq \mu_{\max} - \alpha_{\min} V_{k-1}(\tilde{x}_{k-1}) \quad (59)$$

is fulfilled to guarantee the boundedness of \tilde{x}_k :

$$E\{\|\tilde{x}_k\|^2\} < \frac{p_{\max}}{p_{\min}} E\{\|\tilde{x}_0\|^2\} (1 - \alpha_{\min})^k + \frac{\mu_{\max}}{p_{\min}} \sum_{i=1}^{k-1} (1 - \alpha_{\min})^i. \quad \square$$

Remark 1. Matrices Λ_k and Γ_k are unknown instrumental diagonal matrices introduced to evaluate the error which is due to the UT technique. From (58), it is shown that the stability of the algorithm depends on the magnitudes of Λ_k and Γ_k hence different Λ_k and Γ_k may change the values of α_{\min} and μ_{\max} . However, although different Λ_k and Γ_k may change the value of α_{\min} , according to (53), the last item in (49) will remain negative and the relationship shown in (59) will not be changed so long as the matrix \hat{Q}_k is positive definite. In other words, if $\hat{Q}_k \geq \hat{q}_{\min} I$ is fulfilled, an upper bound on $E[V_k(\tilde{x}_k) | \tilde{x}_{k-1}] - V_{k-1}(\tilde{x}_{k-1})$ can be obtained and the estimation error will remain bounded even for bad approximation to the nonlinear model.

Remark 2. To ensure the stability of the UKF, matrices \hat{Q}_k need to be positive definite. On the grounds of (19), as $\Delta P_{k|k-1}$ and $\partial P_{k|k-1}$ may be not positive definite matrices, an extra additive matrix ΔQ_k should be introduced to modify the UKF slightly so that $\hat{Q}_k \geq \hat{q}_{\min} I$ be satisfied always. Obviously, if ΔQ_k is sufficiently large, condition (35) can always be fulfilled. This means that the UKF can tolerate high order error introduced during the UT by enlarging the noise covariance matrix. On the other hand, the precision of the algorithm also is related to the value of \hat{Q}_k .

Remark 3

Condition (33) is closely related to the observability property of the general nonlinear system (1) as the related discussion in the next Section IV shows.

IV. NONLINEAR SYSTEM OBSERVABILITY IS SIGNIFICANT FOR THE UKF DESIGNS

In this section, the related observability property of the general nonlinear system is more closely discussed in conjunction with the stochastic stability of the UKF. For this purpose, firstly the following results on the observability rank condition [39], [13], [40] for the general nonlinear discrete-time system in the form (1) are recalled.

Lemma 4.1: The general nonlinear system given by (1) satisfies the nonlinear observability rank condition at $x_k \in R^r$, if the nonlinear observability matrix

$$U(x_k) = \begin{bmatrix} \frac{\partial h}{\partial x}(x_k) \\ \frac{\partial h}{\partial x}(x_{k+1}) \frac{\partial f}{\partial x}(x_k) \\ \frac{\partial h}{\partial x}(x_{k+r-1}) \frac{\partial f}{\partial x}(x_{k+r-2}) \frac{\partial f}{\partial x}(x_k) \end{bmatrix} \quad (60)$$

has full rank r at x_k .

For the proof of this theorem we make use of some auxiliary results. First we recall the uniform observability of linear time-varying systems [41].

Lemma 4.2: Consider time-varying matrices $\Lambda_k F_k$, $\Gamma_k H_k$, $\forall k \geq 0$ and let the observability Gramian be given by

$$M_{k+1,k} = \sum_{i=k}^{k+l} \Phi_{i,k}^T H_i^T \Gamma_i^2 H_i \Phi_{i,k} \quad (61)$$

for some integer $n > 0$ with $\Phi_{k,k} = I$ and

$$\Phi_{i,k} = \Lambda_{i-1} F_{i-1} \cdots \Lambda_k F_k \quad (62)$$

for $i > k$. Then matrices $\Lambda_k F_k$, $\Gamma_k H_k$, $\forall k \geq 0$ are said to satisfy the uniform observability condition, if there are real numbers m_{\min} , $m_{\max} > 0$ and an integer $l > 0$, such that the following inequality holds:

$$m_{\min} I \leq M_{k+1,k} \leq m_{\max} I \quad (63)$$

Lemma 4.3: Consider the measurement covariance \hat{P}_k for $\forall k \geq 0$, and let the following conditions hold:

(1) There are real numbers $q_{\min}, q_{\max}, \hat{q}_{\min}, \hat{q}_{\max}, \hat{r}_{\min}, \hat{r}_{\max} > 0$ such that the matrices Q_k , \tilde{Q}_k and \hat{R}_k are bounded by

$$q_{\min} I \leq Q_k \leq q_{\max} I \quad (64)$$

$$\hat{q}_{\min} I \leq \tilde{Q}_k \leq \hat{q}_{\max} I \quad (65)$$

$$\hat{r}_{\min} I \leq \hat{R}_k \leq \hat{r}_{\max} I \quad (66)$$

(2) Matrices $\Lambda_k F_k$, $\Gamma_k H_k$ satisfy the uniform observability condition.

(3) The initial condition matrix P_0 is positive definite.

Then there exist positive real numbers $p_{\max}, p_{\min} > 0$ such that \hat{P}_k is bounded via

$$p_{\min} I \leq \hat{P}_k \leq p_{\max} I \quad (67)$$

for every $k \geq 0$. Now it is possible to state the other main result.

Theorem 4.1: Consider the general nonlinear stochastic systems represented by (1) and the UKF as described by (2)-(9). Assume there are real numbers $q_{\min}, q_{\max}, \hat{q}_{\max}, \hat{q}_{\min}, \hat{r}_{\min}, \hat{r}_{\max} > 0$ with

$$q_{\min} I \leq Q_k \leq q_{\max} I \quad (68)$$

$$\hat{q}_{\min} I \leq \tilde{Q}_k \leq \hat{q}_{\max} I \quad (69)$$

$$\hat{q}_{\min} I \leq \hat{Q}_k \leq \hat{q}_{\max} I \quad (69)$$

$$\hat{r}_{\min} I \leq \hat{R}_k \leq \hat{r}_{\max} I \quad (70)$$

for $k \geq 0$, such that the following conditions hold:

(1) The general nonlinear system given by (1) satisfies the observability rank condition for every vector $x_k \in R^r$.

(2) The nonlinear uncertainty functions $f(\bullet)$ and $h(\bullet)$ are twice continuously differentiable with respect to their independent variables in x , and also $(\partial f / \partial x)(x) \neq 0$ holds for every vector $x_k \in R^r$.

Then the estimation error \tilde{x}_k given by (10)-(11) is exponentially bounded in the sense of mean square.

Proof: In the proof of Theorem 4.1 relies on the use of Theorem 3.1. It is shown here that the inequalities (68)-(70) and the stated Conditions 1 and 2 in Theorem 4.1 together with the observability results given by Lemmas 4.1 and 4.3 imply the Conditions (34)-(36) in Theorem 3.1. Namely, it can be seen at once that (34)-(36) and (68)-(70) coincide; in other words, inequalities (34)-(36) are satisfied if inequalities (68)-(70) are.

Suppose functions f_i , h_i are the components of f and h , respectively. Since f and h are twice differentiable for every independent variable in vector x according to Assumption 2 and R^r is compact, the Hessian matrices of f_i and h_i are bounded with respect to the spectral norm of matrices. It is therefore that constants k_f and k_h are given by

$$k_f = \max_{1 \leq i \leq q} \sup_{x \in R^r} \|Hess f_i(x)\|, \quad k_h = \max_{1 \leq i \leq m} \sup_{x \in R^r} \|Hess h_i(x)\|. \quad (71)$$

Concerning the remaining conditions of Theorem 3.1, it is sufficient to ensure these conditions hold one time-step in advance. Now it is important to notice that constants f_{\min} , f_{\max} , h_{\min} , h_{\max} , λ_{\min} , λ_{\max} , γ_{\min} , γ_{\max} , p_{\max} , p_{\min} in (29)-(33) can be chosen independently of the time k . This in turn means the boundedness of \tilde{x}_k and x_k implies the desired bounds on Λ_k , F_k , Γ_k , H_k and \hat{P}_k , which are needed. Then from Theorem 3.1 the boundedness of \tilde{x}_{k+1} for the next time step is readily obtained. By repeating this procedure again the bounds on Λ_{k+1} , F_{k+1} , Γ_{k+1} , H_{k+1} and \hat{P}_{k+1} are obtained, and therefore on \tilde{x}_{k+2} as well. Continuation of this proving strategy yields the desired result.

In order to establish the bounds on Λ_k , F_k , Γ_k , H_k , and \hat{P}_k not that the cases $0 \leq k < r$ and $k \geq r$ can be treated separately. This is due to the fact that that finite steps are needed to set up the uniform observability condition.

Firstly, the initial finite step cycle $0 \leq k < r$ is considered. By considering the boundedness of x_k , \tilde{x}_k , and therefore of \hat{x}_k , Λ_k , F_k , Γ_k , H_k , it follows that $\hat{P}_{k+1} > 0$ if $\hat{P}_k > 0$. Now taking the minimum and maximum eigenvalue of \hat{P}_k and the maximum singular value of Λ_k , F_k , Γ_k , H_k for $0 \leq k < r$, the bounds (29)-(33) for $0 \leq k < r$ are derived.

Secondly, the case of the setup steps $k \geq r$ is considered next. It should be noted that neither any eigenvalue of \hat{P}_k converges to zero nor any of the matrices Λ_k , F_k , Γ_k , H_k , \hat{P}_k diverges. The boundedness

$$p_{\min} I \leq \hat{P}_k \leq p_{\max} I \quad (72)$$

follows according to Lemma 4.3 by utilizing the boundedness of \hat{x}_i for $r \leq i \leq k$ in a given region $\|\tilde{x}_i\| \leq \varepsilon$, $\varepsilon > 0$. Moreover, the norm boundedness of Λ_k , F_k , Γ_k , H_k also follows from the continuity of $\partial f / \partial x$ and $\partial h / \partial x$, the compactness of R^r , and the fact that estimated samples $\hat{x}_k \in R^r$. By means of these arguments, Theorem 3.1 can be readily applied which terminates the proof. \square

The obtained results of this section and of the preceding one showed the estimation error of the discrete-time Unscented

Kalman Filter remains bounded provided the general nonlinear system (1) to be observed satisfies the appropriate conditions without the requirements of: (i) a sufficiently small initial estimation error; and (ii) sufficiently weak noise. The latter precisely mark are the major benefits of the new results in this paper.

V. ILLUSTRATIVE NUMERICAL EXAMPLE AND SIMULATION RESULTS

To illustrate the significance of the derived theorems and the respective conditions, in this section the UKF is applied to a relevant example system. The error behaviour of the discrete-time unscented Kalman filter in both versions, the basic UKF and the modified MUKF, is then verified by numerical simulations.

The following general nonlinear stochastic example system of the class (1) described with the model functions f and h

$$f(x_k) = \begin{bmatrix} x_{1,k} + \tau x_{2,k} \\ x_{2,k} + \tau(-x_{1,k} + (x_{1,k}^2 + x_{2,k}^2 - 1)x_{2,k}) \end{bmatrix} \quad (73)$$

$$h(x_k) = \exp(c - x_{1,k}) \quad (74)$$

is considered. From (73) and (74) it is readily calculated:

$$F_k = \frac{\partial f}{\partial x}(\hat{x}_k) = \begin{bmatrix} 1 & \tau \\ -1 + 2\hat{x}_{1,k}\hat{x}_{2,k} & 1 + (\hat{x}_{1,k}^2 + 3\hat{x}_{2,k}^2 - 1)\tau \end{bmatrix}, \quad (75)$$

$$H_k = \frac{\partial h}{\partial x}(\hat{x}_k) = \begin{bmatrix} -\exp(-\hat{x}_{1,k}) & 0 \end{bmatrix}. \quad (76)$$

This section is written in two parts up. The first one is devoted to verify the theoretical result in Section III, while the second one to verify the result in Section IV. The following initial data have been chosen: $Q_k = I\tau$, $R_k = 1/\tau$, $P_0 = I_2$, $\hat{P}_0 = I_2$, $x_0 = [0.8 \quad 0.2]^T$. The sampling time chosen is $\tau = 0.001$, and the executing time steps are $k = 10^4$. Matrices G_k and D_k as well as the initial value \hat{x}_0 have been chosen for each indicated particular case as shown in Table I.

TABLE I. INITIAL VALUES AND NOISE-WEIGHTING MATRICES FOR THE NUMERICAL SIMULATION

	<i>Small initial error and small noise</i>	<i>Large noise</i>	<i>Large initial error</i>
\tilde{x}_0	$[0.5 \quad 0.5]^T$	$[0.5 \quad 0.5]^T$	$[2.3 \quad 2.2]^T$
G_k	$\sqrt{10^{-5}} I$	$\sqrt{10^{-3}} I$	$\sqrt{10^{-5}} I$
D_k	$\sqrt{10}$	$\sqrt{10}$	$\sqrt{10}$
ΔQ_k	$diag([0.015^2 \quad 0.02^2])$	$diag([0.015^2 \quad 0.02^2])$	$diag([0.015^2 \quad 0.02^2])$ $diag([0.018^2 \quad 0.4^2])$

The numerical simulations results according to Theorem 3.1 of Section III are discussed first. In order to fulfil the

assumption of Theorem 3.1 as shown in (29)-(36), the extra matrix $\Delta Q_k = \text{diag}([0.015^2 \ 0.02^2])$ is designed by experiment and added in the UKF. Notice that deliberately a case with strong process noise and large initial error has been explored.

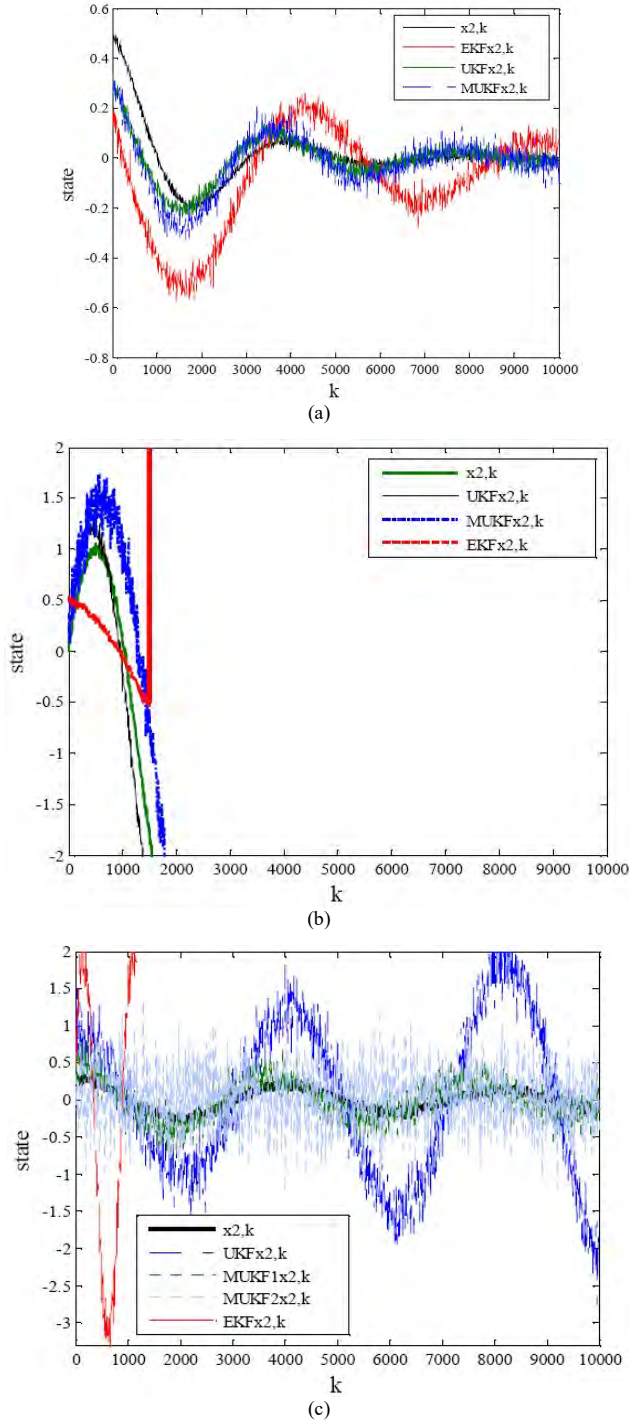


Figure 2. The state component $x_2(t)$ and its estimation with the EKF and the UKF for the example system: (a) small initial error and small noise; (b) large noise; (c) large initial error.

The relevant simulation results are depicted in Figures 2 and 3. In these figures, sample results for the unknown state $x_{2,k}$ and $\hat{x}_{2,k}$ the estimated state as well as for the estimation error are plotted versus the discrete time.

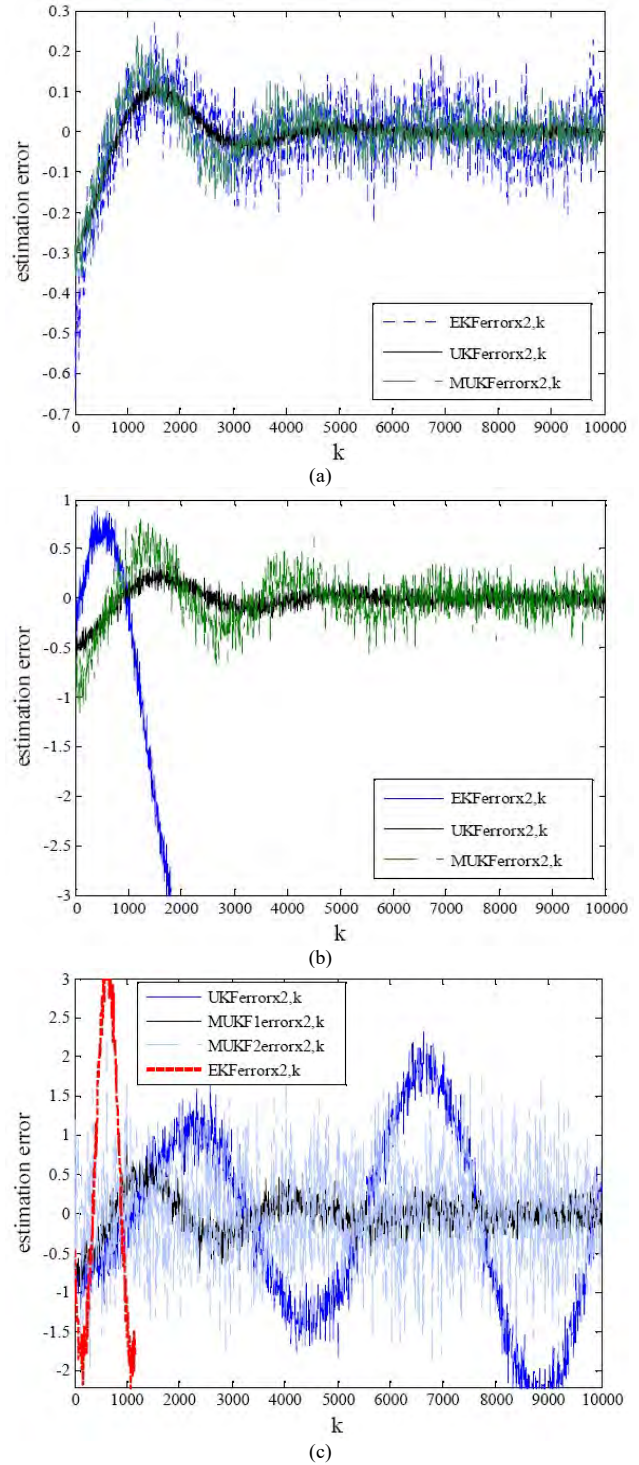


Figure 3. The estimation error $\xi_2(t)$ with the EKF and the UKF for the example system: (a) small initial error and small noise; (b) large noise; (c) large initial error.

The same simulation procedure is applied to this nonlinear model. There have been generated 10 trajectories and each of the nonlinear filters is performed for 10 times. The error standard deviations of the UKF and the EKF with different choices of $\Delta Q_k = \text{diag}([0.018^2 \quad 0.4^2])$.

It can be seen in Figures 1(a) and 2(a) that in the case of small initial error and small noise, the estimation error of the UKF and the MUKF is considerably smaller than that of the EKF. When the large noise is introduced to act (initial error kept unchanged), the estimation error of both the UKF and the MUKF remains bounded as simulation results in Figures 1(b) and 2(b) show. For the case when large initial error is taken into consideration (noise is kept unchanged), as it can be seen in Figures 1(c) and 2(c), the estimation error of EKF and UKF appear to be divergent. In contrast, the MUKF still can achieve the estimation error to remain well bounded.

The plots verify that provided a sufficiently large valued matrix ΔQ_k is chosen and added, the discrete-time MUKF appears to be rather robust. However, setting very large valued matrix ΔQ_k make the estimation error divergent for the other two filters, as shown in Figures 1(c) and 2(c). In addition, the EKF appears to be an efficient estimator for this application, but its error standard deviation obviously.

For the treatment of the original nonlinear system case we turn to the theoretical results of Section IV. To fulfil the Assumption 1 in Theorem 4.1 the general nonlinear system has to satisfy the observability rank condition for every $x_k \in R^2$. By using (60) and (78), (79) it is obtained

$$U(x_k) = \begin{bmatrix} -\exp(-\hat{x}_{1,k}) & 0 \\ -\exp(-\hat{x}_{1,k}) & -\exp(-\hat{x}_{1,k})\tau \end{bmatrix}$$

Therefore, the calculation yields $\text{rank}U(x_k) = 2$.

For $x_k \in R^2$, Assumption 1 holds. With (73), (74), and (75) it can be easily checked that Assumption 2 is also fulfilled. The simulation results are depicted in Figures 1 and 2, where sample paths for the unknown state $x_{2,k}$ and the estimated state $\hat{x}_{2,k}$ as well as for the estimation error $\tilde{x}_{2,k}$ are plotted versus k . In the case of large measurement noise or large initial error, the estimation error remains bounded, which is due to the extra additive matrix ΔQ_k . Moreover, the EKF appears to be an efficient estimator for this application, but its error standard deviation obviously.

Apparently, the simulation results on the explored example confirm that the discrete-time UKF and the appropriate conditions for the stability of the discrete-time modified UKF are all effective. Thus, the theory is well supported by the application example of a category of inherently nonlinear dynamical system under observation and control.

VI. CONCLUSIONS

The error dynamics behaviour of the Unscented Kalman Filter when applied to estimation problems for nonlinear stochastic discrete-time systems of general type has been thoroughly investigated. It was shown in Section III that the estimation error is bounded in the mean square under certain conditions, which have been derived. According to some of the standard results on the boundedness of stochastic processes, the stability of the UKF can be ensured without the requirement of small initial estimation error by means of the appropriate choice of an additive positive definite matrix ΔQ_k . However, this matrix has to be designed by empirical tests; the only guidance is that it should be sufficiently large valued. Nonetheless, if this additional positive definite matrix ΔQ_k is set too large valued then the standard deviations may become significant, which degrades the estimation quality. Therefore the design effort for matrix ΔQ_k may be seen as a trade-off between the requirements for stability and for accuracy. As shown in Section IV, the condition established in Section III can be reduced to a nonlinear observability rank condition of the plant model, which could be checked in advance. The UKF demonstrated considerably high performance under the worst initial conditions through numerical examples and simulations, sample results of which are given in the previous section.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of the P.R. of China (Grant 60274009) and Specialized Research Fund for the Doctoral Program of Higher Education (Grant 20020145007), and also in part by Ministry of Education & Science of the Republic of Macedonia (Grant, 14-3154/1-17.12.2007).

REFERENCES

- [1] R. E. Kalman, "A new approach to linear filtering and prediction problems." *ASME Transactions Pt. D Journal of Basic Engineering*, vol. 82, pp. 35-45, 1960.
- [2] R. E. Kalman and R. Bucy, "New results in linear filtering and prediction." *ASME Transactions Pt. D Journal of Basic Engineering*, vol. 83, pp. 95-108, 1961.
- [3] R. E. Kalman, "On general theory of control systems." In *Proceedings of the First IFAC World Congress*, Moscow, RU. Moscow, USSR: Academy of Sciences of the USSR, 1960.
- [4] R. E. Kalman, "Physical and mathematical mechanisms of instability in nonlinear automatic control systems." *ASME Transactions Pt. D Journal Basic Engineering*, vol. 79, pp. 553-566, 1957.
- [5] G. M. Dimirovski, N.E. Gough and S. Barnett, "Categories in systems and control theory." *International Journal of Systems Science*, vol. 8, no. 9, pp. 1081-1090, 1977.
- [6] Y.-W. Jing, C. Bing, S.-Y. Zhang and G. M. Dimirovski, "Decentralised observer based stabilization of nonlinear interconnected systems," *Control & Decision*, vol. 2, no. 3, pp. 256-259, 1997.
- [7] G. M. Dimirovski, Y.-W. Jing, "Towards hybrid soft-computing approach to control of complex systems." In *Proceedings of the 2002 IEEE International Conference on Intelligent Systems*, Varna, BG. Sofia, BG: SAI of Bulgaria and the IEEE, Piscataway, NJ, vol. 1, pp. 47-54, 2002.

- [8] J. D. Stefanovski and G. M. Dimirovski, "A new approach to static output stabilization of linear dynamic systems." *International Journal of Systems Science*, vol. 37, no. 9, pp. 643-662, 2006.
- [9] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York, NY: Academic Press, 1972.
- [10] P. K. Rajasekaran, "Optimal linear estimation of stochastic signals in the presence of multiplicative noise." *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES27, pp. 462-468, 1971.
- [11] T. J. Tarn and Y. Rasis, "Observers for nonlinear stochastic systems." *IEEE Transactions on Automatic Control*, vol. 21, pp. 441-448, 1976.
- [12] E. Yaz and A. Azemi, "Observer design for discrete and continuous nonlinear stochastic systems." *International Journal of Systems Science*, vol. 24, pp. 2289-2302, 1993.
- [13] Y. Song and J. W. Grizzle, "The extended Kalman filter as a local asymptotic observer for discrete-time nonlinear systems." *Journal of Mathematics of Systems, Estimation & Control*, vol. 5, pp. 59-78, 1995.
- [14] T. D. Kolemisevska-Gugulovska, G. M. Dimirovski, C. Popovska and N. E. Gough, "Non-linear Kalman filter in simulation modelling of lake water level dynamics." In *Proceedings of the UKCC International Conference on Control*. London, UK: IEE Publication No 455, vol. II pp. 1294-1299, June 1998.
- [15] R. Konrad, G. Stefan, Y. Engin and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter." *IEEE Transactions on Automatic Control*, vol. 44, no. 8, pp. 1636-1638, 1994.
- [16] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter." *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714-728, 1999.
- [17] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the continuous-time extended Kalman filter." *IEEE Proceedings Pt. C Control Theory & Applications*, vol. 147, no. 1, pp. 45-52, 2000.
- [18] S. J. Julier, J. K. Uhlmann and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems." In *Proceedings of the 1995 American Control Conference*, Seattle, WA, USA. Piscataway, NJ: the AACC and the IEEE, pp. 1628-1632, 1995.
- [19] S. J. Julier and J. K. Uhlmann, "A consistent, unbiased method for converting between polar and Cartesian coordinate systems." In *Proceedings of Aero Sense: The 11th International Symposium on Aerospace / Defense Sensing, Simulation and Controls*, Orlando, FL, USA. New York, NY: the AAAI, pp. 110-121, 1997.
- [20] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems." In *Proceedings of Aero Sense: The 11th International Symposium Aerospace / Defense Sensing, Simulation and Controls*, Orlando, FL, USA. New York, NY: the AAAI, pp. 54-65, 1997.
- [21] S. J. Julier, J. K. Uhlmann and H. F. Durrant-Whyte, "A new approach for the nonlinear transformation of means and covariances in filters and estimators." *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477-482, 2000.
- [22] T. Lefebvre, H. Bruyninckx and J. De Schutter, "Comment on 'A new method for the nonlinear transformation of means and covariances in filters and estimators.'" *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1406-1408, 2002.
- [23] S. J. Julier, "The scaled unscented transformation." In *Proceedings of the 2002 American Control Conference*, Anchorage, AK, USA. Piscataway, NJ: the AACC and the IEEE, pp. 4555-4559, 2002.
- [24] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation." *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401-422, 2004.
- [25] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation." In *Proceedings of Adaptive Systems for Signal Processing, Communications, and Control Symposium*. London, UK: the IEE, pp. 153-158, 2000.
- [26] B. Ristic, A. Farina, D. Benvenuti and M. S. Arulampalam, "Performance bounds and comparison of nonlinear filters for tracking a ballistic object on re-entry." *IEEE proceedings of the Radar Sonar Navigation*, vol. 150, no. 2, pp. 65-70, 2003.
- [27] Y. Jing, J. Xu, G. M. Dimirovski, and Y. Zhou, "Optimal nonlinear estimation for aircraft flight control in wind shear." In *Proceedings of the 2009 American Control Conference*, St. Louis, MO, USA. Piscataway, NJ: the AACC and the IEEE, pp. 3813-3818, June 2009.
- [28] K. Xiong, H. Y. Zhang and C. W. Chan, "Performance evaluation of UKF-based nonlinear filtering." *Automatica*, vol. 42, no. 2, pp. 261-270, 2006.
- [29] J. Xu, T. Kolemisevska-Gugulovska, X. Zhaneg, Y. Jing and G. M. Dimirovski, "UKF based nonlinear filtering for parameter estimation in linear systems with correlated noises," in *Proceedings of the 17th IFAC World Congress*, Seoul, Korea. Seoul, KO: the ICORS and the IFAC, pp. 8413-8436, July 2008.
- [30] J. Xu, G. M. Dimirovski, Y. Jing, and C. Shen, "UKF design and stability for nonlinear stochastic systems with correlated noises," in *Proceedings of the 46th IEEE Conference on Decision & Control*, New Orleans, LA, USA, pp. 6226-6231, December 2007.
- [31] J. Xu, Y. Jing, G. M. Dimirovski, and Y. Ban, "Two-stage unscented Kalman filter for nonlinear stochastic systems in the presence of unknown random bias," in *Proceedings of the 2008 American Control Conference*, Seattle, WA, USA. Piscataway, NJ: the AACC and the IEEE, pp. 3530-3535, June 2008.
- [32] Y. Zhou, J. Xu, Y. Jing and G. M. Dimirovski, "Unscented Kalman-Bucy filter for nonlinear continuous-time systems with multiple delayed measurements." in *Proceedings of the 2010 American Control Conference*, Baltimore, MD, USA. Piscataway, NJ: the AACC and the IEEE, pp. 5302-5307, June 2010.
- [33] J. Xu, S. Wang, G. M. Dimirovski, and Y. Jing, "Stochastic stability for the continuous-time unscented Kalman filter," in *Proceedings of the 47th IEEE Conference on Decision & Control*, Cancun, Yucatan, Mexico. Piscataway, NJ: the IEEE, pp. 5110-5115, December 2008.
- [34] K. Plarre and F. Bullo, "On Kalman filtering for detectable systems with intermittent observations." *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 386-390, 2009.
- [35] S. Kluge, K. Reif, and M. Brokate, "Stochastic stability of the extended Kalman filter with intermittent observations." *IEEE Transactions on Automatic Control*, vol. 55, no. 2, pp. 514-518, 2010.
- [36] L. Shi, M. Epstein and R. M. Murray, "Kalman filtering over a packet dropping network: A probabilistic perspective." *IEEE Transactions on Automatic Control*, vol. 55, no. 3, pp. 594-604, 2010.
- [37] L. Li and Y. Xia, "Stochastic stability of unscented Kalman filter with intermittent observations." *Automatica*, vol. 47, 120-128, 2011.
- [38] R. G. Agniel and E. I. Jury, "Almost sure boundedness of randomly sampled systems." *SIAM Journal of Control*, vol. 19, pp. 372-384, 1971.
- [39] H. Nijmeijer, "Observability of autonomous discrete time nonlinear systems: A geometric approach." *International Journal of Control*, vol. 36, pp. 867-873, 1982.
- [40] E. D. Sontag, "On the observability of polynomial systems I: Finite time problems." *SIAM Journal of Control & Optimization*, vol. 17, pp. 139-151, 1979.
- [41] B. D. O. Anderson and J. B. Moore, "Detectability and stabilizability of time-varying discrete-time linear systems." *SIAM Journal of Control & Optimization*, vol. 19, pp. 20-32, 1981.

Complex Multi-networks with Faulty Inter-network Connections: Synchronization via Novel Pinning-node Control*

To the loving memory of late Prof. Siying Zhang, Academician of the Chinese Academy of Sciences

Yuanwei Jing, Guanrong Chen, Peng Shi, Georgi M. Dimirovski

Abstract—A synchronization control is derived by using the pinning control theory in this paper. A sufficient stability condition and a pinning control scheme are given by which the nodes in the same network can achieve synchronization via Lyapunov's general stability theory of nonlinear systems. A multi-network system structure consisted of two coupled dynamical networks is explored as an illustrative simulation example. Numerical simulation results demonstrate the proposed controller can achieve stabilized synchronization for the multi-network system despite the fault. In addition, it is shown the especially designed pinning control scheme is more effective than the randomly created pinning control scheme.

Abstract—Dynamical multi-networks; control; faulty inter-network connections; nonlinear nodes systems; pinning control theory; stabilized synchronization.

I. INTRODUCTION

Nowadays, multi-networks can be seen everywhere in the real world. In-depth studies of complex networks give real insight into the interdependence, cooperation and/or competition between real-world networks. It appears, however, these have been solely anticipated as in [1] until clearly seen in the 2001 study of Strogatz [18] and subsequently by Lü et al. [11]. Nonetheless, multi-networks existed widely in both nature and human society, e.g. communication and transportation networks, power grids, societal and social relations networks to name a few. Naturally, it is necessary to understand deeper the topology, property and functionality of such complex networks [2, 3, 15]. Therefore, the study of multi-networks bears important significance for science and technology as well as society developments. Recently developed pinning control theory by

Chen and his fellow coworkers [4, 11], which include several feasible strategies, has been well established to solve synchronization control problems in complex dynamical networks. Typically in most of the literature there is assumed nodes in different communities (sector regions) have the same dimensions when investigating both cluster and network synchronizations. In general, the nodes in multi-networks can be quite different. As shown in [11, 12], failure of nodes in a network not only affect its own network but also affects all other networks through the interconnection nodes in multi-networks. And the change in another of the networks will affect that particular network, and this process may occur repeatedly, which results in multiple networks to lose their synchronization. Therefore, it is of considerable practical and also theoretical significance, to study the inter-network coupling faults in order to retain the synchronization of the network and reduce the impact of the network connection failures to the entire network.

In [23], which was aiming at studying the heterogeneous community structure of complex networks, the synchronizing controller is designed to achieve clustering synchronization. In [3], a sufficient condition for the synchronization of network is given, and the complex network is brought into the state by designing pinch controller without assuming the external coupling matrix was irreducible and symmetric. In [6], [13], [14], [19] and [25], nonlinear coupling network is considered, and the criterion of cluster synchronization is obtained by pinned the *inter-act* nodes and the *intra-act* nodes with zero input. Furthermore in [14], [22], [24], the coupling between the inter-act nodes and intra-act nodes for the complex networks encompassing multiple communities is represented separately hence the model is more realistic. In [19], [24], different dimensions of nodes in different network are taken into account and a dimension matching matrix to overcome different dimensions.

Recently article [15] contributed a modified multi-network mathematical model that takes into account for the dynamic equations of nodes in different networks, and thus distinguishes inter-acting and intra-acting networks for the multi-network coupling system. The next section presents an outline of this model. This follow up paper is based on that article by Ren and co-authors [15]. The controller, designed by using a scheme from theory of pinning control, via obtaining sufficient conditions for stabilized synchronization of the multi-network. Novel scheme of node selection and the validity correctness of conclusions have been verified by multiple computer simulations. The fault problem among

*This research has been generously supported by Chinese National Natural Science Foundation (grants, 61473073; 61104074) during a couple of years, and also in part by Science Fund for SCIE articles in 2020 of Dogas University.

Yuanwei Jing (ywjjing@mail.neu.edu.cn) is with College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, P. R. China.

Guanrong Chen (eegchen@cityu.edu.hk) is with the City University of Hong Kong, Hong Kong SAR, P.R. China

Peng Shi (peng.shi@adelaide.edu.au) is with the College of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, SA 5005, Australia.

Georgi M. Dimirovski (dimir@feit.ukim.edu.mk) is with the Doctoral School of Faculty of Electrical Engineering and Information Technologies (FEIT), SSs Cyril and Methodius University, Karpos 2, 18 Rugjer Boskovik Str., MK-1000 Skopje, R. N. Macedonia; Correspondence Author.

different networks is also considered and some future research is pointed out in the concluding remarks.

II. ON MULTI-NETWORK REPRESENTATION ESSENTIALS VERSUS PINNING CONTROL STRATEGY

Consider r dynamical networks, suppose that the k -th network is composed of N_k nodes which is n_k dimension. Note $i=1,2,\dots,N_k, k=1,2,\dots,r$, i -th node in the k -th network is described by

$$\begin{cases} \dot{x}_i^k(t) = f^k(x_i^k(t)) + \sum_{j=1, j \neq i}^{N_k} a_{ij}^{kk} (x_j^k(t) - x_i^k(t)) \\ + \sum_{l=1, l \neq k}^r \sum_{j=1}^{N_l} (m_{ij}^{kl}(t) + a_{ij}^{kl}) (\Gamma^{kl} x_j^l(t) - x_i^k(t)). \end{cases} \quad (1)$$

Here $t \in R^+$, $x_i^k(t) = (x_{i1}^k(t), x_{i2}^k(t), \dots, x_{in_k}^k(t))^T \in R^{n_k}$ is the state vector of node i in the k -th network. $f^k(\cdot): R^{n_k} \rightarrow R^{n_k}$ is a nonlinear vector-valued function denoting dynamic and behavioral characteristics of the nodes. $A_{kk} = (a_{ij}^{kk}) \in R^{N_k \times N_k}$ is the coupling configuration matrix denoting the topological structure and the coupling strength of the k -th network. If there is a connection between node i and node j ($i \neq j$), $a_{ij}^{kk} > 0$; otherwise, $a_{ij}^{kk} = 0$. $A_{kl} = (a_{ij}^{kl}) \in R^{N_k \times N_l}$ is the external connection denoting the topological structure and the coupling strength from the l -th to the k -th network. If there is a connection between node i and node j , $a_{ij}^{kl} > 0$; otherwise, $a_{ij}^{kl} = 0$. $m_{ij}^{kl}(t)$ is the fault signal denoting connection problem between the l -th and the k -th network. Γ^{kl} is the dimension-transformation matrix.

$$\Gamma^{kl} = \begin{cases} \begin{bmatrix} I_{n_l} \\ 0 \end{bmatrix}_{n_k \times n_l} & n_k > n_l \\ \begin{bmatrix} I_{n_k} & 0 \end{bmatrix}_{n_k \times n_l} & n_k \leq n_l \end{cases} \quad (2)$$

It is further assumed the network failures occurred between the l -th and k -th network in the multi-network (1). Let in addition suppose the network (1) possesses the property

$$a_{ii}^{kk} = - \sum_{j=1, j \neq i}^{N_k} a_{ij}^{kk} - \sum_{l=1, l \neq k}^r \sum_{j=1}^{N_l} (m_{ij}^{kl}(t) + a_{ij}^{kl}), i=1,2,\dots,N_k, k=1,\dots,r.$$

Then, it can be shown network (1) is described as follows:

$$\begin{cases} \dot{x}_i^k(t) = f^k(x_i^k(t)) + \sum_{j=1}^{N_k} a_{ij}^{kk} x_j^k(t) + \sum_{l=1, l \neq k}^r \sum_{j=1}^{N_l} (m_{ij}^{kl}(t) + a_{ij}^{kl}) \Gamma^{kl} x_j^l(t) \\ i=1,2,\dots,N_k, k=1,2,\dots,r. \end{cases} \quad (3)$$

The adopted definitions and assumptions are presented next as follows.

Definition 1: Suppose equations (4) are established for each node, then network (1) has achieved cluster synchronization:

$$\lim_{t \rightarrow \infty} \|x_i^k(t) - s^k(t)\| = 0, i=1,2,\dots,N_k, k=1,2,\dots,r, \quad (4)$$

where $s^k(t)$ is a solution of an isolated node in the k -th network. Here $\dot{s}^k(t) = f^k(s^k(t))$, and $s^k(t)$ may be an equilibrium point, a periodic orbit, or even a chaotic orbit.

Definition 2: Suppose $A = (a_{ij}) \in R^{m \times n}$, $B = (b_{ij}) \in R^{p \times q}$, then

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} \in R^{mp \times nq},$$

is a Kronecker product of A and B , and it satisfies:

1/. $(A+B) \otimes C = A \otimes C + B \otimes C$; 2/. $(A \otimes B)^T = A^T \otimes B^T$.

Assumption 1: There exists a constant $\theta_k > 0$, such that

the nonlinear function $f^k(\cdot)$ satisfies

$$(x-y)^T (f^k(x) - f^k(y)) \leq \theta_k (x-y)^T (x-y), \quad \forall x, y \in R^{n_k} \quad (5)$$

Assumption 2: Coupling fault among networks and changes of coupling connection are bounded. Namely,

$0 < \underline{m}_{ij}^{kl} \leq m_{ij}^{kl}(t) \leq \bar{m}_{ij}^{kl}$, $m_{ij}^{kl}(t)$ satisfies the update rate:

$$\dot{m}_{ij}^{kl}(t) = -\beta e_i^k(t)^T \Gamma^{kl} s^l(t), \quad i=1,2,\dots,N_k,$$

$j=1,2,\dots,N_l, k,l=1,2,\dots,r$, β is a positive constant,

\underline{m}_{ij}^{kl} and \bar{m}_{ij}^{kl} are the upper and lower bound of $m_{ij}^{kl}(t)$.

Assumption 3: Define two sets $V = \{i_1, i_2, \dots, i_N\}$ and

$V_{pin} = \{i_1, i_2, \dots, i_l\}$ as the sets of total nodes and of the

selected pinned nodes for the controlled network (3), respectively. All nodes in $V \setminus V_{pin}$ are accessible from the

pinned node set V_{pin} , i.e., for any node $i \in V \setminus V_{pin}$, we can

always find a node $j \in V_{pin}$, such that there is a directed path from node j to node i .

Remark 1: Assumption 3 shows that there are no isolated nodes in the intra-acting networks.

The synchronization controller is designed via the pinning control principle and scheme. Thus, the i -th node in the k -th network is described by

$$\begin{cases} \dot{x}_i^k(t) = f^k(x_i^k(t)) + \sum_{j=1}^{N_k} a_{ij}^{kk} x_j^k(t) + \sum_{l=1, l \neq k}^r \sum_{j=1}^{N_l} (m_{ij}^{kl}(t) + a_{ij}^{kl}) \Gamma^{kl} x_j^l(t) + u_i^k(t) \\ i=1,2,\dots,N_k, k=1,2,\dots,r. \end{cases} \quad (6)$$

Inspired by the recent advances in pinning control theory [10], [15], [17] pinning control scheme with the following controllers is adopted:

$$\begin{aligned} u_i^k(t) &= -\sum_{j=1}^{N_k} h_{ij}^{kk}(t) s^k(t) - d_i^k e_i^k(t), \\ \dot{h}_{ij}^{kk}(t) &= e_i^k(t)^T s^k(t), i \in \phi_k, \\ u_i^k(t) &= -d_i^k e_i^k(t), i = \tilde{\phi}_k - \phi_k, \end{aligned} \quad (7a)$$

Here $h_{ij}^{kk}(t)$ is one-dimensional variable, d_i^k is feedback control gain satisfying

$$\begin{cases} d_i^k > 0, i = 1, \dots, l_k, \\ d_i^k = 0, i = l_k + 1, \dots, N_k, \end{cases} \quad (7b).$$

In addition, for the case of multi-networks, the concept of inter-act node and intra-act node is important. Node i is called the *inter-act node* if i belongs to ϕ_k , while i is said the *intra-act node* if i belongs to $\tilde{\phi}_k - \phi_k$. It is thus implied that inter-act node can receive information from the other clusters/networks where as intra-act node can only exchange information among nodes in the same cluster/network. Therefore, let now suppose the synchronization error of node i in the k -th network is described as:

$$e_i^k(t) = x_i^k(t) - s^k(t). \quad (8)$$

The error system for complex network (6) can be obtained as

$$\begin{aligned} \dot{e}_i^k(t) &= f^k(x_i^k(t)) - f^k(s^k(t)) + \sum_{j=1}^{N_k} a_{ij}^{kk} e_j^k(t) + \sum_{l=1, l \neq k}^r \sum_{j=1}^{N_l} (m_{ij}^{kl}(t) + a_{ij}^{kl}) \Gamma^{kl} e_j^l(t) \\ &+ \sum_{j=1}^{N_k} a_{ij}^{kk} s^k(t) + \sum_{l=1, l \neq k}^r \sum_{j=1}^{N_l} (m_{ij}^{kl}(t) + a_{ij}^{kl}) \Gamma^{kl} s^l(t) + u_i^k(t) \end{aligned} \quad (9)$$

where $i = 1, 2, \dots, N_k, k = 1, 2, \dots, r$.

Lemma 1 [5]: If node i is an intra-act node, namely, $i \in \tilde{\phi}_k - \phi_k$, then the following result holds

$$\sum_{j=1}^{N_k} a_{ij}^{kk} s^k = 0, \sum_{l=1, l \neq k}^r \sum_{j=1}^{N_l} (m_{ij}^{kl}(t) + a_{ij}^{kl}) \Gamma^{kl} s^l(t) = 0, \quad (10)$$

for all $i = 1, 2, \dots, N_k, k = 1, 2, \dots, r$.

Lemma 2 [5]: For a symmetric matrix $M = (m_{ij})_{N \times N}$ and diagonal $D = \text{diag}(d_1, \dots, d_q, 0, \dots, 0)_{N \times N}$, $i = 1, 2, \dots, q$ ($1 \leq q \leq N$) let

$$M - D = \begin{bmatrix} A - \tilde{D} & C \\ C^T & M_q \end{bmatrix} \quad (11)$$

where M_q is the minor matrix of M by removing its first q row-column pairs, and let $A = (a_{ij})_{q \times q}$, $a_{ij} = a_{ji} = m_{ij}$, $i, j = 1, 2, \dots, q$, $C = (c_{ij})_{q \times (N-q)}$, $c_{ij} = m_{ij}$, $i = 1, 2, \dots, q$, $j = q+1, \dots, N$, $M_q = (m_{qj})_{(N-q) \times (N-q)}$, $m_{qj} = m_{qj} = m_{i+q, j+q}$, $i, j = 1, 2, \dots, N-q$. If furthermore

$d_i > \lambda_{\max}(A - CM_q^{-1}C^T)$, then $M - D < 0$ is equivalent to $M_q < 0$.

Lemma 3 [8] (Gerschgorin Disc Theorem): Let $A = (a_{ij})_{n \times n}$ be a complex matrix and let $R_i(A) = \sum_{j=1, j \neq i}^n |a_{ij}|$, $1 \leq i \leq n$

which denote the deleted absolute row sums of A . Then all the eigenvalues of A are located in the union of n discs

$$G(A) = \bigcup_{i=1}^n \{z \in \mathbb{C} : |z - a_{ii}| \leq R_i(A)\}. \quad (12)$$

Remark 2: If A is a real symmetrical matrix, it follows from Lemma 1 that the eigenvalues λ of A satisfy

$$\min_{1 \leq i \leq n} (a_{ii} - R_i(A)) \leq \lambda \leq \max_{1 \leq i \leq n} (a_{ii} + R_i(A)) \quad (13)$$

Lemma 4 [9]: Assume that A, B are $N \times N$ Hermitian matrices, and let $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_N, \beta_1 \geq \beta_2 \geq \dots \geq \beta_N$, and $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_N$ be the eigenvalues of A, B , and $A + B$, respectively. Then, the inequality relationship

$$\alpha_i + \beta_N \leq \gamma_i \leq \alpha_i + \beta_1 \quad (14)$$

holds true.

III. MAIN NOVEL RESULTS

A sufficient stability conditions for multi-networks with coupling fault among networks under pinning control strategy derived in [7, 11, 13, 14, 17] based on Lyapunov stability theory [8, 9, 12, 16, 25].

Theorem 1: Under the Assumptions 1-3, the controlled complex network (6) with the controller (7) with the below given pinning node scheme can reinforce and realize the desired synchronization if

$$\Theta - D + A^s < 0, \quad (11)$$

where

$$D = \text{diag}(d_1^1 I_{n_1}, \dots, d_{N_1}^1 I_{n_1}, d_1^2 I_{n_2}, \dots, d_{N_2}^2 I_{n_2}, \dots, d_1^r I_{n_r}, \dots, d_{N_r}^r I_{n_r}),$$

$$\Theta = \text{diag}(\theta_1 I_{N_1 \times n_1}, \dots, \theta_r I_{N_r \times n_r}), A^s = \frac{A + A^T}{2},$$

$$A = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \dots & \tilde{A}_{1r} \\ \tilde{A}_{21} & \tilde{A}_{22} & \dots & \tilde{A}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{A}_{r1} & \tilde{A}_{r2} & \dots & \tilde{A}_{rr} \end{bmatrix}, \tilde{A}_{kk} = \bar{A}_{kk} \otimes I_{n_k}, \tilde{A}_{kl} = \bar{A}_{kl} \otimes \Gamma^{kl},$$

$(\bar{A}_{kl})_{ij} = \bar{m}_{ij}^{kl} + a_{ij}^{kl}, (\bar{A}_{kk})_{ij} = a_{ij}^{kk}, i, j = 1, 2, \dots, N_k, k = 1, 2, \dots, r$ is satisfied.

Proof. See [15].

The Specifically Designed Pinning-Node Scheme

(1) All *inter-act* nodes are subject to control and the total number is denoted as q_0 .

(2) The nodes with diagonal elements of matrix G that are

bigger than zero are controlled in all the remaining nodes, and the total number is denoted as q_1 .

(3) All remaining nodes are sorted by the row sum of matrix descending order, and let $q = q_0 + q_1$.

(4) Calculate whether the matrix G_q is negative definite. If G_q is not negative definite, let $q = q + 1$, go to step (3), otherwise, by (27) to calculate feedback control gains. Stop.

IV. SOME RESULTS ON A BENCHMARK MULTI-NETWORK

A multi-network of complex dynamic systems with two directional networks is taken as a benchmark example for numerical simulation in order to clearly show the theoretical results presented here. The topological structure of this multi-network is shown in Figure 1.

The first network contains six nodes, which are defined to represent the four-dimensional hyper-chaotic Lorenz systems. The second network is defined as having four nodes that represent three-dimensional Rossler dynamic systems. Thus, Figure 1 depicts a fairly complex multi-network (network of networks) in which inter-linking faults may occur.

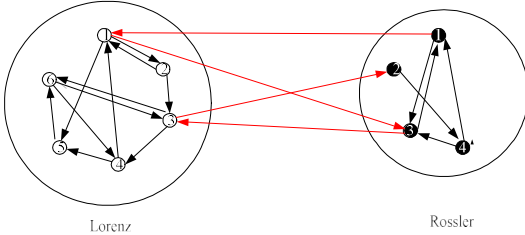


Fig.1 Topological structure of the considered multi-network

The coupling matrix of the network is given as follows:

$$A = \begin{bmatrix} -80 & 30 & 0 & 0 & 40 & 0 & 0 & 0 & 10 & 0 \\ 10 & -60 & 50 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -110 & 20 & 0 & 80 & 0 & 10 & 0 & 0 \\ 70 & 0 & 0 & -90 & 20 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -60 & 60 & 0 & 0 & 0 & 0 \\ 0 & 0 & 70 & 30 & 0 & -100 & 0 & 0 & 0 & 0 \\ \hline 10 & 0 & 0 & 0 & 0 & 0 & -50 & 0 & 40 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -60 & 0 & 60 \\ 0 & 0 & 20 & 0 & 0 & 0 & 60 & 0 & -80 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 50 & 0 & 40 & -90 \end{bmatrix}$$

The hyper-chaotic Lorenz systems are described by means of the model

$$\begin{aligned} \dot{x}_1(t) &= 10(x_2(t) - x_1(t)) + x_4(t), \\ \dot{x}_2(t) &= 28x_1(t) - x_2(t) - x_1(t)x_3(t), \\ \dot{x}_3(t) &= x_1(t)x_2(t) - 8/3x_3(t), \\ \dot{x}_4(t) &= -x_1(t)x_3(t) + 13/10x_3(t). \end{aligned}$$

And the Rossler systems are described by means of the model

$$\begin{aligned} \dot{x}_1(t) &= -x_2(t) - x_3(t) \\ \dot{x}_2(t) &= x_1(t) + 0.2x_2(t) \\ \dot{x}_3(t) &= x_1(t)x_3(t) - 5.7x_3(t) + 0.2 \end{aligned}$$

Since chaotic attractors of the hyper-chaotic Lorenz systems and the Rossler systems appear in a bounded region the computer simulations are feasible. Thus by numerical simulations, one can find that there exist some constants for the first network $M_{11} = 25, M_{12} = 25, M_{13} = 45, M_{14} = 180$, such as $|s_{11}| \leq M_{11}, |s_{12}| \leq M_{12}, |s_{13}| \leq M_{13}, |s_{14}| \leq M_{14}$.

For Assumption 1 the parameters can be calculated using the following method to the first network as an example. One has to consider

$$\begin{aligned} & (x_i - s^1)^T (f^1(x_i) - f^1(s^1)) = \\ & = e_i^T (10e_{i2} - 10e_{i1} + e_{i4}, 28e_{i1} - e_{i2} - x_{i1}x_{i3} + s_{11}s_{13}, x_{i1}x_{i2} - \\ & \quad - s_{11}s_{12} - 8/3e_{i3}, 1.3e_{i4} - x_{i1}x_{i3} + s_{11}s_{13})^T \\ & \leq -10e_{i1}^2 - e_{i2}^2 - 8/3e_{i3}^2 + 1.3e_{i4}^2 + (38 + M_{13})|e_{i1}e_{i2}| + M_{12}|e_{i1}e_{i3}| \\ & \quad + (1 + M_{13})|e_{i1}e_{i4}| + M_{11}|e_{i3}e_{i4}| \\ & \leq \left(-10 + \frac{\varepsilon_1(38 + M_{13})}{2} + \frac{\varepsilon_2M_{12}}{2} + \frac{\varepsilon_3(1 + M_{13})}{2}\right)e_{i1}^2 + \left(-1 + \frac{38 + M_{13}}{2\varepsilon_1}\right)e_{i2}^2 \\ & \quad + \left(-\frac{8}{3} + \frac{M_{12}}{2\varepsilon_2} + \frac{\varepsilon_4M_{11}}{2}\right)e_{i3}^2 + \left(1.3 + \frac{1 + M_{13}}{2\varepsilon_3} + \frac{M_{11}}{2\varepsilon_4}\right)e_{i4}^2 \end{aligned}$$

where $\varepsilon_i (i = 1, 2, 3, 4)$ are arbitrary positive constants. By choosing $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = (1, 0.4, 0.6, 1.2)$, one could obtain $\theta_1 = 50.3$ such that Assumption 1 holds. Similarly, for the second network one could choose $\theta_2 = 38.1$.

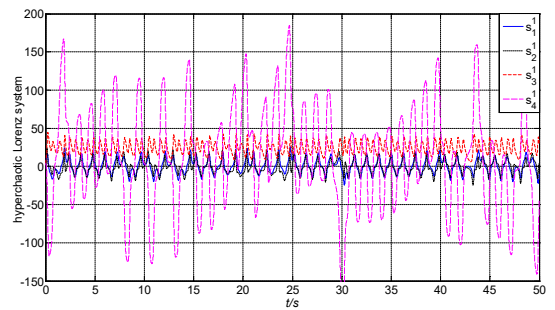


Fig.2 State evolution curves of the hyper-chaotic Lorenz systems with $s_{10} = (7, 6, 5, 4)^T$ into the given multi-network

Initial values of the states of the hyper-chaotic Lorenz systems and of the Rossler system are $x_0 = (7, 6, 5, 4)^T$ and $x_0 = (-10, 0, 10)^T$, respectively. State evolution curves of hyper-chaotic Lorenz systems are shown in Figure 2 while the state evolution curves of Rossler systems are shown in Figure 3. They are the final synchronization state of those two directional networks within the given multi-network.

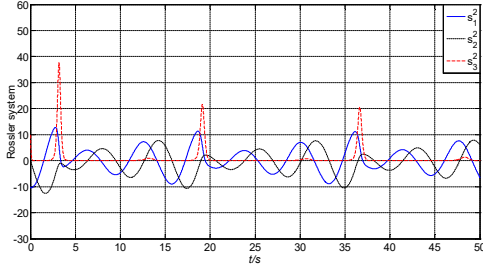


Fig.3 State evolution curves of the Rossler with $s_{20} = (-10, 0, 10)^T$ into the given multi-network

The evolution curves of the state synchronization errors for the nodes of the two types of complex networks are shown in Figure 4 and Figure 5 for the case study without the controller being applied. As can be seen from the figures, only on the grounds of the coupling effect, each node in the network cannot achieve synchronization. Thus it appears, the proposed pinning control strategy is thus indispensable if the synchronization within the multi-network is to be achieved.

A. Simulation results under specifically pinning scheme

Considered coupling fault among networks is emulated by means of step signal in the multi-network model here. By choosing $m_{13}^{12}(t) = 10 * \varepsilon(t-1)$, $m_{32}^{12}(t) = 15 * \varepsilon(t-1)$, $m_{11}^{21}(t) = 20 * \varepsilon(t-1)$, $m_{33}^{21}(t) = 25 * \varepsilon(t-1)$, the respective fault curve is shown in Figure 6.

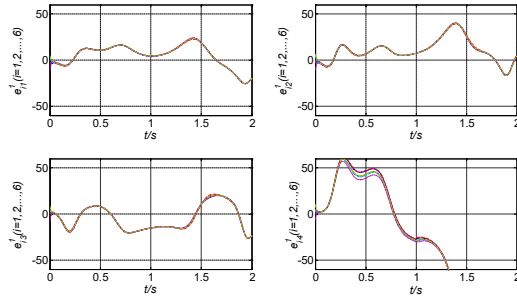


Fig.4 Error curve $e_i^1 \in R^4, i=1, 2, \dots, 6$ of the first uncontrolled network in the multi-network

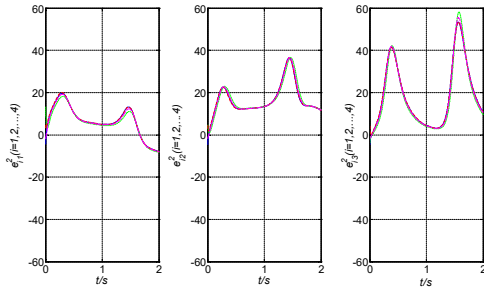


Fig.5 Error curve $e_i^2 \in R^3, i=1, 2, \dots, 4$ of the second uncontrolled network in the multi-network

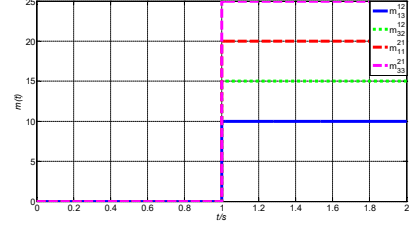


Fig.6 Evolution curve emulating the inter-network fault

Substituting the simulation values into (2.21) and (2.22), and the nodes are selected according to the node selection scheme:

- (1) The inter-act nodes 1 and 3 in the first network and the inter-act nodes 1 and 3 in the second network are controlled. The total number of inter-act nodes is denoted as $q_0 = 4$.
- (2) In the remaining nodes, the node that diagonal element of the matrix G is more than zero is controlled, which is the node 6 in first network. The number of nodes is $q_1 = 1$.
- (3) All remaining nodes are sorted by the row sum of matrix descending order, let $q = q_0 + q_1$.
- (4) G_q is negative definite through calculating when the nodes 1, 3 and 6 in the first network and the nodes 1 and 3 in the second network are controlled. Theorem 1 is satisfied by (27) to calculate feedback control gains, that are

$$\begin{cases} d_i^1 = 200, i=1, 3, 6 \text{ and } d_i^2 = 200, i=1, 3 \\ d_i^1 = 0, i=2, 4 \\ d_i^2 = 0, i=2, 4 \end{cases}$$

Under this specifically derived pinning control scheme, the state error curves of the nodes of the first network and the second network are shown in Figure 7 and Figure 8.

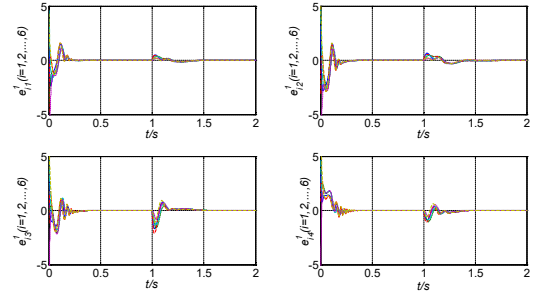


Fig.7 Error curve $e_i^1 \in R^4, i=1, 2, \dots, 6$ of the first controlled network with the proposed special pinning control strategy

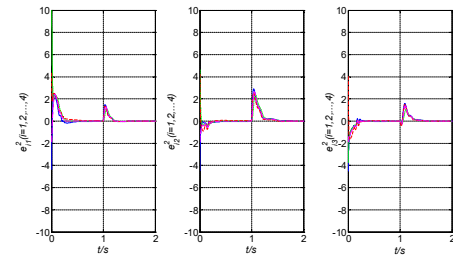


Fig.8 Error curve $e_i^2 \in R^3, i=1, 2, \dots, 4$ of the second controlled network with the proposed special pinning control strategy

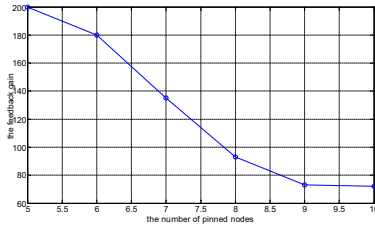


Fig. 9 Relationship between the feedback gain and the number of pinning nodes taken in

It can be readily seen from these figures with simulation results the nodes in the first network reach synchronization about 0.5 s before without failure in the role of the pinning controller. The first network reaches synchronized state again about 1.7 s when the failure is applied at first second node. The second network nodes reach synchronization about 0.6 s without failure before. The second network reaches synchronized state again about 1.8 s when the failure is applied at the first second. Relationship between feedback gain and the number of pinned nodes taken according to Eq. (27) can be calculated (Fig. 9). Apparently, increase of the number of feedback nodes causes feedback gains decrease.

V. CONCLUDING REMARKS

In this follow up paper to [15] a novel synchronization controller was designed by using the theory of pinning control scheme [7-9, 11, 12, 20-22], for which sufficient conditions for multi-network synchronization are derived. The selection node scheme is given and the correctness is verified by simulations of the Lorenz-Rossler benchmark multi-network. An innovated mathematical model of multi-networks that takes into account the dynamic equations of nodes in different networks was used. It involves distinction of inter-acting and intra-acting networks relative to multi-network system coupling structure. The dimension matching matrices are used to deal with the different dimensions of nodes. Through the disc theorem, the selection scheme of pinned node is constructed and the formula of feedback gain is given. The problem of permanent fault of coupled nodes among different networks was also observed. The future research is envisaged into exploring traffic flow prediction [5] within such inter-network faulty multi-networks.

ACKNOWLEDGMENT

Georgi Dimirovski acknowledges crucial contributions by Tao Ren and his collaborators [15] to this research topic and useful consultancy in writing this paper up.

REFERENCES

- [1] G. M. Dimirovski, N. E. Gough, S. Barnett, "Categoris in systems and control theory." *International Journal of Systems Science*, vol. 8, no. 9, pp. 1081–1090, 1977.
- [2] T. Dimitrova, L. Kocarev, "Graphical models over heterogeneous domains and multi-level networks." *IEEE Access*, vol. 6, pp. 69682–69701, 2018.
- [3] T. P. Chen, X. W. Liu, W. L. Lu, "Pinning complex networks by a single controller." *IEEE Transactions on Circuits and Systems Pt. I: Regular Papers*, vol. 54, no. 6, pp. 1317–1326, 2007.
- [4] Z. Fan, G. Chen, "Pinning control of scale-free complex networks." In *Proceedings of the IEEE International Symposium on Circuits and Systems*. Piscatawy, NJ: The IEEE, 2005, vol. 1, pp. 284–287.
- [5] Y. Han, Y. Jing, K. Li, G. M. Dimirovski, "Network traffic prediction using variational mode decomposition and multi-reservoir Echo state network." *IEEE Access*, vol. 7, pp. 138364–138377, 2019.
- [6] C. Hu, H. J. Jiang, "Cluster synchronization for directed community networks via pinning partial schemes." *Chaos, Solitons & Fractals*, vol. 45, no. 11, pp. 1368–1377, 2012.
- [7] C. Li, G. Chen, "Synchronization in general complex dynamical networks with coupling delays." *Physica A*, vol. 343, pp. 263–278, 2004.
- [8] X. Li, X. Wang, G. Chen, "Pinning a complex dynamical network to its equilibrium." *IEEE Transactions on Circuits and Systems Pt. I: Regular Papers*, vol. 51, no. 10, pp. 2074–2087, 2004.
- [9] Z. Li, G. Chen, "Global synchronization and asymptotic stability of complex dynamical networks." *IEEE Transactions on Circuits and Systems Pt. I: Regular Papers*, vol. 53, no. 1, pp. 28–33, 2006.
- [10] R. Q. Lu, W. W. Yu, J. H. Lü, Yu X. H., "Synchronization on complex networks of networks." *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 21, pp. 2100–2118, 2014.
- [11] J. Lü, X. Yu, G. Chen, D. Cheng, "Characterizing the synchronizability of small-world dynamical networks." *IEEE Transactions on Circuits and Systems Pt. I: Regular Papers*, vol. 51, no. 4, pp. 787–796, 2004.
- [12] J. Lü, G. Chen, "A time-varying complex dynamical network model and its controlled synchronization criteria." *IEEE Transactions on Automatic Control*, vol. 50, no. 6, pp. 841–846, 2005.
- [13] Q. Ma, J. W. Lu, "Cluster synchronization for directed complex dynamical networks via pinning control." *Neurocomputing*, vol. 101, pp. 354–360, 2013.
- [14] T. Ren, G. M. Dimirovski, S. Liu, Q. Zhang, "Cluster synchronization of nonlinearly coupled directed network via pinning control strategy." In: *Proceedings of the 15th IEEE International Conference on Control and Automation*, IEEE-ICCA Piscatawy, NJ and Singapore, SG: CSC Singapore and the IEEE, July 2019, pp. 1–8.
- [15] T. Ren, S. X. Sun, R. R. Wang, X. Y. Cheng, G. M. Dimirovski, "Synchronization for multi-networks with two types of inter-network coupling faults: Pinning control effects." *IET Control Theory & Applications*, vol. 14, is. 11, pp. 1497–1507, July 2020.
- [16] D. D. Siljak, "Dynamic graphs." *Nonlinear Analysis: Hybrid Systems*, vol. 2, pp. 544–567, 2008.
- [17] Q. Song, J. D. Cao, "On pinning synchronization of directed and undirected complex dynamical networks." *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 3, pp. 672–680, 2010.
- [18] S. H. Strogatz, "Exploring complex networks." *Nature*, vol. 410, no. 8, pp. 268–276, March 2001.
- [19] G. Wang, Y. Shen, "Cluster synchronization of directed complex dynamical networks with non-identical nodes via pinning control." *International Journal of Systems Science*, vol. 44, no. 9, pp. 1557–1586, 2013.
- [20] X. F. Wang, G. Chen, "Synchronization in scale-free dynamical networks: robustness and fragility." *IEEE Transactions on Circuits and Systems Pt. I: Regular Papers*, vol. 49, no. 4, pp. 54–62, 2002.
- [21] X. F. Wang, G. Chen, "Complex networks: small-world, scale-free and beyond." *IEEE Circuits and Systems Magazine*, vol. 3, no. 1, pp. 6–20, 2003.
- [22] X. F. Wang, G. Chen, "Pinning control of scale-free dynamical networks." *Physica A*, vol. 324, pp. 166–178, 2004.
- [23] Z. Y. Wu, "Cluster synchronization in colored community network with different order node dynamics." *Communications in Nonlinear Science & Numerical Simulation*, vol. 19, no. 4, pp. 1079–1087, 2014.
- [24] H. Y. Yao, S. G. Wang, "Cluster projective synchronization of complex networks with nonidentical dynamical nodes." *Chinese Physics B*, vol. 21, no. 11, art. 110506, 2012.
- [25] J. Zhou, J. Lu, J. Lü, "Adaptive synchronization of an uncertain complex dynamical network." *IEEE Transactions on Automatic Control*, vol. 51, no. 4, pp. 652–656, 2006.



ETAI 4: E-HEALTH

Insieme: A Unifying Electronic and Mobile Health Platform

Primož Kocuvan, Erik Dovgan, Tine Kolenik, Matjaž Gams
Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia
{primož.kocuvan, erik.dovgan, tine.kolenik, matjaz.gams}@ijs.si

Abstract—This paper describes the current state of the electronic and mobile health platform, Insieme. The first prototype of the platform is already finalized showing the advantages of the new approach to electronic and mobile health. The final goal is to create a free and open e-health platform which will connect users who seek information about specific medical issue (e.g., diagnosis, health services, health products, and other health-related info), and the providers of that information. Information providers are doctors, nurses, call centers, and organizations and companies that offer health-related products. The architectural design of the platform consists of a web-based interactional user interface, the virtual assistants, and the Rocketchat communication platform for text-based messaging between call centers, experts in health domain, and the virtual assistants.

Keywords: *electronic and mobile health, connecting platform, e-health, virtual assistants*

I. INTRODUCTION

Current healthcare systems face severe problems. One reason is aging and therefore more people deal with chronic diseases [4, 5]. This represents a huge impact on the stability of the healthcare system in countries that already cannot provide more nurses and doctors per patient. Also, healthcare cannot provide greater amount of hospital beds for patients, which was observed in the last pandemic of covid-19. Because of this, additional mechanisms are needed to provide the service even at the current level, primarily the electronic health.

Due to huge amount of the information on the Web, one of the main issues for the patients is how to get timely, correct and precise information about their health problem, e.g., where to get additional help, how can he or she treat the disease, and most importantly how to prevent it. We propose a solution in the form of an electronic and mobile health platform, which builds upon and improves the existing solutions (e.g., [7, 8]) with the experience obtained from the previous projects.

E-health platforms aim at solving various challenges such as market failures of eHealth models, low number of doctors in rural areas, or increasing number of elderly people. Due to heterogeneity of their purpose, their functional requirements and thus functional design vary significantly. Their common goal is to mediate transactions between various (supply-side and demand-side) users. In addition, components around the

platform may exist and they have to be independent of the platform. The platforms must facilitate the use and the creation

of such components, for example, new eHealth applications by third party providers. However, there also exist eHealth platforms that are data sharing platforms only. They collect the data instead of mediating transactions between users [18]. The proposed platform is mainly a data sharing platform, although transactions, i.e., conversations between users are also supported. Currently, the platform is in its design phase, which means we have not performed any quantitative analysis yet.

The remainder of this paper is structured as follows. In the second section, we briefly present services and how can we search through the system. The third section presents the call centers and the interaction with them using Rocketchat. The fourth section describes the virtual assistants, while the fifth section presents the intelligent cognitive assistant. Finally, the sixth section summarizes our work.

II. ARCHITECTURAL DESIGN

The platform consists of three main parts: Services and search, call centers, and bot. The main component which is at most important is "Search and services". For storing the services, we are using the Postgres database. We are using the elasticsearch search engine which we integrated with Django framework. Essentially we divided the work into two main categories. The first is the design of the web portal (developed with Django) and the second category is integrating the Rocketchat communication platform into the web portal. We are using Rocketchat for communication between patients and doctors. Also, we wrote a bot from scratch that gives information about services, waiting queues, etc. The bot is connected with the Rocketchat platform. The rough architecture of the three main parts is shown in Fig 1.

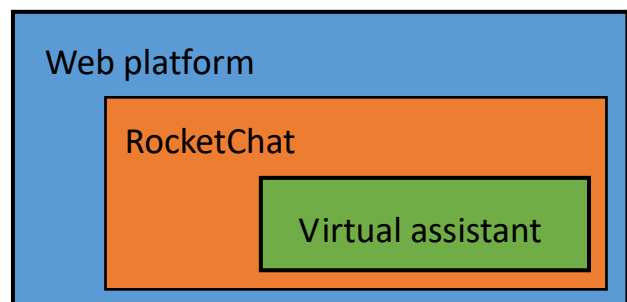


Fig. 1. Rough architecture of the platform

Welcome to Insieme, the Electronic and Mobile Health (EMH) platform!

The Insieme platform was developed within the cross-border project ISE-EMH and connects Italian and Slovenian partners. On this platform you can get online information and human support in the field of electronic and mobile health (EMH).

Fig. 2. Search functionality of the Insieme platform.

III. SERVICES AND SEARCH

Service is a textual description of a health-related subject or topic with

- the key information, i.e., a short description about the service,
- one or many links which point to service's website(s),
- one or more references to phone, mail or multimedia where addition information can be found.

Services are added to the platform by the medical personnel, experts in the related medical field, or companies which offer medical products. Essentially, the platform's most basic function is a modern health-based aggregator of services.

For example, an oncology service is represented by a short description about the types of cancer, and contains links to the Institute of Oncology, support groups for oncology patients, etc.

Services are divided into medical fields, and each field has its subfields. Current medical fields in the platform are:

- heart and blood vessels,
- dermatology,
- related platforms,
- first aid,
- ageing,
- sexually transmitted diseases,
- infections, coronavirus,
- mental illness and cognitive impairment,
- waiting queues,
- respiratory diseases.

Fig. 3. Available call centers for online help.

TABLE I. EXAMPLE OF INVERTED INDEX DATA STRUCTURE

Word	Service/Field/Subfield
oncology	1,2,5
cancer	2,5,10
dermatitis	14,18,19
...	...

Besides searching the data on the platform, the user can additionally find health information through call centers or with the virtual assistants. Call centers and virtual assistants are described more in detail in the following sections.

Fig. 3. Search result for First aid.

To find a service the user can either navigate the web page or use the search functionality. The search is implemented with Elasticsearch [9, 10, 11]. Elasticsearch is a highly scalable open-source solution for finding specific search string in a huge set of data very fast. Elasticsearch engine works very differently than other database retrieval systems. The core of the engine is the so-called inverted index. For each service and also each medical field and subfield, the inverted index creates a set of rows for each new word in their description. As the result, the inverted index is a data structure as a table, where each word represents a key where we can find the particular word (see Table 1). When we search for a particular string, we lookup the key which returns a set of services in which we find the string.

Information discovery has been further enhanced by combining Elasticsearch and the Trigram similarity [12]. Trigram similarity is mathematically defined as (1).

$$TrigramSimilarity = \frac{|SearchTrigram \cap TargetTrigrams|}{|TargetTrigrams|} \quad (1)$$

Fig. 2. shows the search functionality of the Insieme platform. Besides text search, the search supports various filter options such as filtering by medical fields and subfields. Fig. 4. shows an example of returned results. The search engine therefore enables users to independently, without any human intervention, find what they are looking for.

IV. COMMUNICATION WITH CALL CENTERS AND MEDICAL EXPERTS

The Insieme platform enables call centers and medical experts to register and define their field(s) of expertise. This enables the Insieme platform to connect the users with the experts and call centers. Users can communicate with call centers and medical experts through the Rocketchat communication tool. Rocketchat is an open-source communication platform that has been used in related platforms such as the E-Tourist Information System [13]. In the Insieme platform, the Rocketchat tool has been enhanced with the autocomplete functionality that enables medical experts to search for the services when communicating with the users, and to insert the found services into the conversation. Consequently, the communication between experts and users is faster and more effective. For each field and subfield, the available experts are listed on the web page. Fig. 5 shows an example of a conversation between a user and a medical expert.

Another important functionality of the Insieme platform is the support for the call centers. Call centers can include several experts and are represented as one (Rocketchat) group. The call centers whose at least one operator is available, are listed on the web page as shown in Fig. 3. Call centers are different from using the search engine as they connect users with humans who can give them more detailed and complex help than the search engine.

V. VIRTUAL ASSISTANTS

Virtual assistant (VA) technology is one of the most successful AI fields. Its prominence has been achieved through advancements in artificial intelligence (AI), showcased by Google, Apple mobile phones and Amazon VAs [1, 2]. VA is a complex virtual agent which takes input, processes it, and returns a logical and reasonable answer [3]. It uses natural language processing and understanding (NLP, NLU) to generate the answer. Chatbots are typically created for specific tasks and are predictable while VAs, on the other hand, are not very predictable, because they don't follow a precise algorithm for processing and generating answers.

The Insieme platform includes a hybrid conversational ecosystem of virtual agents. It consists of various assistants the user can contact according to his/her needs. The agents have elements of a chatbot and also a VA. An example of a chatbot is the Assistant for waiting queues, which enables users to check the waiting queues in the Slovenian healthcare

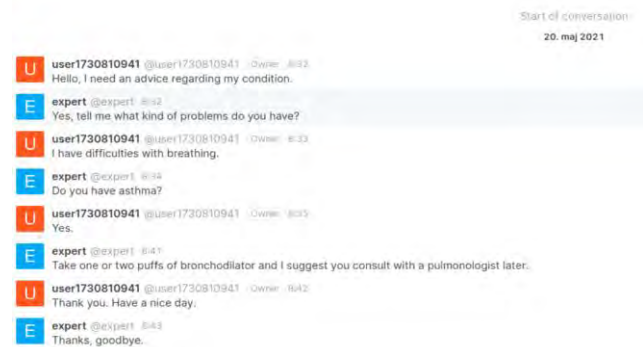


Fig. 5. An example of conversation between a user and an expert.

system. To this end, it searches through the Slovenian National Institute for Public Health system publicly available as a web application [6]. More advanced virtual assistants in the Insieme platform include the Intelligent cognitive assistant described in the following section. Fig. 6 shows the greeting message of the conversational ecosystem, which includes also the list of the currently integrated virtual agents. Virtual agents engage users in conversation, which can on the one hand have a better user experience role than a search engine, but it can also act in a way where the dialogue form helps the user with their problems, e.g. mental health problems (see next Section).

VI. INTELLIGENT COGNITIVE ASSISTANT FOR ATTITUDE AND BEHAVIOR CHANGE IN MENTAL HEALTH

The Insieme conversational ecosystem is being upgraded with an advanced intelligent cognitive assistant (ICA) for attitude and behavior change in mental health, which represent an important technology in the field of digital mental health [14, 15]. The ICA will include a Cognitive architecture (CARCH) which is based on the Theory of mind (ToM) model (see Fig. 7). This model forms the basis of the ICA because it is responsible for understanding thoughts and feelings. The goal of the cognitive assistant will be to detect stress, anxiety, and depression (SAD) level of the user, and create a textual personalized response for him or her to help that person feel better and lower their anxiety and depression levels.

The Intelligent cognitive assistant will collect linguistic input from the users using Rocketchat. The collected data will then be processed to infer the person's affect and to build appropriate user models. These user models will consist of data connected with mental health, such as questionnaire scores, and of cognitive and personality data. For the latter, we intend to use the Big Five Personality Traits questionnaire [17] to model the user. This will enable the ICA to personalize the responses to each individual by selecting the best strategies for specific people with specific mental health issues. Even more, ICA will continuously check the users' mental health and if it is deteriorating regardless of the help ICA is giving, ICA will



Fig. 6. Greeting message from the Insieme conversational ecosystem of virtual agents.

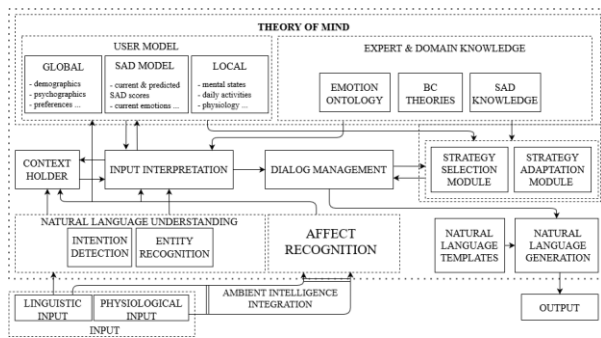


Fig. 7. The Theory of Mind (ToM) model.

change the strategy of help. This will make the ICA strongly adaptive to unforeseen circumstances that the users will find themselves in. The help will be provided in the form of text messages, utilizing motivational methods and cognitive behavioral theory techniques. This will be contextualized through persuasive frameworks, such as Cialdini's Principles of Persuasion [16], which can be used to personalize the messages to better influence individuals.

VII. CONCLUSIONS AND FUTURE WORK

This paper outlines the work on the development of the Insieme platform. The platform builds on the decades of previous research and development in the field of electronic help, including the cooperation in the EkoSMART system. The main functionality of the Insieme platform is to gather essential information such as services and experts in the medical fields for the purpose of providing the users the wanted information efficiently and effectively. Besides, it enables the users to directly communicate with a wide range of experts and call centers. The platform is available in three languages: English, Italian, and Slovene.

In the next steps, we will integrate the Intelligent cognitive assistant that is being developed, in the platform. We are also aiming at including additional data to the platform, such as publicly available datasets and scientific papers. This will enable to users to get better insight into the medical fields and learn details on the interesting medical issues. A more challenging task would be the support for collection of physiological data for users that have appropriate wearable devices and are willing to share these data. Based on such data, experts could enhance their advices, and intelligent cognitive assistant could better model the users. The created functional ecosystem of health experts, intelligent system modules for user interaction, and intelligent cognitive assistant technology (e.g., for mental health) offers a comprehensive platform that can serve as a successful digitalization of health services and therefore expand the eHealth options.

ACKNOWLEDGMENT

The paper was supported by the ISE-EMH project funded by the program V-A Italy-Slovenia 2014-2020. This work was also funded by the Slovenian Research Agency (research core funding No. P2-0209).

REFERENCES

- [1] T. Kolenik, M. Gjoreski, and M. Gams "PerMEASS – Personal Mental Health Virtual Assistant with Novel Ambient Intelligence Integration," CEUR, vol. 2820, no. 6, 2020, pp. 8–12.
- [2] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," Journal of Biomedical Informatics, vol. 73, 2017, pp. 159–170.
- [3] J. Oakley. Intelligent cognitive assistants (ICA), https://www.nsf.gov/crssprgm/nano/reports/ICA2_Workshop_Report_2_018.pdf, accessed 2021-05-20.
- [4] Challenges to European healthcare systems at a glance. <http://www.bff.com/>, accessed: 2021-05-20.
- [5] C. L. Storto and A. G. Goncharuk. "Performance measurement of healthcare systems in Europe," Journal of Applied Management and Investments, vol. 6, no. 3, 2017, pp. 170–174.
- [6] Waiting queues in Slovenian healthcare: <https://cakalnedobe.ezdrav.si/>, accessed 2021-05-20.
- [7] E-health <https://zvem.ezdrav.si>, accessed 2021-05-20.
- [8] Be healthy (Slovene: Bodi zdrav) <https://bodizdrav.net/>, accessed 2021-05-20.
- [9] A. Andhavarapu, Learning Elasticsearch, Packt Publishing Ltd, 2017.
- [10] C. Bhadane, H. A. Mody, D. U. Shah, and P. R. Sheth, "Use of Elastic Search for Intelligent Algorithms to Ease the Healthcare Industry," International Journal of Soft Computing and Engineering, vol. 3, no. 6, 2014, pp. 222–225.
- [11] V. A. Zamfir, M. Carabas, C. Carabas, and N. Tapus, "Systems monitoring and big data analysis using the elasticsearch system", in Proc. of 22nd International Conference on Control Systems and Computer Science (CSCS), 2019, pp. 188–193.
- [12] A. Niewiadomski and A. Akinwale, "Efficient similarity measures for texts matching," Journal of Applied Computer Science, vol. 23, no. 1, 2015, pp. 7–28.
- [13] G. Grasselli, "e-Tourist 2.0: an Adaptation of the e-Tourist for the AS-IT-IC Project," in Prof. of Information Society, 2018, pp. 20–22.
- [14] T. Kolenik and M. Gams, "Intelligent Cognitive Assistants for Attitude and Behavior Change Support in Mental Health: State-of-the-Art Technical Review," Electronics, vol. 10, no. 11, 2021, article no. 1250.
- [15] T. Kolenik and M. Gams, "Persuasive Technology for Mental Health: One Step Closer to (Mental Health Care) Equality?," IEEE Technology and Society Magazine, vol. 40, no. 1, 2021, pp. 80–86.
- [16] R. Cialdini, "Pre-Suasion: A Revolutionary Way to Influence and Persuade", Simon and Schuster, 2016.
- [17] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german," Journal of Research in Personality, vol. 41, no. 1, 2007, pp. 203–212.
- [18] M. Benedict, H. Herrmann, and W. Esswein, "eHealth-Platforms - The Case of Europe," in Proceedings of the Medical Informatics Europe, 2018, 5 pages.

A System for Automatic Detection of Major Depressive Disorder Based on Brain Activity

Daniela Janeva^{1,2}, Silvana Markovska-Simoska², Branislav Gerazov¹

¹Faculty of Electrical Engineering and Information Technologies Ss. Cyril and Methodius University

²Macedonian Academy of Sciences and Arts

Skopje, Macedonia

danielajaneva@hotmail.com

Abstract—MDD or major depressive disorder is a psychiatric disorder which is present in today's society on a large scale. The complexity of the symptoms may lead to misdiagnosis. Automatic detection of depression is state of the art problem which would objectify the diagnosis. The electroencephalogram (EEG) as a medical device, records the human brain activity in real time. Besides all of the physiological characteristics, EEG signals portray the emotional brain activity in real time, which motivates the idea of creating a system for automatic depression detection based on EEG signals. For the methods proposed in this paper a database of 30 healthy controls (HC) and 34 depressed subjects (MDD), is used. For each of the subjects the brain activity of 3 different states is recorded i.e., Eyes Open (EO), Eyes Closed (EC) and doing a visual stimuli task (TASK). From the total number of 22 electrodes in the EEG, we have analyzed only the signals from channels F3 and F4. Those channels are located on the frontal lobe and are associated with the frontal alpha asymmetry which has been stated as a biomarker for depression. A dataset of extracted features of F3 and F4 is created and machine learning algorithms for classification are then applied. The highest achieved accuracy of 98.5% is obtained with the RF (Random Forests) classification model.

Keywords—Major depressive disorder - MDD; electroencephalogram - EEG; brain activity; frontal alpha asymmetry; feature extraction; classification; machine learning; Random Forests;

I. INTRODUCTION

Depression is a widely used term, most often in the context of describing certain negative feelings. Melancholia and sadness are often misinterpreted and termed as depression. In general, depression represents the depressive disorders, which are group of diseases with a specific diagnosis that need a proper medical treatment. The major depressive disorder (MDD) is characterized by a long-term depressive mood or loss in interest in things that were previously interesting to the individual [1]. Because the depressive feelings are very common, it is important to distinguish between sadness as a feeling and depression as a medical condition. MDD is manifested in a constant feeling of sadness, apathy, poor concentration, impact on diet, loss of desire and interest in doing things that used to be fun. The individual becomes very moody and does not enjoy life. Loss of concentration and energy is another common symptom for this disorder. Feelings of hopelessness and despair could potentially evoke suicidal thoughts. This disorder could occur without any apparent reason, as well as a posttraumatic consequence of a bad experience. Depressed patients may experience cognitive

association and in severe cases paranoia and delusion [2]. According to statistics from the WHO >300 million per year are registered to be suffering from major depressive disorder [3]. Around 800.000 individuals per year, lose their life because of suicide caused by depression [4]. Therefore, early diagnosis of MDD is of significant importance and could potentially save many lives. Nowadays, the research in this field is focused on the brain in order to understand underlying mechanisms and biomarkers of depression. The most common diagnosis of depression is a structured interview conducted by psychologist or psychiatrist [5]. There are several different standardized clinical tests for diagnosis of depression. Each of the standardized test contributes to a bias in the decisions for further treatment. The interview is conducted by a psychiatrist and the diagnosis depends subjectively on doctors' decisions and experience. Furthermore, depressed individuals are less likely to openly talk about their thoughts and feelings. Therefore, finding an effective method for detecting depression is a state-of-the-art significant research problem.

With today's acquisition technology, psychological data driven research, for automatic diagnosis of mental illness opens a new window for development of objective and accurate automatic tools for detection of depressive disorders. This approach would provide fast and accurate results, which is crucial when it comes to mental illnesses. For that purpose, the electroencephalogram (EEG) is the ideal device to be the core of such a system. It measures the brain activity and besides all types of physiological data, it reflects the emotional brain activity in real time. EEG signals are recording the spontaneous and rhythmic electrical activity of neurons in the brain, from the scalp as a surface area [6]. Depression as a mental illness manifests abnormal brain activity as previous research suggests [7]. EEG is an effective way of recording and analyzing brain activity given its ease of use, availability, cost and sufficient resolution. The measured values are correlated with different states of brain activity in rest.

In this paper we propose a system for automatic detection of depression based on EEG. In section II, we present the applied methodology. We start by explaining the dataset. After that we present the applied signal pre-processing step followed by feature extraction. The feature dataset is separated as presented in II-D. We enclose the methodology section with the application of machine learning algorithms for classification and analysis of the obtained results. In section III we conclude the applied framework and deploy the best model. At the end we demonstrate the utilization of the proposed system.

II. METHODOLOGY

The proposed system for depression detection based on brain activity, underlies the methodology represented on Fig 1. Raw EEG signals were pre-processed, followed by the step of feature extraction. A feature space, significant for MDD was created. The feature dataset was then divided into training and testing data. ML classification algorithms were applied on the training data and evaluated on the testing data. The best model was then deployed for further predictions.

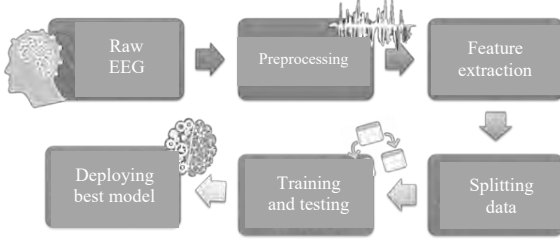


Fig 1. Applied methodology

A. Data

In this paper the database consists of 64 subjects, 34 of which are diagnosed with major depressive disorder and 30 healthy control subjects [16]. The mean time duration of the EEG for the subjects is 15 minutes. The acquisition system consists of 22 electrodes placed according to 10-20 standard international system as shown on Fig 2. The dataset consists of raw data stored in EDF (European Data Format) [8]. For each of the subjects 3 states are recorded: EC (Eyes Closed), EO (Eyes Open), and TASK. The sampling rate of the device is 256 Hz. MDD significance is mostly distinguishable in the frontal lobe in the brain, in terms of EEG, it is represented by the activity in F3 and F4 i.e., the electrodes of the left and right hemisphere accordingly Fig 2. For further steps, only the activity from these two electrodes is taken into consideration.

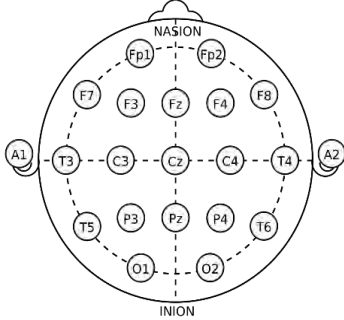


Fig 2. International 10-20 positioning system¹

B. Pre-processing

For loading, signal pre-processing, and modeling the data with ML algorithms the Python programming language is used. Specifically, working with EEG signals is done with MNE module which is an open-source module for exploring, visualizing and analyzing human neurophysiological data [9]. The preprocessing includes normalization of the signal's amplitude, and noise filtering.

1) Normalization

The amplitude of the EEG signals is measured in microvolts. The first step in the pipeline is to normalize the amplitude in range from -1 to +1. This is done with formula 1.

$$x_n[n] = \frac{x[n]}{||x[n]||_{max}} \quad (1)$$

In equation (1), $x_n[n]$ is the normalized signal, $x[n]$ is the raw EEG signal, and $||x[n]||_{max}$ represents absolute maximum amplitude.

2) Artefacts

Previous research has shown that the energy in the spectrum of EEG signals which is due to artifacts is significant in differentiating between control and depressed individuals [10]. For that purpose, the artifact removal procedure was not applied in our methodology.

3) Filtering

For further feature extraction, EEG signals were filtered with a notch filter of 50Hz to remove the influence of the city network. Furthermore, a bandpass FIR filter was applied in the range of 0.5 Hz to 60 Hz, which removed the DC component, and included the valuable range for feature extraction [11][15]. Fig 3. Shows the power spectral density (PSD) before signal filtering. Fig 4. shows the corresponding PSD after filtering of the signals. For signal filtering and computing PSD, we used the MNE module.

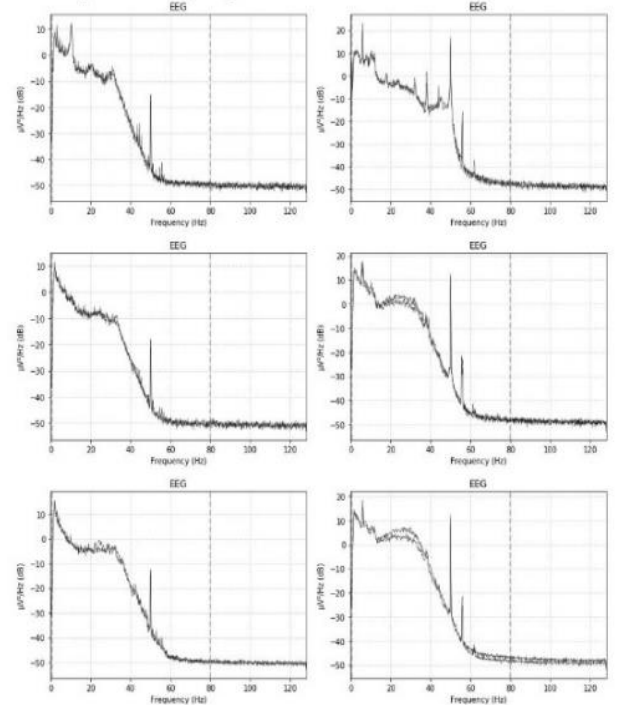


Fig 3. Pre-filtering PSD of EEG signals of F3 and F4 electrodes; Right HC, Left column MDD, EC, EO and TASK respectively at each row for column

¹ 10-20 system (EEG) - Wikipedia

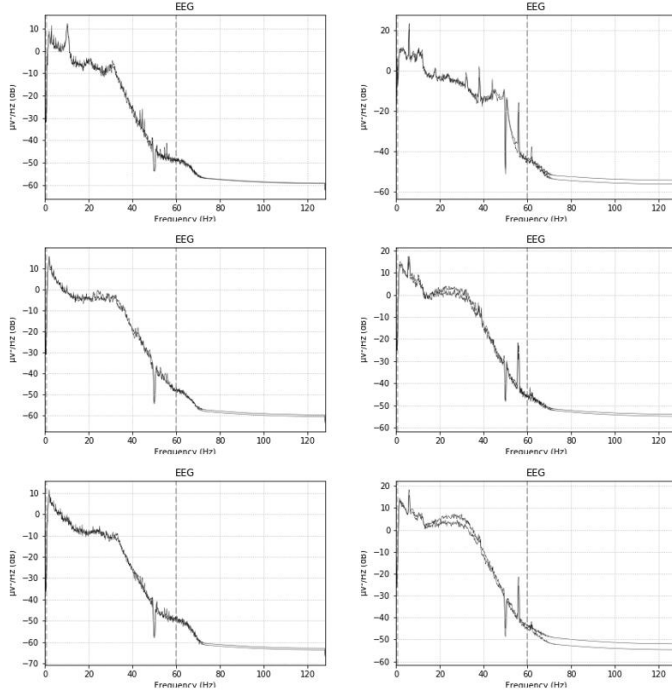


Fig 4. Post-filtering PSD of EEG signals of F3 and F4 electrodes; Right column HC, Left column MDD, EC, EO and TASK respectively at each from top.

C. Feature Extraction

Significant differences in the PSD on Fig 3 and Fig 4. can be seen in the alpha frequency range between HC and MDD. The total time duration per session of the EEG signals is 300s. Due to the unstationary nature of the EEG signal, the features are extracted in time windows of 30s. For each window segment, a power spectral density is estimated with Welch method as shown on Fig 5 [12].

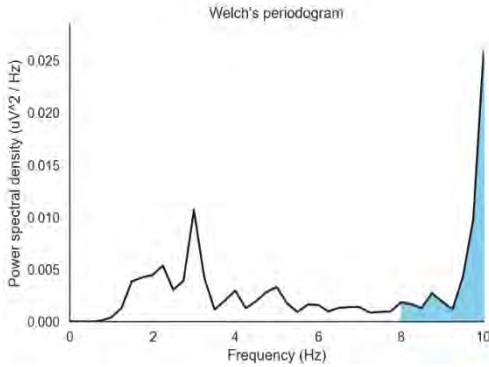


Fig 5. Welch periodogram. Blue area represents alpha band power [8-10Hz].

According to Welch periodogram for each time segment, considering both electrodes F3 and F4 whose placement is shown on Fig 2., the following features have been computed:

- Absolute power of the window selected segment
- Absolute power per band.
- Relative power of each band.
- Frontal Alpha Asymmetry
- Spectral entropy
- Spectrum mean value

The total number of features is 31. For each feature 4652 values are generated. The index of frontal alpha asymmetry is calculated by formula (2) [13].

$$FAA = \ln \left(\frac{WF_4}{WF_3} \right) \quad (2)$$

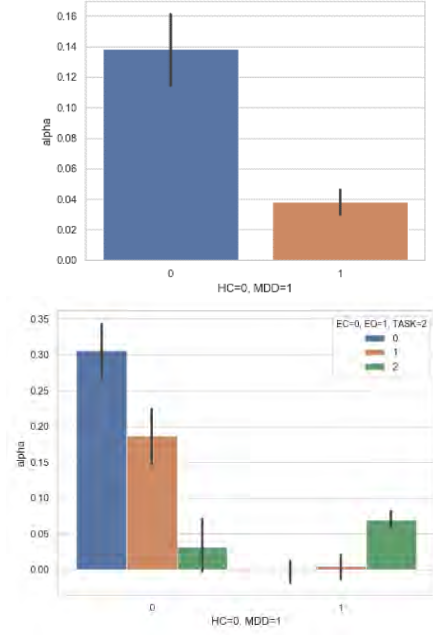


Fig 6. Histograms of frontal alpha asymmetry in HC and MDD – up, different states - down.

D. Data separation

Data was separated 80% for training and 20 % for testing. For each model hyperparameters were tuned with grid search Cross validation algorithm. The number of folds used for cross validation for each model is 5. Models with best hyperparameters are fit to the dataset.

E. ML classification algorithms

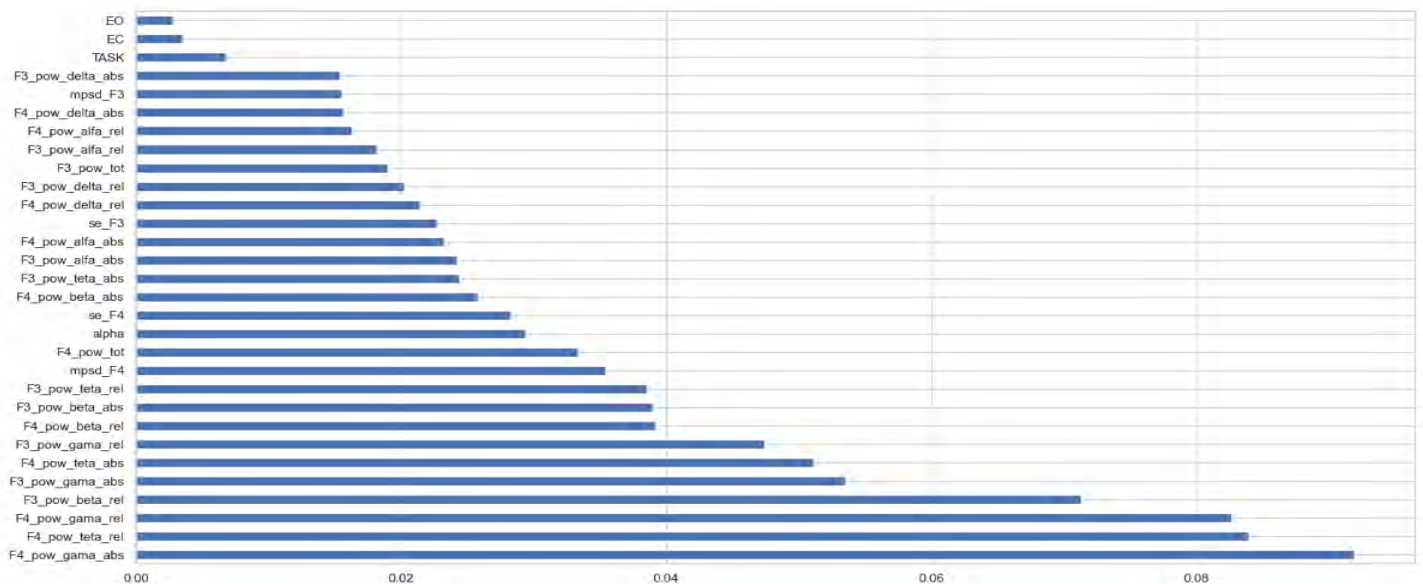
The feature training dataset was trained with several different models. The hyperparameters were tuned with grid search cross validation algorithm. The scikit learn package was used for training and testing the models [14]. For each model, the accuracy and F1 score are shown in the table 1.

Table 1: Accuracy and F1 score for model evaluation

Model	Acc Score [%]	F1 Score [%]
Support Vector Machines	88.51	89.42
Logistic Regression	89.26	90.18
Random Forest	98.5	98.59
MLP	91.94	92.58
Stochastic Gradient Decent	87.65	92.58
Decision Tree	95.92	96.15
KNearestNeighbor	96.56	96.74
Gaussian Naive Bayes	82.6	85

	Actual 0	Actual 1
Predicted 0	45.54%	1.40%
Predicted 1	2.04%	51.02%

mean_fit_time	std_fit_time	mean_score_time	std_score_time	parameters	split0_test_score	split1_test_score	split2_test_score	split3_test_score	split4_test_score	mean_test_score	std_test_score	rank_test_score
0.96	0.06	0.022	0.02	{'criterion': 'gini', 'max_features': 'auto', 'n_estimators': 100}	0.978	0.977	0.979	0.974	0.983	0.978	0.0031	6
4.75	0.14	0.101	0.01	{'criterion': 'gini', 'max_features': 'auto', 'n_estimators': 500}	0.979	0.977	0.978	0.974	0.982	0.978	0.0027	10
9.48	0.18	0.199	0.02	{'criterion': 'gini', 'max_features': 'auto', 'n_estimators': 1000}	0.983	0.975	0.978	0.973	0.982	0.978	0.0040	8
0.83	0.03	0.022	0.01	{'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 100}	0.987	0.977	0.975	0.974	0.977	0.978	0.0048	11
4.05	0.06	0.102	0.02	{'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 500}	0.983	0.977	0.975	0.974	0.982	0.978	0.0037	9
8.16	0.12	0.207	0.01	{'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 1000}	0.985	0.975	0.981	0.974	0.981	0.979	0.0039	4
1.52	0.01	0.022	0.01	{'criterion': 'entropy', 'max_features': 'auto', 'n_estimators': 100}	0.981	0.975	0.981	0.974	0.985	0.979	0.0039	3
7.58	0.09	0.098	0.01	{'criterion': 'entropy', 'max_features': 'auto', 'n_estimators': 500}	0.981	0.977	0.982	0.971	0.981	0.978	0.0039	7
15.2	0.31	0.204	0.02	{'criterion': 'entropy', 'max_features': 'auto', 'n_estimators': 1000}	0.982	0.977	0.981	0.977	0.981	0.979	0.0022	1
1.30	0.02	0.022	0.01	{'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 100}	0.985	0.979	0.977	0.974	0.982	0.979	0.0038	2
6.47	0.18	0.099	0.01	{'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 500}	0.985	0.975	0.974	0.970	0.982	0.977	0.0054	12
12.5	0.07	0.194	0.01	{'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 1000}	0.985	0.977	0.977	0.974	0.982	0.979	0.0040	5



III. CONCLUSION

In this paper we propose a system for detection of major depressive disorder based on EEG signals. Our approach underlies a feature extraction procedure followed by application of classification machine learning algorithms. Highest accuracy score of 98.5% was achieved with Random Forests, and the best model was deployed for further usage. With the help of the graphical user interface shown on Fig 9, clinicians could effortlessly input an .edf file to confirm the diagnosis for their patients, and proceed with further treatment.

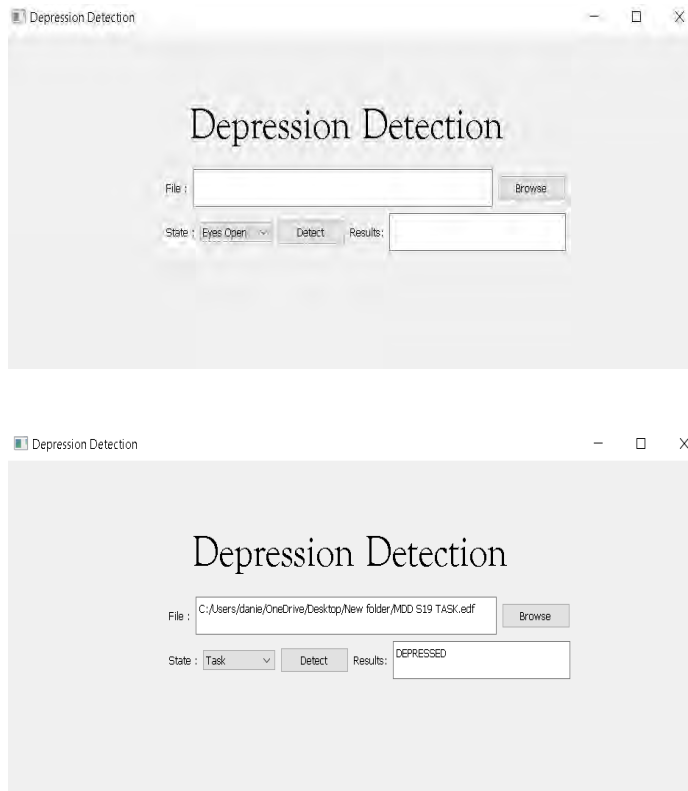


Fig 9: Graphical User Interface for loading EDF data files and detecting depression based on best model in Table 2.

REFERENCES

- [1] B., R. H., and G. Agam. "Major depressive disorder." *New England Journal of Medicine* 358.1 (2008): 55-68.
- [2] M.E.P Seligman, Hellessness: on Depression, Development and Death, Times Books/Henry Holt & Co, 1975.
- [3] "World Health Organization," [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs369/en/>. [Accessed on May 2020].
- [4] W. H. Organization, "Mental Health: new understanding, new hope," World Health Organization, The World Health Report, 2001.
- [5] V.C.Pangman, J.Sloan, L.Guse, "An examination of psychometric," *Applied Nursing Research*, volume 13, no. 4, pp. 209-213, 2000.
- [6] L., D.B., Emotions and the electroencephalogram, 1950.
- [7] K. J. Tenke Ce, "The stability of resting frontal electroencephalographic asymmetry in depression.," *Psychophysiology*, vol 41, no. 0, pp. 269-80, 2004.
- [8] K. Bob, et al. "A simple format for exchange of digitized polygraphic recordings." *Electro-encephalography and clinical neurophysiology* 82.5 (1992): 391-393.
- [9] A. Gramfort, M. Luessi, E. Larson, D.A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, 2013. doi:10.3389/fnins.2013.00267.
- [10] Stern, John M. Atlas of EEG patterns. Lippincott Williams & Wilkins, 2005.
- [11] Sanei, Saeid, and Jonathon A. Chambers. EEG signal processing. John Wiley & Sons, 2013.
- [12] Welch, Peter. "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms." *IEEE Transactions on audio and electroacoustics* 15.2 (1967): 70-73.
- [13] B., Benny B., et al. "Approach the good, withdraw from the bad—a review on frontal alpha asymmetry measures in applied psychological research." *Psychology* 4.03 (2013): 261.
- [14] P., F., et al. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research* 12 (2011): 2825-2830.
- [15] B. Gerazov, Biomedical Electronics, Skopje, 2016, pp.43-56
- [16] W. Mumtaz, "MDD Patients and Healthy Controls EEG Data (New)," 2016.

Predicting Trends and Anomalies in Daily Activities

Vito Janko

Department of intelligent systems
Jožef Stefan Institute
Ljubljana, Slovenia
vito.janko@ijs.si

Mitja Luštrek

Department of intelligent systems
Jožef Stefan Institute
Ljubljana, Slovenia
mitja.lustrek@ijs.si

Abstract—In this work, we analyzed behavioral data of 99 elderly users (their physical activity, sleep patterns, mental health) as part of the WellCo European project, whose objective was to develop a health and wellbeing coach. We tracked how this type of data changes from week to week and then tried to predict future data as well as find anomalies in the current data. This task was accomplished using a combination of statistical modeling and machine learning models. Machine learning models outperformed the baseline model in almost all cases, and the anomalies found matched well the human intuition of the concept.

Keywords—*machine learning; sensor data; coaching; elderly; anomalies, pervasive health*

I. INTRODUCTION

Ageing trends in Europe show that in the coming decades, the number of elderly (aged over 65 years) people will increase, reaching 28.7% of the EU by 2080 [1]. Moreover, thanks to the advances in technology, health care and pharmacological treatments, life expectancy has also notably increased. Despite this increase in life expectancy, the quality of life in these last years may be poor due to chronic illnesses such as heart disease, cancer, stroke, and diabetes. The most risk for preventable chronic conditions is accounted for by unhealthy behaviours like poor diet, physical inactivity, smoking, etc.

This could be partially mitigated by encouraging and raising awareness about healthier lifestyle choices. This was exactly the goal of WellCo [2] a European funded project – where we developed a personalised health and wellbeing coaching system. This system would collect data from a smartwatch and a smartphone worn by the user and from validated questionnaires – and then analyse it and propose helpful recommendations. Following these recommendations could help the users to adopt healthier behaviour choices. This system was successfully tested and validated on 99 users from three different countries in a 6-month pilot.

A part of the WellCo project – that we focus on in this paper – was to analyse the behavioural trends of its users in regards to their physical activity, sleep patterns and mental health. The goal was to predict future behavior in these areas and identify anomalies in their current behavior. Understanding the future trends could help us to send them personalized recommendations about their progress and how long until they reach their desired goals. Detecting anomalies, on the other hand, could help us identify some current issues the user may

be experiencing. These two tasks were done by a combination of statistical modeling and machine learning. Similar tasks were done in related work with a variety of methods, e.g., by making semantic rules using hierarchical clustering [3] or using Hidden Markov Models [4].

The paper is structured as follows. In Section II we briefly describe the data used: how it was selected, pre-processed and aggregated to a weekly basis. In Section III we describe the methodology, both for anomaly detection and behavior prediction. The results are presented in Section IV, alongside a short description on how the generated outputs would be used in practice. Finally, we discuss the results and conclude in Section V.

II. DATASET

During the pilot, 99 users carried a smartphone with the WellCo application and wore a smartwatch (either TicWatch or Withings). Both devices were constantly collecting sensor data in addition to occasionally providing the user with a questionnaire to fill. The collected data ranged from motion sensor data like acceleration and orientation, physiological data like heart rate, pressure and body weight, and finally data about the user's social behaviors (phone activity, number of text messages/calls, etc.). In some cases, the data was not collected by wristband, but entered manually to the application (e.g., weight). Some of the data was already partially processed into “higher-level” concepts like sleep duration and the number of steps taken.

In this paper we focused on only three categories of data: user's physical activity, sleep patterns and mental health. These three were chosen as they were the most semantically representative of the user's behavioral trends and were sampled with sufficient frequency (at least one value per week) to enable modelling and anomaly detection. Each of these data categories was represented by different data streams, but only one stream per data modality was chosen for further analysis.

A. Physical activity

The WellCo system collected the daily number of steps taken by each user, the distance travelled and the estimated number of calories burned. These three metrics were highly correlated so we picked only one to avoid redundancy. Since the user's daily goals were most often expressed in the number of steps taken, we chose this metric for the task.

B. Sleeping patterns

The main information provided by the smartwatch on the user's sleep patterns were the duration of sleep, start of sleep and end of sleep. Each of these three data points was provided for each day – although in practice more than half the data was missing, probably due to the users not using the devices late and/or early in the day.

In addition, this data was very prone to false positives in sleeping detection and often reported sleeping periods of 12 hours or more, which we believe to be inaccurate.

While the start and end of sleep were provided, we focused mainly on its duration as it is more important for users' wellbeing. The start and end of sleep could in theory be used to detect a "drift" that could indicate a change in sleeping patterns, but the data proved too unreliable for the task.

C. Mental health

Mental health information was available from weekly validated questionnaires. These questionnaires had 11 questions relevant to the mental health, like the two example questions listed below:

- Over the last 2 weeks, how often have you been feeling nervous, anxious or on edge?
- In the last month, how often have you felt that you were unable to control the important things in your life?

D. Pre-processing

The data as originally presented was very volatile and varied wildly from day to day (example data in Figure 1). This was expected for physical activity as, for example, the number of steps one performs per day can differ significantly due to a variety of factors – time available, weather, different chores to do, mobility restrictions due to COVID-19 etc. The sleep patterns (start, end, duration of sleep) were expected to be more consistent, but they were still changing due to measurement errors. To elaborate, if the device (smartphone or smartwatch) that was measuring the sleep was not picked up immediately in the morning, the device might mistakenly believe the user is still sleeping. The reverse happens if the user stopped using the device before going to bed (or used it while doing a very stationary activity, e.g., reading) – the device's idleness could be mistaken for the beginning of sleep.

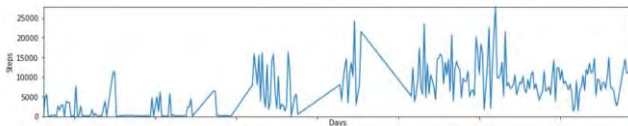


Fig. 1. The number of steps taken by one user on each day.

As seen from Fig. 1, the number of steps varies from day to day, but there can still be overall trends: in this case the user increased their average number of daily steps somewhere in the middle of the trial period. To detect such trends, we used a seven-day moving average filter to smooth the daily values. In addition, we grouped the data into Monday-to-Sunday weeks and calculated a few statistics for each week: the average value, standard deviation, the number of times the value is greater

than the user's average value (e.g., how many days the user was active).

The aggregation could have been performed with a different granularity, e.g., averaging every three days, five days etc. However, there is a good reason to believe that the seven-day average is the most suitable for observing the changes in behavior trends over time, as most people naturally have different routines for different days of the week. For example, if someone goes on longer walks during the weekends than during the weekdays, a three-day averaging would constantly observe an alternating pattern of improvement and regression of the number of steps taken – one that does not actually indicate a long-term change of the user behavior. Mental health does not naturally occur in weekly patterns, however, the questionnaires were given to the users on a weekly basis, so the same type of aggregation still applies.

To validate the weekly-pattern hypothesis for the physical activity, we took the number of steps taken by each user. The appropriately trimmed data-series were concatenated one after another and from the resulting series we then created a Fourier transform to get an insight into the frequency components of the dataset. If a certain behavior pattern occurs periodically, then the frequency corresponding to that period has a bigger magnitude. The results of this procedure are seen in Fig. 2, from which it is clear that most behavior patterns related to the number of steps happen with the periodicity of one week. The same pattern is not as obvious with sleeping data, but there was still a bias in slightly longer sleep times during the weekend.

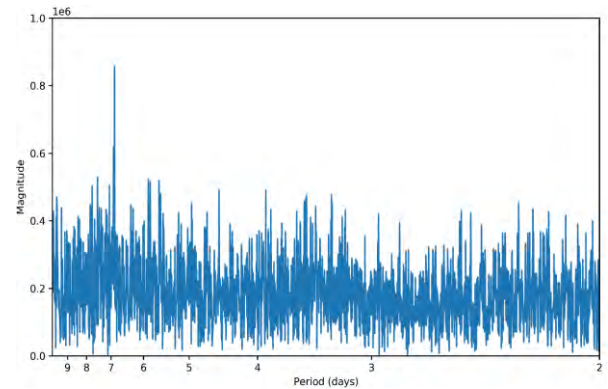


Fig. 2. The Fourier transform of the daily number of steps taken (of all users). The frequencies are labelled based on the number of days they represent.

Two extra steps were done as a part of the pre-processing: the first was to clean the sleeping data by removing data points where the sleep was recorded to last less than one hour, and change data where more than twelve hours of sleep were detected to twelve hours (measurements with these values were visually outliers). The second was to create a single metric from the questionnaires. Each of the questions had a discrete answer on the scale of 0 to X (X usually being 3, but it varied from question to question) with larger numbers indicating a potential problem. All the answers were first normalized based on their max value, then from each set of answers two metrics were calculated: the mean value and the max value. The mean value indicates the overall mental state based on the questionnaire, while the max value can help identify specific

potential problems. Since the mental health data had a lot of missing values (the users did not fill the questionnaire each week), we back-propagated known answers to missing past data. Interpolating the missing values would also be a reasonable alternative.

III. METHODOLOGY

A. Behavior anomaly detection

Detecting anomalies in the user behavior was done by comparing the user's behavior – i.e., the calculated statistics – of the current week to the previous weeks. We defined “deviation” as the difference between the two, and “anomaly” as an uncommonly large deviation (both terms will be more formally defined in this section). It is of note that the behavior was never compared to any baseline number (e.g., if the sleep time is less than 6 hours then declare it an anomaly). This is due to the large differences in behavior between the users, as seen in the distribution in Fig. 3. One can observe that what is considered a low number of daily steps for one user may be a very high number for another.

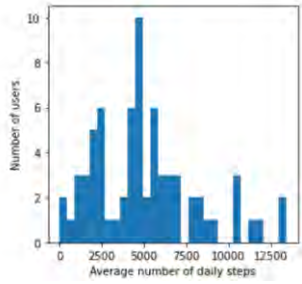


Fig. 3. The distribution of users based on their average number of steps. Similar distributions can be observed for sleep duration and mental health.

We tested three comparisons for defining deviations: (1) the current week against the previous one, (2) the current week against the last n weeks and (3) the current week against all previous weeks. The distribution of deviations of the first type is shown in the left graph of Fig. 4 for the “number of steps” modality. This distribution resembles a Gaussian distribution (also plotted in the same figure) and is roughly the same for all three analyzed modalities. The problem with this approach is that it is “too local” and after any “positive” deviation, a “negative” one would be detected even if the user's actual behavior became completely average. The opposite problem occurs when comparing the current week against all previous weeks: after the user's trend is changed, deviations could wrongly be detected for many following weeks.

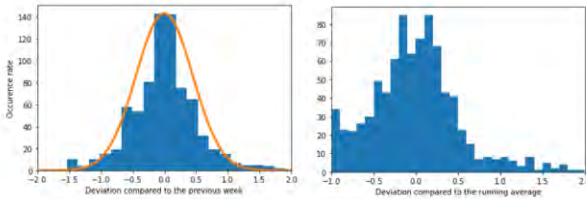


Fig. 4. The distribution of deviations compared to the users last n weeks ($n = 1$, which is best modeled with a simple ratio and $n = 4$, calculated with Equation 1). For the first case we also show a fitted Gaussian curve.

For comparing against the average of last n weeks we used Equation 1. In this equation, x represents the time-series of one of the data modalities (e.g., the average number of steps made in a week), m represents the current week's index and n represents the number of previous weeks to use in the comparison.

$$deviation = \frac{x_m - \frac{1}{n} \sum_{i=1}^n x_{m-i}}{\frac{1}{n} \sum_{i=1}^n x_{m-i}} \quad (1)$$

We tested different values of n and empirically determined that $n = 4$ provided a sensible compromise between the problems described in the previous paragraph and best corresponds to the intuitive notion of a deviation and consequently an anomaly.

The distribution of anomalies generated by Equation 1 is shown in the right graph of Figure 4. It still loosely follows the Gaussian distribution, but with its left end flattened at -1 . This happens due to the values of analyzed modalities being always positive and thus being unable to exceed this value given Equation 1.

Positive numbers indicate an increase compared to the average, while negative numbers indicate a decrease. An anomaly is then detected using the following steps:

- The deviation is calculated for each week and each user using the past data. The distribution of these deviations is then calculated and stored (Figure 4).
- A threshold is determined for a deviation to be considered an anomaly. This is done by first deciding on roughly how many data points we want to fall into the anomaly category (e.g., 5%) and then using the distribution to find a cut-off that defines a tail with that exact probability density.
- Each data modality was considered an anomaly in one direction only. For the number of steps, only a low number counted – as making “too many” steps was considered to be beneficial. For the sleep duration, we again only took the “sleep is too short” direction. While sleeping too long is not desirable and could be considered an anomaly in principle, doing this was not feasible due to the high number of times sleep was falsely recorded as “too long”. Finally, for the mental health, only high values (e.g., high amount of stress) were considered for obvious reasons.
- Deviation is calculated for the current week and compared to the pre-calculated threshold value in the desired direction. If the threshold is exceeded, the current week is considered an anomaly in respect to the analyzed statistic.

B. Behavior trend prediction

The goal of this section was to determine if the user's behavior in the current week can be determined from their past behavior. We tackled this problem by using machine learning models – the features were the statistics calculated from the

past weeks, while the classes were different aspects of the current week's behavior that we were interested in.

1) Prediction targets

We have chosen three different targets for the prediction, each connected to a different aspect of behavior. These were *a)* the previously defined anomalies, *b)* the user's average performance and *c)* the retention of their progress. Each target variable was defined for the daily number of steps taken, sleep length and mental health – resulting in nine combinations and thus prediction targets in total (e.g., one combination could be “anomalies in the daily number of steps”).

Anomalies: In Section III.A, we defined behavior anomalies and described how to detect them as they happen. It would be useful, however, to be able to predict them in advance in order to be potentially able to prevent them. Thus, the first prediction target are the anomalies that happen in the current week.

Average performance: The next point of interest that could be predicted is the actual value of the observed data modality. If a user made 2000 steps on average in the previous week, 1000 steps in the week before that, how many will they make on average this week? Knowing this would enable us to estimate how many weeks would be required for the user to reach their set goals on a consistent basis and appropriately encourage them in the process. The data was normalized for each user so that “1” represents their average performance.

Progress: The last target of interest, connected with the last one, was if the observed data modality will increase in the next week at all – regardless of the increase size. Knowing that the user's progress is likely to stop in the next week could enable us to send a timely message of encouragement that would help reinforce the positive habit.

2) Features

The features used for prediction were the mean and standard deviation of each of the three data modalities taken for each of the four weeks before the week being predicted. The difference between the week's mean value was also taken as a feature, in addition to the anomalies detected in each of the three previous weeks for each data modality. This results in 32 different features. It is of note that no data from the week being predicted was taken.

To improve the prediction results, we employed feature selection. Our feature selection method worked as follows: for each prediction target we ranked the features based on their importance (mutual information indicator [5]). We then selected the best feature, ran the predictor, and noted the accuracy for the classification problems and mean absolute error (MAE) for the regression ones. The second feature was added and the predictor was run again. If the accuracy (or MAE) improved, the feature was kept, otherwise it was removed. Then the third feature was added, and so on for the rest of features. This was repeated for each predictor (three predictors were tested – Section III.B.3). This way every predictor/prediction target combination had a different feature set.

For a given modality, the best features were primarily of the same modality (e.g., when predicting sleep duration for the

next week, the selected features were the sleep durations for the previous weeks and deviations between them. While this was not surprising, it leads to the conclusion that there was no strong connection between the data modalities.

3) Predictors

Out of the nine prediction targets, three were continuous and six were discrete. For the continuous prediction targets we chose three regressors to test: Linear Regression, Support Vector Machine (SVM) and Random Forest (RF). For the discrete prediction target we chose three classifiers: Logistic Regression, SVM and RF classifier. All predictors were implemented in the scikit-learn toolkit [6].

IV. RESULTS

A. Anomaly detection

To calculate the anomalies, one must – as mentioned previously – decide on the percentage of deviations that should be counted as anomalous. Since the distribution does not show any natural cut-off point, we manually chose the cut-off threshold of 5% (the biggest 5% of deviations were marked as anomalies.)

Using this threshold, we calculated the distribution of the anomalies among the users, and found them unevenly distributed (an example in Fig. 5). Only 29% of users ever experienced a physical activity anomaly, 25% a sleep anomaly and 38% a mental state anomaly.

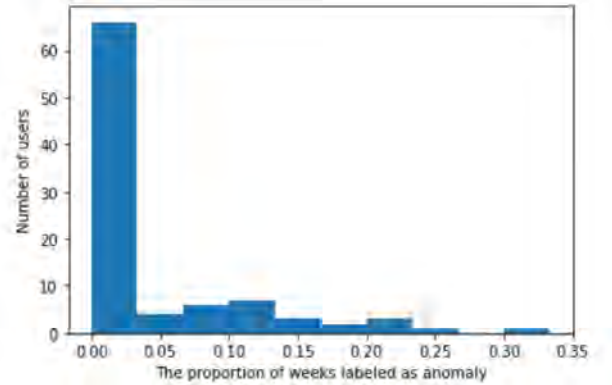


Fig. 5. The distribution of physical-activity anomalies across the users.

An example of a user's behavior trend, its weekly average and anomalies are shown in Fig. 6.

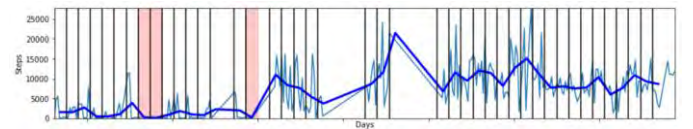


Fig. 6. The number of steps made by one user. The data is separated into weeks, with the weekly average emphasized in dark blue and anomalies highlighted in red. Wider gaps between weeks are due to missing data.

Finally, anomalies from different sensor modalities did not seem related to each other, as might be expected if they were due to illness or some other general disruption of the user's life. To empirically show this, we calculated the correlation

coefficient between the anomalies for each data modality – and no correlation exceeded 0.08.

B. Behaviour trend prediction

We tried predicting all three prediction targets using three different predictors. All tests were done using 10-fold cross validation, for each combination of the target variable and predictor. The tests were additionally repeated both using all the features and using features selected by the method described in Section III.B.2. For the continuous targets (*average performance*), the performance was measured using MAE, while accuracy was used for the discrete targets.

The results for each combination are shown in Table I. The baseline performance – predicting the average value or the majority class – is also noted. The anomalies are excluded as they could not be predicted – at least not better than the baseline classification. For the other two prediction classes the baseline case was usually exceeded (Progress class had the largest margin of difference). In some cases, the feature selection improved the results, in others the same results were achieved with significantly fewer features (five on average across the prediction problems).

TABLE I. MAE FOR THE AVERAGE VALUE CLASS AND PREDICTION ACCURACY [%] FOR THE PROGRESS CLASS. THE BEST RESULTS IN EACH ROW ARE BOLD IF THEY EXCEED THE BASELINE: ALWAYS PREDICTING MAJORITY/AVERAGE.

Metric	Class	Feature set	Linear/ Logistic regression	SVM	RF	Baseline
MAE	Average value (steps)	All	0.386	0.357	0.400	0.369
		Selected	0.360	0.344	0.395	0.369
	Average value (sleep)	All	0.095	0.114	0.097	0.133
		Selected	0.088	0.096	0.097	0.133
	Average value (mental)	All	0.115	0.074	0.065	0.248
		Selected	0.115	0.074	0.065	0.248
Acc.	Progress (steps)	All	64.7	63.2	59.8	52.5
		Selected	64.7	63.2	60.5	52.5
	Progress (sleep)	All feature	69.7	70.9	73.3	70.9
		Selected	69.7	70.9	73.3	70.9
	Progress (mental)	All	94.1	96.3	95.5	96.3
		Selected	94.1	96.3	95.9	96.3

C. Integration

To use this methodology in the WellCo system, we ran the described methods every week on the WellCo application server and calculated the characteristics of the previous week for each user. Based on the results a personalized recommendation would be given. An example for the case where physical activity is deteriorating.

[username], I hope you are well but I have detected you are not moving too much these days. Being active prevents many common diseases such as heart disease and diabetes.

V. CONCLUSION

In this work we described how we used sensor data and questionnaire answers to detect anomalies in the user's physical activity, sleep duration and mental health. In addition, we tried predicting these three types of data for the next week.

There was no “ground truth” to determine if the anomalies detected are the real ones, but they visually (when the data is plotted) matched the intuitive notion of an anomaly. They can be defined using an arbitrary threshold, and can be thus adjusted depending on the task. An important result was that the anomalies of different modalities were not connected and happened independently from each other. It is also interesting to note that the majority of users (ca. 70%) did not experience any physical-activity anomaly, and similarly for the remaining two modalities.

Predicting future behavior, one week in advance, proved more challenging. Anomalies in particular were unsurprisingly hardest to predict in advance (almost by definition). The results suggest that the user's behavior is predictable to a degree, but that there is not sufficient data available to us in this study to fully capture it. Alternatively, one could argue that human behavior will always remain partially unpredictable.

It is important to note that the average performance in any of the three modalities generally did not change much over longer periods of time. Thus, we did not get many examples of a steady progress or regression that would let us estimate the speed of such change. Analyzing this data also revealed a lack of long-term trends in the user behavior – in almost all cases their mean (of steps taken, sleep duration, etc.) remained static during the trials with very frequent oscillations around that mean. This could be the result of the COVID-19 epidemic that limited the mobility of the users. In most cases the user's progress in one week is immediately followed by a regression in the next one. In addition, it shows that it is rare to keep improving for more than 3 weeks.

ACKNOWLEDGMENT

The work in this paper was funded by the WellCo project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 769765.

REFERENCES

- [1] Aging population, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing, Accessed 10.06.2021
- [2] Wellco project site, <http://wellco-project.eu/>, Accessed: 10.06.2021
- [3] Hoque, Enamul, et al. "Holmes: A comprehensive anomaly detection system for daily in-home activities." 2015 International Conference on Distributed Computing in Sensor Systems. IEEE, 2015.
- [4] Forkan, Abdur Rahim Mohammad, et al. "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living." Pattern Recognition 48.3 (2015): 628-641.
- [5] Vergara, Jorge R., and Pablo A. Estévez. "A review of feature selection methods based on mutual information." Neural computing and applications 24.1 (2014): 175-186.
- [6] Sci-kit learn, <https://scikit-learn.org/stable/>, Accessed: 10.06.2021

Finding efficient intervention plans against Covid-19

Second place at the XPRIZE Pandemic Response Challenge

Nina Reščič^{1,2,*}, Vito Janko^{1,*}, David Susič^{1,2,*}, Carlo De Masi^{1,*}, Aljoša Vodopija^{1,2,*}, Matej Marinko^{1,3,*}, Tea Tušar¹, Erik Dovgan¹, Matej Cigale¹, Anton Gradišek¹, Matjaž Gams¹, Mitja Luštrek¹

¹ Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia

² Jožef Stefan International Postgraduate School, Ljubljana, Slovenia

³ Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia

*Authors contributed equally

~Corresponding author: nina.rescic@ijs.si

Abstract— Covid-19 has so far affected every country in the world. The Non-Pharmaceutical Interventions (NPIs) by governments have proven themselves quite effective at stopping the spread of infections, but when applied in a very strict and long-lasting manner could have devastating consequences for the economic and social well-being of the population. XPRIZE and Cognizant organized the \$500,000 XPRIZE Pandemic Response Challenge, where the participants were tasked to find good trade-offs between the costs and benefits of NPIs. This paper describes the solution by the team JSI vs Covid that placed second and won a \$250,000 prize. The described solution uses an SEIR model to predict the spread of the infections, with the model parameters being dynamically changed based on active NPIs using machine learning. It then uses multi-objective optimization to find the desired trade-offs between NPI strictness and effectiveness.

Keywords— Covid-19; countermeasures; Non-Pharmaceutical Interventions; epidemiological models; multi-objective optimisation

I. INTRODUCTION

The Covid-19 pandemic has negatively affected the whole world, with the virus spreading extremely fast. While non-pharmaceutical interventions (NPI) like closing schools and cancelling public events have proven effective at containing the pandemic [1, 2], they come with a large cost to the economy, and with a quality of life decrease for the general population. Policy-makers were thus given a challenging task of balancing the spread of the pandemic with the socio-economic costs of the NPIs. This task was made even harder due to how unprecedented the situation is and due to lack of reliable data about the exact extent to which the NPIs affect the spread of the virus.

As time goes on, however, more and more data about the pandemic becomes available (e.g., the number of infections in each country) and one could use artificial intelligence to 1) understand the effect of various NPIs, and 2) propose sensible intervention plans based on historical evidence and not just based on the intuition of policy makers. This was the exact idea of the XPRIZE: Pandemic Response Challenge [3] where two

hundred research teams from all around the world competed to achieve the two previously mentioned tasks and to win the \$500,000 prize purse (sponsored by Cognizant [4]). In this paper we describe our submission to this competition, how we tackled both problems and ultimately ended up as being one of the two winners [3].

First, in Section II we describe in more detail the two tasks given by the competition, then in Section III we describe our methods, and show the results in Section IV. Finally, we conclude in Section V.

II. THE PANDEMIC RESPONSE CHALLENGE

The competition was split into two phases. In the first one the “Prediction” phase the competitors had to predict the number of infections for 236 regions, given the NPIs that were in place in these regions (regions correspond to most countries in the world and some regions inside countries such as US states. To ensure fairness, the submitted models were tested each day after submission, for months, and were given actual NPIs in place at that time period and had to predict the number of infections from the submission date on. The models could use any other additional data if it was provided before the submission date.

In the second “Prescription” phase, the competitors had to create intervention plans for different situations (different countries and time periods) for two months in advance. There were 12 possible NPIs to pick from an OxCGRT database [5], each with different levels of strictness. An intervention plan could consist of any combination of these, and could change from day to day. For example, a possible intervention plan would be to use strict NPIs at the beginning, but gradually lower the stringency as time moves on and the predicted number of infections’ fall. The prescribed intervention plans obviously could not be tested in real life so their quality was assessed based on two criteria, the predicted number of infections and the socio-economic cost. The prediction was made by the “standard predictor” provided by the organizers [6]. The socio-

economic cost of each NPI was provided by the organizers during the evaluation phase the submitted prescriptor was required to work with any cost provided. This mimics the real-life application of policy makers providing their own custom costs, fitted to the needs of their country. Each competitor could prescribe up to ten different intervention plans with different trade-offs between the two criteria. An intervention plan was considered better than another if it dominated it, which means that it was better on one criterion and not worse on the other. A solution (intervention plan) is said to be nondominated if there is no solution that dominates it. In a favourable case, the ten proposed plans should be spread all along the Pareto front - the image of all nondominated solutions.

III. METHODS

In Section III.A and Section III.B we describe our methods for the “Predictor” and “Prescription” phase of the competition, respectively.

A. Predictor

The goal of the “Predictor” phase was to predict the number of infections for each country/region for each day, months in advance, given the NPIs in that country/region. We did so by using a SEIR epidemiological model that was improved so that its “spreading rate” parameter β can dynamically adapt to the changes to the NPIs. The mapping between the NPIs and the spreading rate parameter was done using machine learning.

1) Datasets

We worked with two datasets. The first dataset consisted of daily reported infections and was used to fit the SEIR model (Section III.A.2). This data was collected from the Oxford Covid-19 Government Response Tracker (OxCGRT) database [5]. The second dataset was used for the β prediction model. While the main factors affecting the speed of the spread are the active NPIs, we collected the data on all other conditions we believed could be affecting it. We started with a dataset of 93 static (one per country) features such as development, culture, health, etc., which were extracted in our previous research on Covid-19 [7]. We then added “dynamic” features that could change day-by-day, namely the weather (temperature, humidity, etc.) and different NPIs collected from the OxCGRT database.

All the data was used for fitting the epidemiological model (Section III.A.3), but only a subset of 108 countries for training the machine learning model (Section III.A.4). The inclusion criteria for a country to be part of the training set were: sufficient data for that country, and negative correlation between NPI stringency and number of infections. The latter condition was due to some countries having inadequate testing and thus inaccurate data.

2) SEIR epidemiological model

One of the most commonly used approaches to predict the number of new daily infections are the epidemiological models. We used the standard SEIR model that uses *Susceptible*, *Exposed*, *Infected* and *Removed* compartments (FIGURE 1).

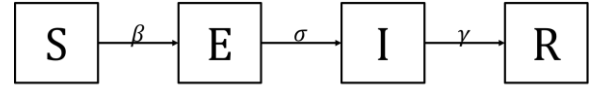


FIGURE 1: Scheme of SEIR model

The model uses parameters β , σ and γ that determine the transition probabilities from one (compartment) state to another as shown in the system below.

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dE}{dt} &= \frac{\beta SI}{N} - \sigma E \\ \frac{dI}{dt} &= \sigma E - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

The β parameter (infection rate) was fitted based on the historical data using least square error, while the σ (incubation period) and γ (recovery rate) were set as static values based on the values found in the literature. The fitted β parameter would then be used as the prediction target for the machine learning model (Section III.A.3).

3) Fitting β

Since the β s are constantly (and sometimes drastically) changing over time in ways that cannot be modelled using a simple function, we decided to split the data from each region into intervals and fit them separately. These intervals were created in two different ways:

- depending on NPIs – if they changed by more than a predetermined threshold value a new interval is started, and
- depending on the infection trends – the intervals were created so that the number of infections were either rising or falling on each interval. The fitting was done separately on each set of intervals, and then we determined for each region which fitting gave better prediction accuracy and used those β s as the ground truth for the machine learning models in the next step.

4) Predicting β

Fitted β s from previous subsections were then used to create a machine learning problem where the goal was to predict β s from the features. To predict future infections, a sequence of β s were calculated from future data (e.g., NPIs) and then inserted into the SEIR model.

5) Other Components

The submitted pipeline works as follows:

1. If the NPIs in the given region have not changed after submission, take the last pre-calculated β for that region and use it in the SEIR model to predict the daily number of infections.
2. If the NPIs in the given region have changed, first use the features for that region (mainly the provided NPIs) in

conjunction with the created machine learning model to calculate the β parameter for each day. Then use calculated β in the SEIR model to predict the daily number of infections.

3. In all cases, when β changes the transition is made smoother by using exponentially weighted averages of the β values.
4. A linear model was used for β prediction. We performed feature selection with 153 collected features and those that showed by far the strongest correlation with β were the NPIs for the (t-14)-th day.

An exception to this procedure was made in roughly 40 regions where the SEIR model was not a good fit on the historical data, usually due to the too low number of daily infections. In these cases, we took the last month of known data for that region and then found other regions in the past that exhibited a similar infection pattern, i.e., their number of infections closely matched, when normalized for the population. We looked at what happened in those regions after the inspected period, and used this to predict the future in the original period. Since the predictions made in step 3 are expected to be of length less than 180 days, we used the standard procedure to create the remainder of the predictions.

B. Prescriptor

In the second competition phase we had to create intervention plans for different time periods and for different regions. Such intervention plans should have good trade-offs between the stringency of the interventions and the projected infections that result from them. Such problems are commonly tackled with multi-objective evolutionary algorithms (MOEAs) that imitate biological evolution to search the space of possible intervention plans, evaluate them in terms of their stringency and the number of infections, and find plans with good trade-offs between the objectives. We were facing a time constraint as well - we only had 6 hours to evaluate 235 regions (90 seconds on average per region). We used the NSGA-II [8] algorithm for the task. The intervention plans were represented as vectors, where the i -th variable represents what is the aggregated socio-economic costs of all NPIs to be used on the i -th week.

We decided to optimize with the granularity of one week instead of one day for two reasons: 1) it is unrealistic to expect real-life policies to change with a higher frequency and 2) the quality of the solutions did not substantially improve when using a smaller granularity. It is of note that this granularity parameter is adjustable if our system would be used in practice and a decision maker would so desire.

To evaluate such a vector of socio-economic costs during the optimisation process, they are first expanded so that each variable represents one day, then for each day the NPIs are selected so that they do not exceed the cost for that day and so that they are as effective in reducing the number of infections as possible. The effectiveness of every NPI combination according to the “standard predictor” was precomputed in advance, so that the selection in the previous step can be done

with no computational overhead. Finally, the resulting matrix of NPIs for each day is sent to the standard predictor. The optimisation could in theory directly use the matrix representation where each value represents the presence (and strictness) of each NPI, but we have empirically evaluated that this only increases the search space and thus search time, without providing better solutions.

The resulting intervention plans, made by the described optimisation process, provided great trade-offs but the method turned out to be computationally too expensive, as each call to the standard predictor needed a few seconds for evaluation – and each region needed roughly 10,000 evaluations for the optimisation process to converge. Given the time constraint this process was much too slow.

We thus developed two methods derived from this standard multi-objective optimisation approach, and then combined them at the end.

1) Pre-computed plans

Our first method was to compute several plans in advance, and then for a specific region during the competition evaluation select the plan that is the most appropriate for the current situation in that region. The criteria for being “most appropriate” were the following: the desired length of prescription (we pre-calculated plans for 90-, 75-, 60- and 45-days, and selected the one closest to the desired length), infection trend (infections raising, falling, stable, raising fast, falling fast), and size of the country/region (small, large). For each of the listed combinations, ten prescriptions were pre-calculated and could be used for a given region during the competition. Since the socio-economic costs were still unknown at the time of the submission, our pre-calculated intervention plans only specified the maximum socio-economic costs for each day – which is in any case the natural representation of our optimisation process. Then during the evaluation, when actual costs for each NPIs were given, we selected the most effective NPIs as previously described.

2) Optimisation with the SEIR model

The second method used similar optimisation, but with two exceptions: 1) this optimisation was not done in advance, but directly for the country/regions and time intervals of interest, and 2) a fast surrogate model was used instead of the standard predictor. Surrogate model is a technique often used in optimisation where a computationally expensive model (in our case the standard predictor) is replaced by a simpler model that still returns similar results but is much faster. In our case the surrogate model was the same as epidemiological SEIR described in Section III.B with two differences. First, its parameters were fitted to the standard predictor’s outputs instead of ground-truth infections. Second, the code was rewritten in Cython (static compiler for Python) to be faster. It still used the same pipeline of first using NPIs to determine the β parameter for each day and then dynamically changing that parameter during the SEIR evaluation.

3) Full prescriptor pipeline

The submitted pipeline works as follows:

1. For each region we first retrieved the data for the three weeks leading to the start date. This data is either stored in a historical file or is computed with the standard predictor if data is in the future. Based on this data, the country and prescription length, we chose the pre-computed plans described in Section III.B.1.
2. In the edge case where infection data always equals 0, we prescribe no interventions. Otherwise, we also run the optimisation described in Section III.B.2 to create new prescription plans from scratch.
3. Both pre-computed and surrogate optimisation return intervention plans of comparable quality, each having their strengths and weaknesses. The latter uses a surrogate model instead of the “real” one, and is done with severe time constraints, while the former optimizes for different (although similar) time/region combinations than the target ones.

We discovered, however, that combining both results frequently increases the overall quality of the obtained Pareto front approximation. Having the twenty solutions – ten from each method – we finally select ten best ones for the final submission. This step was done using the greedy Hypervolume Subset Selection (gHSS) method [9]. This approach finds an approximate solution to the hypervolume subset selection problem. In our case, the objective is to obtain the subset of ten solutions that maximize the hypervolume in the objective space. Large hypervolume values result in large dominated areas, therefore, solutions selected by gHSS are expected to dominate a large number of competitors’ solutions.

IV. RESULTS

A. Predictor

1) Predicting β

Feature selection was run, however none of the features turned out to be selected more often than others. On the selected features we trained three models (linear regression, decision tree regressor and random forest regressor) with the target being β . Linear regression performed the best by far, but none of the models built on the selected features performed better than the model trained only on the NPIs. Thus, for the final model we selected linear regression and trained it only on the NPI data with the delay of $(t - 14)$ days, since this delay turned out to have the strongest negative correlation with the normalized β s.

2) Competition Performance

The competition organizers calculated general mean average errors (MAE) and MAE by region to evaluate the predictors of every competitor. The full list of results can be found on [10], although the team names are anonymised. Our submission floated between fourth and first place, depending on the day of evaluation, and landed in second place on the last day of the evaluation, which meant we qualified for the second round. A sample prediction can be seen in FIGURE 2. The only consistent prediction error our method was doing (also visible on the same

figure), was not taking holidays into account as the testing rate dropped significantly during such periods.

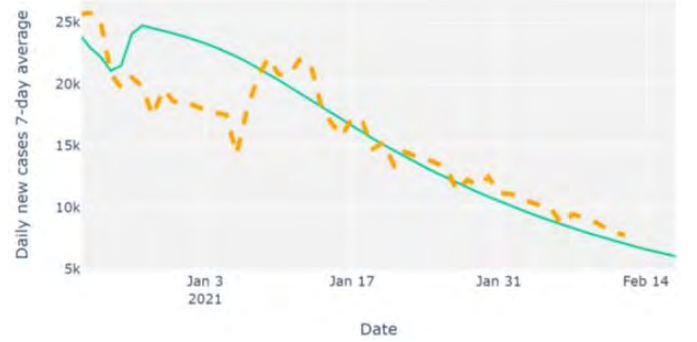


FIGURE 2: (Orange) The number of infections reported in Germany. The numbers were smoothed by using a 7-day moving average. (Green) The predicted number of infections in the same country/period.



FIGURE 3: (Top) A prescribed plan for each week, where we list the maximum NPI cost for each week. (Bottom) A prescribed plan where each column represents one week, and each row is the intensity of a different NPI.

B. Prescriptor

Our system had to prescribe 10 intervention plans for each country/region for different time intervals and different socio-economic costs. The full list of results can be found at [11]. Sample prescriptions in the objective space are shown in FIGURE 3. The main criteria for the competition was the so-called domination count: a solution scored a point for each other solution it dominated. The points achieved by the top 10 teams are listed in TABLE 1. Numerically, we were the best performing team in the competition (while the numerical results were anonymized, the structure of our prescription was easily recognizable among the results) and when this was combined with the “qualitative score” of the judges, we landed in second place.

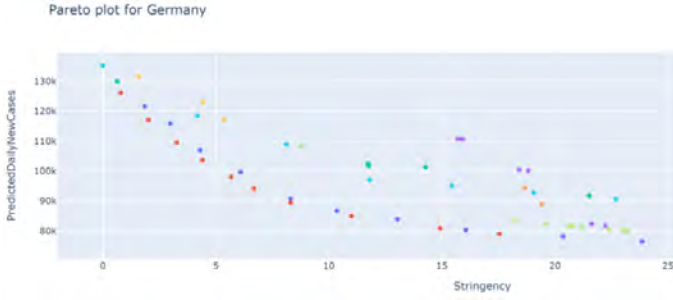


FIGURE 4: Prescribed plans from different teams for Germany. Each dot represents a different trade-off between the predicted number of cases and the NPI cost aggregated socio-economic cost. Some plans have low costs and a high number of infections, others vice versa. Our submission is represented by blue dots, and is visibly one of the two best ones.

TABLE 1: The domination count of the 10 best performing teams. Our submission is bolded.

Rank	Domination count
1	515247
2	490819
3	458146
4	435691
5	396968
6	313141
6	313141
8	288694
9	148766
10	134391

V. CONCLUSION

The XPRIZE: Pandemic Response Challenge focused on the development of data-driven AI systems to predict COVID-19 infection rates and prescriptions of intervention plans that regional governments, communities, and organizations can implement to minimize harm when reopening their economies.

While the problem of predicting new infections has been addressed many times, the real innovation of the competition was to find a way to prescribe NPI plans in such a way that both the number of infections and the stringency of the plans are the lowest possible. Our key insight when designing the predictor was to use machine learning to enhance the classical SEIR epidemiological model. This allowed us to dynamically adapt to the changes in NPIs as they were happening. On the other hand, the key insight for the prescriptor was to use MOEA methodology which is not common in this domain and then to

find ways for making it less computationally expensive. The latter was done with a combination of surrogate model usage, computing sample prescriptions in advance, using weekly granularity for the optimisation and clever solution representation. Representing solutions using the overall stringency (rather than individual interventions) lead to far more effective optimisation (due to search-space reduction) and consequently better intervention plans in a reasonable time.

Another important issue to explore is how to present such methods and their outputs to decision-makers. We developed a prototype web application that could be used for such a purpose, but collaboration with actual decision-makers is necessary to test and improve it.

ACKNOWLEDGMENT

This research was funded by Slovenian Research Agency (research core funding No. P2-0209 (B)).

REFERENCES

- [1] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker and e. al., “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe,” *Nature*, 2020.
- [2] S. Moore, E. M. Hill, M. J. Tildesley, L. Dyson and M. J. Keeling, “Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study,” *The Lancet Infectious Diseases*, vol. 21, no. 6, pp. 793-802, 2021.
- [3] XPRIZE. [Online]. Available: <https://www.xprize.org/challenge/pandemicresponse>. [Accessed 20 June 2021].
- [4] Cognizant. [Online]. Available: <https://www.cognizant.com/>. [Accessed 20 June 2021].
- [5] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar and H. Tatlow, “A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker),” *Nature Human Behaviour*, p. 529–538, 2021.
- [6] R. Miikkulainen, O. Francon, E. Meyerson, X. Qiu, D. Sargent, E. Canzani and B. Hodjat, “From Prediction to Prescription: Evolutionary Optimization of Nonpharmaceutical Interventions in the COVID-19 Pandemic,” *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 2, pp. 386-401, 2021.
- [7] V. Janko, G. Slapničar, E. Dovgan, N. Reščič, T. Kolenik, M. Gjoreski, M. Smerkol, M. Gams and M. Luštrek, “Machine Learning for Analyzing Non-countermeasure Factors Affecting Early Spread of COVID-19,” *Preprints*, 2021.
- [8] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197, 2002.
- [9] A. P. Guerreiro, C. M. Fonseca and L. Paquete, “Greedy hypervolume subset selection in low dimensions,” *Evolutionary Computation*, vol. 24, pp. 521-544, 2016.
- [10] XPRIZE, “XPRIZE Pandemic Response Challenge: Phase1 Results,” [Online]. Available: <https://phase1.xprize.evolution.ml/>. [Accessed 20 June 2021].
- [11] XPRIZE, “XPRIZE Pandemic Response Challenge: Phase2 Results,” [Online]. Available: <https://phase2.xprize.evolution.ml/>. [Accessed 20 June 2021].

Machine Learning Based Anomaly Detection in Ambient Assisted Living Environments

Ana Cholakoska, Valentin Rakovic,
Hristijan Gjoreski, Marija Kalendar
Faculty of Electrical Engineering and
Information Technologies
Ss. Cyril and Methodius University
Skopje, R. North Macedonia

{acholak,valentin,hristijan,marijaka}@feit.ukim.edu.mk

Bjarne Pfitzner, Bert Arnrich
University of Potsdam
Hasso Plattner Institute
Potsdam, Germany
bjarne.pfitzner@hpi.de
bert.arnrich@hpi.de

Abstract - Improving the security of the Internet of things is one of the most important and critical issues facing the modern world. With the rapid development and widespread use of the Internet of things, the ability of these devices to communicate securely without compromising their performance is a major challenge. The majority of these devices are limited in power and ability to perform complex computer calculations. This is where anomaly and intrusion detection systems come into play. In this paper, various machine learning algorithms are applied to effectively detect anomalies in such networks. The results obtained show great accuracy and precision (97%), as well as short execution time.

Keywords - Internet of things, machine learning, security, anomaly detection, ambient assisted living

I. INTRODUCTION

The need to facilitate and automate processes is leading to the rapid development of the Internet of Things – part of the fourth industrial revolution. Millions of different devices from multiple manufacturers connected in various ways work together to provide a variety of functions for home, medicine, industry, infrastructure, transportation, and more. However, this diversity poses many problems, mainly related to privacy and security[1,2,6].

Cyber-attacks in such networks are no exception. Depending on the attack surface available, an attack could vary from gaining unauthorized access to a device to power outages, resulting in potentially significant financial and economic losses. The prevailing security threats are especially important in smart homes, utilized for Ambient Assisted Living(AAL).[3]

Using AAL, patients can be monitored around the clock and their clinical outcome may be improved, while reducing costs and optimizing healthcare productivity. Additionally, doctors can easily communicate with patients and have the chance to detect diseases earlier. Nevertheless, the increasing deployment of Internet-connected devices for AAL environments, puts users at significant risk, as personal and health related information become remotely accessible.

Most of the existing studies and frameworks aimed at detecting network anomalies do not explicitly address traffic generated by devices that fall under the Internet of

Things[9,11], even fewer work with a relevant data set of an Ambient Assisted Living environment. More recently, machine and deep learning algorithms have been implemented in this area, which have been shown to give good results in detecting such anomalies in networks[10-14].

This paper examines the detection of anomalies in AAL environments by utilizing machine learning algorithms. To the authors' knowledge this is the first work that focuses on ML-based anomaly detection in IoT-centered AAL environments. Section 2 describes the current situation and some of the existing solutions. Section 3 shows the work methodology - selecting the data set and features and processing the data. Sections 4 and 5 provide an overview of the results and a discussion, as well as a conclusion.

II. MACHINE LEARNING FOR ANOMALY DETECTION

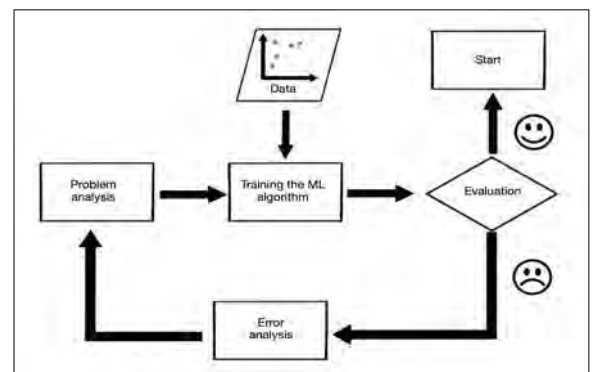


Figure 1: Problem solving process using machine learning

Due to the complexity of modern cyber attacks, the expectations on the modern intrusion detection systems are huge. However, in IoT networks, many specific challenges arise. These devices are constrained, in large numbers and generate heterogeneous data. Therefore, intrusion detection systems that require significant amount of computing or memory usage are not suitable for these networks. So, researchers are inclined to machine and deep learning algorithms to help improve the intrusion detection and prevention process.

Many approaches to tackle these anomaly detection problems have been made and show promising results. Hodo et al.[4], Nobakht[8] and Hossein et al.[18] use multi-layer perception (MLP), logistic regression, support vector machine (SVM) and artificial immune system (AIS) to detect only one type of the following threats: denial of service (DoS), unauthorized access and botnet attacks.

Some approaches, like Golmah[12] and Tanpure[14] propose a hybrid IDS with a classification model to detect abnormal behavior. The first one uses an SVM and the second one uses K-means and Naïve Bayes to group and classify data. It can be seen that such combination of algorithms improves the accuracy of anomaly detection.

However, these approaches don't use data generated from real IoT environments and mostly rely on the KDD dataset[9], which is a dataset that focuses primarily on four types of attacks: DoS, probing, user to root (U2R) and remote to local (R2L). It does not include newer and specific attacks to IoT networks. Also, a machine learning approach for intrusion detection in Ambient Assisted Living environments has not been proposed yet.

III. METHODOLOGY

This paper proposes a system for detecting attacks by differentiating anomalies from normal data flow in AAL network traffic. The goal of the system is to detect an anomaly when an attack occurs, i.e., the assumption is that the network behavior will deviate from the normal pattern of behavior and this way the anomaly can be detected[7,15]. To do so, four commonly used machine learning algorithms for real-time anomaly detection: Naïve Bayes (NB), Random Forest (RF), AdaBoost (AB) and K Nearest Neighbors (KNN) have been trained and compared. The classification task is binary, where the anomalies are labeled as 0, while the normal data flow is labeled as 1.

A. Data set selection

To evaluate the approach, a publicly available data set - "IoT Intrusion Dataset 2020" was used, created by a group of researchers from the University of Ontario, Canada[16]. The data set was obtained by simulating network traffic in an AAL network using a smartphone, a security camera, a voice recognition speaker, and several computers. The data set consists of network packets monitored at different time intervals. Table 1 shows the ratio of normal to packets belonging to a particular type of attack. Most anomaly packets refer to the most common attacks that can occur in an IoT network: Mirai botnet, Denial of Service, Scan Port OS, and Man In The Middle (MITM) [17]. Each packet has certain characteristics - package length, source address, destination address, source port, destination port, etc.

B. Data processing and feature analysis

Firstly, preprocessing of the data set was performed. It was purged of zero values, infinite values and incompatible data types. Additionally, feature selection was performed using Random Forest algorithm. The empirical analysis showed the

five most informative features and the analysis was continued using only these five features: Flow Duration, Fwd Packet Length Std, Flow IAT Max, Flow Bytes/s, Fwd Packet Length Min.

Table 1: Packet classification

Packet type	Number of packets
Normal	229140
Mirai-UDP Flooding	183554
Mirai-Hostbruteforce	121181
DoS-Synflooding	59391
Mirai-HTTP Flooding	55818
Mirai-Ackflooding	55124
Scan Port OS	53073
MITM ARP Spoofing	35377
Scan Hostport	22192
Bot	1966

IV. RESULTS AND DISCUSSION

A. Work environment

The following configuration was used to train and test the refined data set: MacBook Pro with eight-core M1 chip, 8-core graphics card, 16-core Neural Engine and 8 GB RAM. The M1 chip is the first chip system designed by Apple based on the ARM architecture. This chip has the ability to learn 15 times faster than its predecessors, which use Intel architecture. [5]

B. Train and test data set

As can be seen from Table 1, 229140 packets are marked as normal, while the other packets are marked as packages that contain an anomaly (attack). 80% of the data are used in the training of algorithms, while 20% of the data are used for testing and evaluation. In order to be able to compare the performance of the models and find the algorithm that is best for this problem, it is necessary to use several objective statistical indicators - accuracy, precision, sensitivity and F1 score.

Table 2: Results of performed experiments

Algorithm	Accuracy	Precision	Sensitivity	F1 score	Execution time
Naïve Bayes	0.78	0.88	0.62	0.63	0.57
Random Forest	0.91	0.93	0.85	0.88	2.25
Ada Boost	0.94	0.93	0.91	0.92	12.49
K Nearest Neighbors	0.97	0.96	0.95	0.96	17.0054

C. Results obtained and discussion

The experiment was performed ten times for each of the selected algorithms. Table 2 shows the average values of the results obtained.

As can be seen, the RF algorithm, as well as the AB algorithm provide high accuracy and precision (91-94%), but the sensitivity of RF (85%) as well as the F1 result (88%) are lower than those of AB (91%, 92%). In terms of execution time, it can be noted that the AB algorithm runs significantly longer (12.4 seconds) than RF (2.24 seconds) and NB (0.5 seconds).

The KNN algorithm has the longest execution time (17 seconds), but with this algorithm the highest percentages of accuracy (97%), accuracy (96%), F1 score (96%), as well as sensitivity (95%) can be observed. The NB algorithm runs the fastest, but it has the lowest results for all other statistics.

In general, what can be said is that KNN as an algorithm, despite the longer execution time, still gives results that are closest to 100% in terms of accuracy, precision, F1 result and sensitivity. It should also be noted the RF algorithm, which despite the lower sensitivity scores and F1 score, still shows significant accuracy and precision with much shorter execution times. This is especially important in real-time systems.

V. CONCLUSION AND FUTURE WORK

In this paper, research has been done to detect anomalies in IoT-based AAL environment with the help of several machine learning algorithms. Using the characteristics of the data set, the KNN algorithm obtained an accuracy of 97% at an average execution time of 17 seconds. 93% accuracy was also obtained with an average execution time of 2.2 seconds on the RF algorithm.

As future work, these models can be expanded with additional features, in order to improve the previously obtained results. Additionally, other machine learning and deep learning algorithms could be investigated. As a further means of increasing the amount of data and improving the diversity of considered attacks, federated learning could be incorporated.

These approaches are expected to improve the results for classifying anomalies in the Internet of Things in real time.

ACKNOWLEDGMENT

This work has been supported by the WideHealth project - European Union's Horizon 2020 research and innovation programmender grant agreement No. 95227.

REFERENCES

- [1] K. Kimani, V. Oduol, K. Langat, "Cyber Security Challenges for IoT-based Smart Grid Networks", *International Journal of Critical Infrastructure Protection*, vol.25, 2019.
- [2] D. Satria, H. Ahmadian, "Designing Home Security Monitoring System Based Internet of Things(IoTs) Model", *Jurnal Serambi Engineering*, vol.3, 2018.
- [3] V. Venkatesh, V. Vaithyanathan, P. K. Murali, P.R. Chelliah, "A secure Ambient Assisted Living (AAL) environment: An implementation view.", *International Conference on Computer Communication and Informatics (ICCCI)*, 2012.
- [4] E. Hodo, X. Bellekens, A. Hamilton, P. Dubouilh, E. Iorkyase, C. Tachtatzis, R. Atkinson, "Threat analysis of IoT networks Using Artificial Neural Network Intrusion Detection System", *3th International Symposium on Networks, Computers and Communications (ISNCC)*, 2016.
- [5] T. Wilder, A. Bender, "Apple unleashes M1", <https://www.apple.com/newsroom/2020/11/apple-unleashes-m1/>, 2020.
- [6] A. Mishra, A. Dixit, "Resolving Threats in IoT: ID Spoofing to DDoS," *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-7, 2018.
- [7] M. M. Shurman, R. Khrais, A.R. Yateem, "IoT Denial-of-Service Attack Detection and Prevention Using Hybrid IDS", *2019 International Arab Conference on Information Technology (ACIT)*, pp. 252-254, 2019.
- [8] M. Nobakht, "IoT-NetSec: Policy-Based IoT Network Security Using OpenFlow", *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2019.
- [9] M. Tavallae, E. Bagheri, W. Lu, A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pp. 1-6, 2009.
- [10] M.A. Ferrag, L. Maglaras, S. Moschoyiannis, H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study". *Journal of Information Security and Applications*, vol.50, 2019.
- [11] N. Moustafa, J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," *2015 Military Communications and Information Systems Conference (MilCIS)*, Canberra, pp. 1-6, 2015.
- [12] V. Golmah, "An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM", *International Journal of Database Theory and Application*, vol. 7, No. 2, pp. 59-70, 2014.
- [13] S. A. Hajare, "Detection of Network Attacks Using Big Data Analysis", *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 4, issue 5, pp. 86-88, 2016.

- [14] S.S. Tanpure, G. D. Patel, Z. Raja, J. Jagtap, A. Pathan. "Intrusion Detection System in Data Mining using Hybrid Approach.", International Journal of Computer Applications, 2016.
- [15] J. Veeramreddy, V.V.R, Prasad, K. M. Prasad, "A Review of Anomaly based Intrusion Detection Systems", International Journal of Computer Applications, pp.28-35, 2011.
- [16] I. Ullah, Q.H. Mahmoud, "A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks", Goutte C., Zhu X. Advances in Artificial Intelligence, Canadian AI 2020, Lecture Notes in Computer Science, vol 12109, Springer, Cham, 2020.
- [17] I. Ahmad, R. Ziar, M. Niazy, S.Khan, "Survey on IoT: Security Threats and Applications", Journal of Robotics and Control (JRC), vol.2, 2020.
- [18] H.R. Zeidanloo, F. Hosseinpour, P. Najafi, "Botnet detection based on common network behaviors by utilizing Artificial Immune System(AIS)", 2nd International Conference on Software Technology and Engineering, 2010.

Investigating Presence of Ethnoracial Bias in Clinical Data using Machine Learning

Bojana Velichkovska¹, Hristijan Gjoreski¹, Daniel Denkovski¹, Marija Kalendar¹, Leo Anthnoy Celi², Venet Osmani³

¹Ss. Cyril and Methodius University, Faculty of Electrical Engineering and Information Technologies, Skopje, R. N. Macedonia

²Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

³Fondazione Bruno Kessler Research Institute, Trento, Italy

{[bojanav](mailto:bojanav@feit.ukim.edu.mk), [hristijang](mailto:hristijang@feit.ukim.edu.mk), [daniield](mailto:daniield@feit.ukim.edu.mk), [marijaka](mailto:marijaka@feit.ukim.edu.mk)}@feit.ukim.edu.mk, leoanthnoyceli@yahoo.com, vosmani@fbk.eu

Abstract— An important target for machine learning research is obtaining unbiased results, which require addressing bias that might be present in the data as well as the methodology. This is of utmost importance in medical applications of machine learning, where trained models should be unbiased so as to result in systems that are widely applicable, reliable and fair. Since bias can sometimes be introduced through the data itself, in this paper we investigate the presence of ethnoracial bias in patients' clinical data. We focus primarily on vital signs and demographic information and classify patient ethnoraces in subsets of two from the three ethnoracial groups (African Americans, Caucasians, and Hispanics). Our results show that ethnorace can be identified in two out of three patients, setting the initial base for further investigation of the complex issue of ethnoracial bias.

Keywords—ethnoracial bias, clinical data, vital signs, machine learning

I. INTRODUCTION

Machine learning (ML) research is becoming increasingly focused on addressing complex healthcare problems and as a result establishing ML systems for clinical decision support, including diagnostic, prognostic, and risk prediction. In this regard, ML applied to healthcare has already shown significant results [1] [2], which may develop beyond recommendations of clinical actions, towards full-scale assistance such as autonomous triage and patient stratification.

However, the issue of bias in ML research is gaining increasing attention since the results are as reliable as the process is objective. This means that generated results should be unbiased throughout all the stages of the ML process: data collection, data preparation (from data selection to data preprocessing), model configuration, and model training and validation. There are two important aspects to consider: bias can be inherent in the data used in the research or stem from the ML methodology used in the research. Whether the bias is introduced from the data itself or in the development methodology, it presents a significant challenge in terms of trustworthiness of the models and worse can lead to unfair decision making, potentially harming disadvantaged groups, including gender, races and ethnicities.

An important hindrance for increased application of ML models are bias conflicts which must be addressed. An opinion piece [3] states a “silent curriculum” in medical practice teaches students to differentiate between patients based on their race, saying, “among two patients in pain waiting in an emergency

department examination room, the white one is more likely to get medications and the black one is more likely to be discharged with a note documenting narcotic-seeking behavior”.

Biased medical practice results in ethnoracially unfair medical trials that produce datasets biased towards the majority population, e.g., imbalanced datasets with dominant representation of one ethnorace over the others [4] [5], or datasets obtained entirely from one ethnoracial group. The study in [6] shows that, even though ethnorace influences response to cancer treatments and outcomes, no ethnoracial statuses are recorded in majority of patients, and in cases of recorded ethnorace the highest represented ethnorace in melanoma, breast and lung cancer trials are White people (25.94%), followed by Asians (4.97%), and African Americans (1.08%), resulting in overrepresentation. Additionally, melanoma is one of the deadliest skin cancers known, yet melanoma datasets have shown underrepresentation of different ethnoraces [7].

Working with biased datasets can influence development and produce biased ML applications. There have been many reports of detected racial bias in medical ML applications. The study in [8] shows patients being assigned a risk score depending on their skin color; namely, Black patients which are placed in the same risk category as a subset of White patients, health-wise had considerably worse symptoms. To add to the severity of the problem, the ML algorithm reduced the number of Black patients which should have been assigned additional care by more than half. Another example is an algorithm for diagnosis of diabetic retinopathy showing poor performance in populations living outside of the location where it was developed [9].

Analysis of racial bias in ML applications can also be performed by observing the model's performance over different ethnoraces [10]. In [11] the authors present their investigation into the performance of three severity scoring systems in four ethnoraces, focusing on hospital mortality; their results show all three models overestimated mortality across all ethnoraces, however, they conclude that severity scores have statistical bias since the overestimated mortalities are most notable with Hispanic and Black patients.

From our investigation it appears there are no existing analysis that focus on detection of racial bias in clinical data itself and hence this is the focus of this paper. Typically, when approaching an ML problem, data preprocessing almost always includes removal of features which could potentially introduce

faulty or prejudicial bias in the results, such as gender and ethnicity. However, vital signs, including blood pressure, heart rate and oxygen saturation, are used in clinical data research, and considered unbiased. They are therefore presumed free of information which could be prejudicial in any way. From here, we raise the question, “Is it possible for data, supposedly stripped of biased information, to still incorporate bias in the results?”. Therefore, our aim is finding out if we can detect bias in clinical data, that is, identify ethnicity and race based on vital signs and demographic information only.

The rest of the paper is organized as follows. In section II we describe the dataset, our data preparation process, and our approach to the problem. In section III we present our results, whereas in section IV we discuss them. Section V concludes this paper.

II. METHODOLOGY

A. Dataset

The dataset used for this paper, eICU Collaborative Research Database (eICU-CRD) [5], contains extensive records of 200,859 patient admissions in the ICU. In this research we focus on the patient’s general information (PGI) and patient’s vital signs measurements (PVS), and are listed along with their units in Table I. During the data preparation process, we consider patients under the age of 89, where all PGI of interest and PVS are present (so called complete case analysis). The PGI we use are patient’s age, height, admission weight, and discharge weight, whereas the PVS we use are statistical features (mean, minimum, maximum, variance, and standard deviation) extrapolated from the patient’s heart rate, oxygen saturation, respiration, and blood pressure (systolic, diastolic, and mean) within the first 24 hours of admission. Patients with undocumented or undefined ethnicity, and patients missing PGI were excluded from this analysis.

After applying the selection criteria, the dataset contained small pools of patients from certain ethnic/racial groups (such as Asian and Native American). Therefore, for our analysis we selected only the three predominant ethnicities present in the remaining dataset, namely Caucasian, African American and Hispanic. The distribution of patients per ethnic/racial group is given in Fig. 1. As the chart shows, the resulting dataset is highly imbalanced, in favor of Caucasians. The African American and

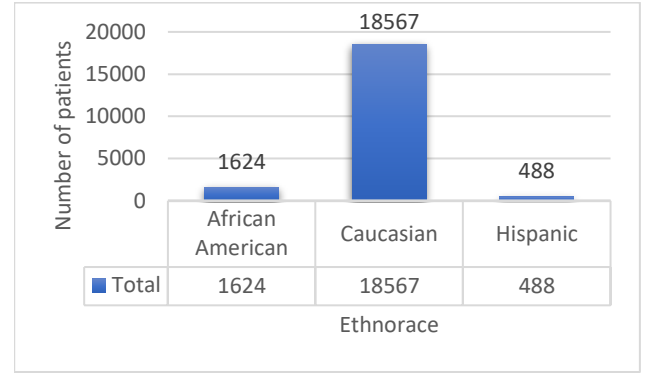


Fig. 1 Patient number per ethnicity

Hispanic ethnicities have a significantly lower number of patients, with the Hispanic group being the smallest.

B. Model development and validation

We classify patients’ ethnicities in subsets of two from the three ethnicities we have, i.e., we have three distinct comparative tests, Caucasian vs African American, African American vs Hispanic, and Caucasian vs Hispanic patients. For each comparative test, we evaluate the performance of two ML algorithms, namely logistic regression (LR) and XGBoost. LR predicts the probability of dichotomous target variables. XGBoost is an efficient implementation of the gradient boosted trees, which makes predictions by ensembling new models to correct the errors made by existing models, until no further improvements can be made. The evaluation of the models is performed internally using stratified five-fold cross validation. In five-fold cross validation, four folds are used for the model training, whereas the remaining fold is used for testing the model’s performance. We use stratified cross validation in order to maintain the original classes’ distribution.

The assessment of our models is performed by computing the area under the receiver operator characteristic curve (AUC-ROC), the area under the precision-recall curve (AU-PRC), and additional metrics, including positive predictive value (PPV), negative predictive value (NPV), F1 score, and recall for each model. The AUC-ROC curve shows the trade-off between true positive rate (TPR) and false positive rate (FPR). TPR (also known as recall and sensitivity) is the proportion of samples correctly predicted as positives out of all positive observations. FPR is the proportion of samples incorrectly predicted as positives out of all negative observations. Classifiers with curves closer to the top-left corner have better performance compared to classifiers with a curve closer to the 45-degree diagonal. The AU-PRC shows the trade-off between precision or PPV and TPR. PPV represents the proportion of samples predicted correctly as positives out of all samples predicted as positives. NPV represents the proportion of samples correctly predicted as negatives and all samples predicted as negatives. F1 score is the harmonic mean of precision and recall.

III. EXPERIMENTAL RESULTS

Table II shows the difference between each pair of ethnicities (classes) in our dataset. The ratios provided are important indicators when analyzing the performance of our models, since the AU-PRCs obtained are influenced by data

TABLE I DATASET: VARIABLES AND UNITS

Variable	Unit
Patient’s general information	
Age	years
Height	cm
Admission Weight	kg
Discharge Weight	kg
Patient’s vital signs	
Heart Rate	bpm
Oxygen Saturation	%
Respiratory Rate	insp/min
Blood Pressure (systolic, diastolic, mean)	mmHg

TABLE II PATIENT NUMBER RATIO BETWEEN MAJORITY AND MINORITY CLASSES IN EACH COMPARATIVE TEST

Dataset: Class Distribution		
Majority class	Minority class	Patient Number Ratio
Caucasian	African American	11.42
African American	Hispanic	3.32
Caucasian	Hispanic	37.90

ratio. We can see that the highest data imbalance occurs between Caucasians vs Hispanics, whereas the lowest data imbalance can be seen in African Americans vs Hispanics.

Our results consist of two sets: first, using imbalanced data in the training process, and second, using balanced data in the training process. In both cases we maintained the original class ration for the test data.

A. Imbalanced train data

Initially, we used an imbalanced dataset, both for training and testing the models. All three comparative tests showed the model was biased in favor of the majority class. Since the ratio between African Americans and Hispanics is the lowest imbalance ratio in our dataset, we decided to show the performance of the ML models trained on the original distribution of data from these two classes.

TABLE III shows confusion matrices taken from a random fold for LR [12] and XGBoost [13], and for both classifiers most of the patients are classified as part of the majority class. Therefore, these results clearly illustrate that the imbalance in the data made the models biased towards the majority class. Caucasian patients are 11.42 times more than African American patients and 37.9 times more than Hispanics, which further accentuates the bias towards Caucasians, when using the imbalanced dataset for training.

B. Balanced train data

Since the imbalanced dataset proved to be biased towards the majority class and resulted in models placing most of the samples in the dominant class, understandably there was low model performance. In order to give the models a learning chance, we decided to correct the imbalance present in the dataset, during the training process. To achieve this, we randomly under sampled the majority class (using RandomUndersample [14]) in the training data, as to balance the train dataset. We train the classifiers with the balanced data, while the test data for each fold keeps the original distribution in order to evaluate the performance of each model on real data.

TABLE III AFRICAN AMERICAN VS HISPANIC - CONFUSION MATRICES. TAKEN FROM A RANDOMLY SELECTED FOLD.

African American vs Hispanic				
Models	Logistic Regression		XGBoost	
	Predicted			
Actual	316	9	304	21
	96	2	87	11

Using the balanced train data, we illustrate results from three comparative binary classifications between each two ethnoraical groups – firstly, we have Caucasians vs African Americans, next we have African Americans vs Hispanics, and lastly, we have Caucasians vs Hispanics.

The results of each comparative test for each classifier are summarized in TABLE IV, TABLE V, and TABLE VI. 95% confidence intervals are provided in the brackets. Additionally, the AUC-ROC and AU-PRC are illustrated in Fig. 2, where the 95% confidence intervals for the classifiers are shown in their corresponding coloring with lower opacity.

IV. DISCUSSION

From the results, it can be observed that XGBoost performed best in classifying Caucasians vs African Americans, whereas the other two comparative tests give weaker results. For both African Americans vs Hispanics and Caucasians vs Hispanics, XGBoost shows significant similarity in the AUC-ROC curve. On the other hand, LR has the worst performance for Caucasians vs African Americans, and the best performance for Caucasians vs Hispanics. This outcome is understandable, because LR is known to operate well even with small sample sizes, which is the case in the last comparison. However, from the confidence intervals for both classifiers along all comparisons we can see that XGBoost has a narrower range around the estimate, which means that the estimate provided by XGBoost is more stable compared to the estimate given by LR.

Observing the additional metrics obtained in the experiments we can see that throughout all of them the confidence intervals are narrower for XGBoost.

The PPVs for all the experiments are low. However, the values for the recall (which show us the number of correctly

TABLE IV CAUCASIAN VS AFRICAN AMERICAN - RESULTS. CONFIDENCE INTERVALS PROVIDED IN BRACKETS.

Metric	LR	XGBoost
AUC	0.583 [0.552 – 0.614]	0.726 [0.703 – 0.749]
PPV	0.121 [0.114 – 0.128]	0.138 [0.134 – 0.142]
Recall	0.558 [0.501 – 0.615]	0.640 [0.599 – 0.681]
NPV	0.944 [0.938 – 0.950]	0.954 [0.950 – 0.958]
F1	0.199 [0.186 – 0.212]	0.227 [0.220 – 0.234]

TABLE V AFRICAN AMERICAN VS HISPANIC - RESULTS. CONFIDENCE INTERVALS PROVIDED IN BRACKETS.

Metric	LR	XGBoost
AUC	0.671 [0.621 – 0.721]	0.675 [0.634 – 0.716]
PPV	0.317 [0.292 – 0.342]	0.310 [0.291 – 0.329]
Recall	0.601 [0.535 – 0.667]	0.594 [0.530 – 0.658]
NPV	0.836 [0.816 – 0.856]	0.833 [0.811 – 0.855]
F1	0.414 [0.379 – 0.449]	0.407 [0.377 – 0.437]

TABLE VI CAUCASIAN VS HISPANIC - RESULTS. CONFIDENCE INTERVALS PROVIDED IN BRACKETS.

Metric	LR	XGBoost
AUC	0.737 [0.693 – 0.781]	0.648 [0.636 – 0.660]
PPV	0.045 [0.040 – 0.050]	0.044 [0.043 – 0.045]
Recall	0.602 [0.524 – 0.680]	0.627 [0.616 – 0.638]
NPV	0.985 [0.982 – 0.988]	0.985 [0.984 – 0.986]
F1	0.083 [0.073 – 0.093]	0.081 [0.080 – 0.082]

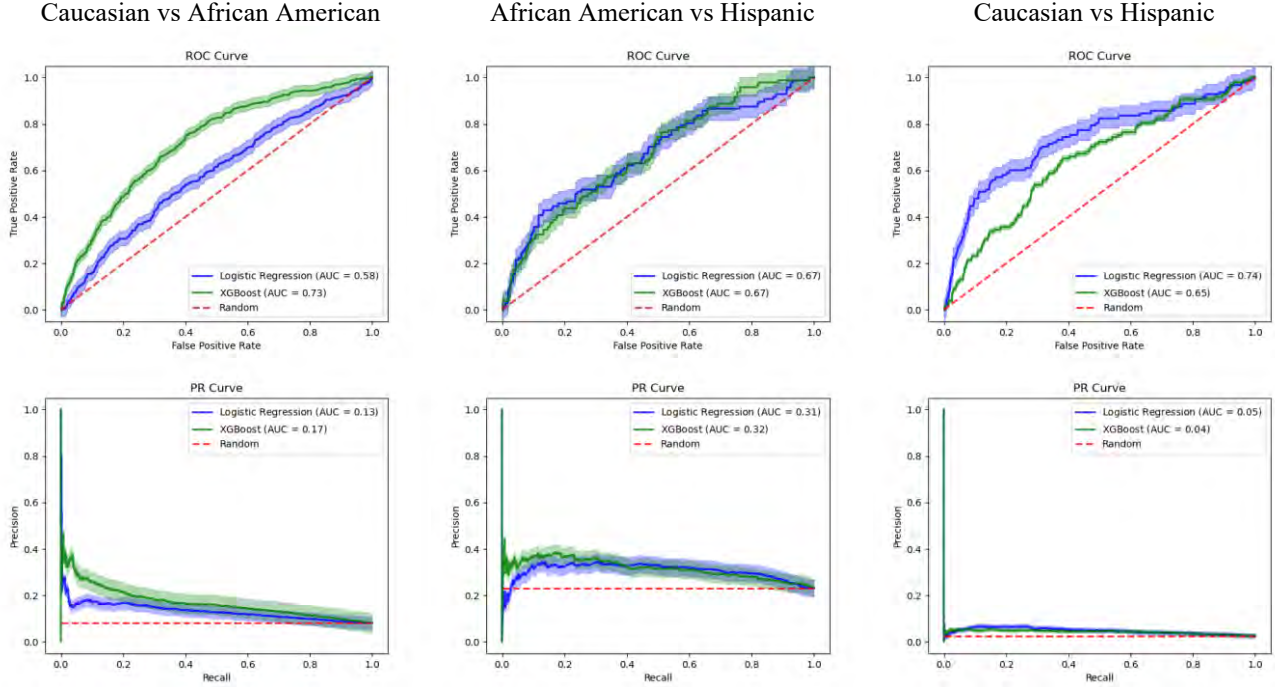


Fig. 2 Performance on each model across three comparative tests. The top row represents AUC-ROC, the bottom row represents AU-PRC. Confidence intervals are shown with lower opacity.

returned patients divided by the number of patients which should have been returned) show that on average two thirds of patients are correctly classified, as is further illustrated with the randomly selected confusion matrices provided in TABLE VII.

These results show that patient’s general information and vital signs could include ethnoracial bias. Perhaps this bias arises from the bias present in medical practice, e.g., Black or Hispanic patients admitted to the ICU might be in a worse condition than White patients. Another potential reason can be similarities in general information and biological markers (e.g., height, weight, heart rate) of patients that represent an ethnoracial group.

However, these results are not conclusive. Ethnoraces can be difficult to identify due to interracial marriages, and we cannot claim with certainty that Caucasians “misclassified” as African Americans, are not biracial or even multiracial. Furthermore, the

eICU dataset consists of patients with various diagnosis, e.g., rhythm disturbances, pneumonia, aneurysms, and so on, and every diagnosis influences different vital signs in different ways. Additionally, the results are obtained on a small number of African American and Hispanic patients, which might not give an accurate representation on these ethnoracial groups.

V. CONCLUSION

With the increased number of ML applications in medicine it is important to ensure the developed models are unbiased and perform correctly in spite of a patient’s ethnorace. Since bias can be introduced through data, we investigated the presence of ethnoracial bias in clinical data; more specifically, we analysed general information and vital signs of patients from three ethnoraces to determine whether ML models can detect biological markers representative of an ethnorace. We compared the performance of two ML algorithms in comparing two by two ethnoraces in balanced train data. Our results show that two out of three patients in all experiments are placed in the correct ethnorace; however, the sample size of the observed ethnoraces as well as the fluid concept of ethnorace indicate the need for further investigation.

ACKNOWLEDGMENT

Part of the study was supported by WideHealth project - Horizon 2020, under grant agreement No 95227.

REFERENCES

- [1] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., de Marvao, A., Dawes, T., O’Regan, D.P., Kainz, B., Glocker, B. and Rueckert, D. (2018). Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation. *IEEE Transactions on Medical Imaging*, 37(2), pp.384–395.

TABLE VII CONFUSION MATRICES FOR ALL EXPERIMENTS. TAKEN FROM A RANDOMLY SELECTED FOLD.

	Caucasian vs African American		African American vs Hispanic		Caucasian vs Hispanic	
Logistic Regression						
	Predicted					
Actual	2308	1406	211	114	2430	1284
	124	201	53	45	35	62
XGBoost						
	Predicted					
Actual	2429	1285	206	119	2398	1316
	109	216	45	53	37	60

- [2] McCoy, A. and Das, R. (2017). Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Quality*, 6(2), p.e000158.
- [3] Brooks, K.C. (2015). A Silent Curriculum. *JAMA*, 313(19), p.1909.
- [4] Gutman, D., Codella, N.C.F., Celebi, E., Helba, B., Marchetti, M., Mishra, N. and Halpern, A. (2016). Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). *arXiv:1605.01397 [cs]*. [online] Available at: <https://arxiv.org/abs/1605.01397>
- [5] Pollard, T.J., Johnson, A.E.W., Raffa, J.D., Celi, L.A., Mark, R.G. and Badawi, O. (2018). The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific Data*, 5(1).
- [6] Guerrero, S., López-Cortés, A., Indacochea, A., García-Cárdenas, J.M., Zambrano, A.K., Cabrera-Andrade, A., Guevara-Ramírez, P., González, D.A., Leone, P.E. and Paz-y-Miño, C. (2018). Analysis of Racial/Ethnic Representation in Select Basic and Applied Cancer Research Studies. *Scientific Reports*, 8(1).
- [7] Das, S. (2021). Automated Bias Reduction in Deep Learning Based Melanoma Diagnosis using a Semi-Supervised Algorithm. [online] Available at: <https://doi.org/10.1101/2021.01.13.21249774>, 2021.
- [8] Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, [online] 366(6464), pp.447–453. Available at: <https://science.sciencemag.org/content/366/6464/447.full>.
- [9] W. D. Heaven, "Technology Review," 2020 April 27. [Online]. Available at: <https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/>.
- [10] Noseworthy, P.A., Attia, Z.I., Brewer, L.C., Hayes, S.N., Yao, X., Kapa, S., Friedman, P.A. and Lopez-Jimenez, F. (2020). Assessing and Mitigating Bias in Medical Artificial Intelligence. *Circulation: Arrhythmia and Electrophysiology*, 13(3).
- [11] Sarkar, R., Martin, C., Mattie, H., Gichoya, J.W., Stone, D.J. and Celi, L.A. (2021). Performance of intensive care unit severity scoring systems across different ethnicities in the USA: a retrospective observational study. *The Lancet Digital Health*, 3(4), pp.e241–e249.
- [12] Peng, C.-Y.J., Lee, K.L. and Ingersoll, G.M. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1), pp.3–14.
- [13] Chen, T. and Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*.
- [14] Hasanin, T. and Khoshgoftaar, T.M. (2018). The Effects of Random Undersampling with Simulated Class Imbalance for Big Data. In: *IEEE International Conference on Information Reuse and Integration for Data Science*. pp.70–79.

Is Web Transforming Our Minds and Where is Our Civilisation Going to?

Matjaž Gams

Institut »Jožef Stefan«

Jamova 39, Ljubljana, Slovenia

matjaz.gams@ijs.si, <https://dis.ijs.si/mezi/>

Abstract—This paper analyses civilization dangers that Bill Gates and Elon Musk warn about: pandemics, demographic and AI. In addition, the undesired societal changes especially in relation to the mind tampering and the Web are observed. The analysis about longevity of human civilization seems to indicate that we humans on our own are not smart enough to find a solution. Luckily, there is a possible solution: to create and cooperate with superintelligence.

Keywords: civilization dangers, artificial intelligence, demographics, Web, media objectivity

I. INTRODUCTION

Before the COVID-19 pandemics, practically nobody thought it could happen even though Bill Gates warned about it five years ago at the TED talk with a title: “The next outbreak? We are not ready”ⁱ. Yet, pandemics COVID-19 happened in 2020 and we were indeed not ready. It should be noted that Bill Gates is one of the richest and most influential people in the world. Why is humanity so blind for the forthcoming major dangers that scientists and visionaries present reasonable arguments for? Can we learn anything from this tragic event?

Elon Musk is probably the greatest technological genius and visionary of current times. He mass-introduced electric cars (Tesla) to our planet and is planning several space activities including Starlink and going to Mars to avoid human extinction in case anything catastrophic happens to our planet. Most often he mentions three major civilization dangers: artificial intelligence (AI), demographics and environment. These and related civilization dangers and major shifts in recent years are the major topics of this paper.

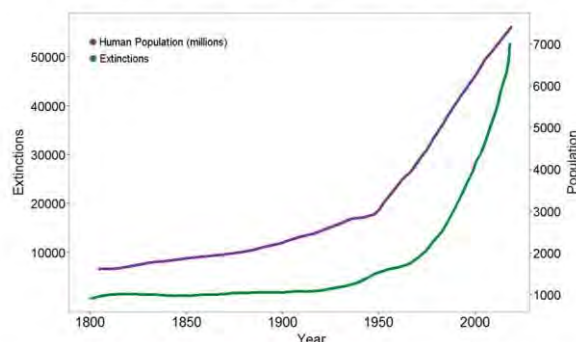
ⁱThe TED talk is available at:

https://www.ted.com/talks/bill_gates_the_next_outbreak_we_re_not_ready

II. HUMAN PROGRESS THROUGH DEMOGRAPHIC CHANGES

When analyzing recent progress of human civilization, one must first note the demographic changes (problematic), and second the growth of the knowledge, science, technology and the information society (positive). In terms of demographic changes, in the 20th century there was a clear exponential growth after World War II. Figure 1 indicates not only the exponential growth of the human population, but also the exponential growth of the animal extinctions – we humans are already overpopulating our planet causing havoc among animals and plants and the environment.

Humans & The Extinction Crisis



Data source: Scott, J.M. 2008. *Threats to Biological Diversity: Global, Continental, Local*. U.S. Geological Survey, Idaho Cooperative Fish and Wildlife, Research Unit, University Of Idaho.

Fig. 1. Growth of human population and animal extinction.

With fertility (i.e. number of children born per woman) around 5 as in the 20th century, the human population would reach 1 person per m² in 13 generations. In 40 generations there would mathematically be 1 person per each kilogram of our planet, which is obviously impossible [1]. At the same time, animal extinction is 100x faster than it was a century ago; in the last 40 years there are 50% animals less; a study in Germany showed that in 27 years there are 75% less

flying insects. Therefore, the population growth is unsustainable for a larger period of time [2].

On the other hand, latest demographic studies [3] predict that the world population growth will stop in 2065, therefore the halting mechanisms in our society are already long triggered and effective. The paper was supported by the Melinda and Bill Gates foundation. Bill Gates as probably the most important person whose life in recent decades is dedicated to saving our planet and our civilization, is also intensively working on the demographic issues, in particular how to stop the exponential growth of the human population in particular in specific areas like Africa. However, the tide is turning, the number of newborns in the world is not growing anymore for the last two decades and the human population mainly increases due to longer life span. The new danger not becoming evident, even more surprisingly - it still causes disbelief and public rejection [4] is first of all the danger of extinction of smaller nations and languages, thus causing globalization, unification and corresponding reduction of world cultures, languages and overall richness of human civilization. The one world speaking one language would be much poorer in the mental, cognitive and cultural sense. In addition, some studies indicate that the danger of civilization collapse would increase significantly with growing globalization. Currently, there are still major blocks such as USA and China, but one global village would be a potential disaster since all civilizations inevitably face saturation and decline [5]. In case of many sub-civilizations, other take the lead after the major ones stumble. In case of one global civilization only, there is no one to reboot human progress.

III. THE PREDICTED TIMESPAN OF THE HUMAN TECHNOLOGICAL CIVILIZATION

Several authors [6][7] tried to predict the longevity of the human technological civilization, e.g. the one that is technologically at the level of sending data to the universe. Civilizations inevitably leave some energy traces because of their activities and it must be detectable. Scientists have for decades performed more and more advanced studies to find other civilizations, they investigate planets, habitable planets and signs of life. No sign of life was detected elsewhere except potentially Mars. No civilization was detected and from the first "Where are they" proclaimed by Fermi in 1950, 4 years before the death of Alan Turing, the father of digital computers [7][8] till now no trustful and repeatable sighting was reported. It seems that we are alone in our galaxy, or at least in our part of our galaxy. That leads to two hypotheses:

1. either we are indeed the first (or one of the first) technological civilization or
2. civilizations are of limited time span.

Unfortunately, the probability of the first hypothesis given the age of our galaxy is irrelevant compared to the second option.

The studies performed at the Jozef Stefan Institute [5][6] indicates that most likely the longevity of our technological civilization is around 1.000 to maybe 10.000 years. Most likely, we will destroy ourselves since humans are on Earth for millions of years and no major destruction was detected observable as major extinction in that period of time. Historical finding indicate only five major extinctions happened in the history of our planet, the first one 440 million years agoⁱⁱ and therefore the Earth is quite a safe planet. More likely, the growing powers combined of humans with irresponsible governing will cause the collapse of our civilization. Or, we will just sadly change our orientations towards internal issues such as hedonism not going to other planets and other stars. Another danger lurks in the globalisation and mind laundering as presented in the next section.

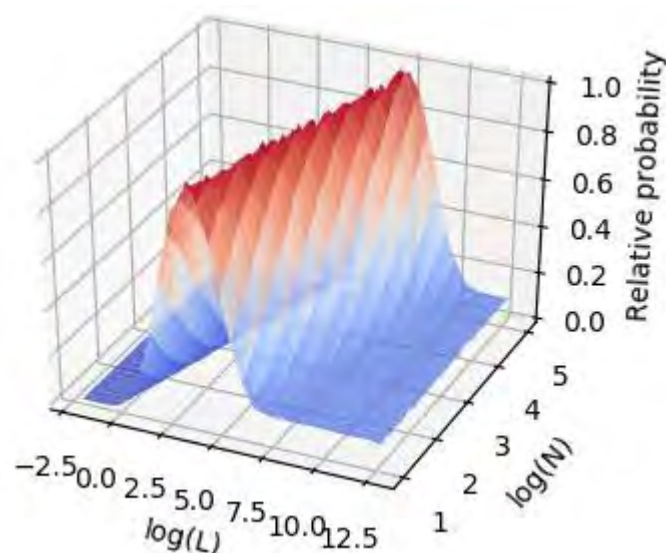


Fig. 2. Predictions of human longevity (L) depending on the number of civilizations in our galaxy (N) with one of the models designed by the author's group.

IV. EXPLOSION OF FAKE NEWS, MYTHS AND MIND TAMPERING

With the Donald Trump's candidacy for USA president, the fake news epidemics of the news media erupted even before the fake news about COVID-19. If one remembers only the false or even conspiracy theories promoted by mass media, e.g. mask recommendations by the major world health organizations: first they were supposed to be irrelevant and was no use of wearing masks. Second, they became recommended, then some even proposed two, one on top of the other. Another example: the COVID-19 virus was proposed by Donald

ⁱⁱhttps://www.mdpi.com/journal/geosciences/special_issues/mass_extinctions

Trump to be manufactured in China. Mass media denoted the idea false, ideological and even racist and forcefully rejected any study in that direction. As of the end of May 2021, when this paper is being written, Facebook deleted all posts about laboratory-generated COVID-19 virus. However, about this time several information previously censured reappeared at least making the possibility of human-generated virus as or even more likely than the natural source. It is interesting to read these lines in the proceedings of the conference in Autumn 2021 and compare it to the current positionsⁱⁱⁱ. Even before the pandemics, in the era of Donald Trump, around half of all Americans (49%) indicated in the questionnaires that the media is very biased, according to the trustful Gallup / Knight poll “American Views 2020: Trust, Media and Democracy,” polling 20,000 Americans [9]. The majority of Americans believed that the media are becoming propaganda tools; 74% of them believed that the writing of the media is directed by the owners, which is 5% more than in 2017. They were of the opinion that biased reporting is purposeful.

The consequences of mass media deliberate propaganda are being highlighted by commercials, and the Web. Altogether they lead to a conclusion that a large part of population, and particular the younger population, is under strong mental and cognitive influence affecting their behavior. The effectiveness of these mechanisms might still not exceed the effects of the strongly censured media in the dictator’s regimes, but is becoming close. In the tests performed by the author, the majority of younger population believed that women in Slovenia earn significantly less, 10-20%, whereas statistical data shows that the income per an hour of work is practically identical, and similarly for the current pensions where women live seven years more on average. Statistical data also indicates that Indian Women in USA learn more than white males [10]. Also, majority of younger population in our studies thought that Donald Trump is not above average intelligent. How one person can believe that the president of any country, yet USA, can be of average intelligence, is not clear, unless the mass brain laundering comes into effect. And third, when these brain-washing effects (i.e. Coercive Persuasion) on the polled students are revealed to them, strong negative feelings are demonstrated – indicating that these objectively false ideas have penetrated deeply into the system of beliefs of the individual minds of young smart people. A disclaimer is needed: the purpose of the paper is not to allude to any political orientation or ideology; rather, the turning of young free-will students into followers of a specific globalist ideology is against the preferences of an advanced civilization. To make things worse, the online social networks in studies in recent years turned extremely harmful emotionally and in relation to science: they

enhance herd instincts through seemingly innocent mechanism like “likes”, encouraging the worst in the masses, attacks on dissidents, polarization, violence, revolution; increase in (self) violence, suicide, hatred, depression. AI recommendation algorithms unfortunately in the service of capitalism profits and other negative motivations confirm nonsensical ideas and again unfortunately, our brains against supercomputers with AI algorithms cannot defend in particular if the brains are young and super sensitive. Please note that AI primarily helps [11] and in general substantially enables the human progress so some negative misuse should not be messed with the overall contribution.

The dark side of the online media was revealed in public for example by The Digital Dilemma and Worldnet. The landmark report The Digital Dilemma was introduced by the Academy’s Science and Technology Council. It showed that several online mechanisms are in effect negative for individuals and society, with likes in social networks an example. The inventors expected that the likes will promote kindness and positive emotions all over the network, rewarding good ideas and positive feelings. In reality, it was the contrary: netizens behaved surprisingly primitive and performed massive brutal attacks on particular victims sometimes chosen randomly. The medieval which-hunt as if re-emerged from the dark centuries ago.

Second, the effect on an average netizen was negative in terms of feelings and loss of sense of reality. Several studies showed that the number of hours daily spent on digital networks was directly proportional to the amount of negative thoughts in a retrospective way, including thoughts of a suicide. Secondly, the number of people believing in proven wrong, unscientific opinions significantly increased. For example, the number of conferences about flat Earth is on constant rise in recent years. In Slovenia, a record percentage of population does not trust the vaccines, according to The Lancet paper, which puts Slovenia among the 5 worst countries in the world. In the last two years, some of our studies have revealed the sources and mechanisms of these negative effects.

V. DISCUSSION AND CONCLUSIONS

There are several dangers to the current progress of human civilization and the COVID-19 crises indicated that the dangers are not fictive, but lurking in the near future. Scientists need to study them in detail and perceive the dangerous scenarios in advance enabling humans to prevent or at least decrease the danger when they (some will probably inevitable) happen.

Among the dangers often mentioned by visionaries by philanthropists like Bill Gates and technological super geniuses like Elon Musk, the list includes: ecology, demography, AI, social eclipse and alike. We have briefly discussed some of them in this paper.

While the internet remains one of the best and most democratic media in the world and most helpful for the

ⁱⁱⁱ Note that the team from the author’s department won the second place at the worldwide XPRIZE competition for best COVID-19 measurements and 250.000 dollars award: <https://dis.ijs.si/?p=436>

progress of human civilization, many visionaries warned that eventually some will use the power of the online media for their own (dark) purposes. It is still not clear whether these negative side-effects are emergent or there are influential individuals or societies behind them or both.

Studies of longevity indicate that there are around 1.000 to 10.000 years of growth of human civilization ahead of us, and then a civilization collapse will inevitably happen. That seems confirmed by the statistics and the lack of contact with other civilizations. Yet, there is a potential solution that is rarely mentioned in the media: with the appearance of superintelligence, i.e. artificial intelligence superior to the human mind, there is a reasonable possibility that in synergy between AI and human AI we will conquer our galaxy and not struggle to survive.

REFERENCES

- [1] M. Gams, J. Malačič (ed.), *Bela knjiga slovenske demografije*. Ljubljana: Jozef Stefan Institute, 2019.
- [2] E. Kolbert, *The Sixth Extinction: An Unnatural History*. New York: Henry Holt and Company, 2014.
- [3] S. E. Vollset et. al., "World population likely to shrink after mid-century, forecasting major shifts in global population and economic power," *The Lancet*, vol 396, pp. 1285-1306, October 2020.
- [4] M. Gams, Presentation at a council meeting on demography in the National Council on demography, 2018, <https://www.youtube.com/watch?v=A4rai9zoNg0>.
- [5] B. Šircelj, L. Blatnik Guzelj, A. Zavrtanik Drglin, M. Gams, "Expected human longevity," In *Cognitive Science*, vol. II., T. Strle, T. Kolenik, Markič, Olga, Eds. Ljubljana: Information Society 2019, Jožef Stefan Institute, 2019, pp. 61-65.
- [6] A. Marinko, K. Golob, E. Jemec, U. Klun, M. Gams, "A new study of expected human longevity," In *Cognitive Science*, vol. II., T. Strle, J. Černe, Markič, Olga, Eds. Ljubljana: Information Society 2020, Jožef Stefan Institute, 2020, pp. 38-41.
- [7] J. O. Engler and H. von Wehrden. "Where is everybody?" an empirical appraisal of occurrence, prevalence and sustainability of technological species in the universe," *International Journal of Astrobiology*, vol. 6, pp. 499-505, January 2018.
- [8] N. Herzfeld, "Where Is Everybody?" Fermi's Paradox, Evolution, and Sin," *Theology and Science*, vol. 3, pp. 366-372, June 2019.
- [9] Knight Foundation, "American Views 2020: Trust, Media and Democracy," 2020, kf.org/usviews20.
- [10] The Economic Times, "Indian-American women earn more than white men in US: Report," 2016, https://economictimes.indiatimes.com/nri/nris-in-news/indian-american-women-earn-more-than-white-men-in-us-report/articleshow/50484732.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst.
- [11] T. Kolenik, M. Gams, "Persuasive technology for mental health: step closer to (mental health care) equality?" *IEEE technology & society magazine*, vol 1, pp. 80-86, March 2021.



ETAI 5: COMMUNICATION NETWORKS – 5G

Evaluation of Distributed NFV Infrastructures for Efficient Edge Computing in 5G

Gjorgji Ilievski

IT Department, Cyber Security
Makedonski Telekom AD
Skopje, Macedonia

Pero Latkoski

Faculty of Electrical Engineering and Information Technologies
Ss. Cyril & Methodius University
Skopje, Macedonia

Abstract—5G penetration in practice means increased demand for improvements on any possible level: network latency and bandwidth, computing, storage, security, etc. Bringing the resources at close proximity to the service consumers is one of the options for such improvements. 5G architecture is combined with Network Slicing and Network Function Virtualization (NFV) to segment the network and to instantiate the network functions. Distribution of the NFV Infrastructure (NFVI) is a key for the 5G Radio Access Network requirements for latency under 1ms. Designing the NFVI according to the specific needs, means that even some parts of the 5G Core network can be moved towards the edge, allowing offload from the devices at central datacenters and minimizing the backhaul traffic.

Our work is focused on distribution of the NFVI and we are analyzing two possible scenarios: distribution of the data plane while keeping the Management and Orchestration (MANO) of the NFV at central location and full distribution scenario where all components are distributed across multiple locations. We've made analytical models of the distributed NFVI and we are investigating the network packet sojourn time. We draw conclusions on the parameters that are impacting the network packet sojourn time and we compare the two possible distribution scenarios. The analysis is important for choosing the right architecture given different service scenarios.

Keywords— *Analytical Model, Distributed, MANO, NFV, SDN.*

I. INTRODUCTION

Today the digital industry is changing rapidly on a daily basis. Everyone must transform and adapt quickly to the changes. This continuous change comes with a cost boosting operational and capital expenses (OPEX and CAPEX) [1]. Virtualization was a first step in the series of evolutionary steps in the field, followed by cloud services, both private and public, going into a direction of economy of scale [2]. Software Defined Networking (SDN) is fully embedded into the virtualization scenarios allowing cloud providers with network abstraction, separation of control and user plane, to run network functions as virtual machines on a commodity hardware. That has improved the expenses, but public cloud providers have to make huge investments to move towards the direction of edge computing. Local providers and telecommunication providers can fill the gap, building edge network infrastructure with high availability, scalability, security, low latency and bandwidth. 5G Radio Access Network (RAN) combined with Network

Functions Virtualization (NFV) emerged as a novel network architecture concept which complements the SDN technology, creating a highly agile and dynamic environment [3], and chaining Virtual Network Function (VNF) elements to create end-to-end communication services.

5G mobile networks are already implemented in many parts of the world. They will allow the expansion of the Internet of Things (IoT) with massive machine to machine communication, ultra-reliable low-latency communications (URLLC) [4] which calls for data plane latency of less than 1ms and enhanced mobile broadband with significantly faster data speeds and greater capacity [5].

Centralized solutions are dependent on the connectivity with a central site, while distributed environments provide more scalability and adaptability, in the same time providing network transport offload, minimizing the backhaul traffic. To implement this, the location of the application processing and the network orchestration used in NFV, such as 5G core and virtualized RAN (vRAN), must be carefully considered [6]. Our investigation of the possible NFVI distribution scenarios dives into that direction.

The advantages and disadvantages of both a centralized and distributed control plane in the SDN architectures have been widely researched [7], [8], [9], [10]. The SDN and NFV technologies are complementary. The standardization of the NFV technologies has been guided by ETSI ISG NFV [11]. NFV significantly impacts the emerging 5G networks.

Due to the previous, this paper focuses on the analytical modeling of different NFV environments based on OF and analyses the factors that influence the delay in the packets within the environment. We are investigating two different scenarios: distributed architecture of the data plane with centralized Management and Orchestration (MANO) environment, and distributed environment where every location has its own data plane and MANO elements.

Analytical modeling, mostly on SDN environments based on OpenFlow (OF) architecture has been widely used for performance evaluation [12], [13]. Different parameters and use cases are analyzed to bring out novel algorithms and architectures that improve system performance [14], [15]. Some works concentrate on the controller performance in the SDN control plane [16] while others are focused on the data plane

and the SDN switching [17]. In our work we are focused on the NFV architecture where SDN is taken as only a part of the overall NFV environment.

Other authors have also explored the performance of systems that are based on SDN and NFV [18], [19]. It comes naturally that the queuing theory is used as an analytical modeling basis [20]. In most of the cases M/M/1 queueing model is used [13], [14], although different authors have explored other queueing models, like M/M/m [21] or G/G/m [22].

To the best of our knowledge, there is no publication that focuses on the distributed NFV environments from an analytical point of view, while at the same time taking into consideration both the packet delay caused by the NVF environment and the delay caused by the distance of the distributed environments. We consider the subject very important for the future development of NFV services. Our research can help system architects to evaluate the positive and negative sides of NFV distribution.

We have made a MATLAB [23] based simulation of our analytical models to show the main factors that contribute to the NFV network latency. We consider this to be important in real time scenarios allowing network architects to anticipate the latency in an OF NFV environment.

II. ANALYTICAL MODELS

In this section we are presenting two analytical models of an NFV system. Classical queuing theory principles are used for the mathematical analysis.

Similar to [24] we are modeling the data plane as Jacksonian Network with all switches being Jackson's servers. Every distributed location has one switch. Although in practice, in a single location a single switch can be branched to multiple underlying switches, for simplicity we will assume only one switch per location. We assume that at each node the arrival processes are considered to be Poisson processes. The service time of packets are assumed to follow exponential distributions. The arrivals at different nodes and their respective service times are independent of each other. The network of queues has reached balanced-state and the utilization of all of the queues is less than one. With these assumptions we can model all the nodes as M/M/1 systems with infinite queue sizes.

The system is OF based. Every controller is connected to a VNF manager which is directly connected to the element managers (EM) and VNFs. The VNF manager performs the job of starting, scaling and continuing the work of the VNFs.

We assume that the Virtual Infrastructure Manager (VIM) does not contribute to the sojourn time. We consider that the underlying infrastructure always has available resources. The NFV Orchestrator is also not considered to contribute to the sojourn time of packets due to its role in the NFV MANO.

A. Distributed data plane with single controller and VNF manager

Figure 1 shows a distributed data plane on multiple locations

with central controller and NFV MANO. Such environment is practical when it is necessary that some networking elements of the service are closer to the consumer.

We have k distributed location and k switches. We denote λ_i as the arrival rate of new packets at switch i where: $i \in \{1, 2, 3, \dots, k\}$

The full notation and explanations are given in Table I.

We calculate the utilization with the following formula:

$$\rho = \frac{\Gamma}{\mu} - \text{utilization} \quad (1)$$

The balance equation for switch i is given a sum from the arrival rate of the packets from outside and the arrival of packets from the other switches:

$$\gamma_i = \lambda_i + \sum_{j=1, j \neq i}^k (p_{ij} \cdot u_{ij}) \gamma_j \quad (2)$$

Where γ_i is net input for switch i .

We use u_{ij} to take into consideration the service chains in which there is no connection between switches i and j .

$$u_{ij} = \begin{cases} 1 & \text{if update from } j \text{ involves } i \\ 0 & \text{if update from } j \text{ doesn't involve } i \end{cases} \quad (3)$$

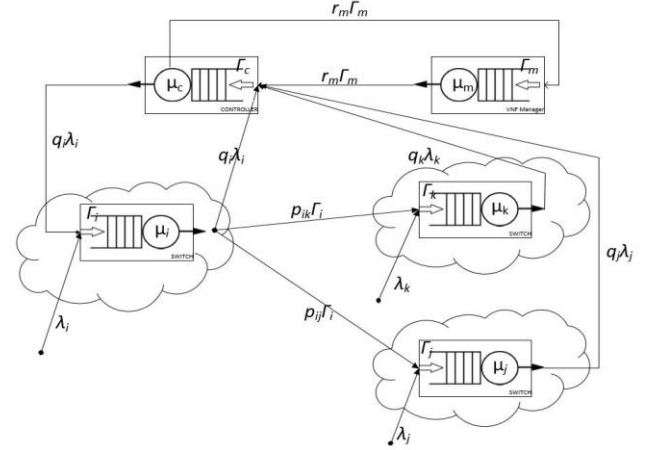


Fig. 1: NFV Model with distributed data plane and central controller and NFV MANO

The total arrival rate at the switch i is the sum of the net input to the switch and the packets that arrive to the switch from the controllers [17].

$$\Gamma_i = \gamma_i + q_i \cdot \lambda_i + \sum_{j=1, j \neq i}^k (q_j \cdot u_{ij}) \lambda_j \quad (4)$$

The total arrival rates at the VNF manager and the controller are:

$$\Gamma_m = \sum_{i=1}^k q_i \cdot \lambda_i \cdot r_i \quad (5)$$

$$\Gamma_c = \sum_{i=1}^k q_i \cdot \lambda_i + \sum_{i=1}^k q_i \cdot \lambda_i \cdot r_i \quad (6)$$

With this we can calculate the average number of flows and the average sojourn time of packets in the switches, the controller and the VNF manager:

$$E[N_i] = \frac{\rho_i}{1 - \rho_i} \quad (7)$$

$$E[T_i] = \frac{1}{\mu_i - \Gamma_i} \quad (8)$$

$$E[N_c] = \frac{\rho_c}{1 - \rho_c} \quad (9)$$

$$E[T_c] = \frac{1}{\mu_c - \Gamma_c} \quad (10)$$

$$E[N_m] = \frac{\rho_m}{1 - \rho_m} \quad (11)$$

$$E[T_m] = \frac{1}{\mu_m - \Gamma_m} \quad (12)$$

The average time a packet spends in the system, without the time spent in the links between distributed environments is:

$$E[T_d] = \frac{1}{\sum_{i=1}^k \lambda_i} \cdot \left[\sum_{i=1}^k E[T_i] + E[T_c] + E[T_m] \right] \quad (13)$$

The transmission and the propagation latency are given with:

$$L_{Tij} = \frac{B_{ij}}{Bw_{ij}} \quad (14)$$

$$L_{Pij} = \frac{d_{ij}}{s} \quad (15)$$

The propagation latency in practice is much smaller than the transmission latency.

The latency from all the switches to the controller is:

$$E[T_{NC}] = \sum_{i=1}^k q_i \cdot (L_{Tic} + L_{Pic}) \quad (16)$$

The latency from all the switches to switch i is:

$$E[T_{Ni}] = \sum_{j=1, j \neq i}^k p_{ij} \cdot (L_{Tij} + L_{Pij}) \cdot u_{ij} \quad (17)$$

Latency for a chained service for the transport between the switches is:

$$E[T_{NS}] = \sum_{i=1}^k \sum_{j=1, j \neq i}^k p_{ij} \cdot (L_{Tij} + L_{Pij}) \cdot u_{ij} \quad (18)$$

The mean packet latency for a network packet traveling in a chained service is the sum of the latency of the path it takes between switches (in different distributed environments) and the latency caused by the need to send packets to the controller.

$$E[T_N] = E[T_{NC}] + E[T_{NS}] \quad (19)$$

Now we have the average packet latency caused by the architecture (13) and the average packet latency caused by the transport of packets in the links between the distributed locations (19), we can calculate the total average packet latency in this distributed environment:

$$E[T_{tot}] = E[T_d] + E[T_N] \quad (20)$$

B. Fully distributed environment

Our model for full distributed environments is given in Figure 2. Every environment has a data plane with a single switch, a controller, and a VNF Manager dedicated for that environment. As the underlying physical infrastructure has to be distributed, every location must also have VIM.

All controllers can communicate to the switch in their own location, sending packet-out messages, thus changing the flow tables at their own location. Every controller is fully aware of

the network infrastructure and all services in the NNFV environment. In case of a change initiated from one controller, that controller sends the information to all other controllers. Every location has an independent arrival rate and every service can start from any location. A service can span in one or multiple locations.

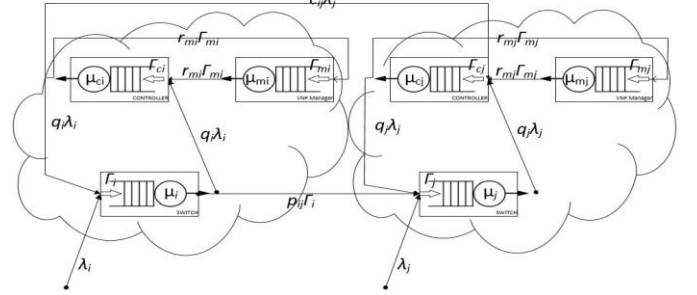


Fig. 2: Fully distributed NNFV Model

This environment is recommended if there is a need that VNF elements are close to the end consumers, but most of the services are using VNF elements mainly in its native location. As it will be shown later in this document, the links between distributed locations have high impact on the sojourn time of the packets and if the service spans in 3 or more distributed environments, the transport latency in the network dominates in the packets sojourn time.

The average time a packet spends in the system that is fully distributed, with controller and VNF manager at every location, without the time spent in the links between distributed environments is:

$$E[T_d] = \frac{1}{\sum_{i=1}^k \lambda_i} \cdot \left[\sum_{i=1}^k E[T_i] + \sum_{i=1}^k E[T_{ci}] + \sum_{i=1}^k E[T_{mi}] \right] \quad (21)$$

$$\begin{aligned} E[T_d] &= \frac{1}{\sum_{i=1}^k \lambda_i} \\ &\cdot \left[\sum_{i=1}^k \frac{1}{\mu_i - (\gamma_i + q_i \cdot \lambda_i + \sum_{j=1, j \neq i}^k (c_{ij} \cdot u_{ij}) \lambda_j)} \right. \\ &+ \sum_{i=1}^k \frac{1}{\mu_{ci} - (q_i \cdot \lambda_i + q_i \cdot \lambda_i \cdot r_i + \sum_{j=1, j \neq i}^k c_{ji} \cdot \lambda_j)} \\ &\left. + \sum_{i=1}^k \frac{1}{\mu_{mi} - (q_i \cdot \lambda_i \cdot r_i + \sum_{j=1, j \neq i}^k c_{ij} \cdot \lambda_j \cdot r_j)} \right] \end{aligned} \quad (22)$$

Now, let's calculate the average time a packet spends in the links connecting the distributed location. This gives the latency caused by all the controllers with their communication among each other:

$$E[T_{NC}] = \sum_{i=1}^k \sum_{j=1, j \neq i}^k c_{ij} \cdot (L_{Tij} + L_{Pij}) \cdot u_{ij} \quad (23)$$

The latency from all the switches to switch i is:

$$E[T_{Ni}] = \sum_{j=1, j \neq i}^k p_{ij} \cdot (L_{Tij} + L_{Pij}) \cdot u_{ij} \quad (24)$$

Latency for a chained service for the transport between the

switches is:

$$E[T_{NS}] = \sum_{i=1}^k \sum_{j=1, j \neq i}^k p_{ij} \cdot (L_{Tij} + L_{Pij}) \cdot u_{ij} \quad (25)$$

The mean packet latency for a network packet traveling in a chained service is sum of the latency of the path it takes between switches and the latency caused by the communication among the controllers.

$$E[T_N] = E[T_{NC}] + E[T_{NS}] \quad (26)$$

We can calculate the average packet sojourn time, using (22) and (26).

$$E[T_{tot}] = E[T_d] + E[T_N] \quad (27)$$

TABLE I. Notation explanation

Symbol	Parameter Name
λ_i	arrival rate at switch i
q_i	probability of packet going from switch i to controller
p_{ij}	probability of packet going from switch i to switch j
r_m	probability of packet going from controller to VNF manager
$\mu_i; \mu_c; \mu_m$	service rate (switch i, controller c, VNF manager m)
$\Gamma_i; \Gamma_c; \Gamma_m$	total arrival rate (switch i, controller c, VNF manager m)
$N_i; N_c; N_m$	number of packets (switch i, controller c, VNF manager m)
$T_i; T_c; T_m$	sojourn time at the (switch i, controller c, VNF manager m)
$E[N]$	mean value for number of packets in M/M/1 queue
$E[T]$	mean value for the sojourn time of packets in M/M/1 queue
L_T	transmission latency
L_P	propagation latency
T_N	time spent in network
T_{NC}	time spent in network for packets going to the controller
T_{NS}	time spent in network for packets going between switches
$E[T_{NC}]$	mean value for the sojourn time for packets to the controller
$E[T_{NS}]$	mean time for the sojourn time for packets between switches
B_{ij}	number of packets on link from i to j
Bw_{ij}	bandwidth on link from i to j
d_{ij}	distance from i to j
s	speed of data in link

III. PERFORMANCE EVALUATION

To evaluate the proposed systems with the analytical models described in the previous sections, we have developed simulation scripts in Matlab. We will discuss the environments separately and at the end we will compare the two proposed distributed environments.

A. Distributed data plane with single controller and VNF manager evaluation

This distributed environment is served by a single controller and single VNF manager while the data plane is distributed on multiple locations, with a single switch on every physical location. Similar to [13] we assume that 4% of the packets need to go to the controller from the switches, in the OF network, for the flow tables to be updated and 4% of the packets that reach the controller need to go to the VNF Manager.

Also, our evaluation assumes that the physical links between the distributed locations have speeds of 10Gbps and the distance between physical locations is 100km. The link speed

assumption is done due to the fact that this link speed is most often use in practice, while the physical distance is taken as a minimal distance for separate datacenters, but this parameter has very low impact on the overall latency.

First, we evaluate how is the packet sojourn time affected by the number of physical locations in the system, when there are steady arrival rates at the switches of $\lambda=50000$ pkts/sec, controller's service rate of 90000pkts/sec and VNF Manager's service rate to be 95000 pkts/sec.

Packet sojourn time relative to the probability of packets being send to a different distributed location, with 5, 10 and 15 distributed locations is given in Figures 3, 4 and 5 respectfully. It can be seen that the packet sojourn time is rising faster as the number of distributed locations is rising. We can see that in a case of 15 distributed locations, even with a probability of about 16% that a packet will go to a different distributed location and an arrival rate of 12000pkts/sec the system collapses and it cannot serve the incoming packets. In these cases, switches with higher service rates have to be used or a different architecture that will not use as much distributed locations has to be implemented. In our simulations the switches have service rate of 90000pkts/sec.

B. Fully distributed environment evaluation

In a fully distributed environment every location has a switch, a controller and a VNF manager.

Figure 6 shows the packet sojourn time relative to the number of distributed environments where we take a fixed probability of 4% that a packet coming to a switch will need to be pushed to the controller, that a packet from the controller will need to be pushed to the VNF manager, and that a packet will go through a link to another distributed location (when the service spans to multiple locations). As expected, the time rises with the number of distributed locations.

To check the delay caused in the system by different probabilities that the packet will go to other distributed location, we have made simulations with 5, 10 and 15 different locations. The results are given in figures 7, 8 and 9. Similarly to the scenario with single controller, the packet sojourn time rises with the rise of the probability of the packets in an NFV based service to go between different distributed locations.

C. Distributed environments comparison

Now we can compare the two different scenarios of distributed environments. We are evaluating the two different systems in a scenario where we check the packet sojourn time relative to the probability that a service is spanned across multiple distributed locations, thus we are changing the probability that the packet will go through a link from one switch on one location, to another in a different location. The probabilities for packet-in messages, from the switches to the controller and from the controller to the VNF manager are fixed at 4%. Figures 10, 11 and 12 show the dependence of the packet

sojourn time from the probability of packets between switches in case of 5, 10 and 15 locations respectively.

We can see from figure 7 that with small number of locations the architecture with central Nfv MANO performs better, regardless of the probability of packets sent to other distributed locations, although the higher the probability, the performances are similar, and the central controller scenario is approaching to the fully distributed environment scenario.

But when we analyze the 10 locations scenario, we can see that for lower probabilities of packets sent to other distributed locations, the architecture with central Nfv MANO performs better. But as the probability rises, this approach gets worst and the packet sojourn time grows exponentially while in the fully distributed scenario the packet sojourn time is stable. In the case of central Nfv MANO the system becomes unstable much sooner than in a case of a fully distributed environment.

In the 15 locations scenario, the average sojourn time goes higher above the average sojourn time in system with distributed Nfv MANO. It can be seen visually on figure 12.

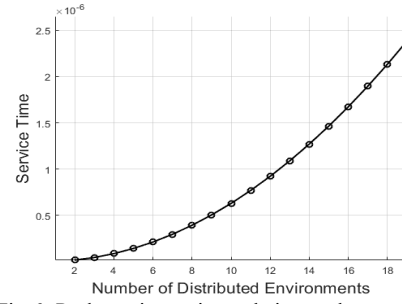


Fig 6: Packet sojourn time relative to the number of distributed env. in a fully distributed scenario and static probability of 4%

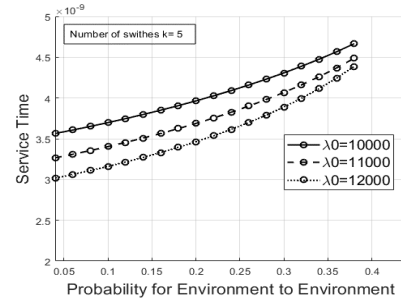


Fig 7: Service on 5 locations in a fully distributed scenario

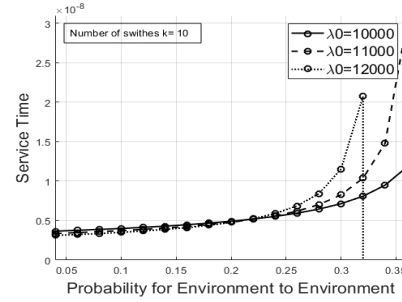


Fig 8: Service on 10 locations in a fully distributed scenario

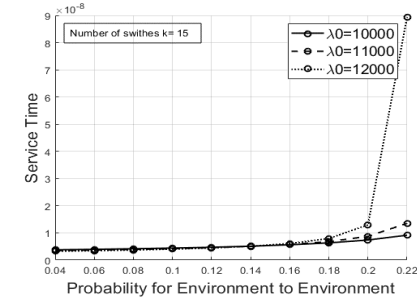


Fig 9: Service on 15 locations in a fully distributed scenario

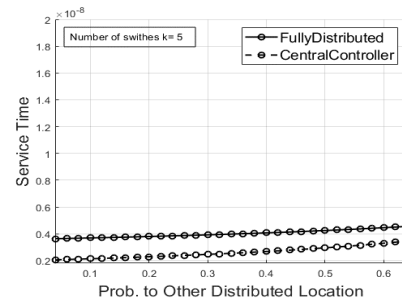


Fig 10: Packet sojourn time comparison with 5 locations

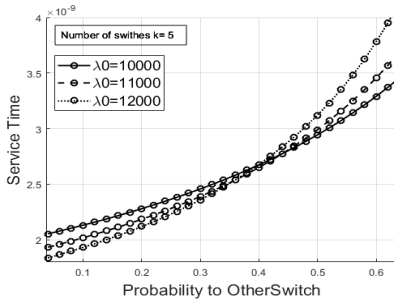


Fig 3: Packet sojourn time on 5 distributed locations

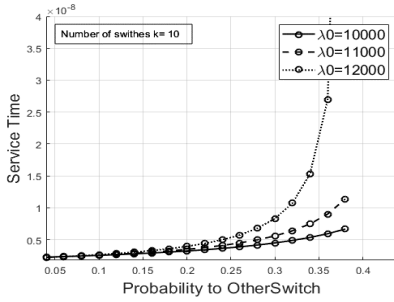


Fig 4: Packet sojourn time with service on 10 distributed locations

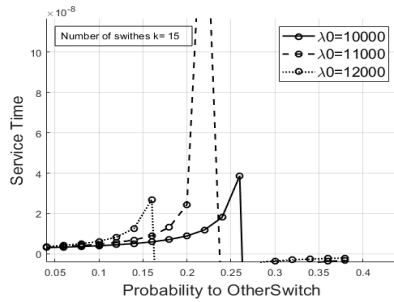


Fig 5: Packet sojourn time with service on 15 dist. locations

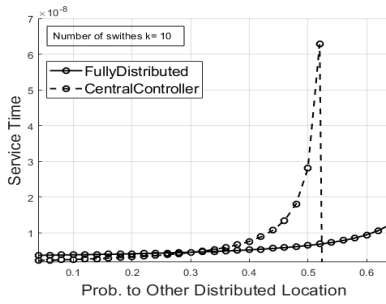


Fig 11: Packet sojourn time comparison with 10 locations

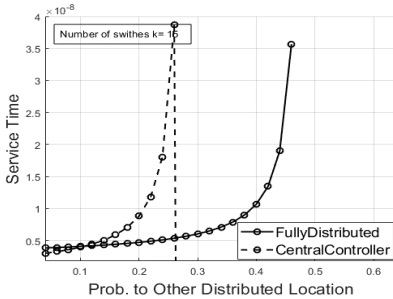


Fig 12: Packet sojourn time comparison with 15 locations

IV. CONCLUSION

In this paper we have made analytical models of two NFV architectures: a distributed data plane with a central NFV MANO and a fully distributed environment with data plane and NFV MANO elements on all locations. We have modeled the data plane as Jacksonian Network with all switches being Jackson's servers. We've set up the assumptions and we modeled all the nodes as M/M/1 systems. The equations for our models calculate the average packet sojourn time in the systems, taking into consideration the delay caused by the architectural elements and the links connecting the distributed locations.

Simulation scripts in Matlab were developed and we have evaluated the performance of the systems under different conditions. We have compared the two different distributed environments to see which one performs better under different conditions.

Although theoretical, our analysis is a good basis for preparing an adequate architecture in practice having in mind different types of services that can run onto the system. By analyzing the services and the positive and negative sides of the proposed scenarios, the best approach can be chosen in practice.

REFERENCES

- [1] J. Reis, M. Amorim, N. Melao, P. Matos, "Digital Transformation: A Literature Review and Guidelines for Future Research.", WorldCIST, 2018, pp 411-421
- [2] J. Reis, M. Amorim, N. Melao, Y. Cohen, M. Rodrigues, "Digitalization: A Literature Review and Research Agenda.", LICIEOM, 2020, pp 443-456

- [3] D. Huang, A. Chowdhary, S. Pisharody, "SDN and NFV: From Theory to Practice.", CRC Press, 2018.
- [4] M. Eiman, "Minimum Technical Performance Requirements for IMT-2020 Radio Interface(s).", Presentation. 2018.
- [5] F. Yousaf, M. Bredel, s. Schaller, F. Schneider, "NFV and SDN - Key Technology Enablers for 5G Networks." IEEE Journal on Selected Areas in Communications. 2017
- [6] Ericsson Publishing, "Edge computing and 5G - Harnessing the distributed cloud for 5G success", [Online], 2020
- [7] S. Panev, P. Latkoski, "Handover analysis of openflow-based mobile networks with distributed control plane." Computers & Electrical Engineering. 2020.
- [8] T. Issa, Z. Raoul, A. Konate, J.C. Adepo, B. Cousin, A. Olivier, "Analytical load balancing model in distributed open flow controller system." Sci Res Eng, 2018, pp. 863-875.
- [9] F. Bannour, S. Souihi, A. Mellouk, "Distributed SDN Control: Survey, Taxonomy and Challenges." IEEE Communications Surveys & Tutorials, 2017
- [10] A. Aissioui, A. Ksentini, A. Gueroui, "An efficient elastic distributed SDN controller for follow-me cloud," 2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications, Abu Dhabi, 2015, pp. 876-881
- [11] <https://www.etsi.org/technologies/nfv>
- [12] M. Jarschel, S. Oechsner, D. Schlosser, R. Pries, S. Goll, P. Tran-Gia, "Modeling and Performance Evaluation of an OpenFlow Architecture", 2011
- [13] M. Jarschel, O. Østerbø, A. Chilwan, K. Mahmood, "Modelling of OpenFlow-based software-defined networks: The multiple node case." IET Networks, 2015
- [14] S. Zhihao, K. Wolter, "Delay Evaluation of OpenFlow Network Based on Queueing Model." 2016
- [15] M. Iushchenko, V. Shuvalov, B. Zelentsov, Boris. "Modeling the Software Dependability For a SDN/NFV", 2019
- [16] B. Xiong, X. Peng, J. Zhao, "A Concise Queueing Model for Controller Performance in Software-Defined Networks." Journal of Computers. 2016, pp. 232-237
- [17] W. Saied, N. Ben Youssef, A. Saadaoui, A. Bouhoula, "Deep and Automated SDN Data Plane Analysis." International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2019
- [18] C. Bouras, A. Kollia, A. Papazois, "SDN & NFV in 5G: Advancements and challenges." 20th Conference on Innovations in Clouds, Internet and Networks (ICIN), 2017
- [19] M. Iushchenko, V. Shuvalov, B. Zelentsov, "Modeling the Software Dependability For a SDN/NFV Controller", 2019 International Multi-Conference on Engineering, Computer and Information Sciences, 2019.
- [20] S. Azeem, S. Sharma, "Advanced Technologies Involved In Network Convergence: SDN & NFV." International Journal of Development Research, 2019
- [21] K. Sood, S. Yu, Y. Xiang, H. Cheng, "A general QoS aware flow-balancing and resource management scheme in distributed software-defined networks." IEEE Access, 2016.
- [22] J. Prados, P. Ameigeiras, J. Ramos, P. Andres-Maldonado, J. Lopez-Soler, "Analytical modeling for Virtualized Network Functions". IEEE ICC Workshops, 2017.
- [23] MATLAB:R2019a, Natick, Massachusetts, The Mathworks
- [24] C. Sarkar, SK. Setua, "Analytical model for openflow-based software-defined network." Progress in computer, analytics and networking; 2018. pp. 583-592.

Particle Swarm Optimization (PSO) based Resource Allocation for Device to Device Communication for 5G Network

Wisam Hayder Mahdi

Department of Communication Engineering,
University of Diyala, Engineerin College
Diyala, Iraq
wisam_haider@uodiyala.edu.iq

Necmi Taşpınar

Department of Electrical and Electronics Engineering
Erciyes University, Engineering Faculty
Kayseri, Turkey
taspinar@erciyes.edu.tr

Abstract - Device-to-device communication system (D2D) is an evolution-release 12 (LTE) technology built over the long term. For devices that are based on LTE, direct connection can be allowed when they are located nearby via a D2D is a peer-to-peer connection, which allows this. Improvement of the performance, energy consumption, spectral quality, and latency that's done by using D2D communication. This is seen as an increase and decrease in two cases, the first is to increase the spectrum quality of the network and energy usage, while the second case is to reduce the traffic offload on the base station and also reduce the transmission delay. Within this work and in order to achieve a PSO- based resource allocation system when using the fifth generation communication networks, we have described the overall numerical result of achieving this system. Using MATLAB, the results are analyzed, so that each program will be run 100 times, and then an average is taken to plot the graph for the simulation. Under the base station coverage area, the systems are deployed randomly. Half-duplex and full duplex D2D communications are supported by the network, so that the user reuses the block of uplink resources from the CU users.

Keywords: *Device-to-Device (D2D), Long term evaluation (LTE), Particle swarm optimization (PSO), Quality of Service (QoS), Resource Allocation (RA).*

I. INTRODUCTION

The proximity of mobile devices invested by communications from Device-to-Device (D2D) for the purpose of exchanging information, which takes place through direct links, without the need to request routing, which is through the base station. Improving the efficient use of the spectrum that the D2D communications underlaying the cellular network promises from by sharing the spectrum of cellular users as well as through a low transmission power which is done by transmission through short-range links. Mitigating the problem of too little or scarcity of spectrum as well as offloading the base station are what can be done and happen when the D2D multicast connection is introduced to traditional cellular networks. Reducing the burden on the base stations, which is done by retransmission of high-rate data that occurs through direct and short-range links, this occurs when D2D users who

act as relay nodes participate [1]. With the predicted explosion in the number of wireless devices and the growing development of multimedia services, it is possible to represent a great challenge to the current cellular networks for the purpose of providing what the user requires [2]. The 5 generation network emerged to meet multiple demands, including the large increase in the demand for mobile data due to the very large number of devices and users, as the 5G network promises high reliability and download speeds of up to 10,000 Mbps [3]. The technology developed for the long-term development of version 12 [LTE] is device-to-device communication technology, which can be described as peer-to-peer communication in which LTE-based devices can communicate directly with each other if they are in close proximity [4]. One of the important features of D2D networks is the reuse of resources allocated to cellular users [5]. The resources in wireless networks that are used by cellular users are allocated to D2D pairs. Allocation of resources can be a challenge by which to maintain the quality of service for wireless networks [6]. High data rate and low latency as well, that's what you want in most group communications, including video streaming and multiplayer games. Then, these applications require an important network feature, which is the multicast feature. The same information that base stations transmit to a group of users at a common rate, this is done by single rate multicast communications. Links that lead to a bottleneck in throughput, especially those that suffer from poor channel conditions and then users cannot receive the relevant source information correctly. Thus, users with better channel quality experience a waste of bandwidth when channel quality differentiation results [7]. D2D users can send signals and receive it directly while these users are staying under control by the base station. Mutual interference occurs due to the sharing of spectrum between D2D users and cell users. One of the greatest challenges for D2D Communications is the interference management that is enabled in cellular networks.. In the absence of effective interference reduction, resulting in communication, cutting, the advantage brought by device to device communications would be removed [8].

A. RELATED WORK

Swarm Intelligence (SI) is one of the techniques of

collective intelligence that mimic the behavior of swarms of organisms that live in groups, and these groups are described as cooperating with each other. There are several examples of Intelligence Swarm algorithms, which we mention as an example and not to limit them Particle Swarm Optimization (PSO) [9], Ant Colony Optimization (ACO) [10], Dragonfly Algorithm (DA) [11], Salp Swarm Algorithm (SSA) [12] and Grey Wolf Optimizer (GWO) [13]. The shortage of resources experienced by wireless networking offers, therefore, leads to the optimal allocation of these resources becoming very important. The resources in the wireless network are initially allocated first to the cellular phone users and then are allocated to the D2D pairs according to their desires. Several different methods have been proposed for the purpose of obtaining an optimal allocation of resources, among these methods are Particle Swarm algorithm, Genetic Algorithm and Fuzzy Logic [14-16]. In a wide range of literature, one of the dominant swarm intelligence algorithms that has been used is the (PSO) algorithm. Originally, such algorithms are designed to solve continuous variable problems. The continuous optimizer can be obtained by converting with the transfer functions where the search space can be represented by binary values [17]. We apply the PSO algorithm to reduce the interference in resource allocation which is based on the resource allocation scheme for the D2D users of 5 G networks. In this scheme, D2D users who have high fitness values are allocated resources to them, and they get high priority in resource allocation.

The paper was established as follows. In Section II, the discussion is about the system model and the problem. In Section III, we give the resource allocation scheme. In Section IV, the simulation results are given. Section V includes conclusions.

II. SYSTEM MODEL AND PROBLEM

Two forms of communication that we can consider a network downlink transmission scenario are direct D2D communication and traditional cellular communication. Cellular contact between the cellular user and BS is between two D2D users without BS interference. Figure 1 shows a downlink scenario in which UE6 is the touch downlink mode, and UE1. UE2 and UE7, UE8 transmit in D2D mode. Within the cell region the other UE3, UE4, and UE5 at cell area. We assume the number of cellular users is M , and the number of D2D pairs is N where $C_m, m=1,2 \dots M$ stands for cellular users, and D_n stands for D2D pair $\{D_{Tn}, D_{Rn}\}$ while D_{Tn} and $D_{Rn}, n=1,2, \dots N$ stands for both D2D transmitter and receiver simultaneously. We also expect each cellular user to be assigned to a single channel, i.e. channels number equals the number of cellular users. Accordingly, assume BS and D2D transmitters with PB and PD capacity.

PB / K is the power of BS transmission allocated to each channel where K has channel number. First of all we find out which form of channel is provided by CU and one or more D2D sets to decide the interference. Channel gain consists of loss of direction depending on distance and fading to a small scale. We took the small-scale device to fad. Path loss to application connections can be calculated as follows: Route loss for BS:

$$PL(d)=128.1+37.6\log(d) \quad (1)$$

Loss of path for the D2D connections (between D_{Tn} and D_{Rn} or between D_{Tn} and C_m)

$$PL(d)=148 + 40\log(d) \quad (2)$$

The distance d is in kilometers here.

Interference for cell-users can be determined as follows:

$$I_{C_m} = \sum_{i \in N} P_D^K G_{D_{T_i}, C_m}^K \quad (3)$$

Here, $G_{D_{T_i}, C_m}^K$, specifies gain of the channel between D2D transmitter off D_{T_i} to cellular users C_m at the k^{th} channel.

At D2D users end, the Interference:

$$I_{D_n} = P_C^k G_{C_m, D_{R_n}}^k + \sum_{n=1, n \neq i}^N P_D G_{D_{T_n}, D_{R_i}}^k \quad (4)$$

Here $G_{C_m, D_{R_n}}^k$ refers to channel gain between D2D users and cellular users at the k^{th} channel and $G_{D_{T_n}, D_{R_i}}^k$ channel gain for the receiver of the accepted D2D pair on the k^{th} channel between other D2D transmitters.

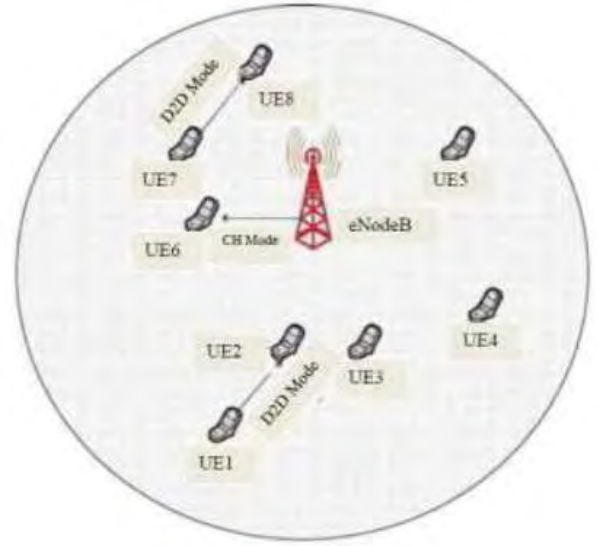


Figure 1: Communication analogy for 5 G Networks

The 5G network that we are dealing with is the high priority for cellular users where the channel is allocated first. Then, D2D pairs are assigned to the channels that are now assigned to cellular users. The total interference is at a minimum, which makes the accommodation of the numbers of D2D users, before which the channel assignment must be in really D2D pairs. For this, at first we have to determine the interference across all users, then the optimization methods are implemented to reduce the interference values due to D2D pairs. Therefore, we must first define interference across all

users, and then implement optimization methods to minimize interference values due to D2D pairs. We noticed both cases, too. First case: One resource block sharing by one CU and one D2D pair (up to the 15th D2D pair). When the number of D2D pairs exceeds the number of channels.

Hence, in our this problem, the objective function is as follows:

$$\text{minimize } \sum_{i=1}^N I_{D_i} \quad (5)$$

III. THE RESOURCE ALLOCATION METHOD

A. Particle Swarm Optimization

In 1995, Kennedy and Eberhart presented PSO as a modern probabilistic method for the first time. The PSO was essentially influenced by the sociological nature aligned with swarms, such as fish schooling or bird flock. In the population the individuals are called the particles. Each of particle is a possible solution of the problem of optimization, and attempts to find the best location by flying in a multi dimensional area.

The PSO is a population-based, global technique of optimization inspired by collective behavior of bird flocks in search of corn. This algorithm is similar to other statistical methods of evolution, such as genetic algorithms.

A population of random solutions initializes the program, and searches for the optima by updating generations. While, various of the genetic algorithms, PSO does not have any evolutionary operators, like crossover and mutation. Possible solutions, called particles, travel through the entire problem space in PSO by following the current ideal particles and investigating the solution space according to the individual and neighborhood dependent particle positions.

The PSO algorithm comprises three significant "best" values: qbest, Jbest, and kbest. And since in the problem space, the particles start tracing their own coordinates, which are related to the optimal solution- fitness value- that they were able to obtain up to this moment. The fitness value is held, as well. That value is called qbest. Another significant value that is monitored by the PSO is the best value available from any neighboring particle. The name for that is Jbest. When a particle considers the entire population as its topological peers, the highest value is the strongest in the world and is called the kbest.

The PSO definition in each iteration consists of modifying each particle's velocity against its qbest and Jbest destinations. Acceleration is determined by a random parameter that produces different random numbers for acceleration against qbest and Jbest destinations.

The PSO algorithm consists of $\mathbf{z}_m = (\mathbf{z}_{m_1}, \dots, \mathbf{z}_{m_T})$ as a vector for m-th particle ($m = 1, \dots, M$) in T -dimension, $\mathbf{v}_m = (\mathbf{v}_{m_1}, \dots, \mathbf{v}_{m_T})$ as velocity for the m-th particle, $\mathbf{p}_m = (\mathbf{p}_{m_1}, \dots, \mathbf{p}_{m_T})$ as the best position of its been found by the particles so far away, and

$\mathbf{p}_m = (\mathbf{p}_{g_1}, \dots, \mathbf{p}_{g_T})$ as the best possible place the entire swarm of particles has been found to date.

The velocity and positioning of the particles in each iteration are modified as follows.

$$\mathbf{v}_{mt}^{u+1} = \mathbf{v}_{mt}^u + c_1 r_1 (\mathbf{p}_{mt} - \mathbf{z}_{mt}^u) + c_2 r_2 (\mathbf{p}_{gt} - \mathbf{z}_{mt}^u) \quad (6)$$

$$\mathbf{z}_{mt}^{u+1} = \mathbf{z}_{mt}^u + \mathbf{v}_{mt}^{u+1} \quad (7)$$

Where u can be represented by the iteration index while the particle index is m . In addition, r_1 and r_2 are independently and homogeneously distributed random values at range of $[0,1]$, and c_1 and c_2 are coefficients acceleration.

B. Fundamental PSO Algorithm

for every particle $i = 1, 2, \dots, S$ **do**

Initialize the position of the particle by a uniform distributed random vector: $\mathbf{x}_i \sim U(\mathbf{b}_{lo}, \mathbf{b}_{up})$

Initialize the best known position of the particle in its initial position: $\mathbf{p}_i \leftarrow \mathbf{x}_i$

If $f(\mathbf{p}_i) < f(\mathbf{g})$ **then**

Update the most suitable place for the swarm: $\mathbf{g} \leftarrow \mathbf{p}_i$

Initialize the velocity of the particle component:

$$\mathbf{v}_i \sim U(-|\mathbf{b}_{up} - \mathbf{b}_{lo}|, |\mathbf{b}_{up} - \mathbf{b}_{lo}|)$$

while on Unfulfilled termination criteria **do**

for every particle $i = 1, \dots, S$ **do**

for every dimension $d = 1, \dots, n$

do Pick randomly generated numbers: $r_p, r_g \sim U(0,1)$

Update the velocity of particle: $\mathbf{v}_{i,d} \leftarrow \omega \mathbf{v}_{i,d} + \phi_p r_p (\mathbf{p}_{i,d} - \mathbf{x}_{i,d}) + \phi_g r_g (\mathbf{g}_d - \mathbf{x}_{i,d})$

Update the most suitable place for particle: $\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{v}_i$

if $f(\mathbf{x}_i) < f(\mathbf{p}_i)$ **then**

Update the particle's best known position: $\mathbf{p}_i \leftarrow \mathbf{x}_i$

if $f(\mathbf{p}_i) < f(\mathbf{g})$ **then**

Update the most suitable place for the swarm: $\mathbf{g} \leftarrow \mathbf{p}_i$

The values \mathbf{b}_{lo} and \mathbf{b}_{up} , the lower and upper search-space boundaries, respectively. The terminating criteria may be the iteration numbers performed, or a solution where there is an effective objective function value found. The professional parameters ω , ϕ_p , and ϕ_g are selected and the action and adequacy of the PSO technique are monitored.

C. Work of the Algorithm

1. The Particle Swarm Optimization algorithm starts with the formation of the original particles (so, particles listen to D2D and CU users SINR) and their assignment of initial velocities.

2. At-particle position, PSO calculates the objective function and decides the best (smallest) function value and the best position.
3. It selects new velocities based on the present velocity, the best locations of the individual particles and their neighbors' best locations.
4. It then changes the particle positions iteratively (the new position is the old one plus the velocity, updated to hold particles within boundaries), the speeds and the neighbours.
5. Iterations continue until the algorithm hits stopping criteria. i.e maximum number of Iteration (MaxIt).
6. Thus, we get the optimized Throughput of D2D and CU Users by using below formula.

The network throughput is measured by using the Shannon Capacity. If the bandwidth of the channel is W and the optimized interference measured is I_d by PSO, the network throughput is determined as follows:

$$\text{Throughput} = B * \log_2 (1 + S/N) \quad (8)$$

The throughput and spectral efficiency of network is computed by equation 8.

IV. SIMULATION RESULTS

Here, we describe simulation results overall in achieving PSO technique that based on the resource allocation system in 5G D2D communication networks. By using MATLAB the results were analyzed where each program will be run 100 times and then the average is taken to plot the simulation graph. The systems are deployed randomly through the coverage area of the BS. The half duplex and full duplex D2D connection is supported by the network so that the user can reuse the block of uplink resources by CU users.

Table-1 illustrates the general simulation parameters and Table 2 illustrates simulation parameters for the Particle Swarm Optimization (PSO).

Table-1: Estimating Parameters of Interference Simulation

Parameters	Values
Radius of Cell	500 meter
Frequency of Carrier	2.1 GHz
D2D User Number	6.0
CU User Number	15.0
Channel Number	15.0
Power Transmission by BS, P_b	78.0 dBm
Power Transmission by Device, P_d	24.0 dBm
Bandwidth of Channel, W	1800 KHz
Device Noise Figure	-116 dBm
Simulation Type	Monte Carlo Simulation

Table-2: Simulation Parameters for Particle Swarm Optimization

Parameters	Values
Number of Decision Variables	10.0
Variables of Lower Bound	10.0
Variables of Upper Bound	10.0
Maximum iteration number	10.0
Size of Population	100.0
Weight of Inertia	1.0
Damping Ratio of Inertia Weight	0.99
Personnel Learning Coefficient	1.5
Global Learning Coefficient	2.0
Global Best Cost	infinite

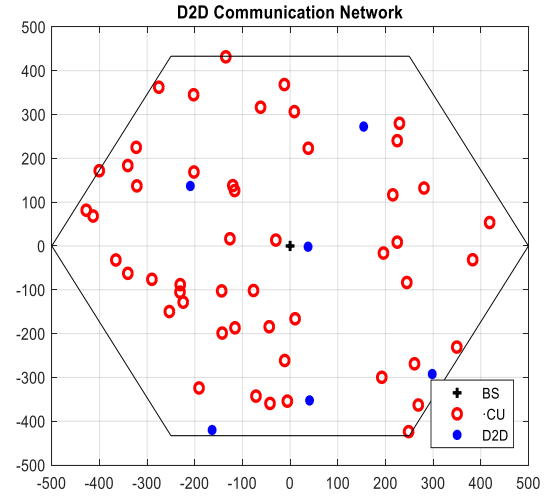


Figure 2: D2D Communication Network

In Figure 2 shows a D2D communication network. D2D communication is possible where the D2D pair is sufficiently far enough from the BS and the transmitter and receiver distance.

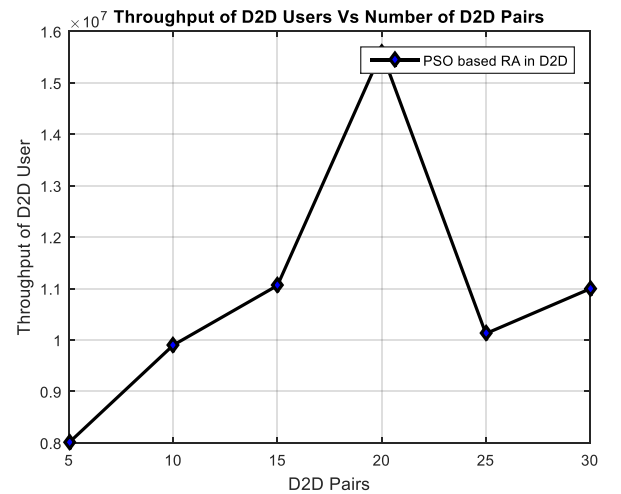


Figure 3: D2D Throughput Vs D2D pairs

Above Figure 3 is a graph showing the D2D Users throughput and the number of D2D pairs. At first, we consider one resource-sharing scenario where CU users share a pair of D2D within a resource block. After 21 D2D pairs are reached, our scheme takes into account a multi-scenario situation in which two pairs of D2D share with CU users in the same resource block which then leads to a decline in throughput. Thus, until 20 D2D pairs, the throughput of the shape is at its maximum and then the throughput deteriorates dramatically until the 25 D2D pairs. Again, the throughput is improved because no matter which pairs of D2D uses the same resource block, its position is changed and increased throughput.

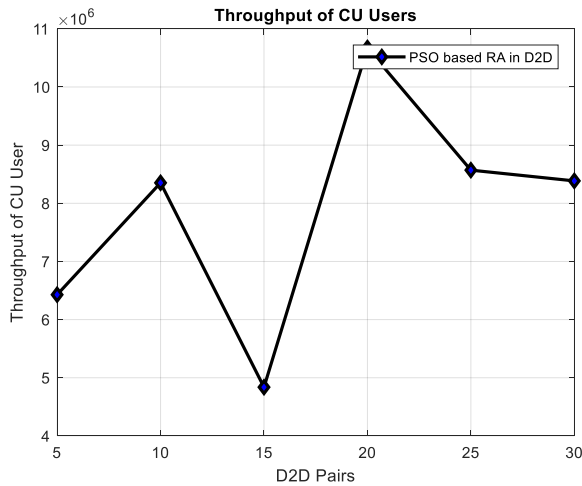


Figure 4: Throughput of CU users

Figure 4 shows the throughput of CU users in the network. We set two realities for CU users (21). Starting with BS, it starts by allocating one channel to each of the 12 users and as it sends 222 pairs of requests for channel assignment, 56 allocates the same channel to user CU which causes interference and degrades the resulting throughput. After 20 pairs, more than one pair of D2D is shared with CU users in the same channel, so starting with the 20 pairs of D2D we will see a significant drop in throughput.

V. CONCLUSION

In this paper, we applied the PSO algorithm to reduce the interference in resource allocation which is based on the resource allocation scheme for the D2D users of 5 G networks. In this scheme, D2D users who have high fitness values are allocated resources to them, and they get high priority in resource allocation. Up to 20th D2D pair our scenario is single resource sharing. After 20th D2D pair we have considered multi resource sharing scheme where both D2D Users and CU Users uses the same resource block.

ACKNOWLEDGMENT

This work was supported by the Erciyes University Scientific Research Projects Coordination Unit (Project No: FDK-2021-10867).

REFERENCES

- [1] Monia Hamdi, Mourad Zaied. Resource allocation based on hybrid genetic algorithm and particle swarm optimization for D2D multicast communications, *Applied Soft Computing Journal* 83 (2019) 105605.
- [2] S.T. Shah, S.F. Hasan, B.-C. Seet, P.H.J. Chong, M.Y. Chung, Device-to-device communications: A contemporary survey, *Wirel. Pers. Commun.* 98 (1) (2018) 1247–1284.
- [3] X. Shen, "Device-to-device communication in 5G cellular networks," *Network, IEEE*, vol. 29, pp. 2-3, 2015.
- [4] K. Doppler, et al., "Device-to-device communication as an underlay to LTE-advanced networks," *Communications Magazine, IEEE*, vol. 47, pp. 42-49, 2009.
- [5] X. Lin, et al., "Spectrum sharing for device-to-device communication in cellular networks," *Wireless Communications, IEEE Transactions on*, vol. 13, pp. 6727- 6740, 2014.
- [6] S. Wen, et al., "QoS-aware mode selection and resource allocation scheme for Device-to-Device (D2D) communication in cellular networks," in *Communications Workshops (ICC), 2013 IEEE International Conference on*, 2013, pp. 101-105.
- [7] M. Hamdi, D. Yuan, M. Zaied, Ga-based scheme for fair joint channel allocation and Power control for underlaying d2d multicast communications, in: *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2017, pp. 446– 451.
- [8] H. Meshgi, D. Zhao, R. Zheng, Joint channel and power allocation in underlay multicast device-to-device communications, in: *2015 IEEE International Conference on communications (ICC)*, 2015, pp. 2937– 2942.
- [9] J. Kennedy, Particle swarm optimization, in: C. Sammut, G.I. Webb (Eds.), *Encyclopedia Of Machine Learning*, Springer US, Boston, MA, 2010, pp. 760–766, http://dx.doi.org/10.1007/978-0-387-30164-8_630.
- [10] M. Dorigo, G.D. Caro, Ant colony optimization: a new meta-heuristic, in: *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, Vol. 2, 1999, pp. 1470–1477, <http://dx.doi.org/10.1109/CEC.1999.782657>.
- [11] S. Mirjalili, Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems, *Neural Comput. Appl.* 27 (4) (2016) 1053–1073, <http://dx.doi.org/10.1007/s00521-015-1920-1>.
- [12] S. Mirjalili, A.H. Gandomi, S.Z. Mirjalili, S. Saremi, H. Faris, S.M. Mirjalili, Salp swarm algorithm: A bio-inspired optimizer for engineering design problems, *Adv. Eng. Softw.* 114 (2017) 163–191, <http://dx.doi.org/10.1016/j.advengsoft.2017.07.002>, URL <http://www.sciencedirect.com/science/article/pii/S0965997816307736>.
- [13] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, *Adv. Eng. Softw.* 69(2014)46–61, <http://dx.doi.org/10.1016/j.advengsoft.2013.12.007>, URL <http://www.sciencedirect.com/science/article/pii/S0965997813001853>.
- [14] W. Gong and X. Wang, "Particle Swarm Optimization Based Power Allocation Schemes of Device-to-Device Multicast Communication," *Wireless personal communications*, vol. 85, pp. 1261-1277, 2015.
- [15] L. Su, et al., "Resource allocation using particle swarm optimization for D2D communication underlay of cellular networks," in *2013 IEEE wireless communications and networking conference (WCNC)*, 2013, pp. 129-133.
- [16] N. U. Hasan, et al., "Network Selection and Channel Allocation for Spectrum Sharing in 5G Heterogeneous Networks," *IEEE Access*, vol. 4, pp. 980-992, 2016.
- [17] M. Mafarja, I. Aljarah, A.A. Heidari, H. Faris, P. Fournier-Viger, X. Li, S. Mirjalili, Binary dragonfly optimization for feature selection using timevarying transfer functions, *Knowl.-Based Syst.* 161 (2018) 185– 204, <http://dx.doi.org/10.1016/j.knosys.2018.08.003>, URL <http://www.sciencedirect.com/science/article/pii/S095070511830399X>.

Investigation of Effect of the Pilot Reuse Factor via Intelligent Optimizations on Energy and Spectral Efficiencies Trade-off in Massive MIMO Systems

Burak Kürşat Gül

Department of Electrical and Electronics Engineering
Erciyes University
Kayseri, Turkey
burak.gul@erciyes.edu.tr

Necmi Taşpınar

Department of Electrical and Electronics Engineering
Erciyes University
Kayseri, Turkey
taspinar@erciyes.edu.tr

Abstract— With the decrease in energy resources, energy saving has become vital nowadays. This situation has led to an increase in the importance given to the efforts to increase energy efficiency in the field of cellular network. However, factors that increase energy efficiency often reduce the spectral efficiency, which is extremely important for a cellular network. With the help of intelligent optimization techniques, it is possible to ascend energy and spectral efficiencies together in a cellular network using Massive MIMO systems. In this paper, three different intelligent optimization techniques have been used to find solutions to the aforementioned problem. In addition, the pilot reuse factor has tried at three different values and the effect of this factor has been examined.

Keywords—massive MIMO; intelligent optimization techniques; energy efficiency; spectral efficiency

I. INTRODUCTION

Energy saving is a vital issue in the field of cellular communication, as it is in every field today. Energy efficiency (EE) [bit/Joule], defined as the amount of data successfully transmitted using unit energy, can be considered as an indicator of how efficiently energy is used in the field of communication [1]. In cellular communication, it is possible to increase energy efficiency by reducing values such as transmission power or the number of active antennas, but in cases where these actions take place, serious decreases are observed in spectral efficiency (SE). Spectral efficiency [bit/s/Hz], which expresses successful complex-valued samples transmitted using unit spectrum, is directly proportional to area throughput (TR). Area throughput is very important factor to avoid density of data traffic. Therefore, both EE and SE decrease are undesirable factors and these terms, which are in conflict with each other, should be kept high together [2].

Massive multi-input multi-output (Massive MIMO) is systems that contain a high number of receivers and transmitters in the cell and can be used in increasing EE [3-5], increasing SE [6-8], improving SE-EE trade-off [9,10]. In addition, there are studies that can determine optimum results on the SE-EE trade-off by using intelligent optimization techniques [11,12].

This study's main target is getting samples of successful combinations of parameters of affecting the SE-EE trade-off by intelligent optimizations. Multi-objective genetic algorithm (MOGA), multi-objective differential evolution algorithm (MODEA) and multi-objective particle swarm optimization (MOPSO) techniques have been used as intelligent optimization techniques and results close enough to true Pareto Front have been obtained. In the simulations, in order to observe the effect of the Pilot reuse factor (f) on the SE-EE trade-off, this parameter has been selected at three various values and analyses have been carried out.

In the rest of paper, information about system of simulations is given in Section 2, simulation results in Section 3, conclusion and discussion about obtained results are given Section 4.

II. SYSTEM MODEL

In a cellular network with the massive MIMO system, the inputs are the users served in the cell and the outputs are the active antennas in the base stations. The number of users served in a cell (K) and the number of active antennas in the base station (M) and transmission power (p) are the parameters that directly affect the SE and EE trade-off. These three parameters have been determined as the independent variables of this study.

The all curves of SE-EE values calculated for all combinations of K , M and p variables in certain ranges are given in Fig. 1. The calculations of this example are represented as in (1) and the results obtained are marked in green. Then all non-dominated solutions have been identified and shown in red.

$$[SE, EE] = \text{Calculate}(K, M, p) \quad (1)$$

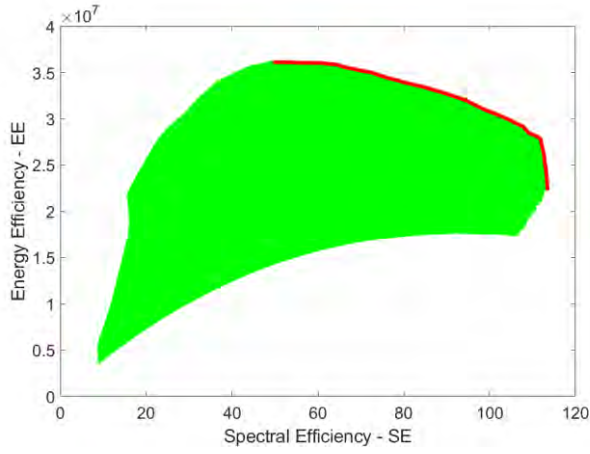


Fig. 1. Example of a true POF.

The curve obtained by combining all non-dominated elements is the true Pareto Optimal Front (POF). Since this curve is not have the regularity, it is not possible to formulate this curve in a simple way. In order to obtain a true POF, SE-EE calculations of all possible combinations must be made. Since it is not necessary to find all solutions through intelligent optimization, they are preferable to solve the problem. The optimizations have been outlined with the pseudo codes in Fig. 2. Although a true POF cannot be created with intelligent optimizations, the results obtained are generally within acceptable proximity to this curve.

```

Initialize the population with random combinations
Calculate the fitness values of population
while ( t < max number of iterations )
    Create a new combination via member i
    Calculate fitness value of new combination 'f(new)'
    if ( f(new) dominate f(i) )
        Replace the new combination with member i
    end if
    if ( rand < Rupdate )
        Create a new random combination and replace it with worst member of population
    end if
end while
Save the non-dominated members

```

Fig. 2. Pseudo code of the main logic of the intelligent optimizations.

By providing $M \gg 1$ and $M > K$ conditions, the effect of inter-cell interference can be minimized. In addition, by increasing the pilot reuse factor, inter-cell interference can be further reduced. However, this situation brings with it the need for more orthogonal Pilot series. The pilot reuse factor, which means the number of orthogonal pilot series used in neighbouring cells in a cellular network system, is exemplified in Fig. 3.

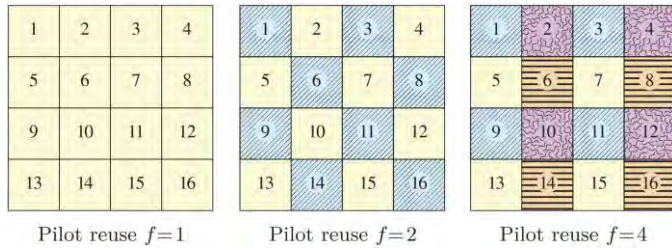


Fig. 3. Example of the pilot reuse factor.

In the examples in the figure, the cells using the same pilot series have been painted with the same pattern.

III. SIMULATION RESULTS

Sample cells have been created in the simulations and main parameters of these simulations are given in Table I.

TABLE I. MAINLY PARAMETERS OF SIMULATIONS

Simulation Parameters	
Parameter	Value
Network layout	Square pattern (wrap-around)
Number of cells	$L = 16$
Cell area	250m x 250m
Channel gain at 1 km	$\gamma = -148.1$ dB
Pathloss exponent	$\alpha = 3.76$
Shadow fading (standard deviation)	$\sigma_{sf} = 10$
Bandwidth	$B = 25$ MHz
Receiver noise power	-94 dBm
Samples per coherence block	$\tau_c = 400$
Pilot reuse factor	$f = 1, 2 \text{ or } 4$

The channels have been selected as correlated Rayleigh fading channels. Among the independent variables of the simulations, K can take values between 10-70 and M 10-100, while p can take values between 50-200 mW. An example of how energy efficiency changes according to K and M values is given in Fig. 4.

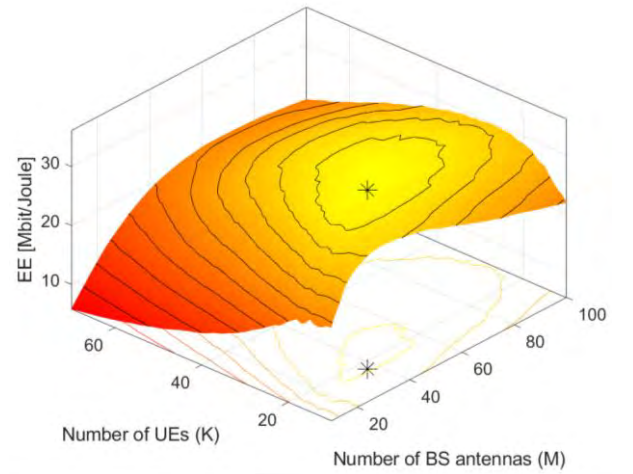


Fig. 4. EE values at various K and M .

For this example, the highest EE value has been calculated at the point $(K, M) = (18, 37)$.

In the studies, the population numbers, which are the common parameters of the MOGA, MOPSO and MODEA techniques, have been determined as 50, and the maximum numbers of iterations have been determined as 100. Table II

shows the basic parameters of the MOGA and their values [13].

TABLE II. PARAMETERS OF MOGA

<i>Parameter</i>	<i>Value</i>
Crossover Percentage	70%
Number of Parents (Offspring)	35
Mutation Percentage	40%
Number of Mutants	40
Mutation Rate	5%
Mutation Step Size	10%

Table III shows the parameters of MOPSO and the values they take [14].

TABLE III. PARAMETERS OF MOPSO

<i>Parameter</i>	<i>Value</i>
Repository Size	80
Inertia Weight	0.5
Inertia Weight Damping Rate	99%
Personal Learning Coefficient	1
Global Learning Coefficient	2
Inflation Rate	10%
Leader Selection Pressure	2
Deletion Selection Pressure	2
Mutation Rate	10%

The parameters of MODEA are given in Table IV [15].

TABLE IV. PARAMETERS OF MODEA

<i>Parameter</i>	<i>Value</i>
Lower Bound of Scaling Factor	20%
Upper Bound of Scaling Factor	80%
Crossover Probability	30%

All SE-EE points in the working range are shown in yellow, true POF in red, while values determined through intelligent optimizations are marked with blue asterisks. The results for the case where the pilot reuse factor is 1 are shown in Fig. 5.

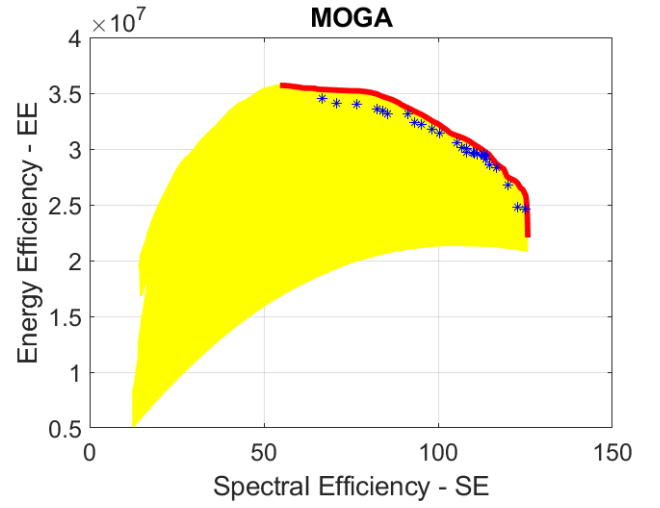


Fig. 5. (a) Comparing true POF and results of MOGA for $f=1$.

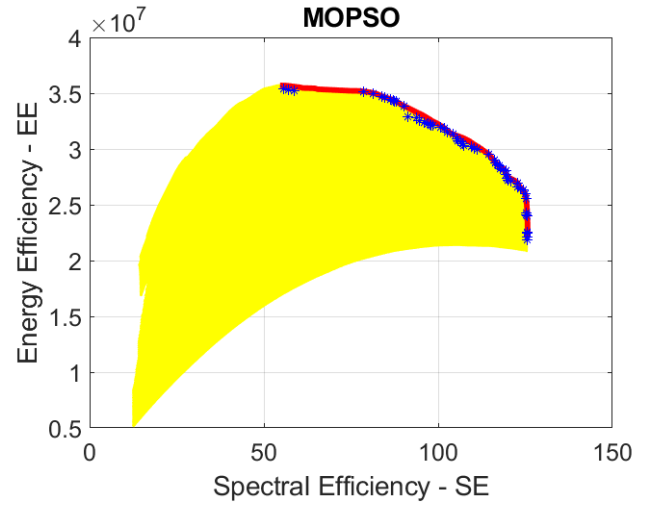


Fig. 5. (b) Comparing true POF and results of MOPSO for $f=1$.

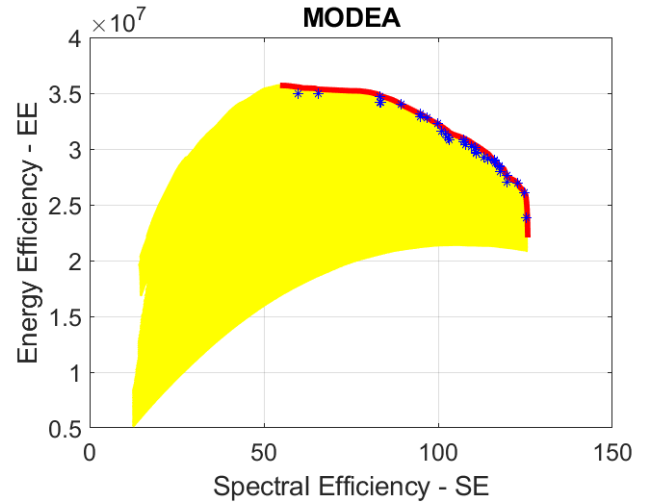


Fig. 5. (c) Comparing true POF and results of MODEA for $f=1$.

Fig. 6 shows the results for $f=2$.

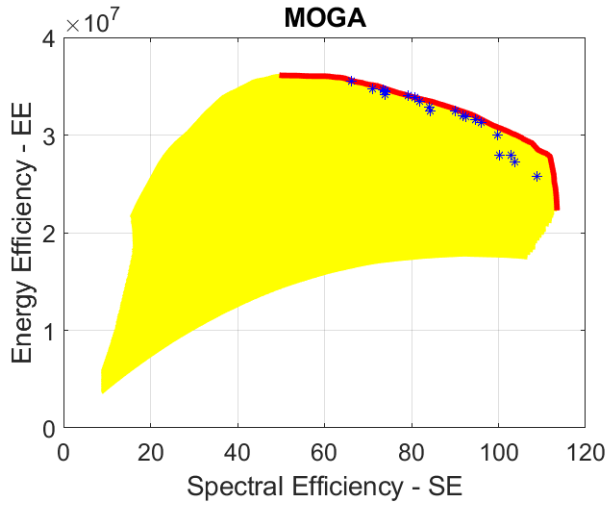


Fig. 6. (a) Comparing true POF and results of MOGA for $f=2$.

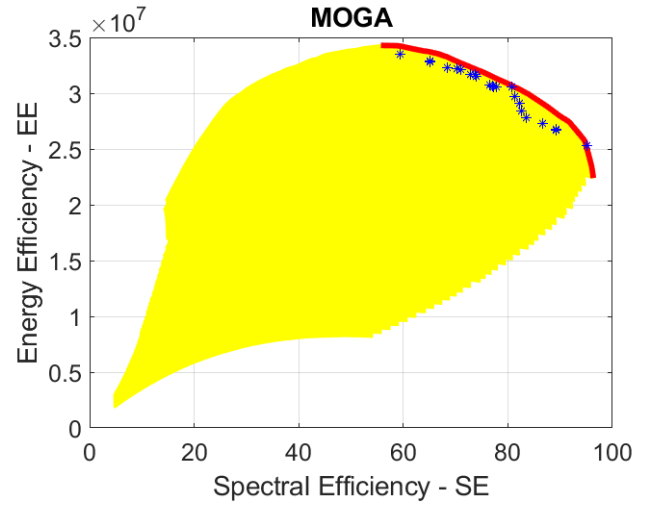


Fig. 7. (a) Comparing true POF and results of MOGA for $f=4$.

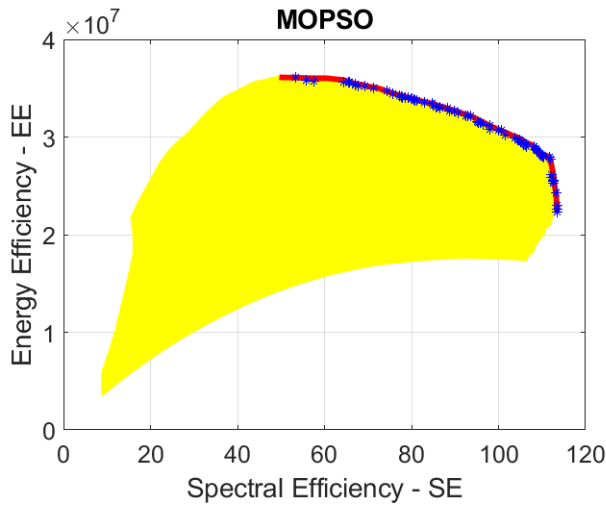


Fig. 6. (b) Comparing true POF and results of MOPSO for $f=2$.

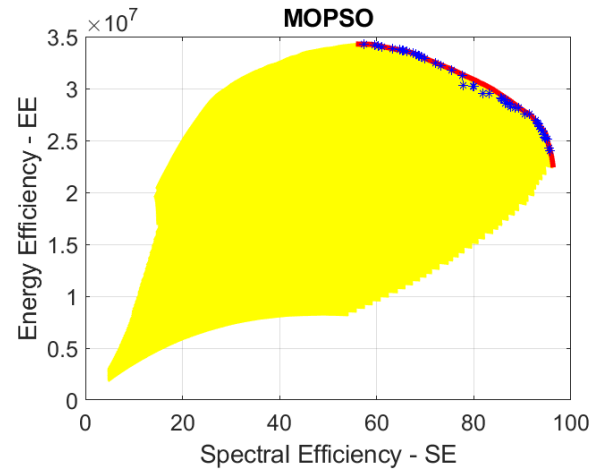


Fig. 7. (b) Comparing true POF and results of MOPSO for $f=4$.

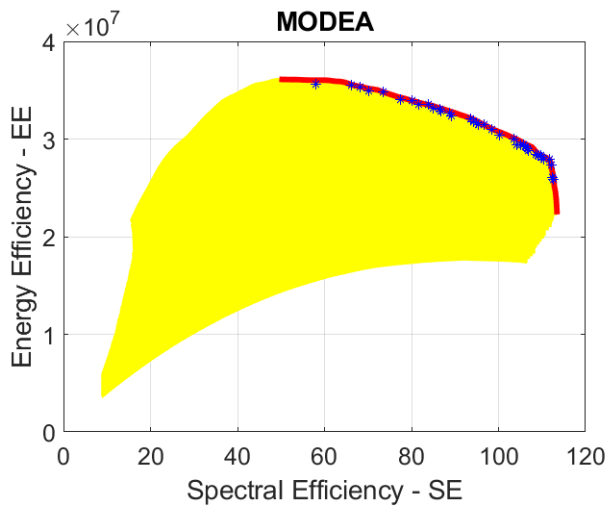


Fig. 6. (c) Comparing true POF and results of MODEA for $f=2$.

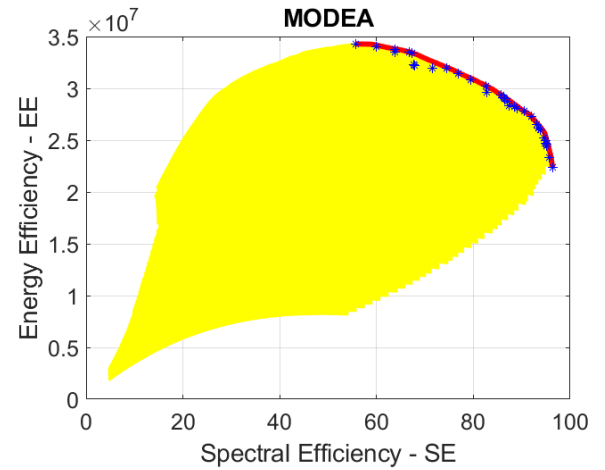


Fig. 7. (c) Comparing true POF and results of MODEA for $f=4$.

The findings obtained in cases when the pilot reuse factor is 4 are given in Fig. 7.

In Fig. 8, the true POF and the results found through the three algorithms for the situation $f = 1$ are shown on the same graph.

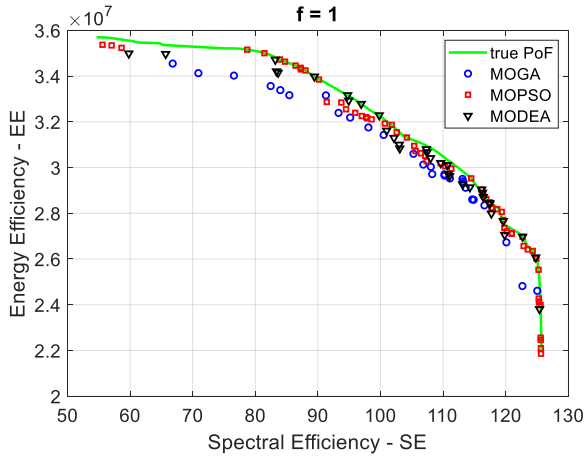


Fig. 8. Comparing true POF and results of all algorithms for $f = 1$.

Fig. 9 shows the results for the $f = 2$ condition.

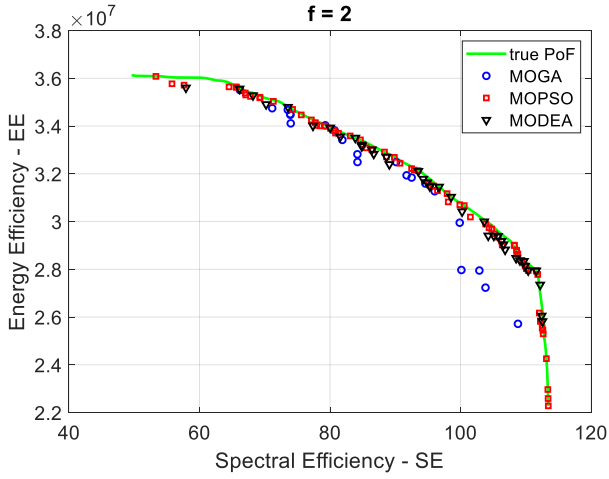


Fig. 9. Comparing true POF and results of all algorithms for $f = 2$.

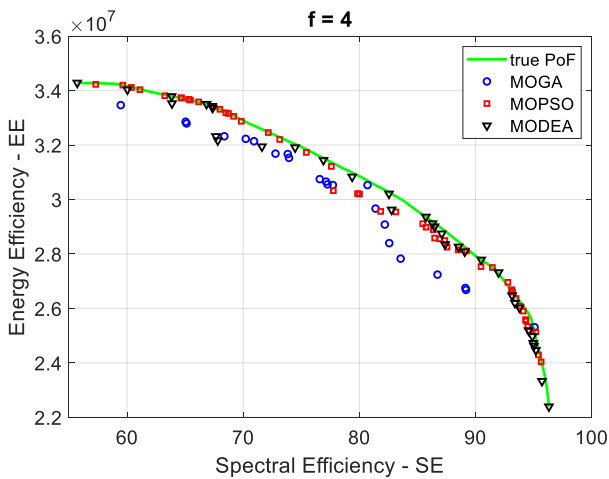


Fig. 10. Comparing true POF and results of all algorithms for $f = 4$.

Fig. 11 shows non-dominated solution numbers found by intelligent optimization algorithms depending on the number of iterations for $f = 2$.

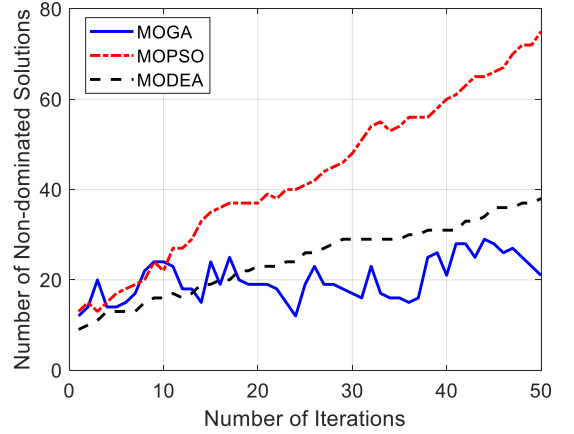


Fig. 11. Number of non-dominated solutions versus number of iterations.

In the optimization results in this example, MOPSO 75, MODEA 38, MOGA 21 detected non-dominated results.

IV. DISCUSSION AND CONCLUSION

In general, it is seen that in all cases the optimizations can achieve results close to true POF, but only in a few cases are successful enough to estimate the Pareto Front curve.

Inverted generational distance (IGD) represents the total Euclidean distance between estimated solutions and the true Pareto front. The smaller this term, the more successful the results. IGD values of each algorithm for the situations that pilot reuse factor equals to 1, 2 and 4 have been given in Table V. In these calculations, the unit of energy efficiency has been determined as [Mbit/joule].

TABLE V. THE IGD VALUES OF ALGORITHMS

Parameter	$f = 1$	$f = 2$	$f = 4$
MOGA	1.62	2.83	1.45
MOPSO	0.61	0.45	0.49
MODEA	0.99	0.99	0.62

Among the algorithms, it is seen that the MOPSO algorithm detects results closer to true POF. It is also seen that the results of the MOPSO algorithm are generally more evenly distributed over the solution set. In addition, it has been observed that when this algorithm is used, a much higher number of sample SE-EE values are obtained compared to other algorithms. In line with this information obtained, we can say that MOPSO is the most successful algorithm among the three algorithms mentioned.

It has been observed that the results are more successful when the pilot reuse factor is equal to four. However, the most successful result has been obtained with MOPSO when the pilot reuse factor is equal to two.

REFERENCES

- [1] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, The Global Footprint of Mobile Communications: The Ecological and Economic Perspective, *IEEE Transactions on Wireless Communications*, 49(8), 2011, pp. 55-62.
- [2] E. Björnson, J. Hoydis, and L. Sanguinetti, Massive MIMO Networks: Spectral, Energy and Hardware Efficiency, *Foundation and Trends in Signal Processing*, 2017, pp. 154-655.
- [3] H. Q. Ngo, E. Larsson, and T. Marzetta, Energy and spectral efficiency of very large multiuser MIMO systems, *IEEE Trans Commun* 61(4), 2013, pp. 1436–1449.
- [4] J. Fan, and Y. Zhang, Energy efficiency of massive MU-MIMO with limited antennas in downlink cellular networks, *Digit Signal Process* 86, 2019, pp. 1–10.
- [5] E. Sharma, R. Budhiraja, K. Vasudevan, and L. Hanzo, Full-Duplex Massive MIMO Multi-Pair Two-Way AF Relaying: Energy Efficiency Optimization, *IEEE Transactions on Communications* 66(8), 8329521, 2018, pp. 3322-3340.
- [6] W. Tan, S. Jin, and C. K. Wen, Spectral efficiency of multi-user millimeter wave systems under single path with uniform rectangular arrays, *J Wireless Com Network*, 181, 2017, pp. 1-13.
- [7] W. B. Hasan, P. Harris, A. Doufexi, and M. Beach, Real-time maximum spectral efficiency for massive MIMO and its limits, *IEEE Access* 6, 2018, pp. 46122-46133.
- [8] H. Pirzadeh, and A. L. Swindlehurst, Spectral efficiency of mixed-ADC massive MIMO, *IEEE Transactions on Signal Processing* 66(13), 2018, pp. 3599-3613.
- [9] Y. Huang, S. He, J. Wang, and J. Zhu, Spectral and Energy Efficiency Tradeoff for Massive MIMO, *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, 2018, pp. 6991-7002.
- [10] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang and X. You, Area Spectral Efficiency and Area Energy Efficiency of Massive MIMO Cellular Systems, *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, 2016, pp. 3243-3254.
- [11] Z. Liu, W. Du and D. Sun, Energy and Spectral Efficiency Tradeoff for Massive MIMO Systems With Transmit Antenna Selection, *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, 2017, pp. 4453-4457.
- [12] Y. Hei, C. Zhang, and W. Song, Energy and spectral efficiency tradeoff in massive MIMO systems with multi-objective adaptive genetic algorithm, *Soft Comput* 23, 2019, pp. 7163–7179.
- [13] T. Murata, and H. Ishibuchi, MOGA: Multi-objective genetic algorithms, *Proc. of 1995 IEEE International Conference on Evolutionary Computation*, 1995, pp. 289-294.
- [14] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, Handling multiple objectives with particle swarm optimization, in *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, 2004, pp. 256-279.
- [15] W. Gong, and Z. Cai, A multiobjective differential evolution algorithm for constrained optimization, *IEEE Congress on Evolutionary Computation*, 2008, pp. 181-188.

Computing on the Edge: A System and Technology Overview

Marija Poposka

Ss. Cyril and Methodius University
Faculty of Electrical Engineering and Information
Technologies
Skopje, Macedonia
poposkam@feit.ukim.edu.mk

Zoran Hadzi-Velkov

Ss. Cyril and Methodius University
Faculty of Electrical Engineering and Information
Technologies
Skopje, Macedonia
zoranhv@feit.ukim.edu.mk

Abstract—The Internet of Things is an emerging concept that unites an enormous number of smart devices continuously producing and exchanging data. Driven by the requirements of the IoT devices and applications, there has been a paradigm shift from centralized mobile cloud computing towards distributed mobile edge computing. Mobile edge computing is characterized by reallocation of the computation functions from the user device to the network edge, usually to the base station. In this way, computation-intensive and time-sensitive applications are offloaded from the user devices that usually have limited resources and capabilities. This paper covers a broad outlook on mobile edge computing, including its architecture, comparison between mobile edge and cloud computing and computation and communication models. Finally, we analyze how mobile edge computing can be applied to three different technologies: Internet of Things, wireless powered communications, and intelligent reflective surfaces.

Keywords—Internet of Things (IoT); cloud computing (CC); mobile edge computing (MEC);

I. INTRODUCTION

The ubiquity of myriads of mobile terminals, sensors and machines in the Internet of Things (IoT) era leads to new challenges and requirements for mobile networks [1]. Since IoT devices have limitations in terms of storage, computational power, battery and speed, resource-intensive, they cannot easily support tasks and applications that require greater memory or computational resources [2]. To overcome this drawback, wireless networks include a computing technology that has been quite relevant for the past decade, i.e. Cloud Computing (CC) [3]. By offloading computation and data to the mobile cloud, resource and computer-intensive applications can be run on small devices with limited resources [4], [5]. However, CC is not acceptable for a great variety of time-critical or real-time applications, due to the exaggerated high latency caused by long propagation distance from the mobile users to the cloud center, since the data is sent to remote servers within the Internet, far away from the users [6]. The main purpose of mobile edge computing (MEC) is to address the problems encountered in CC and to enable competent and consistent integration of cloud computing into the mobile network. Namely, MEC offers storage and processing capabilities at the edge of the mobile network, i.e. base station (BS), but within the radio access network (RAN). MEC functionalities are based on Network function virtualization (NFV) that enables a single edge device (MEC server) to provide computation and storage

power for many mobile devices by creating multiple virtual machines that simultaneously perform different network functions [7]. Since the MEC servers are located close to the end users, the latency is significantly reduced, while the bandwidth is increased, which makes it applicable in latency-critical applications and sets it apart as one of the crucial elements of 5G/6G [8]–[13]. In addition, MEC is really suitable for combining and integrating with other technologies, such as IoT [14], wireless powered communications (WPC) [15], [16], intelligent reflective surfaces [17], [18], in a way that mutually-beneficial relationships are enabled. Besides aforementioned technologies, MEC is constantly applied in current and emerging technologies including healthcare and Internet of Medical Things [19], artificial intelligence and blockchain [20], as well as unmanned aerial vehicle assisted mobile networks [21].

II. REVIEW OF MEC

A. System architecture

Since the MEC layer is located between the mobile devices and the cloud, MEC networks are characterized by three-layer architecture as shown in Fig.1: 1) User Devices Layer, 2) MEC Layer and 3) Cloud/Core Layer. In the lowest layer, myriads of mobile devices, IoT devices and sensors are connected to the core networks mediated by the BS located in their immediate vicinity. The MEC layer is represented by the RAN which includes BSs that have incorporated MEC servers. The aim of this layer is to connect the mobile devices with the core/cloud layer, while optimizing RAN services by pushing intelligence at the base station. Hence, all devices are connected to the

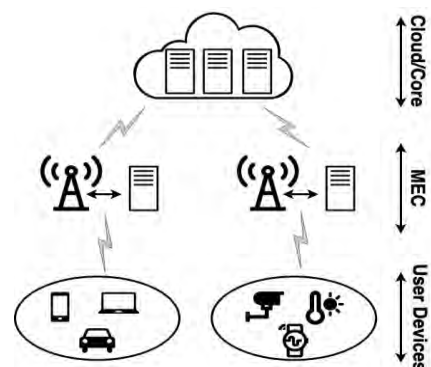


Fig. 1. MEC architecture

nearest MEC server contributing to decentralized network architecture, contrary to CC which is fully centralized. The Core/Cloud Layer connects all BSs and MEC servers with the core network. As the main concern of MEC is to provide network edges with computing and storage capabilities, it is clear that the performance of the MEC network strongly depends on its architecture [9].

B. Advantages of MEC over CC

Although MEC is originally derived from Cloud computing, MEC has few evident advantages worth mentioning:

- **Proximity:** MEC brings computational resources and services closer to the physical location of the mobile user, resulting in improved user experience. This significantly reduces the latency in accessing the services, so numerous new possibilities opened up, such as analyzing the user's behavior in real-time, big-data analytics with better accuracy and leveraging the network context information [10].
- **Low latency:** The overall latency in a mobile network is composed of three parts: first one depends on the propagation distance, the second one depends on the computation capacities of network entities and the last one depends on the information rate. Since MEC servers are located in proximity to the users, propagation distances vary from tens of meters to maximum one kilometer, depending on the cell size, resulting in low propagation latency. On the other hand, propagation distances in CC vary from tens to hundreds of kilometers, as servers may be located across continents. Additionally, until the information reaches its final destination, it has to pass through several different networks, each with different routing and traffic control protocols, adding additional delay. When it comes to computation latency, although CC servers without a doubt dominate with their computational capabilities compared to MEC servers, they serve a much larger number of users than edge servers, thus balancing the gap that occurs in the computation delay. It is this step forward that makes MEC key technology for realizing latency-critical applications and technologies such as Tactile Internet and IoT [11],[14].
- **Context-awareness:** This concept refers to the ability of user devices (UDs) to be aware of the circumstances and environment (i.e. context) that they work under. The goal is to gather real-time data from the UD, such as user activity, location, time and network load, and to interpret it correctly in order to adjust its behavior to the requirement to the particular people or particular situation. In this regard, proximity of UD to MEC servers provides real-time reactivity and responsiveness to the dynamic changes in the environment and contributes towards improving the quality of user experience [12].
- **Energy savings:** One of the main challenges in IoT is to optimize the power consumption by IoT devices,

that have limited storage capabilities and recurrent battery charging or replacement is unfeasible and even impossible. Thanks to the computational offloading, MEC can tackle this problem by prolonging IoT device's battery lives by reducing energy consumption through offloading computation-intensive tasks to MEC servers [13].

C. Computation offloading

Thanks to the computation offloading characteristic of MEC, the energy consumption at UD is reduced and the computation process is speeded up, which reduces the computation delay. The main challenge regarding computation offloading is to decide whether offloading should be done, as well as how much and which part of the data should be offloaded. In this regard, we differentiate three different offloading models: 1) *Local computation*, when no offloading is performed, i.e., the whole computation is done locally at the user device and this model is applicable when offloading is not feasible or does not pay off; 2) *Fully offloading*, when the computation task is relatively simple or highly integrated and has to be fully offloaded to the MEC server and 3) *Partial offloading*, when the computation is partitioned into two parts, one part executed at the user device and the rest is offloaded for the MEC execution. The local execution and fully offloading can also be found in the literature under one name - *binary offloading*. The choice of an appropriate offloading model is constrained by many elements, including: computation capabilities of UD and MEC servers, quality of the radio and backhaul links, application type, task type and user requirements. A task Q is commonly described by the input data size, L (in bits), execution time deadline, τ_d (in seconds) and the computation intensity, X (in CPU cycles per bit) - $Q(L, \tau_d, X)$. By using these three simple parameters, it is possible to evaluate the main system variables that affect the performance of MEC networks, i.e., energy consumption and computation latency. Therefore, the main objective of the works focussed on MEC offloading is minimization of the energy consumption and computation latency or trade-off between them by proposing optimal resource allocation policies [4]. Further to the previous discussion, we will present computational models of UD and MEC servers, with special reference to the computation latency and energy consumption.

- **User Devices:** The CPU handles all computation tasks in UD and is characterized by its clock speed f_m . This parameter directly affects the computation latency in the following way:

$$t_m = \frac{LX}{f_m} \quad (1)$$

showing that higher CPU clock speed contributes to lower computation latency, but at the same time causes an increase in the CPU energy consumption, according to dynamic frequency and voltage scaling technique. The energy consumption of a CPU can be calculated accordingly to [5]:

$$E_m = \alpha LX f_m^2 \quad (2)$$

where α is a constant related to the hardware architecture. From the above we can conclude the need for computation offloading: when the energy consumption required by the task is greater than the energy stored at the battery at a given moment, or when the computation latency prevents a given task from being completed within the execution deadline.

- **MEC servers:** MEC servers serve different UDs in two ways: when the MEC server has sufficient computation capabilities, it allows independent computation for each user device by allocating different VMs and when the MEC server has insufficient computation capabilities, it processes the tasks sequentially, resulting in additional queuing delay. Overall MEC server computation latency for device k consists of CPU computation time and queuing delay and is given by:

$$T_{s,k} = \sum_{i \leq k} t_{s,i}, \quad (3)$$

where $t_{s,j} = \frac{\omega_j}{f_{s,j}}$ is the server computation time for user device j , ω_j is the number of CPU cycles required for the computation offloading and $f_{s,j}$ is allocated CPU cycle frequency for user device j . Let us assume that MEC server K computation tasks, so analogous to (2), the total energy consumption at the MEC server is:

$$E_s = \sum_{k=1}^K \alpha \omega_k f_{s,k}^2. \quad (4)$$

D. Communication Model

In CC communication systems, the channel between the UDs and the server is represented as a bit pipe with constant or random rate. When it comes to MEC, it focuses on small-scale edge services and the aim is to have an efficient air interface to support latency-sensitive applications. Since the wireless fading channels vary randomly in time and frequency, it is important to effectively integrate computation offloading and wireless transmission based on the channel state information. Illustration of fading propagation environment is shown in Fig. 2, where at time m , the complex symbol $x[m]$ is transmitted from the BS. Transmitting through the wireless channel, the signal encounters obstacles on its way that cause the UD to receive delayed replicas of the transmit signal. The received signal at the UD is:

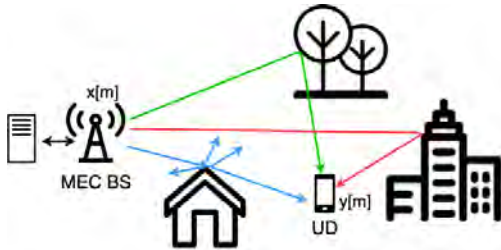


Fig. 2. Wireless communications in fading environment

$$y[m] = \sum_l h_l[m] x[m-l], \quad (5)$$

where $h_l[m]$ is the l th channel filter tap at time m . Problems occur when the channel is in deep fading and there are several ways to tackle this problem: one is to postpone the computation offloading until the SNR achieves sufficient value, to switch to an alternative channel with better quality or to increase the transmit power in order to increase the information rate [7]. The communications in MEC systems commonly occur between UDs and BSs which are co-located with the MEC servers. In case of scarce wireless interface between the UDs and the MEC servers, device-to-device communication enables computation offloading to MEC servers. Additionally, BSs provide access to remote servers through backhaul links, giving MEC servers an opportunity to offload redundant tasks to other small-scale MEC servers or to some large-scale cloud data center.

III. MEC AND OTHER TECHNOLOGIES

In this section, we will review the basic concepts of IoT, wireless powered communications (WPC) and intelligent reflective surface (IRS) and discuss the potential for integrating these technologies with MEC.

A. MEC and WPC

As we mentioned before, extending the battery lifetime and improving the computation capabilities of low-complexity IoT devices are crucial and challenging tasks, and how to tackle these two fundamental performance limitations is a critical research problem. Radio frequency based wireless power transfer (WPT) has emerged as an effective way to prolong finite battery lifetime [15]. WPT uses dedicated energy transmitters to wirelessly charge remote energy-harvesting users (EHUs). Additionally, this paradigm can be jointly combined with wireless communications in order to achieve ubiquitous wireless communications in a self-sustainable way and wireless powered communications have been proposed. If we integrate MEC to this system design, the EHUs can execute their computational tasks locally by themselves or offload all or part of them to the BS. In this way, optimal BS transmit power, CPU frequency, time allocation among EHUs and computation latency can be derived in order to improve system performance [16].

Let's assume we have MEC assisted WPC network consist of single antenna MEC BS and K single antenna EHUs. Each EHU can perform only one of the following computation actions: local computing (mode-0 EHU) or offloading the data to the MEC server (mode-1 EHU). When working in mode-0, the EHU achieves certain computation rate C_k , but when working in mode-1, the EHU's computation rate matches its information rate $C_k = R_k$. Time is divided into TDMA frames of equal duration T , consist of WPT phase of duration $\tau_0 T$, when EHUs harvest energy from the BS that transmit with constant power P_0 and phase of duration $(1 - \tau_0)T$, when EHUs offload data to the BS. For this system model, an optimization problem for maximizing the minimum computation rate of the k th EHU can be proposed, which results in decision criteria for of EHU's

computation mode. In case of mode-0 EHU, the computation time t_k is optimized, while for mode-1 EHU the optimization variables are the offloading time τ_k and the information rate R_k :

$$\max_{C_k} \min_{R_k, \tau_k, t_k} C_k. \quad (6)$$

When EHU operates in mode-0, its computational rate is given by:

$$C_k = \frac{f_k t_k}{\phi T}, \quad (7)$$

where f_k is CPU's computation speed, t_k is computation time and ϕ is the number of computation cycles. On the other hand, when EHU operates in mode-1, the computation rate is:

$$C_k = \frac{B \tau_k}{v_u} \log_2 \left(1 + \frac{\eta \tau_0 P_0 \Omega_k^2}{\tau_k N_0} \right) \quad (8)$$

where B is the communication bandwidth, τ_k is the offloading time for the k th device, η denotes the energy harvesting efficiency ($0 \leq \eta \leq 1$), Ω_k is the gain of the wireless channel between the BS and the k th EHU, N_0 is receiver noise power and, v_u is communication overhead in task offloading.

The constraints regarding this optimization problem may apply to the energy harvesting and offloading time $C1: \sum_k \tau_k + \tau_0 \leq 1, \tau_k \geq 0, \tau_0 \geq 0$; harvested energy for mode-0 EHUs $C2: c_k f_k^3 t_k \leq \eta \tau_0 P_0 \Omega_k^2 T$, where c_k CPU's denotes computation energy efficient coefficient; processors speed and computation time for the mode-0 EHUs $0 \leq t_k \leq T, 0 \leq f_k \leq f_{max}$.

In this way, not only the computation efficiency will be enhanced, but additionally the system fairness will be improved.

B. MEC and IRS

Although MEC brings numerous benefits in terms of reduced latency and energy consumption, it does not face the problems occurring in the wireless channel between the UDs and MEC server. Namely, the wireless environment between UDs and MEC servers is unpredictable and is prone to channel outages, when the channel is in deep fading. Recently, an emerging technology called Intelligent Reflective Surfaces received great attention, since it can improve the propagation environment in a controllable way and enhance the communication between the UDs and the access point. An IRS is a planar array consisting of a large number of passive reflective elements with reconfigurable phase shifts, dynamically controlled by a software controller to reflect the impinging signal in the desired

direction [17]. Driven by the idea of improving MEC performances, authors in [18] proposed RIS-assisted MEC system, where a large IRS is placed close to the UDs to assist their computation offloading to the MEC BS, as shown in Fig. 3. In this way, when the offloading link from the k th UD to the BS, $h_{BS,k}$ is hostile, the computational tasks can be offloaded with the help of IRS through the composite link consist of two parts: from the k th UD to the IRS, $h_{IRS,k}$ and from the IRS to the BS, g . This can reduce the computation latency and provide additional degrees of freedom to further improve the system performances.

IV. CONCLUSION

MEC is a promising, novel network architecture with great potential in offering extended battery lifetime and computation capabilities to resource-constrained UDs and time-sensitive applications. Compared to cloud computing, edge computing will bring data computing and storage closer to the users. In that way, MEC contributes to reduced system latency suitable for real-time IoT applications and extended battery lifetime of low-complexity IoT devices. However, there are still many challenges and open questions in the field of MEC. One of the main concerns is security, since computation tasks are offloaded to the MEC server via wireless medium, which is subject to attacks and interceptions. Another challenging issue is dealing with user's mobility and providing mobility management techniques, as moving and mobile devices can be the reason for link disconnections between the MEC servers and UDs. Last, but not least, scalability is a challenging and important issue, especially in the IoT era. MEC servers should implement load balancing mechanisms in order to ensure service availability, even when the number of UDs is immense.

REFERENCES

- [1] M. R. Palattella *et al.*, "Internet of Things in the 5G Era: Enablers, Architecture, and Business Models," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016, doi: 10.1109/JSAC.2016.2525418.
- [2] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile Edge Computing: A Survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018, doi: 10.1109/JIOT.2017.2750180.
- [3] M. Bahrami, "Cloud Computing for Emerging Mobile Cloud Apps," in *2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, Mar. 2015, pp. 4–5. doi: 10.1109/MobileCloud.2015.40.
- [4] O. Muñoz, A. Pascual Iserte, J. Vidal, and M. Molina, "Energy-latency trade-off for multiuser wireless computation offloading," in *2014 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Apr. 2014, pp. 29–33. doi: 10.1109/WCNCW.2014.6934856.
- [5] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-Optimal Mobile Cloud Computing under Stochastic Wireless Channel," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 9, pp. 4569–4581, Sep. 2013, doi: 10.1109/TWC.2013.072513.121842.
- [6] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Communicating While Computing: Distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 45–55, Nov. 2014, doi: 10.1109/MSP.2014.2334709.
- [7] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Commun. Surv. Tutor.*, vol. 19, no. 4, pp. 2322–2358, Fourthquarter 2017, doi: 10.1109/COMST.2017.2745201.
- [8] S. Safavat, N. N. Sapavath, and D. B. Rawat, "Recent advances in mobile edge computing and content caching," *Digit. Commun. Netw.*, vol. 6, no. 2, pp. 189–194, May 2020, doi: 10.1016/j.dcan.2019.08.004.
- [9] D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust,

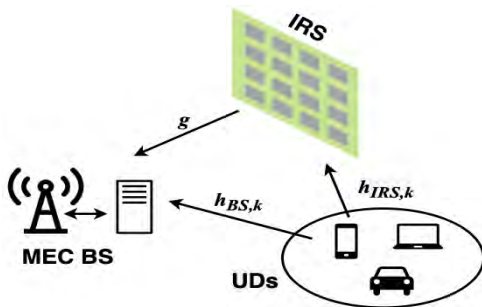


Fig. 3. IRS assisted MEC wireless network

- “Mobile-Edge Computing Architecture: The role of MEC in the Internet of Things,” *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 84–91, Oct. 2016, doi: 10.1109/MCE.2016.2590118.
- [10] S. Josilo and G. Dán, “Selfish Decentralized Computation Offloading for Mobile Cloud Computing in Dense Wireless Networks,” *IEEE Trans. Mob. Comput.*, vol. 18, no. 1, pp. 207–220, Jan. 2019, doi: 10.1109/TMC.2018.2829874.
- [11] G. P. Fettweis, “The Tactile Internet: Applications and Challenges,” *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014, doi: 10.1109/MVT.2013.2295069.
- [12] S. Nunna *et al.*, “Enabling Real-Time Context-Aware Collaboration through 5G and Mobile Edge Computing,” in *2015 12th International Conference on Information Technology - New Generations*, Apr. 2015, pp. 601–605. doi: 10.1109/ITNG.2015.155.
- [13] B. Shi, J. Yang, Z. Huang, and P. Hui, *Offloading Guidelines for Augmented Reality Applications on Wearable Devices*. 2015, p. 1274. doi: 10.1145/2733373.2806402.
- [14] W. Yu *et al.*, “A Survey on the Edge Computing for the Internet of Things,” *IEEE Access*, vol. 6, pp. 6900–6919, 2018, doi: 10.1109/ACCESS.2017.2778504.
- [15] S. Bi, C. K. Ho, and R. Zhang, “Wireless powered communication: opportunities and challenges,” *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 117–125, Apr. 2015, doi: 10.1109/MCOM.2015.7081084.
- [16] F. Wang, J. Xu, X. Wang, and S. Cui, “Joint Offloading and Computing Optimization in Wireless Powered Mobile-Edge Computing Systems,” *IEEE Trans. Wirel. Commun.*, vol. 17, no. 3, pp. 1784–1797, Mar. 2018, doi: 10.1109/TWC.2017.2785305.
- [17] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, “Wireless Communications Through Reconfigurable Intelligent Surfaces,” *IEEE Access*, vol. 7, pp. 116753–116773, 2019, doi: 10.1109/ACCESS.2019.2935192.
- [18] T. Bai, C. Pan, Y. Deng, M. ElKashlan, A. Nallanathan, and L. Hanzo, “Latency Minimization for Intelligent Reflecting Surface Aided Mobile Edge Computing,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, Nov. 2020, doi: 10.1109/JSAC.2020.3007035.
- [19] Md. A. Rahman and M. S. Hossain, “An Internet of Medical Things-Enabled Edge Computing Framework for Tackling COVID-19,” *IEEE Internet Things J.*, pp. 1–1, 2021, doi: 10.1109/JIOT.2021.3051080.
- [20] D. C. Nguyen *et al.*, “Federated Learning Meets Blockchain in Edge Computing: Opportunities and Challenges,” *IEEE Internet Things J.*, pp. 1–1, 2021, doi: 10.1109/JIOT.2021.3072611.
- [21] H. Xu, W. Huang, Y. Zhou, D. Yang, M. Li, and Z. Han, “Edge Computing Resource Allocation for Unmanned Aerial Vehicle Assisted Mobile Network With Blockchain Applications,” *IEEE Trans. Wirel. Commun.*, vol. 20, no. 5, pp. 3107–3121, May 2021, doi: 10.1109/TWC.2020.3047496.

Mobile Edge Computing services with QoS support for beyond 5G Networks – Use Cases

David Nunev¹ Tomislav Shuminoski² Bojana Velichkovska³ Toni Janevski⁴

¹Virtual

Skopje, R.N. Macedonia

^{2,3,4}Ss. Cyril and Methodius University,

Faculty of Electrical Engineering and Information Technologies

Skopje, R.N. Macedonia

davidnunev@yahoo.com¹ tomish@feit.ukim.edu.mk² bojanav@feit.ukim.edu.mk³ tonij@feit.ukim.edu.mk⁴

Abstract—This paper presents a novel research in intelligent multi-access QoS mobile edge computing (MEC) for beyond 5G services. Also, the improved advanced QoS model and architecture for beyond 5G systems and services are proposed. The proposed model combines the most powerful features of both Cloud and Edge computing, independent from any existing and future Radio Access Technology, leading to high performance utility networks with high QoS provisioning for any used multimedia modern service over present and future mobile and wireless networks and systems. Moreover, the proposed architecture will allow applications and network services to be executed at the edge part of the network, giving lower end-to-end delay for the end-user services and applications. Finally, this paper gives an overview of the existing Mobile Edge Computing technologies and several use cases. Undoubtedly, MEC is an innovative network paradigm going beyond 5G to cater for the unprecedented growth of computation demands and the ever-increasing computation quality of user experience requirements.

Keywords—Aggregation; Cloud; Edge Computing; Machine Learning; Quality of Service; Vertical Multi-Homing.

I. INTRODUCTION

The emerging future Mobile Broadband Internet applications require high demands for improved Quality of Service (QoS), which would be supported by services that are orchestrated on-demand and are capable of adapt at runtime, depending on the contextual conditions, in order to provide reduced latency, high bandwidth utilization, high mobility, high scalability, and real time execution with cloud computations. Recent years show increased interest in transferring computing from Clouds towards the network edges or Mobile Edge Computing (MEC). The 5G is an emerging technology that is growing exponentially, supporting many advance services, concepts and networks. Consequently, the 5G is including within this paradigm called Mobile Edge Computing. As 5G is already deploying in many countries round the globe, millions of new devices are deployed and will play part in various present and future networks and architectures that will benefit from the advantages that 5G offers. This means that the number of

services offered by cloud and service providers will increase exponentially and all of that data could be overwhelming even for the cloud's (almost) unlimited resources. On the other side, the existing cloud computing (CC) solutions, cannot completely fulfill and cannot effectively cope with all these requirements and demands. Therefore the MEC concepts and sometimes Fog Computing appeared to resolve these challenges [1]. These concepts distribute computing, data processing, and networking services close to the end users, where computing and intelligent networking can best meet user needs. MEC and Fog Computing provide an infrastructure where distributed edge and user devices collaborate with each other, as well as, with the CC centers, in order to carry out computing, control, networking, and data management tasks. Also, there are significant disparities between MEC and CC systems in terms of computing, data storage, distance to end users and end-to-end latency. MEC has the advantages of achieving lower latency due to the shorter distances, saving energy for the mobile devices, supporting context aware computing and enhancing the privacy and security for mobile applications. Both MEC and CC in the core of their networks are using Network Virtualization, which is a powerful combination of SDN (Software-defined networking) and NFV (Network functions virtualization) infrastructure. In this paper, the user-centric approach is accepted as a basis for our work on Mobile Edge Computing system model, where the future the Mobile Terminals (MTs) would have access to different radio access technologies at the same time and should be able to combine different flows from different technologies using advance QoS algorithms within the Cloud orchestrator for used multimedia services, using vertical multi-homing and multi-streaming performances [2], [3].

II. FUNDAMENTALS AND BACKGROUND OF MEC

The tremendous interest and developments of mobile broadband Internet networks, undoubtedly lead to intensive research works towards advanced mobile and cloud

computing algorithms and frameworks for high level of QoS provisioning in each core and access network. At the first place, the main motivation for our proposed intelligent multi-access QoS provisioning framework could be found in [2-6]. Device-centric multi-RAT architectures, native support of machine-to-machine communications and smarter devices are part of the main trend for 5G [6-8]. Moreover, our framework and design of a novel MT with Mobile Fog CC support is a next step from previous works on adaptive QoS provisioning in heterogeneous wireless and mobile IP networks [5, 6]. Those papers were introducing a framework adaptive QoS provisioning module that provides the best QoS and lower cost for a given multimedia service by using one or more radio access technologies (RATs) at a given time. A key concept that allows highlighting the potential of CC environment is orchestration that aims to coordinate the execution of a set of virtualized services within the same process. The orchestration concept has been widely studied in the context of Web services [9]. Recently it was extended in the CC domain, in order to perform an optimization and management of both physical and virtual resources in complex, federated or multi-cloud environments.

Furthermore we are giving an overview of fundamentals for MEC and MEC in 5G.

The key idea of MEC is in providing an Internet broadband service environment and cloud-computing capabilities at the edge of the mobile network part, within the RAN and in close proximity to MTs. In the foreseeable future, MEC will open up new markets for different industries and sectors by enabling a wide variety of 5G use cases, e.g., Internet of Things / Internet of Everything, Industry 4.0, Vehicle-to-everything (V2X) communication, smart city, Tactile Internet and etc. According to the ETSI [10], [11] white paper MEC can be characterized by some features, namely on-premises, proximity, lower latency, location awareness, and network context information.

Furthermore, there are several use cases for MEC [12]:

- *RAN-Aware video Optimization*: Video is currently taking half of mobile network traffic and set to exceed 70% of traffic over the next couple of years. Providing throughput guidance information is one of the MEC use cases. The proposed solution is to use MEC technology to inform the video server on the optimal bit rate to use given the radio conditions for a particular stream.
- *Video Analysis Service*: Many recognition type application could benefit from the MEC architecture, mostly by the proximity of the computation that is executed at the edge devices. Whenever some video data needs to be analyzed, it can be sent to the MEC server and only needed data can be sent to the centralized cloud. The system benefits of low latency and avoids the problem of network congestion.
- *Augmented Reality Service*: Augmented Reality (AR) is a live view of a real world environment whose elements are supplemented by computer generated inputs such as sound,

video, graphics or other data, A MEC based AR application system should be able to distinguish the requested contents by correctly analyzing the input data and then transmit back the AR data back to the end user.

- *Enterprise and Campus Networks*: In large enterprise organizations, there is a desire to process users locally rather than backhaul traffic to centralized mobile core just so that it can send the data back again. This could be for services as simple as access to corporate intranet, (4k video training to a mass of employees at the same time) or more advanced services such as security policy, location tracking and asset tracking services.
- *IoT Applications*: MEC can be used to process and aggregate the small packets generated by IoT services before they reach the core network. Much of the data generated in a smart building is inherently local and involves D2D communication, so the benefits of this would be from moving the local computing and security, tracking, climate control to the edge servers and process and work with that data closer to the user without significant latency.

Furthermore, there are also many benefits in cooperation of MEC with SDN. The benefits of programmable networks align with the MEC requirements, and the recent form of SDN has the ability to mitigate the barriers that prevent Mobile Edge Computing to reach its full potential. All data flow management, service orchestration and other management tasks are done by the central SDN controller that is transparent to the end-user. Moreover, MEC will fit into the 5G concepts and what specifications have been developed based on the industry consensus. The 3GPP clarified how to deploy MEC in and seamlessly integrate MEC into 5G, which can be illustrated in [10]. Actually, the architecture comprises two parts: the 5G service-based architecture (SBA) and a MEC reference architecture. The network functions defined in the 5G architecture, and their roles can be briefly summarized as: Access and Mobility Management Function (AMF); Session Management Function (SMF); Network Slice Selection Function (NSSF); Network Repository Function (NRF); supports the discovery of network functions and services; Unified Data Management (UDM); Policy Control Function (PCF); Network Exposure Function (NEF); Authentication Server Function (AUSF); User Plane Function (UPF).

The MEC orchestrator (MECO) is the core component of the MEC system level, which maintains information on deployed MEC hosts (i.e., servers), available resources, MEC services, and topology of the entire MEC system. The MECO is also responsible for selecting of MEC hosts for application instantiation, onboarding of application packages, triggering application relocation, and triggering application instantiation and termination. The host level management consists of the MEC platform manager and the virtualization infrastructure manager (VIM). The MEC platform manager carries out the duties on managing the life cycle of applications, providing element management functions, and controlling the application rules and requirements. The MEC platform manager also processes

fault reports and performance measurements received from the VIM. Meanwhile, the VIM is in charge of allocating virtualized resources, preparing the virtualization infrastructure to run software images, provisioning MEC applications, and monitoring application faults and performance. Finally, the MEC host comprises a MEC platform and a virtualization infrastructure. New functional enablers were defined in [13] to integrate MEC into the 5G.

III. INTEGRATED 5G AND MEC ARCHITECTURE AND USE CASES

For supporting the large scale of network connections, 5G uses the tremendous computation and storage resources from remote datacenter and utilizes NFV and SDN technologies to virtualize the network resources for achieving an end-to-end optimized system for service provisioning. However, one issue that 5G network suffers from is the high latency, which could not meet the requirements of the emerging IoT applications. For solving this issue, MEC can be deployed in 5G gNB to eliminate the latency in the core network transmission, enhancing the service provisioning capability of 5G network for small-scale and ultra-low-latency services and application scenarios. As shown in Fig. 1, the future 5G mobile communication network will be a heterogeneous communication network that includes both the centralized Base Station (BS) and multiple distributed BSs. For integrating the MEC and 5G networks, Fig. 1 shows a multi-level computing network that provides edge computing and cloud computing functions. Within this architecture (also presented in [13]), MEC computing resources are allocated in LTE eNB, 5G gNB, super 5G BS, the edge of core networks to provide the computing and storage resources for end-users. To emphasize, there are many benefits of employing MEC into IoT systems, including but not limited to, lowering the amount of traffic passing through the infrastructure and reducing the latency for applications and services.

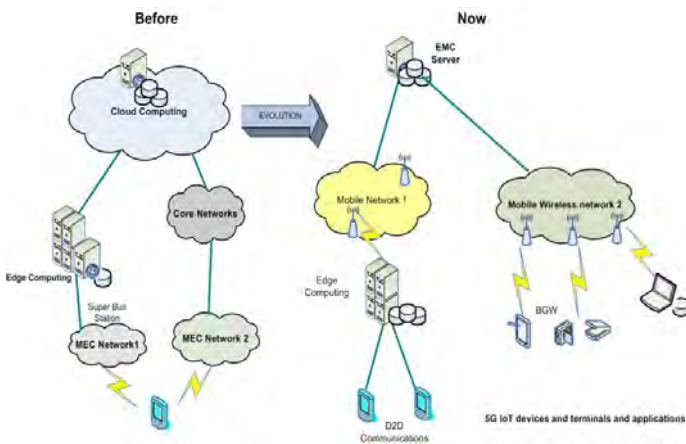


Fig. 1. A 5G combined with MEC.

Among these, the most significant is the low latency introduced by MEC which is suitable for 5G Tactile Internet applications requiring round-trip latency in the millisecond range. MEC technologies are envisioned to work as gateways placed at the middle layer of IoT architecture which can aggregate and process the small data packets generated by IoT services and provide some additional special edge functions before they reach the core network; hence, the end-to-end latency can be reduced.

In addition, based on the context and platforms of MEC, artificial intelligence (AI) or ML (Machine Learning) on the edge can gain the huge benefit to realize distributed IoT applications and intelligent system management, which is now considered as a part of beyond 5G standardization. Inversely, IoT also energizes MEC with mutual advantages. In particular, IoT expands MEC services to all types of smart objects ranging from sensors and actuators to smart vehicles. Integrating MEC capabilities to the IoT systems come with an assurance of better performance in terms of quality of service and ease of implementation.

Furthermore several use cases for MEC and 5G IoT collaboration and integration are summarized.

Use case A: Security, safety, data analytics

Security and safety has become one of the most important verticals for IoT. The developments in technology with ever increasing amount of data from sensors and high resolution video cameras create the need for scalable, flexible and cost-economic solution to analyze the content in real time. MEC can host the analytics applications close to the source and enable increased reliability, better performance and significant savings by processing huge amounts of data locally. Enhanced video analytics enables creating and using rules for different events to trigger alerts and forwarding actions. Real time video analysis can be used to identify and classify objects (person, specific object), create rules for observation areas of interest, define and use event based rules (entering/exiting area, leaving/removing object, loitering) and counting objects (number of people, objects). The solution is flexible to deploy by enabling the video processing and analytics application running at optimum location based on technical and business parameters [12].

Use case B: Vehicle-to-Infrastructure communication

Digitalization of the services is progressing with enormous speed and automotive sector is one area where the new technologies are shaping the whole industry. Self-driving cars have been already demonstrated by both traditional automotive and new internet players. The work on future 5G system is being currently conducted by various organizations globally and the digitalization in the automotive industry is clearly reflected in the use cases and the requirements. IoT is a key driver for the next generation technology and the most of the use cases appear to focus on connected cars. Connected cars is not only about self-driving capability, but many other use cases exist. In

general, all use cases related to smart transportation are of course in strict relation with the already mentioned Internet of Things paradigm. These use cases are not only considered from a theoretical point of view, but early experimental activities are already taking place in these years. MEC is the ideal solution and has been identified as a key component to support these ultra-low latency scenarios as it enables hosting applications close to the users at the edge cloud and therefore providing the shortest path between the applications.

Use case C: Computation offload into the edge cloud

Applications running on MTs may want to offload parts of the computations into the cloud for various reasons, such as availability of more computing power or of specific hardware capabilities, reliability, joint use of the resources in collaborative applications, or saving bitrate on the air interface. The computation offload is particularly suitable for IoT applications and scenarios where terminals have limited computing capabilities, i.e. in those cases where M2M devices have severe low power requirements, in order to guarantee high batteries lifetime. Such offload may happen statically (server components are deployed by the service provider proactively in advance) or dynamically (server components are deployed on demand by request from UE). Also in this use case, applications benefit from low delay provided by MEC.

Use case D: Smart home and smart city

One of the most important use cases of IoT is smart city and its important subset smart home/building. Recently, the MEC contexts and novel 5G technologies have been enabled to emerge the judicious edge big data analysis and wireless access for IoT systems to further improve the urban quality of life for citizen with many aspects including security, privacy, energy management, safety, convenient life, etc...By leveraging the fog-enabled cloud computing environments, the novel implemented smart home systems can reduce 12% utilized network bandwidth, 10% response time, 14% latency and 12.35% in energy consumption. For monitoring and controlling the smart home/buildings, innovative analytics on IoT captured data from smart homes was presented in [14] employing the fog computing nodes. This fog-based IoT system can address the challenges of complexities and resource demands for online and offline data processing, storage, and classification analysis in home/building environment.

For the smart city use cases, the security and privacy aspects were considered in [15] where a blockchain-based smart contract services for the sustainable IoT-enabled economy is proposed for smart cities by employing AI solutions in processing and extracting significant event information at the fog nodes, and then utilizing blockchain algorithms to save and deliver results.

Use case E: Wearable IoT, AR and VR

The newly emerging applications corresponding to

mobile AR, VR, and wearable devices, e.g., smart glasses and watches, are anticipated to be among the most demanding applications over wireless networks so far, but there is still lack of sufficient capacities to execute sophisticated data processing algorithms. To overcome such challenges, the emergence of MEC and 5G techniques would pose the longer battery lifetime, powerful set of computing and storage resources, and low end-to-end latency. Sharing this view, [16] presented Outlet system to explore the available computing resources from user's ambience, e.g., from nearby smart phones, tablets, computers, Wi-Fi APs, to form a MEC platform for executing the offloading tasks from wearable devices. Promising performance achieved by Outlet, e.g., mostly within 97.6% to 99.5% closeness of the optimal performance, has demonstrated the advantage of enabling edge computing technique into wearable IoT systems. Applying MEC on VR devices, [17] presented an effective solution to deliver VR videos over wireless networks minimizing the communication-resource consumption under the delay constraint. This work also demonstrated the interesting tradeoffs among communications, computing, and caching. In [18], a novel delivery framework enabling field of views caching and post-processing procedures at the mobile VR device was proposed to save communication bandwidth while meeting low latency requirement. Impressively, an implementation of MEC concepts over Android OS and Unity VR application engine in [19] enabled to reduce more than 90% computation burden, and more than 95% of the VR frame data being transmitted to MTs by letting MEC servers adaptively store the previous results of VR frame rendering of each user and considerably reuse them for others to reduce the computation load

Use case F: Tactile Internet

Tactile Internet is defined by the ITU as the next evolution of IoT that combines ultra-low latency with extremely high availability, reliability and security. Encompassing human-to-machine and machine-to-machine interaction, Tactile Internet will combine multiple technologies including 5G and MEC, i.e., 5G may be employed for the data transmission with low delay and high reliability while MEC efficiently exploit computing resources close to the end users for better QoE. The applications related to Tactile Internet can be automation, robotics, tele-presence, tele-operation, AR, VR. The following summarizes the recent works focusing on the technical aspects involving to the MEC implementation in Tactile Internet. An energy-efficient design of fog computing networks will support low service response time of end-users in Tactile Internet applications and efficiently utilize the power of fog nodes. The trade-off between the latency and required power was presented and then extended to fog computing networks leveraging cooperation between fog nodes. We can exploit the MEC

systems including cloud, decentralized cloudlets, and neighboring robots equipped with computing resource collaborative nodes for computation offloading in support of a host robot's task execution. MEC based collaborative task execution scheme outperforms the non-collaborative scheme in terms of task response time and energy consumption efficiency. Recently, in [20] designed a hybrid edge caching scheme for Tactile Internet which can reduce latency and achieve better performance in overall energy efficiency than existing ones.

IV. SYSTEM ARCHITECTURE AND MODEL FOR BEYOND 5G MOBILE EDGE COMPUTING SERVICES

The Fig. 2 depicts the system architecture and usage scenario for our proposed intelligent multi-access QoS mobile edge computing framework for beyond 5G services, using heterogeneous environment orchestrated services. First, the main characteristics of our proposed edge MT with incorporated advanced QoS user-centric ML module (AQA) with vertical multi-homing and multi-streaming features are illustrated in [6], and with ML being an essential tool for data intelligence it guarantees improvement of services [21]. The Cloud server placed in the core part of the network is in constant communication with the MEC Radio Access Network (MECRAN) servers in which are placed the multimedia broadband orchestrators which orchestrates the MECs. Moreover, each MT used in the above scenario is multi-RAT node, with several (n) RAT interfaces. The advanced QoS routing algorithm is set within the AQA module on IP layer in both MT as edge device in one side, and the MECRAN server in the another side. Also in the edge devices (MTs) there is a orchestrator agent which collect the QoS parameters of interest and sends to the orchestrator manager in the MECRAN. Here, the QoS parameters of interests are: service price per RAT, MT velocity, MT battery level, MT latency (from MT to MECRAN), detected signals strength, response time, availability, maintainability and etc. If the MECRAN orchestrated-service manager is overloaded, he can send part of his work for processing to the local edge agents in the heterogeneous environment or to the Cloud server in the core or to global Cloud Server Farm. However, part of the optimizations in selecting the most appropriate RAT for a given services/service are done in the service orchestrator agent, but mostly all those optimizations are done in the service orchestrator manager in MECRAN, by starting the AQA module with Machine Learning (ML) algorithm. On a transport layer, the most suitable protocols which are used here are: Stream Control Transmission Protocol (SCTP) [22], Datagram Congestion Control Protocol (DCCP) [23]. Also, on the other end of the connection must be installed SCTP/DCCP on its transport layer in order to have successfully established SCTP/DCCP association.

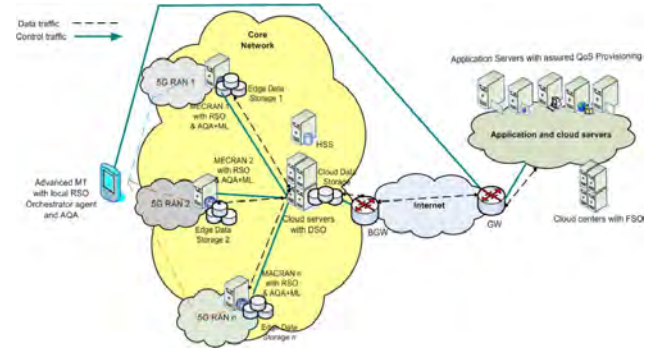


Fig. 2. Sistem architecture and possible beyond 5G scenario.

Furthermore, one of the advantages of our framework is the following: it is defined independently from different RATs, implemented on IP Network Layer. This advanced MT is using Multi-RAT interfaces and is able to provide intelligent QoS management and routing over variety of heterogeneous RATs at the same time. Moreover, the services orchestrator with help of AQA module is able to combine simultaneously several different traffic flows from different multimedia services transmitted over the same or different RAT channels, ML to understand traffic activity, and therefore, improve upon existing services, as well as optimize the chosen traffic flows accordingly to resources. The final aim is in achieving higher throughput, higher access probability ratio and optimally (regarding the resources) using the heterogeneous RAT resources. Our proposed architecture incorporates the above model, which consists of several levels: Regional Service Orchestrator (RSO), Domain Service Orchestrator (DSO) and Global Service Orchestrator (GSO). The RSOs are located at the edges of the MECRANs and at the advanced MTs, enabling semi-autonomous operation of the different Regions. Due to the advanced MT proximity, this provides quicker distribution of the load, lower latency and higher scalability. The DSOs are located in the cloud computing data centers, in the core networks. Each DSO is responsible for their domain/s and supervises the RSOs below. Like that global mechanisms are provided in order to enable intra-domain cooperation between different regions.

At the top of the architecture are located the GSOs, which allow a fruitful interaction between different cloud and fog domains. The GSO enables the management functionality between different cloud and fog domains and, similarly to the DSOs, it should be properly adapted to operate in a global Cloud environment. GSO communicate with other GSOs and like that global mechanisms are provided that enable cooperation among different cloud computing Domains (e.g. under the administration of different authorities). These global mechanisms also enable the creation of a Multi-Domain Mobile Cloud Environment able to support service ubiquity.

The process of establishing a tunnel to the Cloud Server in the core, for routing based on the QoS policies and QoS requirements per service; are carried out immediately after

the establishment of peer-to-peer connection between the MT services orchestrator agent with MEC features and MEC-RAN server on the other side. The MT and MEC-RAN/Cloud server with vertical multi-homing and multi-streaming features and service orchestrator, with ML within, are able to handle simultaneously multiple radio network connections and speed up the transfer of the multimedia services. Moreover, by transmitting each object of each service in a separate stream, the highest level of satisfied end-users is achieved. In that way, by using our proposed MT with AQA with ML algorithm within, instead of creating a separate connection for each object as in TCP, makes use of network capacity aggregation, multi-streaming and multi-homing feature to speed up the transfer of the target multimedia service over separate streams over different RATs. So, all mobile broadband services are going over MEC-RAN and MEC agent in the user's MT (in the downstream direction) and vice versa (in the upstream direction). Also, in comparison with all related works, we must to emphasize that our advanced QoS framework for mobile broadband with MEC and ML is implemented on IP level in the Cloud-servers, MEC-RAN servers and in the edge (MT) sides.

V. CONCLUSION

In this paper overviewed MEC essence, provides existing use cases of MEC with 5G, and proposes a novel beyond 5G framework for MEC for mobile broadband Orchestrated-services in heterogeneous RATs. According to the analysis, our proposed framework with MEC orchestrated-services is expected to perform fairly well under a variety of network conditions and optimally utilized the resources due to the used ML algorithm and MEC processing. In that manner, efficient and QoS-based usage of available mobile resources, plus efficient MEC orchestrated-services performances are most essential for provision of seamless mobile broadband Internet services. The proposed model combines the most powerful features of both Cloud and Edge computing, independent from any existing and future Radio Access Technology, leading to high performance utility networks with high QoS provisioning. Undoubtedly, MEC is an innovative network paradigm to cater for the unprecedented growth of computation demands and the ever-increasing computation quality of user experience requirements. It aims at enabling Cloud Computing capabilities and telecommunication services in close proximity to end users, by pushing abundant computation and storage resources towards the network edges. The direct interaction between MTs and edge servers through wireless and mobile communications brings the possibility of supporting applications with ultra-low latency requirement, prolonging device battery lives and facilitating highly-efficient network operations.

REFERENCES

- [1] Y. Yu (2016), "Mobile Edge Computing Towards 5G: Vision, Recent Progress, and Open Challenges," *China Communications*, vol. *Supplement No. 2*, pp. 89-99.
- [2] Recommendation ITU-T Y.2052 (02/2008): Framework of multi-homing in IPv6-based NGN
- [3] Recommendation ITU-T Y.2056 (08/2011): Framework of vertical multihoming in IPv6-based Next Generation Networks
- [4] Toni Janevski (2009). 5G Mobile Phone Concept. IEEE Consumer Communications and Networking Conference (CCNC) 2009, USA
- [5] Tomislav Shuminoski, Toni Janevski, "Radio Network Aggregation for 5G Mobile Terminals in Heterogeneous Wireless and Mobile Networks", *Wireless Personal Communications*, Volume 78, Issue 2, 2014, Page 1211-1229.
- [6] T. Shuminoski, T. Janevski (2016). "5G mobile terminals with advanced QoS-based user-centric aggregation (AQUA) for heterogeneous wireless and mobile networks", *Wireless Networks*, 22(5), pp. 1553-1570.
- [7] Federico Boccardi et al. (2014). Five Disruptive Technology Directions for 5G. *IEEE Communications Magazine*, Vol. 52, No. 2, pp.: 74-80.
- [8] Boyd Bangerter et al. (2014). Networks and Devices for the 5G Era. *IEEE Communications Magazine*, Vol. 52, No. 2, pp.: 90-96.
- [9] X. Kang et al. (2010), Improving Performance for Decentralized Execution of Composite Web Services, *Proceedings of the 6th World Congress on Services*.
- [10] S. Kekki et al., "MEC in 5G networks," ETSI white paper, no. 28, pp. 1-28, Jun. 2018.
- [11] T. Taleb, S. Dutta, A. Ksentini, M. Iqbal, and H. Flinck, "Mobile edge computing potential in making cities smarter," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 38-43, Mar. 2017.
- [12] Q. Pham et al., "A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art," *IEEE Access*, vol. 8, pp. 116974-117017, 2020.
- [13] "3GPP technical specification group services and system aspects; system architecture for the 5G system," Jun. 2019, 3GPP TS 23.501 v16.1.0.
- [14] A. Yassine, S. Singh, M. S. Hossain, and G. Muhammad, "IoT big data analytics for smart homes with fog and cloud computing," *Future Generation Computer Systems*, vol. 91, pp. 563 - 573, Feb. 2019.
- [15] M. A. Rahman, M. M. Rashid, M. S. Hossain, E. Hassanain, M. F. Alhamid, and M. Guizani, "Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city," *IEEE Access*, 7, pp. 18 611-18 621, Jan. 2019.
- [16] L. Tao, Z. Li, and L. Wu, "Outlet: Outsourcing wearable computing to the ambient mobile computing edge," *IEEE Access*, vol. 6, pp. 18 408-18 419, Mar. 2018.
- [17] X. Yang, Z. Chen, K. Li, Y. Sun, N. Liu, W. Xie, and Y. Zhao, "Communication-constrained mobile edge computing systems for wireless virtual reality: Scheduling and tradeoff," *IEEE Access*, vol. 6, pp. 16 665-16 677, Mar. 2018
- [18] Y. Sun, et al. "Communications, caching, and computing for mobile virtual reality: Modeling and tradeoff," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7573-7586, Nov. 2019.
- [19] Y. Li and W. Gao, "MUVIR: Supporting multi-user mobile virtual reality with resource constrained edge cloud," in *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, Seattle, WA, USA, Oct. 2018.
- [20] J. Xu, K. Ota, and M. Dong, "Energy efficient hybrid edge caching scheme for tactile internet in 5G," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 2, pp. 483-493, Jun. 2019.
- [21] Bin Qian et al. "Orchestrating the Development Lifecycle of Machine Learning-Based IoT Applications: A Taxonomy and Survey", available: <https://arxiv.org/pdf/1910.05433.pdf>
- [22] Shaojian Fu and Mohammed Atiquzzaman (2004). SCTP: State of the Art in Research, Products, and Technical Challenges. *IEEE Communications Magazine*, Vol. 42, pp.: 64-76.
- [23] E. Kohler et al. (2006). RFC: 4340. Datagram Congestion Control Protocol (DCCP). <http://tools.ietf.org/html/rfc4340>. Accessed 22.4.2015.



ETAI 6: INSTRUMENTATION AND MEASUREMENTS

Positional Value Measurement for a Rook and King vs Rook Chess Endgame Algorithm

Adrijan Božinovski
University American College Skopje
Skopje, Macedonia
bozinovski@uacs.edu.mk

Filemon Jankuloski
University American College Skopje
Skopje, Macedonia
filjankuloski@gmail.com

Abstract—In this paper, we will take a look at an algorithm which plays as White in the King and Rook vs King chess endgame. First, we will explore the history of Artificial Intelligence and the evolution of chess machinery over a 70 year time span. Next, we will go over the foundation with which our program was made, and then take a deep look at its structure to better understand how it functions. Moreover, we will discuss the ways in which the program can be improved in the future. Finally, we will conclude the paper by going over the results of the program and Artificial Intelligence's impact in the present.

Keywords—Chess, Algorithm, Artificial Intelligence

I. INTRODUCTION

A. Goal of the Project

Artificial Intelligence (AI) has become increasingly prevalent in our current society. We can find AI being used in many aspects of our daily lives, such as through the use of smartphones and autonomous driving vehicles such as the Tesla [1]. Leading textbooks on AI define it as the study of “intelligent agents”, which can be represented by any device that perceives its environment and takes actions that maximize its chances of achieving its goals [2]. In the context of this paper, the intelligent agent is a chess-playing AI, which implements different kinds of methods and strategies to defeat any opponent in a particular chess endgame scenario. The endgame scenario that we will be exploring is a White Rook and White King, which are controlled by the AI, and a Black King, which is controlled by a human player. Such a program has already been created, and it was the first Macedonian chess program, which was created by Stevo Božinovski in 1969 and it was written in Fortran for the IBM 1130 computer [3]. However, in the case of this paper, the AI is written in Java and uses Netbeans as its IDE. This program and its methods will be showcased in a subsequent section.

B. Evolution of Chess Engines

The subject of AI has been in existence for roughly 70 years, since 1950. Although he did not coin the term “Artificial Intelligence” as such, Alan Turing was the first individual to suggest that human intelligence and machine intelligence are comparable, in his famous 1950 article called “Computing Machinery and Intelligence” [4]. In this article, he explained that if individuals were incapable of making the discernment between a machine and a human being in a teletype dialogue, then it would not be far-fetched to say that a machine is capable of intelligence. The true birth of AI occurred at the Dartmouth College workshop, since it is where the term “Artificial Intelligence” was

coined by John McCarthy [4]. Originally, Dartmouth College was meant to hold a workshop on AI, but due to skepticism and a lack of interest, no more than five people consistently sat through the conference, including McCarthy himself [5]. Despite the initial lack of interest, John McCarthy, Allen Newel, Marvin Minsky, Herbert Simon, and Arthur Samuel were the sole five people who built the foundation for AI to thrive. Fast forward several decades and there is the chess machine history's latest and most revolutionary creation, named AlphaZero, which was released in 2017 by Google. AlphaZero is revolutionary in that it was the first chess machine to utilize a reinforcement learning algorithm by combining the concept of deep learning with the Monte Carlo Tree Search [6]. In fact, by generating and playing through thousands of self-play games, AlphaZero became so adept at chess-playing that it discovered openings which were never conceived of by human players. A full explanation of all aforementioned chess machines and machines listed in Table I is available in [7].

II. THE PROGRAM

A. Foundation of the Program

The algorithm discussed in this project is an improvement on the first Macedonian chess program, made by Stevo Božinovski in 1969. The first program was written in Fortran and was not meant to complete a full game, but rather simulate a scenario where the human player has a Black King, and the computer has a White King and a White Rook. This program used several different methods for specific purposes. “DATSW” was used to plot the chess board, “POTEZ” was used to determine the legality of the human move, “POZIC” was the algorithm which served as both a positive analyzer and a move generator, and “MATIR” which determined if the state of the chessboard was in checkmate [7]. However, in the program shown in this project, the methods are more specialized, the algorithm is arguably more efficient, and the program is incredibly fast. In fact, the program calculates positions so quickly that a one second delay was deliberately put before White would be given a chance to move, so that the program would be more suitable for human players. The program referred to in this paper is also capable of instantaneously resetting the chessboard to the initial position before starting a game, generating new and valid random starting positions, creating an enumeration for the state of the chessboard and its pieces, and calculating the value of the state of the board based on the positions of all pieces on the board and inevitability of a checkmate.

TABLE I. DEVELOPMENT OF CHESS PROGRAMS AND MACHINERY OVER TIME [7]

Name of Program or Machine	Name of Creators	Year of Creation	Algorithms, Methods, or Technology Incorporated
Turochamp	Alan Turing, David Champernowne	1948	Variable lookahead, Two move heuristic, Evaluation of positions based on mobility, piece safety, king mobility, king safety, castling, and more.
NSS	Allen Newell, Herbert Simon, Cliff Shaw	1958	Move generator, Position evaluator, Alpha Beta searching
Mac Hack VI	Richard Greenblatt	1966	Move generator, Position evaluator, Alpha Beta searching, Transposition table
First Macedonian Chess Program	Stevó Božinovski	1969	Move generator, Position evaluator
Chess 4.5	Larry Atkin, David Slate	1975	Move generator, Position evaluator, Alpha Beta searching, Transposition table, Bitboard, Full width search, Iterative deepening
Belle	Ken Thompson	1976	Move generator, Position evaluator, Alpha Beta searching, Transposition table, Lazy and full evaluation, Principal variation splitting
HiTech	Joe Condon	1985	Move generator, Position evaluator, Pattern recognition, Transposition table, Parallel searching, Alpha Beta searching
Fritz	Hans Berliner, Carl Ebeling	1991	Move generator, Position evaluator, Parallel searching, Null move search
Deep Blue	Frans Marsch, Mathias Feist	1996	Move generator, Position evaluator, Alpha Beta searching, Transposition table, Parallel searching, Singular extension
StockFish	IBM Development Team	2016	Move generator, Position evaluator, Iterative deepening, Parallel Search, Transposition Table, Move valuable victim/least valuable aggressor, Null move search, Singular extensions, Futility pruning, Static exchange evaluation
AlphaZero	Tord Romstad, Marco Costalba, Joona Kiiski	2017	Neural networks, Deep learning, Reinforcement learning, Monte Carlo Tree Search, Transposition table, Tensor processing unit

B. Structure of the Program

In Figure 2, a Unified Modeling Language (UML) diagram is shown which showcases the entirety of the program's methods, attributes, and how they relate to each

other. The Board class plays the most important role in the entirety of the program, since it contains a majority of the methods responsible for crucial functionalities.

The method which is responsible for generating White's moves is the "whiteMove()" method. One of the first goals of this method is to use the White Rook to restrict the Black King to the least amount of spaces. In order to accomplish this, the White Rook must be one row above or beneath, or one column to the left or right (i.e., one square diagonally) of the Black King. This row or column which best restricts the Black King is known as the ideal position. In order to understand how to calculate the ideal position, it is crucial to understand how quadrants work.

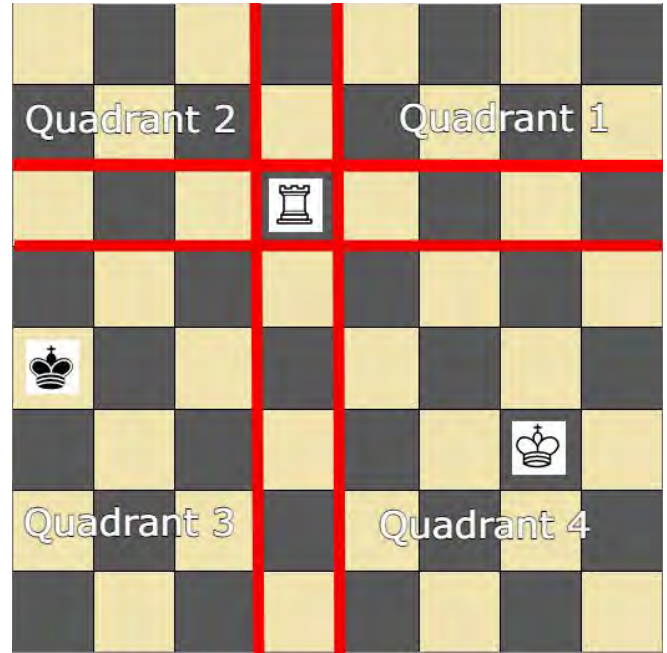


Fig. 1 Quadrant Positioning Inside the Chess Board

Similarly to Figure 1, the chess board is divided into 4 separate quadrants, and each quadrant is relative to the White Rook's position on the board, which is represented by the intersection of the x and y axis. Quadrant 1 is to the upper right of the rook, quadrant 2 is to the upper left, quadrant 3 is to the lower left, and quadrant 4 is to the lower right. Depending on which quadrant the Black King resides in, there are corresponding formulas which the program uses to find how many squares the Black King is restricted to, depending on whether the Rook restricts it from the top, bottom, left, or right. Once the White Rook is in an ideal position, the White King will attempt to come within Knight's distance from the Black King. However, the program does not simply move to any position which is a Knight's distance from the Black King. First, the program retrieves the coordinates of every possible Knight's distance position from the Black King, and calculates the difference between these positions and the White King's coordinates. These differences are then stored in an array and sorted. However, there is more to the program's decision making. The program then traverses this array, and uses the "isFeasible()" method to see if a move is feasible or not. The "isFeasible()" method ensures that moves made by the White Rook or White King are neither out of bounds or in any other way illegal (e.g., if the White King moves into the square of the White Rook), nor within one of the squares protected by the Black King.

The “isProper” method is used to check if a particular Knight’s distance position will result in the “proper sequence”. To explain, a proper sequence is a sequence where the White Rook is in front of the White King and the Black King is in front of the White Rook. This allows for the White King to cut off the Black King, resulting in a check or checkmate. If the White Rook is in

the ideal position and the White King is a Knight’s distance away from the Black King, the program will perform a dead move. A dead move, otherwise known as a waiting move, is a move where the White Rook moves a single square towards or away from the Black King, effectively retaining White’s advantage and eventually forcing Black King into a check or checkmate situation.

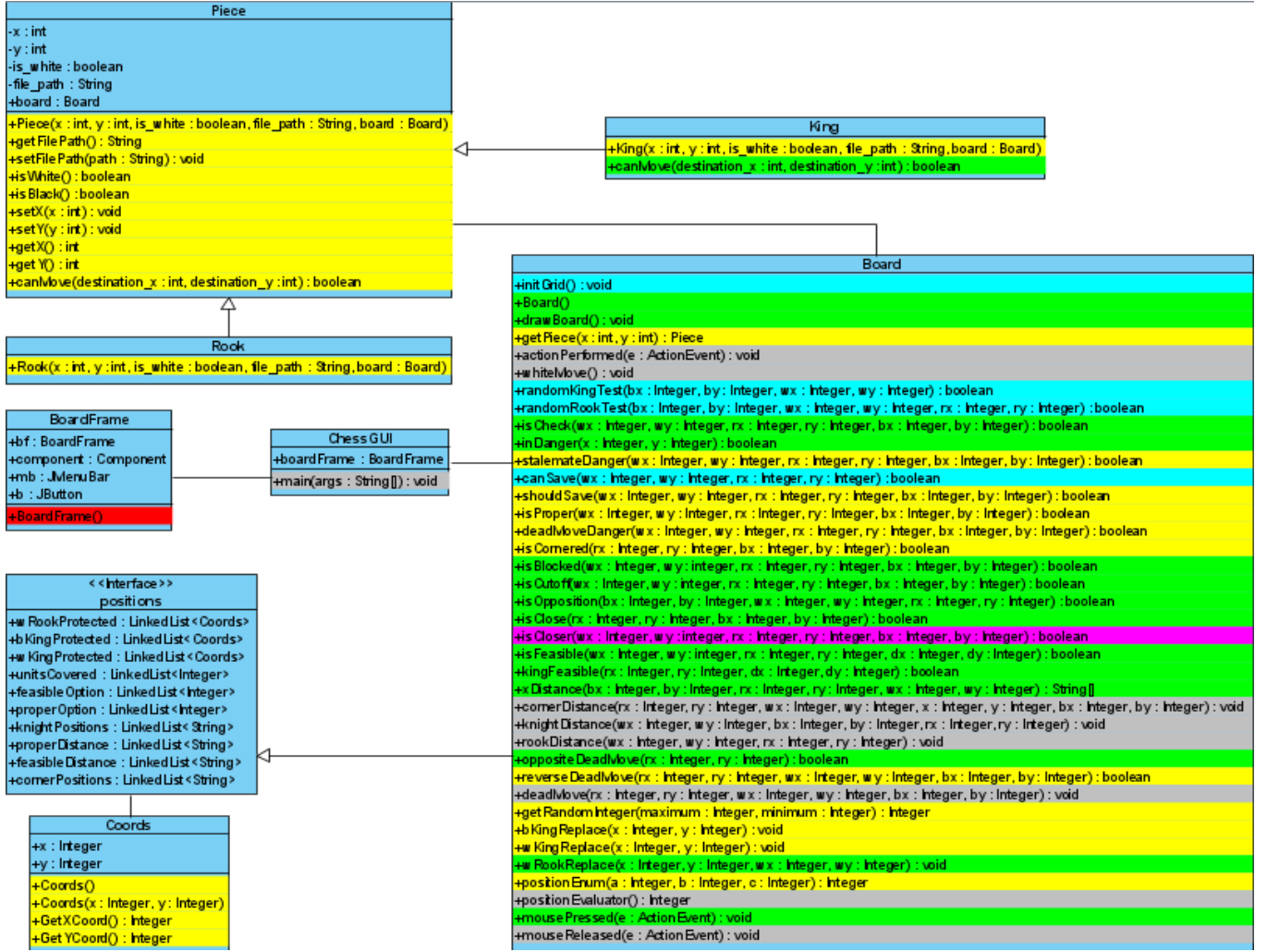


Fig. 2 The UML Diagram of the Program

C. The Position Enumeration and Value Formula

In addition to containing the mechanism for movement and display of the pieces on the board, the program also shows an enumeration for each position and its positional value. The enumeration is calculated such that each piece is given a numerical value based on which field on the chessboard it is placed (from 0 to 63, inclusive). Then, each piece’s value is multiplied either by 1, 64, 64² (in this paper, the pieces which get those factors are the Black King, the White Rook and the White King respectively) and the sum of all those values gives the enumeration of the position shown.

The positional value formula acts as a position evaluator which has 5 elements: “distToEdges” is the added length and height of the squares from the Black King to the edge of the board, “KDistance” is how many squares away the White King is from the optimal Knight’s distance position, “isRookAttacked” represents whether or not the Rook is

attacked, “isBlackToMove” signifies whether it is Black’s turn or White’s turn, and “isCheck” signifies whether or not there is a check. Adding their respective values together gives the positional value, which can be expressed as equation (1). The positional value, or posValue in Equation (1), is meant to show the amount of free squares that the Black King has available to move onto

$$\text{posValue} = \text{distToEdges} + \text{KDistance} + \text{isRookAttacked} + \text{isBlackToMove} + \text{isCheck} \quad (1)$$

For the first criterion, the amount of squares differs greatly depending on whether or not the Black King is in a quadrant or a half. A quadrant typically means a smaller value. However, there are cases where the White King blocks off the White Rook, which enables the Black King to move across 2 quadrants (i.e., a half). When all pieces are on the same row or column, the White King moves, thus leading into a discovered check. A discovered check is when the White King moves out of the way of the White Rook, enabling it to attack the Black King. At this point, it is

Black's move and the human player can choose to go left or right, or up or down depending on whether the discovered check was along the column or the row. This essentially gives the Black King access to two different quadrants, which also increases the number of free squares available to it.

The second criterion is attained by calculating the difference between the "x" and "y" coordinates of the White King and the Knight's distance position, and adding them together. For the third criterion, if the Rook is not attacked, this equates to a value of 0, and if the Rook is attacked, then this equates to a value of 1 (i.e., it gives the Black King one more square to go onto, namely the square that the White Rook is on). For the fourth criterion, if it is Black's turn, the value is equal to 1, otherwise, it is equal to 0 (i.e., the checkmate can only occur when it is White to move - otherwise the Black King still has options to move). Finally, the fifth criterion is obtained by determining whether the state of the check board is in check or not. If it is in check, then the value is -1, otherwise, it is 0 (the amount of squares that the Black King can move onto is decreased by 1 when it is in check, because the square that the Black King is on already is also under attack). Such a selection of values ensures that a position value of -1 shows that a checkmate has been reached.

D. Examples of Positions

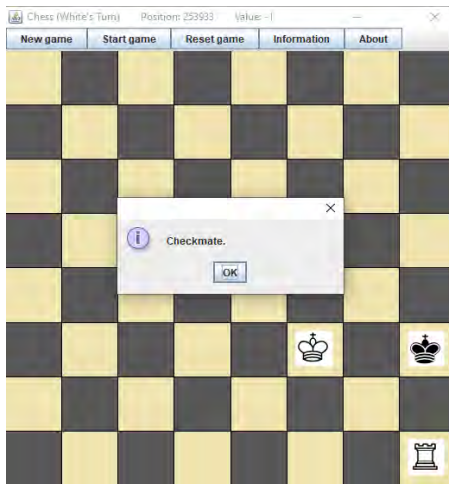


Fig. 3 Board in Checkmate State

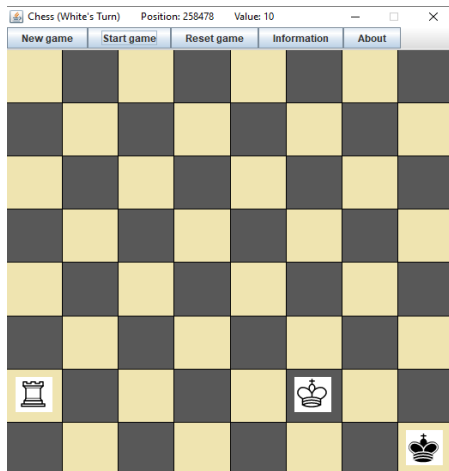


Fig. 4 The White King blocking the White Rook - a necessity to avoid a stalemate

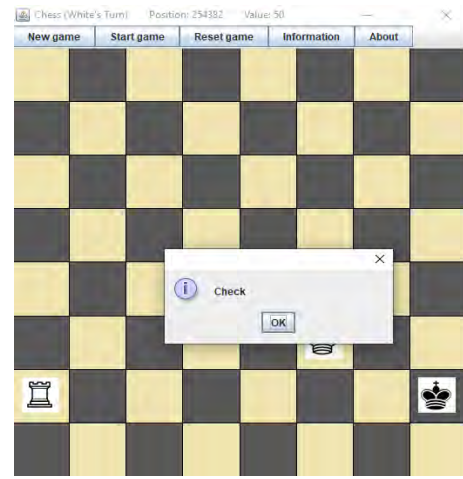


Fig. 5 Discovered Check after having the Black King cross the White Rook, enabling checkmate in subsequent moves

III. DISCUSSION, FUTURE WORK AND PERSPECTIVES

Generally speaking, the positional value formula is meant to represent how close the White pieces are to achieving checkmate, i.e., how many squares are available for the Black King to move onto. As the game progresses, White should tend to decrease the positional value (and thus bring the game closer to checkmate), whereas Black should tend to increase it (and thus bring the game farther away from checkmate). However, there are situations in which this is not the case. One such case is when the White King blocks off the White Rook. In Figure 4, the position value is equal to 10 before the block, but in Figure 5, the value becomes 50. Even though this may be seen as an inconsistency with the formula, this type of play is necessary by White in order to avoid a stalemate. Moreover, the program is currently algorithm-based rather than heuristic-based, and this is something that will be addressed in the future.

The way that chess is played has changed significantly, since the development of chess playing machines and games. Now that these chess engines are available on phones and laptops, there are many children who will be able to learn the game more easily due to easy accessibility of both technology and knowledge. There are also chess machines like AlphaZero that are teaching us new openings in chess theory which have never been seen before. It is fascinating that, originally, it was human beings that were supposed to teach machines how to perform certain tasks. Now, we have machines which are teaching us new ways of playing chess unlike anything that has been seen before. As machines continue to evolve, not just in chess, but in other areas as well, we should stay open minded and use these paragons as examples so that we can improve ourselves too.

IV. CONCLUSION

This paper presents a program that depicts a certain chess endgame scenario, namely White King and White Rook, as played by the computer, versus a Black King, as played by a human player. It also proposes a way to measure the value of a position, indicating how close or how far a position is from reaching a checkmate. Certain positions have been identified which indicate that further work will need to be done in order to convert the program from algorithm-based to AI-based.

ACKNOWLEDGMENT

We would like to give our gratitude to Toni Jankuloski for spending countless hours assisting in debugging the program and thereby improving its performance.

REFERENCES

- [1] Tesla.com. 2021. *Autopilot*. [online] Available at: <<https://www.tesla.com/autopilot>> [Accessed 17 August 2021].
- [2] Russell, S. and Norvig, P., 2003. Artificial Intelligence: A Modern Approach. 2nd ed. Upper Saddle River, New Jersey: Prentice Hall, p.55.
- [3] Božinovski, S., 2016. Cognitive and Emotive Robotics: Artificial Brain Computing Cognitive Actions and Emotive Evaluations, Since 1981. In: ICT Innovations Conference 2016. Skopje, p.11.
- [4] McCarthy, J., 1996. Defending AI Research. New York: Cambridge University Press, p.73.
- [5] Intelligence, A., AI, W. and Europe, C., 2020. History Of Artificial Intelligence. [online] Artificial Intelligence. Available at: <<https://www.coe.int/en/web/artificial-intelligence/history-of-ai>> [Accessed 3 March 2020].
- [6] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T. P., Simonyan, K., and Hassabis, D., 2017. Mastering Chess And Shogi By Self-Play With A General Reinforcement Learning Algorithm. [ebook] Available at: <https://www.researchgate.net/publication/321571298_Mastering_Chess_and_Shogi_by_Self-Play_with_a_General_Reinforcement_Learning_Algorithm>.
- [7] Jankuloski, F. and Božinovski, A., 2020. Chess as Played by Artificial Intelligence. In: 12th ICT Innovations Conference 2020. Skopje: CCIS.

Преглед на безбедносни и сигурносни системи во автомобилската индустрија

Александра Ѓорѓиевска, Маре Србиновска, Мартин Ѓорѓиевски

Факултет за електротехника и информациски технологии

Универзитет „Св. Кирил и Методиј“

Скопје, Р.С.Македонија

gj.aleksandra@yahoo.com, mares@feit.ukim.edu.mk, gj.martin91@gmail.com

Анстракт— Овој труд е насочен кон преглед на современите безбедносни и сигурносни системи во автомобилската индустрија, како и интелигентната платформа од интегрирани сензори на која тие се базираат. Целта на трудот е да се даде кратко објаснување на актуелните безбедносни и сигурносни системи, како што се антиблокирачки систем за заочување, систем на воздушни перничии и сигурносни појаси, систем за итни повици (E-Call систем), како и системите за кражби и сајбер безбедност. Технолошкиот напредок на овие системи претставува основа за постоењето на делумно и целосно автономните возила, а нивната функционалност во целост и директно влијае кон сигурноста и безбедноста на возачите, патниците и сите останати учесници во сообраќајот.

Клучни зборови— безбедносни системи; сигурносни системи; автомобилска индустрија; сензори; автономни возила

I. ВОВЕД

Автомобилската индустрија е во постојан подем и развој, засенувајќи ги постарите технологии со новите, кои доаѓаат на сцена со стремеж за создавање на побезбеден, поефикасен стандард. Со ваквиот брз развој, најголем дел од претходно-механичките системи се заменуваат со електрични системи, што доведува до високо компјутеризирани, полуавтономни или целосно автономни возила. Со цел ваквата платформа од електрични системи да функционира беспрекорно, како и за да се воведат поврзаност на возилата со светот околу нив, сензорите и системите кои тие ги формираат го заземаат централното место во современите возила. Така, во изминатава деценија, сензорите и сензорските системи добиваат сè поголемо значење за автомобилската електроника.

Големиот дел од системите во автомобилите бараат точни, сигурни, безбедни, разновидни и економични сензори за позиција, брзина, визуелна детекција. Тековниот развој во системите за контрола на динамиката на возилата и идните напредни системи за стабилизирање на високо-перформансните возила, бараат и подобри перформанси на инерцијалните

сигнали на динамиката на возилото. Тоа е и една од причините поради кои сензорите се основни компоненти на автомобилските електронски системи за контрола. Притоа, сензорите наменети за автомобилската индустрија мора да потврдат дека ја задоволуваат рамнотежата помеѓу точноста, робусноста, производноста, заменливоста и ниската цена на нивната напредна технологијата. Со тоа, безбедноста и сигурноста на автомобилите и патниците е, во најголем дел, одговорност на сензорите и системите кои тие ги формираат, како во самите автомобили, така и во нивната непосредна околина и мрежна поврзаност, [1].

Во овој овој труд, најнапред е презентирана клучната важност на сензорите како составен дел од безбедноските и сигурноските системи во возилата. Потоа е даден преглед и поделба по категории на најголемиот број сензори кои го овозможуваат функционирањето на автомобилите. Понатаму се разгледуваат и најактуелните сигурносни и безбедносни системи во случај на опасност, како што се: eCall системи (Emergency Call системи, системи за итни повици), системот на воздушните перничии, системите за алармирање и спречување на кражби, како и системи за сајбер безбедност кај полуавтономни и целосно автономни возила и големиот број други системи без кои возилата не би ги исполнувале стандардите на напредната автомобилска индустрија.

II. СЕНЗОРИ ЗА БЕЗБЕДНОСНИ И СИГУРНОСНИ СИСТЕМИ ВО АВТОМОБИЛИТЕ

Брзиот развој во автомобилската електроника е придружен со паралелен развој на сензорите и сензорските системи. Имено, исполнувањето на еколошките барања, системските решенија што обезбедуваат поголема удобност, сигурноските карактеристики како воздушни перничии и програмата за електронска стабилност (ESP, electronic stability program), како и безбедноските системи во кои се вбројуваат системите за алармирање на кражби, доведуваат до развој на интегрирани сензори кои можат да се изработуваат во големи количини. Микроелектромеханичките сензори се пример за

модерна технологија, чиј брз развој е поттикнат од автомобилската индустрија. Нивната употреба, во голема мера влијае и на подобрувањето и оптимирањето на системот за управување со моторниот погон, со што директно учествуваат во намалување на штетните емисии и загадувањето на животната средина. На чекор понапред во размислувањето е и употребата на интелигентните сензорски системи со цел целосно автоматизирање на возилата и патиштата во текот на наредната деценија.

A. Поделба и типови на сензори во автомобилската индустрија

Во моторните возила постојат голем број на сензори и контролни единици, кои може да се разгледуваат како еден вид на нервен систем на возилото. Тие ја вршат функцијата на сетилни органи на возилото и ги претвораат влезните променливи во електрични сигнали, притоа користејќи различни концепти за мерење, во зависност од задачата. Овие сигнали, понатаму се користат во контролните и регулаторните функции од страна на контролните единици во системите за управување на моторот, системите за безбедност, сигурност и удобност.

Автомобилската индустрија бележи голем успех во создавањето на сè поинтелигентни возила, а со тоа автономните возила, кои не многу одамна беа само замисла, сега забрзано го пробиваат својот пат кон нашето секојдневие. Безбедноста кои тие ја нудат, речиси целосно зависи од сензорните системи кои се употребуваат, со цел да се обезбедат информации со екстремно високо ниво на точност, кои потоа се искористуваат за донесување на сложени одлуки во реално време.

Според нивната функција во возилото, сензорите може да се поделат на:

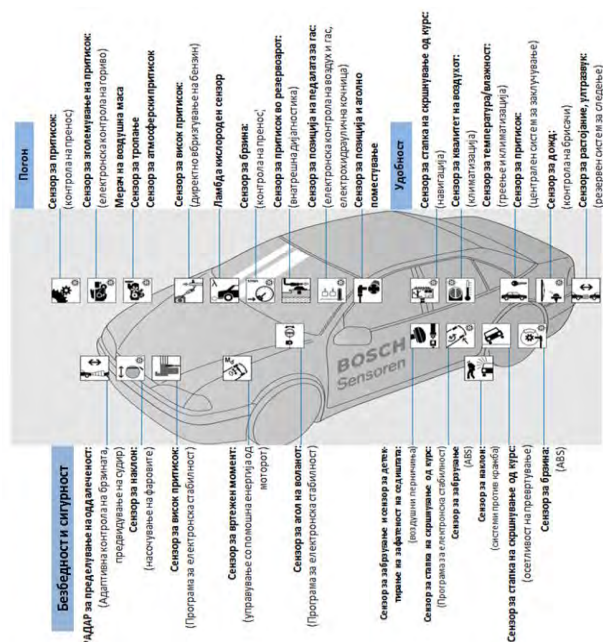
- Погоонски сензори
- Сензори за удобност
- Сензори за безбедност и сигурност

На Слика 1 се прикажани автомобилските системи и сензорите кои припаѓаат на трите горенаведени групи, како и нивната примена [2].

Според нивната местоположба во возилото, сензорните системи може да се поделат во четири групи и тоа:

- Сензори во внатрешноста на возилото (внатрешни сензори):
 - Детектирање на зафатеност (на пример: патник на седиштето)
 - Внатрешни камери
 - Детектирање на загаденост на стаклата
 - Контрола на температурата во возилото

- Безбедност на патниците (на пример: предупредување за безбедносен појас)
- Телематика (Овозможува: поврзување на возилото со Автомобилскиот облак (Vehicular Cloud), со возила од околината и со светот; подобрување на безбедноста на патниците и функционалноста на возилото преку овозможување на специјални услуги како што се: итни повици во случај на опасност, следење на возила, автономно возење) и други.
- Сензори во предниот (надворешен) дел на возилото:
 - Ноќна визуелизација („Night Vision“)
 - Адаптивна контрола на брзината („Adaptive cruise control“)
 - Предупредување и избегнување на судир
 - Препознавање на сообраќајни знаци
 - Препознавање на пешаци
 - Интелигентна контрола на светлата
 - Детекција на временски услови (пример: врнежи од дожд)
 - Превентивно кочење и други.
 - Сензори на страничните делови од возилотот:
 - Детектирање на слепи точки
 - Помош при паркирање
 - Предупредување за преминување на лента
 - Дигитални странични ретровизори



Сл. 1 Автомобилски системи и сензори

- Сензори во задниот (надворешен) дел на возилото
- Камера (ретровизор)
- Помош при паркирање
- Предупредување за судир и други.

На Слика 2 е даден приказ на внатрешната и надворешната околина на дејствување на сензорските системи.



Сл. 2 Околина на дејствување на сензорските системи

Б. Поделба на сигурносните системи во автомобилската индустрија

Сигурносните системи играат клучна улога, не само за возачите и патниците, туку и за пешаците и целокупната околина на патиштата. Во согласност со стандардите за автомобилската индустрија (Automotive Industry Standards, AIS), а врз база на нивната функционалност и начинот на нивна употреба, сигурносните системи се делат на:

- Активни, базирани на технологија која помага и учествува во спречување на судир или други сообраќајни несреќи, како што се Напредните системи за помош/асистенција на возачите (Advanced Driver Assistance Systems, ADAS)
- Пасивни, базирани на технологија која се состои од компоненти на возилото, како што се: воздушни перничии, сигурносни појаси, физичка структура на возилото

1) Активни сигурносни системи

Активните сигурносни системи се одликуваат со неколку клучни карактеристики:

- константно го набљудуваат работењето, функциите и опкружувањето на возилото
- се одликуваат со способност за спречување сообраќајни несреќи и активна помош или асистенција при возењето

- се одликуваат со способност за предвидување на опасни ситуации во сообраќајот и придонесуваат кон избегнување на пречки или несреќи на патот пред тие да се случат
- му овозможуваат поголема контрола на возачот над возилото, како при нормално непречено возење, така и во стресни и опасни ситуации

Според временската рамка во која се појавуваат на сцена, во текот на развојот на автомобилската индустрија, активните сигурносни системи може да се поделат на: прв и втор бран на системи за активна сигурност. Во првиот бран се вбројуваат основните системи за сигурност при возењето, кои одамна се широко распространети во превозните средства и во денешно време, повеќе од 90% од автомобилите на Европските патишта се опремени со истите. Во оваа група сигурносни системи се опфатени:

- Анти-блокирачки систем за заочување (ABS, Anti-lock braking systems): Овој систем помага да се спречи заклучување/блокирање на тркалата на возилото при ненадејно, силно сопирање и му овозможува на возачот да продолжи со управување на возилото.
- Систем за електронска дистрибуција на силата на сопирачките (EBFD, Electronic Brake Force Distribution): Ова претставува технологија на автомобилски сопирачки што автоматски ја менува количината на сила применета на секоја од сопирачките на возилото, врз основа на условите на патот, брзината, вчитувањето итн. Овој систем е секогаш поврзан со ABS системот, што му овозможува да приложи поголем или помал притисок на сопирање на секое од тркалата, со цел да се зголеми моќта на запирање, [4].
- Систем за контрола на електронската стабилност (ESC, Electronic stability control): Овој систем помага да се спречи лизгање на возилото, притоа спречувајќи го возачот да ја изгуби контролата додека свртува на нагла/остра кривина. Технологијата на системот за електронска стабилност може автоматски да ги активира сопирачките за да помогне возилото да се насочи во вистинската насока. Некои ESC системи влијаат и на намалување на моќноста на моторот додека да се врати контролата над возењето. Постојат пет главни компоненти на Системот за контрола на електронската стабилност:
 - Сензори за брзина на тркалото
 - ESC-хидраулична единица со интегрирана електронска контролна единица (ECU, electronic control unit)
 - Сензор за агол на управување
 - Сензор за стапка на скршнување
 - ECU за управување со моторот, за комуникација

Во вториот ран систем за активна сигурност се вбројуваат системите базирани на најсовремените технологии на денешницата, како што се интегрирани сензори и сензорски платформи, радар, камери, GPS (Global Positioning System) и ласери. Во оваа група сигурносни системи се опфатени:

- Автономен систем за итно заочување (AEB, Autonomous emergency braking): Овој систем се одликува со способноста автоматски да започне со сопирање кога сензорите на возилото идентификуваат веројатен судир, или ако судирот е неизбежен, а возачот не презема ништо (или не реагира доволно брзо). AEB системот може да открие потенцијален судир и да ги активира сопирачките за да го избегне или барем да го ублажи неговото влијание. Овој напреден автономен систем користи сензори за следење на близината на возилата во непосредната околина и открива ситуации кога релативната брзина и растојанието помеѓу истите сугерираат дека судирот е неминовен.
- Систем за предупредување при отстапување од коловозна лента (LDW, Lane departure warning): Ова е систем, кој го предупредува возачот ако ја прегазат обележаната лента на коловозот без да го користи индикаторот/трепкачот или доколку возилото излегува од предвидената лента за патување.
- Системи за детекција на поспаност и губење на внимание: Овој систем ја проценува будноста на возачот (на пример, следејќи колку долго вози некој или преку анализирање на начинот на управување со возилото) и го предупредува возачот да паузира кога е потребно.
- Информативен систем за ограничување на брзината (SLI, Speed limit information): Овие системи го известуваат возачот за моменталното ограничување на брзината со прикажување на истото на контролната табла или преку системот за навигација. Тие користат камери за препознавање на знаци на патот и/или користат податоци за ограничување на брзината преземени од системот за навигација.
- Системи за контрола на притисокот во гумите (TPMS, Tyre pressure monitoring systems): Системот го следи притисокот на воздухот на гумите на возилото и ги прикажува овие информации во реално време на возачот, на пример, со помош на предупредувачко светло.

На Слика 3 се прикажани најактуелните активни сигурносни системи во автомобилската индустрија, [3].



Сл. 3 Активни сигурносни системи во автомобилската индустрија

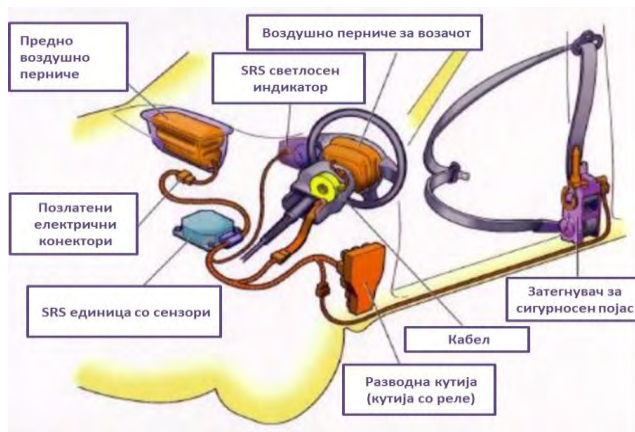
2) Пасивни сигурносни системи

Пасивните сигурносни системи ги заштитуваат патниците и останатите учесници во сообраќајот во случај на сообраќајна несреќа, такашто ги намалуваат, последиците, ударот и/или нивото на повреда при настаната несреќа. Имено, технологијата за пасивна сигурност, уште наречена и „секундарна сигурносна технологија“, е насочена кон ублажување на последиците од несреќата за време на и по ударот, од моментот кога ќе се воспостави првиот контакт.

Пасивните сигурносни системи користат широк спектар на вградени механизми, сензори и функции кои имаат улога во заштитата на патниците и останатите учесници во сообраќајот, а помеѓу нив најактуелни се:

- Дополнителен систем за задржување (SRS): Овој систем е дизајниран за да го задржи телото во случај на ненадејно заочување. SRS системот е дизајниран да работи без какви било активности од возачот или патниците. Причината поради која воздушните перничкиња, како дел од овој систем, се нарекуваат „дополнителни“ е затоа што сигурносниот појас ја претставува примарната линија на заштита на патниците при судир. Со цел, ефикасно работење на воздушните перничкиња, сигурносните појаси треба да се носат постојано. На Слика 4 се прикажани компонентите на SRS системот, [4].
- Сигурносен појас: Сигурносните појаси се системи за ограничување што ги одржуваат патниците правилно поставени за време на несреќа или ненадејно заочување, со што се намалува влијанието на внатрешноста на возилото врз телото на патникот и се спречува исфрлање на патникот од седиштето.

Сигурносните појаси бележат значителен подем во својот развој низ годините, такашто денешните безбедносни појаси се карактеризираат со својство на превентивно затегнување/заочување, имено тие се затегнуваат речиси веднаш во случај удар, со цел да се спречи претерано исфрлување на патниците напред, [5].



Сл. 4 Компонентите на SRS системот

- Воздушно перниче: Воздушните перничиња се карактеризираат со способност за брзо надување при удар (и последователно забавено се издишуваат) за да ги заштитат патниците за време на судир. Тие обезбедуваат меко разделување помеѓу патниците и внатрешноста на возилото за време на несреќата, што може да ги намали, па дури и да ги спречи повредите.
- Зона на деформација: Зоните на деформација контролираат ја одземаат кинетичката енергија при удар. Ова се реализира преку специјално дизајнирани области на возилото кои се деформираат и се распаѓаат за време на несреќа за да се апсорбираат ударите.
- Систем за итни повици (Emergency- Call System, E-Call): Системот за итни повици претставува комбинација од интегриран систем во возило (In Vehicle System, IVS) и соодветна инфраструктура како дел од јавната безбедност (Public Safety Answering Points PSAPs), кој користи информатички и комуникациски технологии. Автономниот интегриран систем испраќа повик до PSAP во случај на несреќа и преку напредната сензорска платформа испраќа податоци за времето и локацијата од местото на несреќата. На тој начин, времето на одговор на службите за итна помош значително се намалува, спасувајќи животи и резултирајќи со помалку сериозни последици од повреди. E-Call системот се одликува со следниве карактеристики:

- Испраќање информации за сообраќајната несреќа до PSAP кога ударот е „почувствуван“ од страна на сензорната мрежа.
- Испраќање информации за сообраќајната несреќа до PSAP кога E-Call копчето е мануелно притиснато од страна на возачот/патникот

- Испраќање информации за локацијата и сериозноста на несреќата
- Воспоставување интеракција со потребните служби

а) Концепт на систем за итни повици

Системот за итни повици, познат како E-Call системи се карактеризираат со способноста да генерираат итен повик, со рачно притискање на копче од страна на патниците на возилата или автоматски преку активирање на сензори во возилото, во случај на сериозна сообраќајна несреќа. Кога е активиран, E-Call системот во возилото воспоставува 112-гласовна врска директно со релевантната PSAP. Дури и во случај кога ниту еден патник не е во состојба да зборува, на пример поради повреди, до PSAP се испраќа „Минимален пакет на податоци“ (Minimum Set of Data, MSD), што вклучува точна локација на местото на несреќата, режим на активирање (автоматски или рачен), идентификациониот број на возилото, временската ознака, како и тековните и претходните позиции. На овој начин, информациите кои се важни за оние што реагираат на итни случаи, до нив стигнуваат што е можно побрзо. Стандардните концепти на податоци што го сочинуваат „Минималниот пакет на податоци“, што треба да се пренесе од возилото до PSAP во случај на несреќа или вонредна состојба, се специфицирани во стандардот EN15722 - Интелигентни транспортни системи – Е-Безбедност – ECall на минимален пакет на податоци (Intelligent transport systems - ESafety - ECall minimum set of data), [6].

Најчестите очекувања од ECall системите се: автономно откривање на несреќи, информирање на службите за итни случаи и пренесување информации како што се локацијата и, доколку е можно, бројот на погодените лица. Иако, ECall иницијативата бара автомобилот/уредот да бидат директно поврзани со 112/PSAP (бесплатна услуга), единствен број за итни случаи во Европа, сепак постојат и други системи, кои може да понудат посебни мрежи за итни случаи или друга дополнителна поддршка.

На Слика 5 се прикажани најактуелните пасивни сигурносни системи во автомобилската индустрија.



Сл. 5 Пасивни сигурносни системи во автомобилската индустрија

В. Безбедносни системи во автомобилската индустрија

Покрај сигурносните системи, во областа на автомобилската индустрија не смее да се запостави и значењето на безбедносните системи. Додека пред само неколку децении беше доволно едноставното заклучување на автомобилите, во денешното време на напредни технологии, се јавува потреба од комплексни системи за безбедност во автомобилската индустрија, како што се напредни системи против кражби (или за детекција на кражби), како и системи за сајбер безбедност кај полуавтономните и целосно автономните возила.

1) Системи за детекција на кражба на автомобили

Досега постоечките системи за детекција на кражба на автомобили се едноставни, како што е звучниот аларм или светлото што трепка. Во овие системи се интегрирани разни сензори, како што се сензори за притисок, вибрации и близина. Меѓутоа, ваквите конвенционални безбедносни системи доаѓаат со низа на недостатоци во поглед на актуелниот развој, како на пример нивните релативно едноставни карактеристики, кои лесно можат да бидат хакирани од страна на современите крадци, поради што и бројот на кражбите на автомобили постојано се зголемува. Современите безбедносни системи се изградени врз основа на прецизни контроли, сигнали и сензори, што овозможуваат да се дизајнира систем за заштита, кој автономно го предупредува корисникот со испраќање на СМС-порака, притоа користејќи GPS, кој овозможува следење на возилото, доколку дојде до пробивање на првиот безбедносен ѕид, како и испраќање на порака до интерниот систем на возилото, со цел негово запирање, [7].

2) Системи за сајбер-безбедност

Современата технологија и компјутеризирање доведуваат до значителен придонес за безбедноста, вредноста и функционалноста на возилото, почнувајќи од контрола на стабилноста до електронско вбризување на гориво, напредните сигурносни системи, па сè до навигација и спречување кражби. Поради тоа што овие системи сè повеќе се потпираат на споделување информации и комуникација во/со/помеѓу возилата, тие се изложени на постојан ризик од сајбер напади. Со зголемување на нивото на поврзаност, современите возила вклучуваат многу функции, заеднички со паметните телефони, како што се мобилни податоци, апликации, гласовни функции, веб-прелистувачи, кодови од преку 100 милион линии, имено константна мрежна поврзаност, која ги прави автомобилите ранливи и подложни на безбедносни проблеми во сајбер-доменот. Ова зголемување на поврзаноста, создава подлога за нови видови „напади“, што дополнително ја нагласува потребата од робусни безбедносни решенија, особено во ваква ера на развој, кога се појавуваат и развиваат поврзани, автономни возила. Така, секоја електронска контролна единица

(ECU) мора на некој начин да биде обезбедена, без разлика дали тоа е преку дополнителни процесори, проверка на кодови, заштита на податоците во мирување и во фаза на пренос или други можности кои се веќе вообичаени за стандардната Интернет-безбедност, [8].

Со цел да се намали опсегот на нападот и да се заштитат најкритичните функции во автомобилскиот систем, најчесто се употребуваат неколку докажано ефективни, современи безбедносни мерки:

- Безбедни комуникации во возилото, кои се базираат на готови криптографски производи за примарни функции како што се внатрешни мрежни поврзувања, откривање на упад или филтрирање податоци;
- Заштита на електронските контролни единици (дури и оние кои не се од клучно значење за работата на возилото) преку употреба на најдобрите практики, како што се бришење интерфејси и услуги што се користат за развој кога автомобилот е подготвен за ослободување во продажба, или со користење на комбинација за размена помеѓу хардверска и софтверска безбедност за максимално искористување на достапното ниво на одбрана од сајбер напади;
- Вршење редовна проверка и истражување за еволуцијата на сајбер безбедноста во слични интелигентни транспортни системи, како што е аеронаутиката, каде што вградените системи и сензорски платформи се споредливи со мрежите во возилата (на пример, поделени оперативни системи, распределба на домени, критичност при транспорт);
- Употреба на технолошката разновидност, со цел спротивставување на безбедносната монокултура (поранлива и поподложна на сајбер напади) и избегнување напад на критични компоненти;
- Одржување на безбедноста со текот на времето, користејќи техники за надзор како што се континуирано управување со ранливоста на хардверот и софтверот, користејќи канали за трансфер на податоци, надвор од стандардниот опсег. Покрај тоа, одржливоста на сајбер безбедноста е и клучен овозможувач на долгорочна заштита базирана на криптографија, [9].

III. ЗАКЛУЧОК

Иако безбедноста на возилото суштински се разликува од сигурноста на возилото и иако овие две области имаат посебни функционални системи, сепак на некој начин, безбедноста на возилото е од суштинско значење за обезбедување на неговата сигурност. Имено, овие две различни групи на системи треба да функционираат заедно и синхронно, со цел да се

овозможи беспрекорна функционалност на автомобилот како целина и на целата поврзана инфраструктура во сообраќајот. Организациските дисциплини што водат до сигурни и издржливи автомобили важат и за нивната безбедност. Всушност, сигурноста, издржливоста и безбедноста мора да започнат со формирање на самиот почеток, во фазата на дизајнирање. Од една страна, дизајнот на сигурносните системи вклучува зони на деформација, воздушни перничња, детекција на близина и системи за автоматско сопирање, додека пак од друга страна, дизајнот на безбедносните системи е насочен кон градење слоеви на заштита, со цел заканата да се изолира пред истата да успее да ги афектира операциите на возилото. Со тоа, целта на таканаречените архитекти на безбедносни системи за автомобилската индустрија е да употребуваат колекција на безбедносни алатки, почнувајќи од шифрирање на критични или приватни податоци, па до изолација на софтверски компоненти по функција, како и комбинација на хардверски и софтверски функции, до тој степен колку што е потребно за да се исполнат целите на трошоците и перформансите и за да се зачува функционалноста на сигурносните системи и возилата во целина.

Со непрекинатиот развој на автомобилската индустрија и со константниот напредок на технологиите и автономните системи кои се составен дел од современите возила, континуирано се појавуваат и нови закани кон безбедносните, а со тоа и кон сигурносните системи за автомобили. Ваквите закани соодветно се решаваат со ефективни решенија кои во многу блиска иднина ќе треба да се стават под приоритетен развој, со цел таканаречениот „Автомобилски облак“, да може беспрекорно да функционира, без да ја загрози целокупната инфраструктура во сообраќајот: возило-возило, возило-сообраќај, возило-животна средина. Секако, новите решенија, како во фазата на развој на системите за возилата, така и во фазата на имплементација и одржување на целокупната мрежа, ќе отворат нови хоризонти во науката и во автомобилската индустрија, со фокус на беспрекорна функција на целосно автономните возила на иднината.

IV. КОРИСТЕНА ЛИТЕРАТУРА

- [1] Dr. S. Sharma, A. Soni, D. Aawarne, Mechanical Engineering Dpt., SAGE University Indore, MP, India, “Automotive Sensors: A Review”
- [2] Prof. Dr.-Ing. Konrad Reif (Ed.)- Bosch Professional Automotive Information: “Automotive Mechatronics-Automotive Networking, Driving Stability Systems, Electronics”, Germany, 2015
- [3] <https://roadsafetyfacts.eu/active-safety-systems-what-are-they-and-how-do-they-work/> , “ACTIVE SAFETY SYSTEMS: WHAT ARE THEY AND HOW DO THEY WORK?”, ACEA – European Automobile Manufacturers Association, 2019
- [4] R. Waghe, Dr. S Y.Gajjal, “Study of Active and Passive Safety Systems and Rearview Mirror Impact Test”, SSRG International Journal of Mechanical Engineering (SSRG-IJME) – volume1 issue 3 July 2014
- [5] <https://roadsafetyfacts.eu/passive-safety-systems-what-are-they-and-how-do-they-work/> , “PASSIVE SAFETY SYSTEMS: WHAT ARE THEY AND HOW DO THEY WORK?”, ACEA – European Automobile Manufacturers Association, 2019
- [6] B. Attila, G. Attila, O. Krammer, S. Hunor, I. Balázs, K. Judit, Z. Szalay, P. Hanák, G. Harsányi, “A Review on Current eCall Systems for Autonomous Car Accident Detection”, Dpt. of Electronics Technology, Budapest University of Technology and Economics, Budapest, Hungary, 15 August 2017
- [7] S.M. Ahmed, H.M. Marhoon, O.Nuri, “Implementation of smart anti-theft car security system based on GSM”, International Journal of Engineering & Technology, 23 March 2019
- [8] Intel Security, “Automotive Security Best Practices, Recommendations for security and privacy in the era of the next-generation car”, 2015
- [9] F.Charbonneau, Dr. M. S. B. Mahmoud, Dr. D. Jackson, France, “Cybersecurity in automotive – How to stay ahead of cyber threats”

Overview of security and safety systems in the automotive industry

Aleksandra Gjorgjievska, Mare Srbinovska, Martin Gjorgjievski

Faculty of Electrical Engineering and Information Technologies

“Ss. Cyril and Methodius” University

Skopje, R.N. Macedonia

gj.aleksandra@yahoo.com, mares@feit.ukim.edu.mk, gj.martin91@gmail.com

Abstract—This paper focuses on an overview of modern security and safety systems in the automotive industry, as well as the intelligent platform of integrated sensors on which they are based. The purpose of this paper is to give a brief explanation of current security and safety systems, such as anti-lock braking system, airbag and seat belt system, emergency call system (E-Call system), as well as theft and cyber-security systems. The technological progress of these systems is the basis for the existence of partially and fully autonomous vehicles, and their functionality fully and directly affects the safety and security of drivers, passengers, and all other participants in traffic.

Keywords— security systems; safety systems; automotive industry; sensors; autonomous vehicles

Design and Evaluation of Collaborative Learning Platform with Integrated Remote Laboratory Environment

Zivko Kokolanski, Bodan Velkovski, Tomislav Shuminoski

Ss. Cyril and Methodius University in Skopje
Faculty of Electrical Engineering and IT
Rugjer Boskovic 18, 1000 Skopje
kokolanski @feit.ukim.edu.mk

Ana B. Kokolanska, Anita K. Mijovska

SETUGS Mihajlo Pupin,
Blagoja Stefkovski bb, 1000 Skopje, Macedonia

Dušan Gleich, Andrej Sarjaš

University of Maribor
Slomškov trg 15, 2000 Maribor, Slovenia

Matjaž Šegula, Matic Podobnik

Kranj School Centre
Kidričeva cesta 555, 3000 Kranj, Slovenia

Zlatko Ruščić, Tibor Kratofil

Technical school Ruder Boskovic in Vinkovci
Stanka Vraza 15, 32100 Vinkovci, Croatia

Abstract— In this paper, the design and evaluation of a computer-based remote virtual laboratory with an integrated collaborative environment in vocational education have been presented. The remote virtual laboratory was implemented as a joined effort of several participating institutions from Universities and Vocational education in the framework of the Erasmus+ project CORELA. The paper summarizes the design and implementation concepts and presents the user experience evaluation results obtained from students from vocational education. The results presented in the paper suggest that the CORELA platform provides a good professional user space for remote laboratory experimentation but also give insight into where the virtual platform could be improved.

Keywords—remote laboratory, collaborative learning, virtual laboratory, vocational education

I. INTRODUCTION

In order to be competitive in the global economy, Europe tends to be most highly skilled region in the world. Recent study shows that the occupational structure of EU employment of the engineering sector tends to shift towards knowledge and skills-intensive jobs from 27.3% in 2007 to 32.4% in 2020 [1]. Industry requires well educated science, technology, engineering and mathematics (STEM) professionals at all education levels. However, at the same time the reduction of professionals in STEM is very high - 30% for the science education and 50% for the engineering. Such technical and educational demands could be met by introducing innovative learning methodologies that include class instructions, practical assignments, laboratory work and extracurricular activities. Vocational education and training (VET) teaching staff have the central role in reaching these goals. On the other hand they must have access to high quality resources to support student's curiosity with state-of-the-art research and developments in STEM, and supported interactive instruments including an experimental laboratory.

The Erasmus+ project CORELA [2] introduces innovative, integrated tailor made platform intended for VET education. The platform improves the remote laboratory concept that has been used in higher education institutions. Usually, this concept is broadly accepted at the technical universities, where remote experiments are implemented to support student's curriculum. Most of the experiments are not widely disseminated and exist side by side with the known remote laboratories such as iLAB (MIT, USA) [3], LabShare (UTS, Australia) [4], VISIR (BTH, Sweden) [5], WebLab-Deusto (Spain) [6]. Moreover, conceptual designs of remote laboratories are presented in numerous research papers [7-9].

In the last few years, simple versions of those laboratory experiments have been proposed to the STEM school teachers' community. However, there is much less experience of implementing such technologies in the secondary vocational education. This project was aimed to channel this effort into the development of internet-based software platform which supports international collaboration through forming groups with pupils from different countries and different cultures while collaborating on remote experiments using remote virtual laboratories (RVL). Unlike in real laboratories, where pupils are often confined to a limited time, closely monitored by a supervisor and without an option to repeat experiment, the RVL platform offer unlimited access and freedom to explore. This can be also extremely important in cases of restricted student mobility such as the COVID-19 pandemic situation that we are currently facing.

This paper elaborates the design and implementation of the virtual remote environment with integrated collaborative learning focusing on the virtual instrumentation design. On the other hand, the paper summarizes the quality control evaluation report that has been done by using a large group of VET students which took part in the practical experiments. We believe that the results reported herein are widely relevant having in mind the international mix of student working groups and VET schools from different partner institutions.

II. COLABORATIVE REMOTE VIRTUAL LABORATORY WITH INTEGRATED COLABORATIVE LEARNING ARCHITECTURE

The innovation of the CORELA RVL platform architecture rises from the idea that globally distributed systems could be interconnected to function concurrently. Such systems are intended to be controlled by international teams of teachers and pupils, also distributed worldwide. The VET students collaborate and communicate through the platform to achieve the required objectives. Remote virtual laboratories, which started their development about two decades ago, are currently seen as the beginning for future advanced global educational systems. They offer a unique opportunity to develop a teaching and learning platform for the development of skills required for efficient collaboration and communication on a local and global scale. Currently there are several RVLs reported worldwide, yet only a few are constructed in such way to allow involved participants to collaborative and operate in real-time. However, a number of other institutions have recognized the advantages of collaborative RVLs and are in the process of redeveloping their RVLs into collaborative learning environment. On the other hand, very little research has been done on the evaluation of collaborative learning in RVLs, especially in secondary education. Among others, this is because a large majority of RVLs are designed as single user laboratories where pupil's collaboration is not possible. This contradicts vocational secondary education practice where pupils normally perform laboratory experiments collaboratively in larger groups. Such approach will not only enhance the employability and career prospects of the VET students by providing them access to sector-specific skills, but at the same time will support their creative potential and increase the innovative capacity of the VET education centers.

The CORELA remote virtual laboratory with integrated collaborative learning consists of two integral parts: a moodle-based platform, and Laboratory Virtual Instrumentation Engineering Workbench (LabVIEW) laboratory environment. The general CORELA architecture is given in Fig. 1.

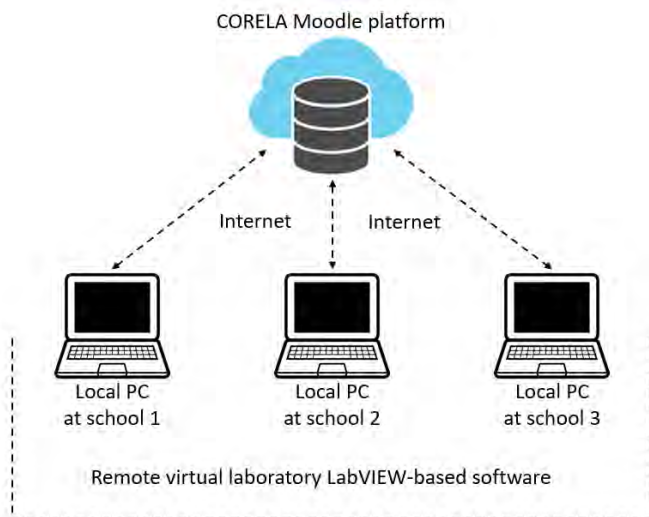


Fig. 1. CORELA remote virtual platform with integrated collaborative learning architecture

The moodle platform contains centralized database hosted on a web-server located in SETUGS Mihajlo Pupin. This platform serves as a data sharing and communication tool between the VET schools. By using the platform, VET teachers can easily distribute educational materials, organize students in groups and deliver assignments. Afterwards the students can realize experiments, work in a collaborative environment and submit reports. However, regarding the CORELA project, two most important aspects of the moodle platform implementation should be highlighted: extended database for communication with the RVL, and tools for application of the new teaching methodology for optimal application of collaborative learning in VET. The available tools for collaborative learning in VET have been defined in accordance with the survey [10] provided in the CORELA project. The survey results showed that students are eager to work in international groups, they prefer real-time chat instead of video or audio communication, and overcome the cultural and language barriers.

The moodle platform database has been extended with additional features in order to couple with the LabVIEW-based RVL platform. It contains additional information regarding the users (login credentials, dedicated user identification, laboratory experiment identification, etc.), as well as specialized data channels for exploitation with the RVL. There are ten easily upgradeable input/output channels in the database associated to the RVL. These channels are polymorphic, meaning that the users can exploit them by using different data types and structures (scalars, arrays, signals, etc.). Once the user is logged-in to the RVL, than he/she can perform continual two-way communication by using the JavaScript Object Notation (JSON) data interchange format that uses human-readable text to store and transmit data objects consisting of attribute-value pairs and arrays. The detailed implementation of the LabVIEW-based RVL is given in the following chapter in this paper.

III. CORELA LABVIEW-BASED REMOTE VIRTUAL LABORATORY

The CORELA remote virtual laboratory is a stand-alone personal computer (PC) software implemented in LabVIEW environment. The program is intended to be installed on each computer at VET institutions and provide direct communication with the CORELA database integrated into the moodle platform. The virtual instrument consists of two main parts: front panel given in Fig.2, and block diagram given in Fig.3. The front panel is the graphical user interface which provide all functionalities of the remote virtual laboratory. It is divided into three main sections: function list (left part in Fig.2), experiment development space (central part in Fig.2), and configuration and log-in user space (right part in Fig.3). Initially, the user must log-in with a valid username and user identification (ID), as well as to provide the laboratory exercise ID. Once the user logs in, than he/she can start developing the laboratory experiment by using the functions given in the function list. There are a lot of predefined functions in the function list, generic ones (mathematics, controls, indicators, probes, etc.), and specialized functions (database connection, signal functions, data acquisition cards, variables, electrical engineering functions, etc.).

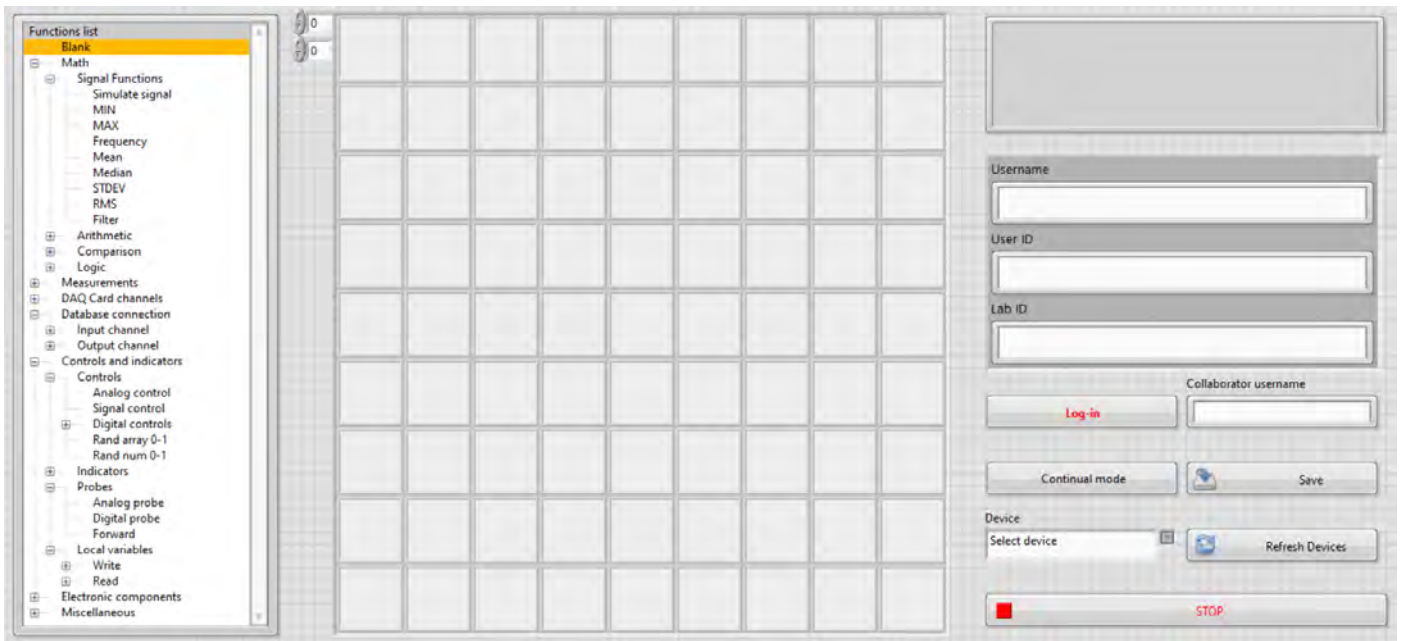


Fig. 2. Front panel of the LabVIEW-based CORELA remote virtual laboratory

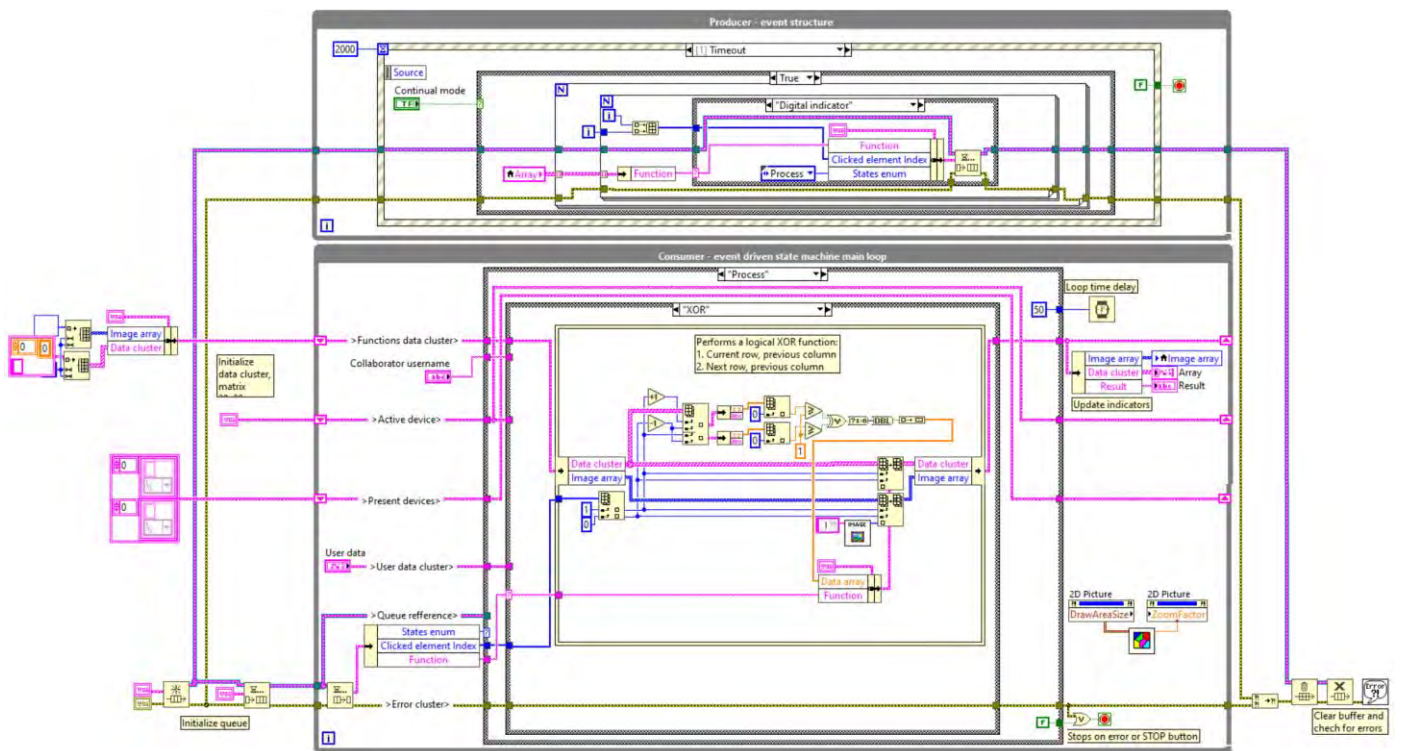


Fig. 3. Block diagram of the LabVIEW-based CORELA remote virtual laboratory

The experiment is realized by selection of particular function from the function list and then placing it in a given square of the experiment development space. There are specific rules for inputs (left side) and outputs (right side) for each function from the functions list that can be analyzed with the implemented help tools. Once the experiment diagram is finished, the user can then execute the program in one of the two regimes: single-stepping, and continual mode.

In the single-stepping mode, a function is been executed as soon as it is placed at the experiment development space. However, if particular input of that function changes later on, than the output of the function will remain unchanged. Such regime is suitable for development of simple calculation tasks, reading single measurements from the data acquisition card, etc. On the other hand, in the continual mode all functions are continually updated with frequency of

approximately 2 Hz. It is obvious that such behavior of the program is suitable real-time applications. Example of a simple exercise in the experiment development space of the RVL platform is given in Fig.4.

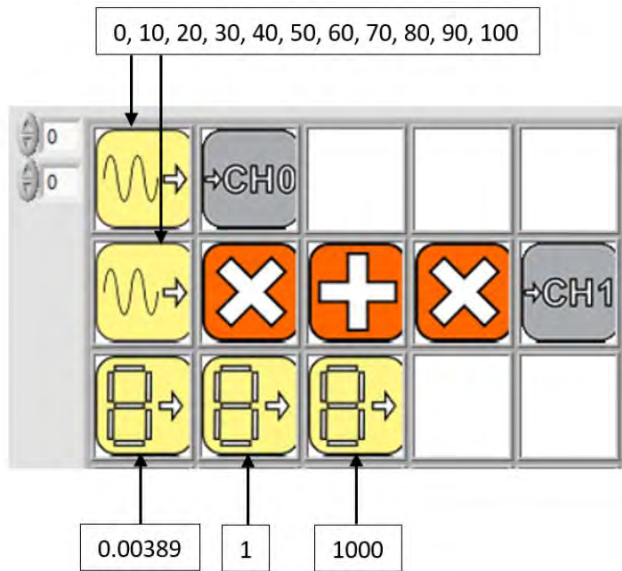


Fig. 4. Realized experiment in the RVL platform, an examination of the transfer function of resistive temperature detector (RTD)

The task of the experiment given in Fig.4 is to simulate the RTD transfer function by using the CORELA RVL platform and compare them with the theoretical transfer function. The Channel 0 in the database contains the input temperature covering entire measurement range from 0 °C to 100 °C with a step of 10 °C. The Channel 1 of CORELA database contains the resistance of the RTD for the predefined temperature set points.

The programming code (block diagram) of the LabVIEW-based CORELA remote virtual laboratory is given in Fig.3. The program is based on so called producer-consumer with integrated event-driven state machine programming architecture. In general, there are two while loops running in parallel: a producer loop (upper part in Fig.3) and consumer loop (lower part in Fig.3). The producer loop is used for registering events, such as clicking on a control button or placing a function on the experiment development space. Once an event is been registered it is automatically processed, and particular message is submitted to the consumer loop by using queues. The program architecture is developed in such a way that it guaranties lossless information. That means that if more than one event occur in a short period of time, the program will buffer all of them in the queue so no information will be lost. The consumer loop process the messages from the queue one by one by using the first-in first-out (FIFO) principle. Each message contains information about the nature of the event and define the state-transition diagram for the state-machine. Later, the state machine executes the programming sequence determined by the message from the producer loop and updates the experiment development space and/or the front panel controls and indicators.

IV. EVALUATION OF CORELA PLATFORM

The quality control of the CORELA remote virtual laboratory with collaborative learning was performed by forming international groups of students from the VET participating institutions in the project (SETU of GS Mihajlo Pupin – North Macedonia, TU Rugjer Boskovic-Croatia and SC Kranj-Slovenia). A total of 33 students from electro-technical program took part in the realization of laboratory exercises by using the RVL platform. Afterwards, the students answered a questionnaire for evaluation of their user experience. The purpose of the questionnaire is to evaluate the students experience in terms of several criteria: collaborative learning, exploitation of the CORELA RVL platform, and evaluation of student satisfaction in conducting remote laboratory experiments.

In total eight questions were related to the concept of collaborative learning in a form of multiple choice answers but also keeping the possibility to add additional explanation. The distribution of used languages in the process of collaborative learning is given in Fig.5.

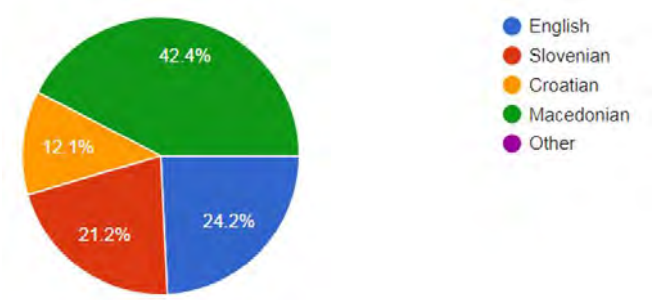


Fig. 5. Used languages in performing collaborative learning by using the CORELA RVL platform

From the results given in Fig.5 it can be seen that all native languages from the VET participating institution and international English language were used for communication between the students. One of the most important conclusion regarding this aspect of the survey is that students are eager to perform collaborative learning in international groups. More than 94% of them didn't experienced any language, religious, cultural, social or any other barriers while working with students from another countries. Moreover, more than 90% of the students think that collaborative learning makes learning easier, and that it makes students who work together achieve more than when they work alone. Such an opinion is of a very high importance because it unites the technical content with the collaborative international experience, which is one of the main goals of the CORELA RVL platform. Regarding the mental process of thinking, students think that collaborative learning increase their motivation to learn and make them express their opinions, argue, debate, negotiate, and increase their knowledge.

Most of the students (74%) found CORELA platform good or very good, whereas the other 26% of them have a satisfactory opinion. Such evaluation suggests that the CORELA platform fulfills the criteria from the user's point of view, but also needs certain improvements. This can be

also confirmed by the quantitative evaluation where the CORELA RVL platform got 4.06 out of 5 score points. During the laboratory experiments, nearly 50% of the students have used a dedicated hardware (NI MyDAQ or Arduino, supported by the platform). These students didn't report any hardware-related problems, but however they suggest that it can be improved to be more intuitive. The overall evaluation of the areas where the CORELA RVL platform could be improved is given in Fig.6.

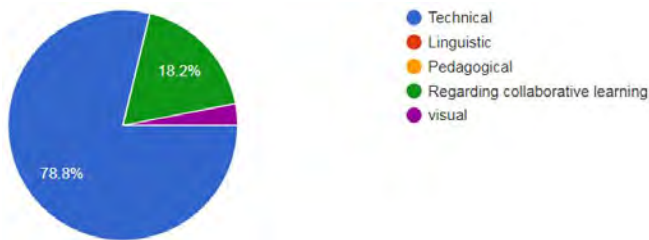


Fig. 6. Areas of improvement for the CORELA RVL platform

From the results given in Fig.6 it is clear that the CORELA RVL platform can, and should, be improved. According to the performed survey, the most improvements should be focused to the technical content, the visual aspect of the program, and the process of building circuit diagrams. Moreover, students suggest that the CORELA moodle platform should also consider using additional tools and techniques for collaborative learning.

V. CONCLUSION

CORELA remote virtual laboratory with integrated collaborative learning introduces innovative, integrated remote virtual laboratory designed to be used by the vocational education centers. The platform proposed in this paper is a step towards diversification and modernization of the teaching methodology in VET.

The paper summarizes the CORELA RVL platform architecture and gives an overview of the process of virtual instrument design. It has been shown that the applied virtual instrument architecture is suitable for implementation of the RVL platform and guarantees lossless information.

The evaluation of the CORELA RVL platform was performed by the VET education centers by forming an international group of students. The conducted survey showed that the students are strongly supporting the collaborative learning aspects implemented in the platform, and find them very inspiring and useful. The user experience of using the RVL platform is well evaluated but also a possible areas of improvements have been identified. It has been suggested that the virtual platform can be improved regarding the technical content, application interaction, and incorporation of more diverse tools for collaborative learning.

ACKNOWLEDGMENT

Authors would like to thanks European Commission to grant Erasmus+ project under grant 2018-1-MK01-KA202-047107.

REFERENCES

- [1] CEDEFOP, Skills supply and demand in Europe: medium-term forecast up to 2020, March 12, 2012
- [2] European Comission, "Collaborative Learning Platform with Integrated Remote Laboratory Environment in VET", Erasmus+ project 2018-1-MK01-KA202-047107, 2018, available: <https://euprojects.mk/maps/report/1098>
- [3] MIT. (2019) ilabs. [Online]. Available: <http://web.mit.edu/edtech/casestudies/ilabs.html>
- [4] RemoteLabs. (2019) Labshare. [Online]. Available: <http://www.labshare.edu.au/>
- [5] Sweden. (2019) Visir. [Online]. Available: <https://www.visir.org/>
- [6] WebLab-Deusto. (2019) Rlms. [Online]. Available: <http://www.weblab.deusto.es>
- [7] S. Frerich, D. Kruse, M. Petermann, A. Kilzer "Virtual Labs and Remote Labs: Practical experience for everyone", 2014 IEEE Global Engineering Education Conference (EDUCON), DOI: 10.1109/EDUCON.2014.6826109, Apr. 2014
- [8] Z. Nedic, J. Machotka, A. Nafalskic "Remote laboratories versus virtual and real laboratories", 33rd Annual Frontiers in Education, DOI: 10.1109/FIE.2003.1263343, Nov. 2003
- [9] Y. Zhang; T. Gao "The Study of Remote Virtual Laboratory Construction", International Conference on E-Product E-Service and E-Entertainment, DOI: 10.1109/ICEEE.2010.5661538, Nov. 2010
- [10] A. B. Kokolanska, A. K. Mijovksa, N. Bozinovska "Methodology Analyses for Collaborative Learning Platform with Integrated Remote Laboratory Environment in Vocational Education and Training", ISBN 978-608-245-424-5, 2019

Error Evaluation in Reactive Power and Energy Measurements Adopting Different Power Theories

Kiril Demerdziev, Vladimir Dimchev

Ss. Cyril and Methodius University in Skopje (UKIM)

Faculty of Electrical Engineering and Information Technologies (FEEIT)

Skopje, Republic of North Macedonia

{kdemerdziev, vladim}@feit.ukim.edu.mk

Abstract – The prevalence of high order harmonics in power systems results in demand for measurement equipment calibration in such non – sinusoidal conditions. This is especially challenging in case of reactive power/energy instruments for 2 reasons: their role in the billing of electrical energy and the fact that the reactive power is not unambiguously defined in case of harmonics. These instruments are based on different measuring principles which provides additional complication for a unified calibration procedures establishment. Because the decomposition of harmonically distorted waveforms is a multivariable problem and in order a valid conclusions to be adopted, instruments' output is supposed to be analyzed from the perspective of different signals' parameters, such as: harmonic order of single components, their amplitude and phase shift, as well as the phase shift of voltages and currents at fundamental frequency.

Keywords – high order harmonics, electricity meter, phase shifts, reactive power.

I. INTRODUCTION

High order harmonics are nowadays highly prevalent in electrical networks and their existence is predominantly a result of non – linear loads such as [1-3]: arc furnaces, welding equipment, lighting installations with discharge lamps and LEDs, battery chargers, rectifiers, etc. The harmonic distortion of voltage and current signals means that demands for accurate measurement in power grids go beyond the instruments' specifications presented for reference sine wave conditions. This is especially important in domain of legal metrology, i.e. when electricity meters are regarded, because of their billing role in the regulated trade of electrical energy. According to EU Directive MID 2014/32/EU [4] “all measuring instruments used for commercial transactions” are supposed to measure the quantity of particular interest with error which will not exceed the maximal permissible error under rated operating conditions. In case of electricity meters, the rated operating conditions, are no longer only pure sinusoidal voltages and currents, but harmonically distorted waveforms as well [5].

In domain of active electricity meters testing, several international standards [6-8], a recommendation [9], and plenty of scientific works exist [10-14], and therefore different examination procedures are established. For example, in EN 50470-3 [8] test signals which possess 10% 5th order voltage harmonic and 40% 5th order current harmonic are presented. Similar limitations are introduced in the 2 test signals presented

in [9], but the voltage and current waveforms are not limited to a single high order component. In the scientific works regarded [10-14], some random test signals are proposed as well.

On the other hand, no such test procedures are proposed, regarding reactive energy meters. In the standard EN62053-23 [15] the accuracy demands for reactive electricity meters are presented, but they are limited to sine wave conditions. The main reason for this is the fact that the term *reactive power/energy* is not unambiguously defined in harmonically polluted environment [5, 16]. Several definitions for reactive power/energy in case of harmonics exist, each one possessing certain advantages and flaws. On the other hand, different meters are based on different measuring principles, all of them provide the same result in case of sinusoidal voltages and currents, but result in a totally different output in case of harmonically polluted systems [5]. A progress with understanding and unification of the reactive power measurement was made with publication of IEEE 1459 standard [17], in which it is stated that the quantity of particular interest for accurate measurement is the fundamental active power, Q_1 . The measurement of this quantity on the other hand does not provide billing equality in terms of harmonic producers penalization and consequently harmonic consumers compensation [18].

In the paper, an analysis of a reactive energy meter's output, which is in compliance with [17], in case of harmonically polluted environment, will be performed. The meter's errors will be calculated in relation to a reference standard (RS), taking into account different definitions of reactive power in case of harmonics. The harmonics' parameters, such as: single components share, their phase shifts and phase angles between the fundamental components, will be taken as influence quantities for determination of the meter's performance.

II. THEORETICAL BACKGROUND

A harmonically distorted voltage or current signal can be mathematically evaluated using a Fourier series as [1-3, 19]:

$$x(t) = \sum_{h=1}^n \sqrt{2} X_h \sin(h\omega t + \alpha_{xh}), \quad (1)$$

where h is the harmonic order, X_h and α_{xh} are the RMS and the phase shift of the component with a frequency h times the

fundamental and n is the maximal harmonic order which is taken into account for practical evaluation. The share of a single harmonic component is usually expressed as a percentage of the fundamental's value, X_1 [9-12]:

$$x_h[\%] = \frac{X_h}{X_1} \cdot 100, \quad (2)$$

while its phase shift is presented in relation to the initial phase shift of a 50 Hz (or 60 Hz) component, at positive zero crossing, α_{x1} :

$$\theta_{xh} = \angle(\alpha_{xh}, \alpha_{x1}) \quad (3)$$

The single phase active power, in case of harmonically distorted voltages and currents is expressed as the mean power in a pre – defined time interval (period), T [9-14]:

$$P = \frac{1}{T} \int_0^T u(t)i(t)dt = \sum_{h=1}^n U_h I_h \cos \varphi_h, \quad (4)$$

and it equals the algebraic sum of the powers obtained from the components at different frequencies. In (4) U_h and I_h are the RMS of the voltage and current of order h and φ_h equals the phase shift between them. The phase shift φ_h equals [20]:

$$\varphi_h = h\varphi_1 + \theta_{ih} - \theta_{uh}, \quad (5)$$

φ_1 being the phase shift between current and voltage at fundamental frequency, while θ_{ih} and θ_{uh} are the phase shifts of the h^{th} order current and voltage harmonics in relation to components at fundamental frequency.

A single phase apparent power of distorted waveforms equals the product of the voltage and current RMS values [19]:

$$S = UI = \sqrt{\sum_{h=1}^n U_h^2} \sqrt{\sum_{h=1}^n I_h^2}, \quad (6)$$

and up to this point of the discussion, the presented equations were correlated to principles which are valid for both sinusoidal and harmonically distorted conditions. If the power triangle of P , Q and S , valid for sine wave signals is taken as a reference, than the reactive power equals [5, 16]:

$$Q = Q_F = \sqrt{S^2 - P^2}, \quad (7)$$

and this equation corresponds to the power theory proposed by Fryze, therefore the reactive power will be labeled as Fryze's power, Q_F . According to Fryze, the current signal, $i(t)$, is separated into 2 components, namely "active" current $i_a(t)$, which is in phase with the voltage and possess the same waveform as $u(t)$, and non – active current, $i_r(t)$, which is the remaining part of the current signal. The Fryze's power theory provides satisfactory explanation of the system's efficiency and can be evaluated by using basic phasor knowledge [5]. The second power definition, regarded in this paper is, the Budeanu's power theory. According to Budeanu the reactive power is presented by using similar equation to the one for active power calculation in non – sinusoidal conditions, (4):

$$Q = Q_B = \sum_{h=1}^n U_h I_h \sin \varphi_h. \quad (8)$$

If the Budeanu definition for reactive power is used, than the remaining power, which exists in the system, beside P and Q_B is called distortion power:

$$D = \sqrt{S^2 - P^2 - Q_B^2}, \quad (9)$$

and it is a result of the mutual interference of voltages and currents at different frequencies.

As stated earlier, IEEE 1459 standard [17] presents a separation principle between fundamental powers and higher frequency components. That is appropriate from the meter's perspective, because the measuring principle of many devices used for reactive power or energy monitoring is based on time or phase rotation of the voltage or current signal for quarter of a period or 90° respectfully. Because only fundamental reactive power can be obtained by time or phase shifting of a voltage or current signal, it is obvious that the measured quantity will be equal to the one calculated as:

$$Q_1 = U_1 I_1 \sin \varphi_1, \quad (10)$$

where U_1 and I_1 are the RMS of voltage and current at 50 Hz and φ_1 is the phase shift between them.

III. MEASUREMENT EQUIPMENT AND PROCEDURE

The experimental part of the work is realized in ISO EN MKC 17025:2018 [21] accredited calibration laboratory, called Laboratory for Electrical Measurements (LEM), which is part of the Faculty of Electrical Engineering and Information Technologies (FEEIT) at Ss. Cyril and Methodius University in Skopje (UKIM). The reference standards in possession of the laboratory are periodically calibrated and maintain international traceability to BIPM [22-23]. For the purposes of this work, LEM's secondary standard, in domain of electric power and energy instruments calibrations, CALMET C300 [20] is used as a RS. It is a three phase low frequency voltage and current source, which is software controlled and possess menus for automatic electricity meters examination. Beside the possibility for sine wave signals generation, the RS reproduces harmonically distorted waveforms as well, which are previously manually set by the user.

The role of a Unit Under Test (UUT) is played by an electricity meter for both active and reactive energy (only reactive energy output is regarded in the paper), Landis+Gyr ZMD405CT44.2407, 3x58 V/100 V, 5A, 50 Hz, accuracy class 1 [24]. Its measuring principle is based on Digital Signal Processing (DSP) and the reactive power is obtained by a 90° phase shift of the measured voltage and its multiplication with an instantaneous value of the current. That being said, the reactive power/energy recording will come to measurement of the fundamental component, Q_1 , (10). The errors will be regarded from the reactive power's, rather than from the reactive energy's perspective in order a simplification of the analysis to be achieved. This simplification is valid, because errors in reactive energy measurement in case of harmonically distorted voltages and currents are *de facto* dominantly errors in reactive power measurement, taking into account that the time factor is not affected by the harmonic distortion of the signals. The connection of the electricity meter to the standard [20] is illustrated in Fig.1. As can be seen from the figure,

UUT's pulse output is connected to a signal conditioning circuit, in which the pulses procession, proportional to the measured energy, are adapted to the pulse input of the RS. In such a configuration a fore mentioned automatic test is performed, by comparison of the energy measured by UUT and the reference energy generated by the standard. The results are presented in a relative error form:

$$\varepsilon = \frac{Q_{UUT} - Q_{C300}}{Q_{C300}} \cdot 100, \quad (11)$$

where Q_{UUT} is a 3 phase reactive power measured by the UUT, and Q_{C300} is a 3 phase reactive power generated by the RS.

The experimental part of the work comprises of 2 measurement procedures which are conducted with test signals similar to those presented in [8], for examination of active energy electricity meters. In both voltage and current waveforms only 5th order harmonics are regarded beside the components at 50 Hz. The signals' magnitudes are held constant throughout the procedures and are equal to the nominal voltage and current of the UUT, 58 V and 5 A, respectively. The share of the 5th order harmonics is fixed in the first measurement procedure, at 10% for $u_5[\%]$ and at 40% for $i_5[\%]$. The phase shift between the 5th order current harmonic and current fundamental, θ_{i5} , is fixed at 60°, while each subprocedure is determined by the change of θ_{u5} , in the interval between 0° and 360°. Subprocedures are comprised of 12 measurement points, corresponding to different phase shifts between fundamental currents and voltages, φ_1 .

The second procedure is also conducted with test signals presented in [8], however single subprocedures are defined by the change of the 5th order harmonic share in the current waveform, $i_5[\%]$, while all other parameters are held constant: $u_5[\%]=10\%$, $\theta_{u5}=0^\circ$, and $\theta_{i5}=60^\circ$. In the procedure, $i_5[\%]$ varies between 20% and 40%, and single measurement points correspond to different values of φ_1 .

From the 2 procedures, 3 measurement data sets are obtained. In each data set errors made by the UUT are regarded in relation to different power definition: fundamental only, Q_1 , Budeanu's, Q_B and Fryze's, Q_F .

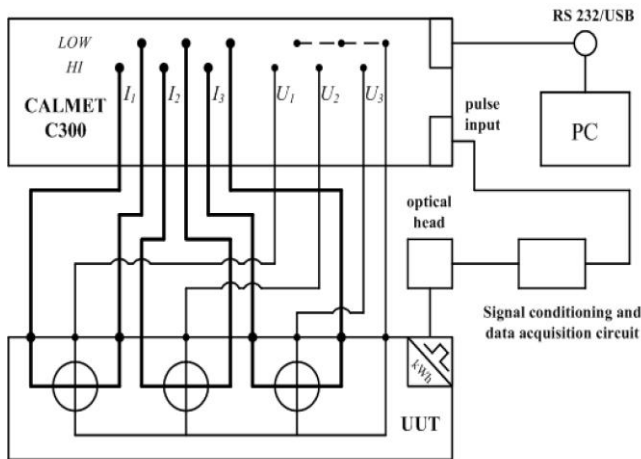


Fig. 1. Connection of a UUT to CALMET C300

IV. MEASUREMENT RESULTS AND DISCUSSION

The first measurement data set represents a comparison between the reactive power measured by the UUT and the fundamental power only generated by the RS. Because the electricity meter [24] measures Q_1 only, errors are labeled as ε_1 , and are expected to be minimal. The measurement results, illustrated in Fig.2 and Fig.3 as relative errors (11), are presented for different harmonic parameter change.

In Fig.2, the error curves comply with different values of the 5th order voltage harmonic phase shift in relation to the angle of the voltage fundamental. As can be seen from the figure, errors when the measured power is compared to the fundamental reactive power generated by the RS [20] are small, and for the most measurement points the UUT's readings are within its declared accuracy class. When the phase shift φ_1 is between 45° and 90°, both in inductive and capacitive range, the measured errors are lower than 1%, nevertheless the initial phase shifts of the 5th order harmonics. When φ_1 is lower than $\pm 45^\circ$, i.e. for lower reactive energy share in the system, the errors are higher and their magnitude can reach up to $\pm 3\%$. Error change in relation to φ_1 follows a sine wave pattern and its period is 5 times smaller than the period of a pure sine wave signal with a frequency of 50 Hz. This phenomena is related to the declared errors in distorted waveforms generation as stated in [20], which can be further evaluated as measurement uncertainty. It is important to be emphasized that the overall uncertainty in the RS's output is a result of both high frequency components and fundamentals, nevertheless only the fundamental reactive power is taken into account.

In Fig.3 different error curves correspond to a different 5th order current harmonic share in the waveform and a sine wave envelope of the error function is once again recorded. It can be concluded that measurement errors are directly proportional to the 5th order harmonic share in the current waveform, being highest when $i_5[\%]=40\%$ and lowest for $i_5[\%]=20\%$. The same conclusion can be derived if the amplitude of the voltage harmonic is variable, and the current harmonic is fixed. From the first measurement data set, it is clear that the meter's actual performance will be within its declared accuracy class even if more than one harmonic component is present in the waveforms of both voltage and current. That is the case because the errors in RS's output performance, which are result of single harmonic components, will eventually tend to cancel one another.

The second measurement data set presents a comparison, between the reactive power measured by the UUT and the reactive power generated by the RS, calculated according to the Budeanu's power definition. If the UUT measures fundamental reactive power only, (10), and the generated power by the RS is calculated as presented in (8), than the relative error, for any measurement point will be calculated, using (11), as:

$$\varepsilon_B = \frac{-3 \cdot U_5 I_5 \sin \varphi_5}{3 U_1 I_1 \sin \varphi_1 + 3 U_5 I_5 \sin \varphi_5} \cdot 100, \quad (12)$$

where the error is labeled as ε_B in order to indicate that it is presented in relation to Budeanu's definition of reactive power.

Taking into account that the fundamental power is much higher than the power component of the 5th order harmonics:

$$U_1 I_1 \sin \varphi_1 \gg U_5 I_5 \sin \varphi_5, \quad (13)$$

the error function will follow a sine wave pattern with a period of $360/5=72^\circ$. Error curves regarding the second data set are presented for different phase shifts of the 5th order voltage harmonic in Fig.4, and for different share of the 5th order current harmonic in Fig.5.

From Fig.4 it can be concluded that, for the concrete settings, the measured value is up to 5% lower or higher than the reactive power presented according to the Budeanu's concept, depending on φ_1 , when this parameter lies between 45° and 90° , both inductive and capacitive. If the results are compared to the results presented on Fig.2 the maximal errors recorded are 5 times higher. A more general conclusion derived from this data set is that for these settings ($u_5[\%]=10\%$,

$i_5[\%]=40\%$, variable φ_5 and φ_1), the reactive power presented according to Budeanu's concept is up to 5 % lower or higher than the fundamental reactive power in the system. For phase shifts φ_1 which correspond to a lower reactive power share in the system, the errors increase even up to 3 times and subsequently deviations up to $\pm 15\%$ are recorded.

From Fig.5, the linear relationship between the errors and the share of the current harmonic is recorded, as mathematically presented in (12). The highest errors for every subset are recorded for $\varphi_1=15^\circ$ i.e. for measurement points which correspond to a low reactive power share in the system and their intensity varies between -4.89% for $i_5[\%]=20\%$ and -9.55% when $i_5[\%]=40\%$. According to (12), similar relationship will be obtained if measurements are regarded for different values of $u_5[\%]$. The error difference between different measurement subsets is more noticeable for lower phase shifts, φ_1 .

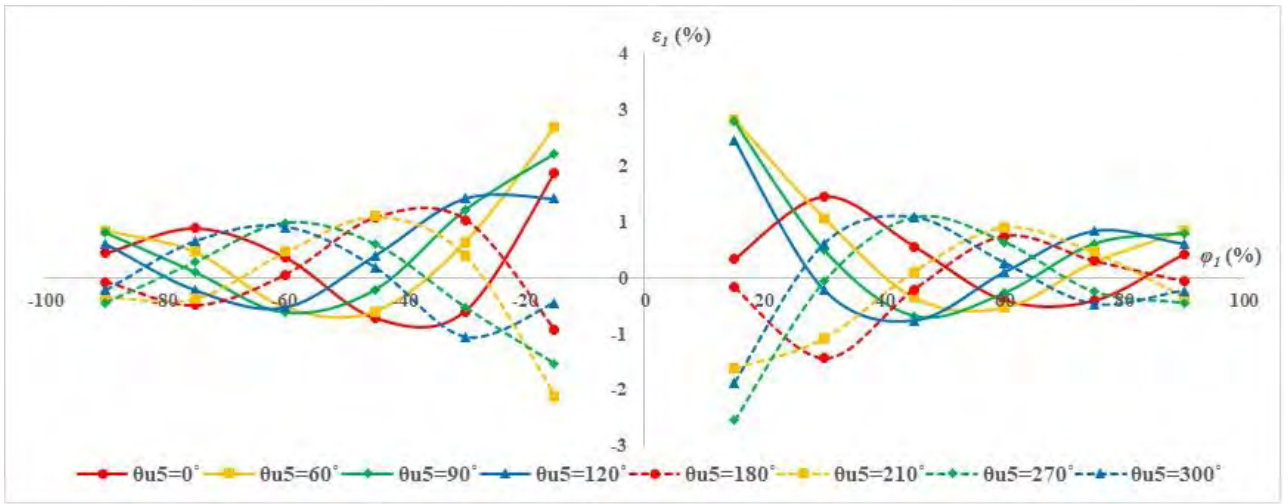


Fig. 2. Error function $\varepsilon_I=f(\varphi_I)$ for different values of θ_{u5} , $u_5[\%]=10\%$, $i_5[\%]=40\%$, $\theta_{i5}=60^\circ$

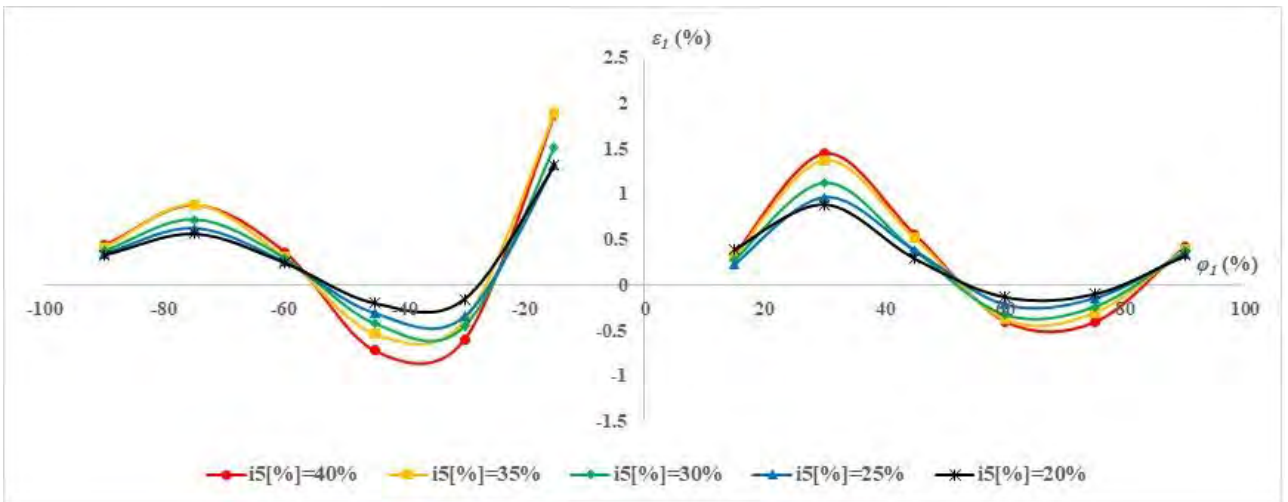


Fig. 3. Error function $\varepsilon_I=f(\varphi_I)$ for different values of $i_5[\%]$, $u_5[\%]=10\%$, $\theta_{u5}=0^\circ$, $\theta_{i5}=60^\circ$

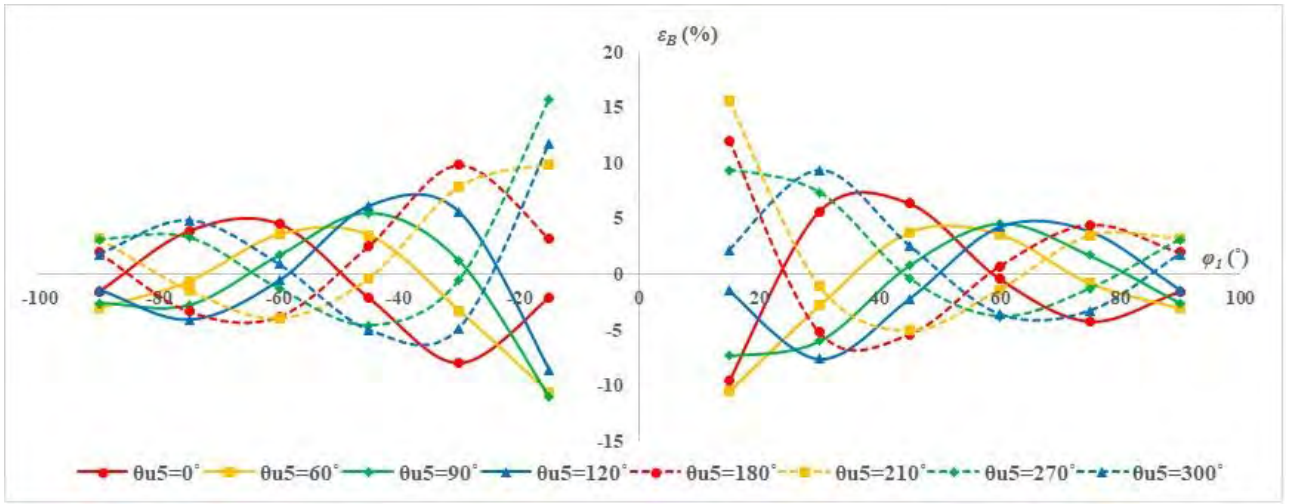


Fig. 4. Error function $\varepsilon_B = f(\varphi_I)$ for different values of θ_{u5} , $u_5[\%]=10\%$, $i_5[\%]=40\%$, $\theta_{i5}=60^\circ$

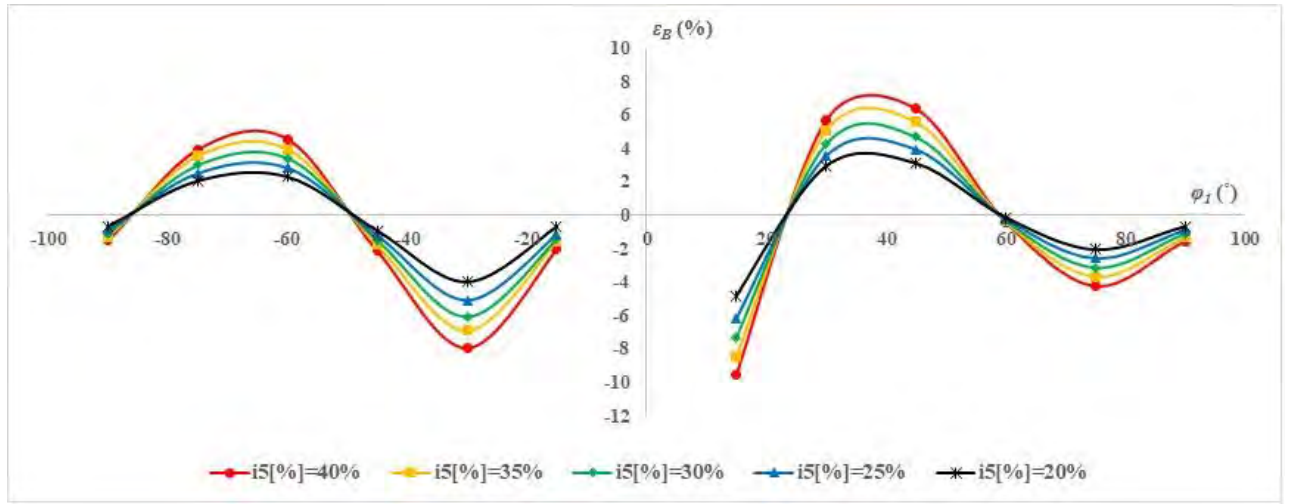


Fig. 5. Error function $\varepsilon_B = f(\varphi_I)$ for different values of $i_5[\%]$, $u_5[\%]=10\%$, $\theta_{u5}=0^\circ$, $\theta_{i5}=60^\circ$

The third data set presents a comparison between the UUT's output and the Fryze's reactive power. If equations (7), (8) and (10) are compared it can be highlighted that this power theory provides the highest amount of reactive power for the same settings, regarding both fundamental and harmonic components. Taking into account the working principle of the UUT, its output is going to be lower than the prescribed power, no matter the value φ_I , or the 5th order harmonic components. The justification is illustrated in Fig.6 and Fig.7.

In Fig.6 the measurement errors are presented in relation to φ_I , and different values of θ_{u5} correspond to different error curves. Error intensity is little dependent on θ_{u5} (and on θ_{i5} accordingly) when φ_I is between 60° and 90° , in both the inductive or capacitive range. In those measurement points, the difference between the power measured by the UUT and Q_F , equals between -6% and -10% for the concrete measurement settings. When $\varphi_I < 60^\circ$, a serious increase in errors is recorded and maximal magnitude between -35% and -55%, depending

on the θ_{u5} are present in the data set. The maximal errors are recorded for the lowest φ_I values, for the concrete test procedures, those are $\pm 15^\circ$

A linear dependence between ε_F and $i_5[\%]$ is observed from the results in Fig.7. On this figure, an additional justification, about the constant errors in case of high reactive power share in the system is presented, i.e. for $\varphi_I > 60^\circ$. A maximal error equaling -55% is recorded in case of 40% share of the 5th order harmonic in the current signal and $\varphi_I = 15^\circ$. On the other hand the maximal error in the measurement subset corresponding to $i_5[\%]=20\%$ results in a maximal error decrease of 20% in the same measurement point. Taking the results presented in Fig.7 it can be concluded that maximal error intensity is correlated to the voltage and current harmonic share in the waveforms in a manner that the change in the relative share of a single harmonic would result in a subsequent equal change in the relative error intensity.

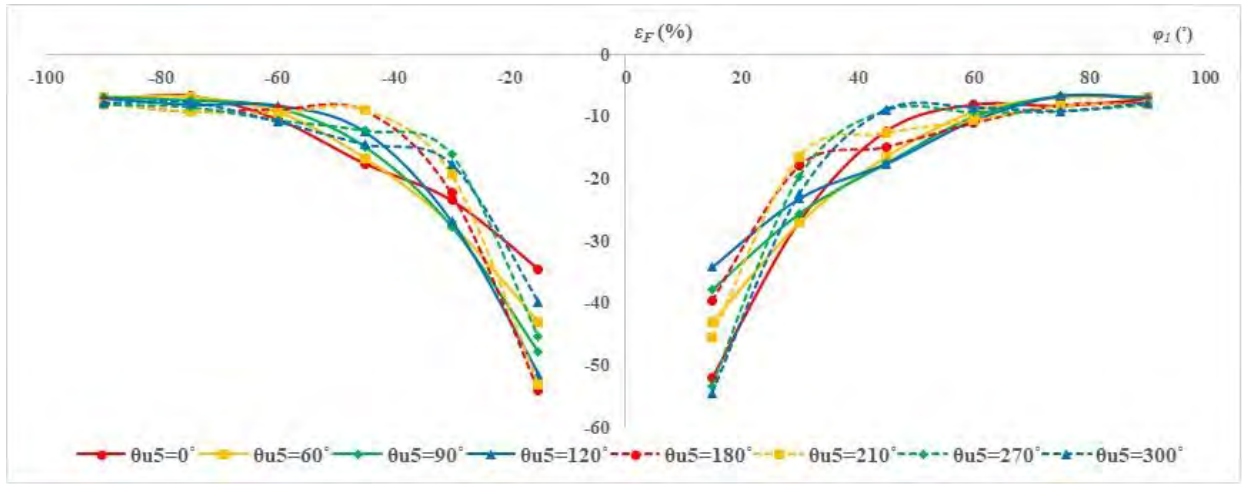


Fig. 6. Error function $\varepsilon_F=f(\varphi_I)$ for different values of θ_{u5} , $u_5[\%]=10\%$, $i_5[\%]=40\%$, $\theta_{i5}=60^\circ$

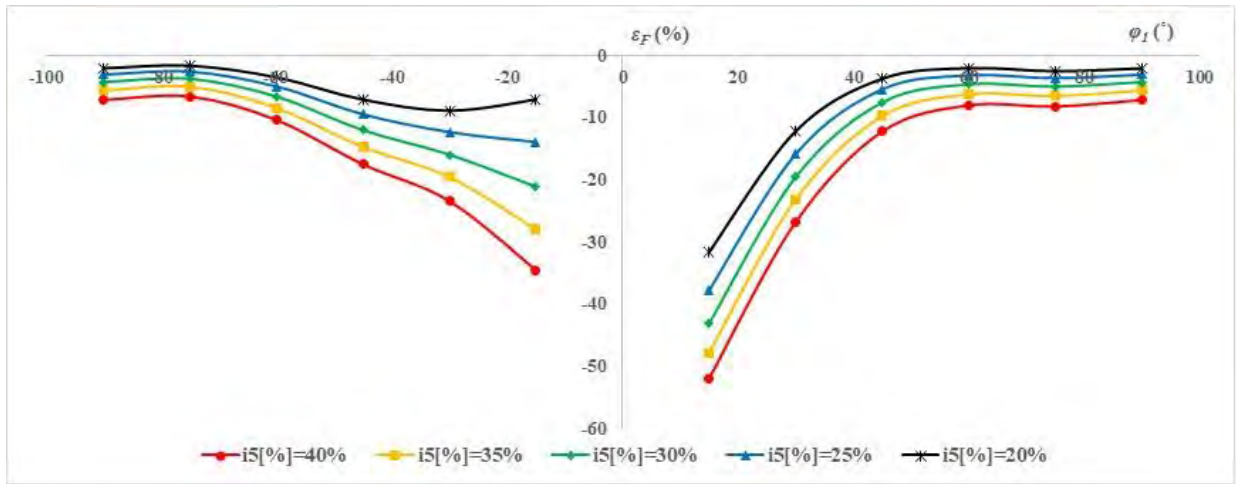


Fig. 7. Error function $\varepsilon_F=f(\varphi_I)$ for different values of $i_5[\%]$, $u_5[\%]=10\%$, $\theta_{u5}=0^\circ$, $\theta_{i5}=60^\circ$

V. CONCLUSION

In the paper an analysis of reactive energy meter's output in relation to different power theories in case of harmonically distorted signals is performed. The analysis is backed up by real measurements performed in an accredited calibration laboratory. Measurement results are presented as relative errors made by the UUT in relation to different harmonic parameters and different power definition existing in practice.

The meter's measurement principle is based on a fixed 90° phase rotation of the measured voltages in order a reactive power (energy) to be obtained. Therefore only the fundamental component of reactive power, Q_I , is recorded. If the UUT is tested with harmonically distorted signals and only the fundamental reactive power generated by the RS is regarded, the measurement errors are within the declared accuracy limits. The errors distribution follows a sine wave pattern in relation to φ_I , because the RS's output is affected by the high order harmonics presence, no matter the fact that only Q_I is taken into account for comparison.

If UUT's output is regarded in relation to the Budeanu power theory the measurement errors are no longer insignificant. The intensity between the measured and the applied value varies with the change of both harmonic amplitudes and phase shifts, as well as phase shifts of fundamental components. As long as φ_I is high enough to provide significant reactive power share in the system errors are low, for the test signals applied are less than $\pm 5\%$. The deviation rises significantly for smaller phase shifts between fundamental components.

The highest errors are recorded when the reactive power is regarded according to the concept proposed by Fryze, taking into account that $Q_F > Q_B$ and $Q_F > Q_I$. UUT's errors are rather constant when φ_I is close to 90° . For lower phase shifts a significant deviation between the measured power and Q_F is recorded, which for the concrete test conditions reaches up to -55% . If error is regarded in relation to $i_5[\%]$, it can be concluded that the change in the harmonic's current share in the system, results in almost equal error change in the measurement points where maximal ε values are obtained.

REFERENCES

- [1] Zobaa, Ahmed F., Ramesh Bansal, and Mario Manana, eds. Power quality: Monitoring, analysis and enhancement. BoD–Books on Demand, 2011.
- [2] Grady, Mack. "Understanding power system harmonics." Austin, TX: University of Texas, 2006.
- [3] Santoso, Surya, et al. Electrical power systems quality. McGraw-Hill Education, 2012.
- [4] European Parliament and of the Council. EU Directive on Measuring Instruments (MID); 2014/32/EU; European Parliament and of the Council: Brussels, Belgium, 2014.
- [5] Cataliotti, Antonio, Valentina Cosentino, Alessandro Lipari, and Salvatore Nuccio. "On the methodologies for the calibration of static electricity meters in the presence of harmonic distortion." In 17th Symposium IMEKO TC 4, 3rd Symposium IMEKO TC 19 and 15th IWADC Work-shop Instrumentation for the ICT Era, pp. 167-172. 2010.
- [6] ENELEC. Electricity Metering Equipment (A.C.)—Part 1: General Requirements, Tests and Test Conditions—Metering Equipment (Class Indexes A, B and C); EN 50470-1:2006+A1:2018; CENELEC: Brussels, Belgium, 2018.
- [7] CENELEC. Electricity Metering Equipment (A.C.)—Part 2: Particular Requirements—Electromechanical Meters for Active Energy (Class Indexes A and B); EN 50470-2:2006+A1:2018; CENELEC: Brussels, Belgium, 2018.
- [8] CENELEC. Electricity Metering Equipment (A.C.)—Part 3: Particular Requirements—Static Meters for Active Energy (Class Indexes A, B and C); EN 50470-3:2006+A1:2018; CENELEC: Brussels, Belgium, 2018.
- [9] "R46-1/2: Active electrical energy meters", OIML, 2012.
- [10] Olencki, Andrzej, and Piotr Mróz. "Testing of energy meters under three-phase determined and random nonsinusoidal conditions." Metrology and Measurement Systems 21, no. 2 (2014): 217-232, 2014.
- [11] Masri, Syafrudin, M. D. Khairunaz, and M. N. Mamat. "Study of electronic energy meter performance under harmonics current condition." In 10th International Conference on Robotics, Vision, Signal Processing and Power Applications, pp. 449-456. Springer, Singapore, 2019.
- [12] Bartolomei, Lorenzo, Diego Cavaliere, Alessandro Mingotti, Lorenzo Peretto, and Roberto Tinarelli. "Testing of electrical energy meters subject to realistic distorted voltages and currents." Energies 13, no. 8 (2020): 2023.
- [13] Morva, György, Vitalii Volokhin, Illia Diahovchenko, and Zsolt Čonka. "Analysis of the impact of nonlinear distortion in voltage and current curves on the errors of electric energy metering devices." In 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 528-533. IEEE, 2017.
- [14] Bartolomei, Lorenzo, Diego Cavaliere, Alessandro Mingotti, Lorenzo Peretto, and Roberto Tinarelli. "Testing of electrical energy meters in off-nominal frequency conditions." In 2019 IEEE 10th International Workshop on Applied Measurements for Power Systems (AMPS), pp. 1-6. IEEE, 2019.
- [15] IEC 62053-2: "Electricity metering equipment - Particular requirements - Part 23: Static meters for reactive energy (classes 2 and 3)", June 2020.
- [16] Barbaro, Pietro Vincenzo, Antonio Cataliotti, Valentina Cosentino, and Salvatore Nuccio. "Behaviour of reactive energy meters in polluted power systems." In XVIII IMEKO World Congress, Metrology for a Sustainable Development, Rio de Janeiro, Brazil, vol. 172, no. 2. 2006.
- [17] IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions, IEEE Std. 1459, 2010.
- [18] Berrisford, Andrew J. "The harmonic impact project—IEEE-1459 power definitions trialed in revenue meters." In 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1-5. IEEE, 2018.
- [19] Rangelov, Y. "Overview on Harmonics in the Electrical Power System" In International Scientific Symposium "Electrical Power Engineering 2014", Varna, Bulgaria, 11-13.09.2014.
- [20] CALMET 300 Three Phase Power Calibrator and Power Engineering Apparatus Testing USER'S MANUAL AND EXTENDED SPECIFICATIONS, Calmet Ltd., Poland, 2013-01.
- [21] EN ISO/IEC 17025 "General requirements for the competence of testing and calibration laboratories", Cenelec, Brussels, 2017.
- [22] Demerdziev Kiril, Cundeva-Blajer Marija, Dimcev Vladimir, Sbrinovska Mare, and Kokolanski Zivko. "Improvement of the FEIT Laboratory of Electrical Measurements Best CMC Through Internationally Traceable Calibrations and Inter-Laboratory Comparisons." In XIV International conference ETAI. 2018.
- [23] Demerdziev, Kiril, Marija Cundeva-Blajer, Vladimir Dimchev, Mare Sbrinovska, and Zivko Kokolanski. "Defining an uncertainty budget in electrical power and energy reference standards calibration." In IEEE EUROCON 2019-18th International Conference on Smart Technologies, pp. 1-6. IEEE, 2019.
- [24] Landis+Gyr, E650 Series 3 User Manual, 2012, <https://www.landisgyr.eu/webfoo/wp-content/uploads/2012/09/D000030108-E650-ZMD300xT-Series-3-User-Manual-en-k.pdf>

Virtual Real Time Power Quality Disturbance Classifier Based on Discrete Wavelet Transform and Machine Learning

Petar Vidoevski, Dimitar Taskovski, Zivko Kokolanski
Ss. Cyril and Methodius University of Skopje
Faculty of Electrical Engineering and Information Technologies
Rugjer Boskovic 18, 1000 Skopje, North Macedonia
pvidoevski@gmail.com, dtaskov@feit.ukim.edu.mk, kokolanski@feit.ukim.edu.mk

Abstract—Power quality has risen in interest in recent years due to the integration of renewable energy sources, the usage of power electronics, increasing power consumption etc. Modern technologies like FPGA, Virtual Instrumentation, Machine Learning algorithms enhanced the power quality monitoring possibilities. In this paper a Virtual Instrument is proposed that classifies Power Quality Disturbances in real time. The classification is done using Random Forest (RF) algorithm and the feature extraction was done using Discrete Wavelet Transform (DWT). Additionally, the classifier was tested with a virtual power quality disturbance generator. From the obtained test results the classifier shows good results. The virtual instruments are developed in LabVIEW and the RF algorithm was implemented in Python.

Keywords—Power Quality; DWT; Machine Learning; LabVIEW; Python, Virtual Instrument

I. INTRODUCTION

Power Quality (in this paper power quality is referred to voltage quality) events classification is one of the most important aspects of power quality monitoring. One way to classify power quality events is by measuring the one-cycle voltage RMS. This method is proposed by IEC 61000-4-30. This classification method is limited, because it only covers voltage dip, voltage swell and voltage interruption [1]-[3]. Recent advances in the computing power of the modern computers allow the easy implementation of Machine Learning algorithms and Signal Processing techniques. Also, the transition from traditional to Virtual Instruments with the help of graphical programming languages like LabVIEW helps in developing affordable, easy programmable, scalable and maintainable instruments.

The development of the classifier consists of two steps:

- Features extraction: is the process of extracting features of the signal with the help of some mathematical tool, and then building dataset of those features
- Building machine learning model: multiple Machine Learning algorithms are trained by the

data set and their respective accuracy is tested, so the best model is chosen for the given dataset.

From the various research in the recent years about this subject multiple solutions are proposed for the feature extraction and the Machine Learning algorithm. As a future extraction method mostly used are: Short Time Fourier Transform (STFT), Gabor Transform, S-Transform, and Wavelet Transform (WT). WT in recent years is the most popular signal processing technique for power quality monitoring. On the other hand there are also quite a few Machine Learning algorithms that are used for the classification process: Decision Tree (DT), Random Forest (RF), K-nearest neighbor, Support Vector Machines (SVM), Neural Networks [1], [3]-[7].

The classifier build in this paper is integrated in a Virtual Instrument build in LabVIEW. For a data acquisition device NI myRIO 1900 card is used. It contains Real Time (RT) processor and a Field-programmable gate array (FPGA) chip. Integrating Python with LabVIEW RT is a perfect instrument for usage in Power Quality monitoring. It supports fast development, scalability and maintainability, large sampling rates, easy development of user interface. Also the integration of LabVIEW with Python helps to perform some operations that are challenging in a graphical programming environment like complicated array operations or Machine Learning.

II. FEATURE EXTRACTION

A. Wavelet Transform

WT is a transform that performs time-frequency analysis of a given signal. WT uses a small wave, wavelet as a window function. The most important part of the WT is the so called mother wavelet ψ . From the mother wavelet, daughter wavelets ψ_{ab} are constructed, that are dilated or are translated across the signal. The translation and dilatation of the signal can be done continuously, Continuous Wavelet Transform (CWT) and discretely, Discrete Wavelet Transform (DWT). In the equation (1) a (real and positive number) is the dilatation parameter and b (real number) is the translation parameter. When changing the values of a , the time-frequency resolution is controlled. Bigger value of a , corresponds to lower frequencies and vice versa.

When using DWT, a and b are depended of the a_0 and b_0 parameters.

$$CWT(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{ab}(t) dt \quad (1)$$

$$\psi_{ab} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

$$DWT[m, n] = \frac{1}{\sqrt{a_0^m}} \sum_{k=-\infty}^{\infty} f[k] \psi\left[\frac{k - nb_0 a_0^m}{a_0^m}\right] \quad (3)$$

$$a = a_0^m, a_0 > 1 \quad (4)$$

$$b = nb_0, b_0 > 0 \quad (5)$$

The usual picked values are $a_0=2$ and $b_0=1$ from which Orthogonal Wavelet Transform is achieved. This transform is implemented via multilevel filter bank called Multi Resolution Analysis (MRA). The filter bank contains low-pass and high-pass filters. The output of the low-pass filter continues to be filtered with another level of low-pass and high-pass filters and so on. In this way information about the both low and high frequencies is obtained. The high frequency components are called detailed coefficients cD and the low frequency components are called approximation coefficients cA . When transiting from one to another level the number of samples of the input signal are cut in half. This method is also known as Wavelet Decomposition. The Wavelet Decomposition is shown on Fig.1. Using this method good frequency and time resolution is achieved. The time frequency characteristics of different types of transforms are shown in Fig.2. [3]-[8].

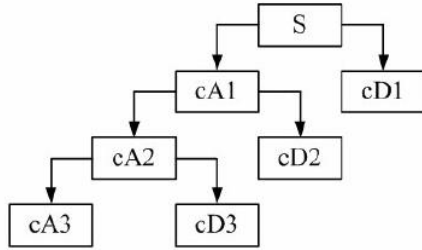


Fig. 1: Wavelet Decomposition

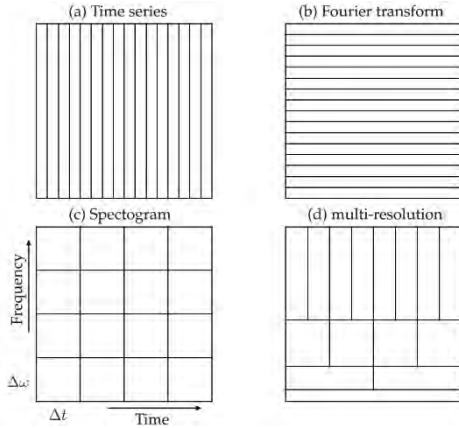


Fig. 2 Time-Frequency resolution for different transforms

B. Feature extraction method

The features are extracted from the well-known mathematical models for power quality disturbances, shown in table 1 [3]-[4]. Around 1000 different disturbances for different parameters of each class are generated (total of 9320 different waveforms), with 20000 samples for each waveform with 100 kHz sampling rate, so it matches the sample number for the data acquisition of the Virtual Instrument that is shown in the next heading. Two datasets are created with and without random Gaussian noise. Random Gaussian noise is added so it can be used for the noise when performing real life measurements.

Table 1: Mathematical models for PQ disturbances

Disturbance	Model	Parameters
Pure Sine	$x(t) = \sin(\omega t)$	$x = 2\pi f$ $f = 50\text{Hz}$
Voltage Sag	$x(t) = [1 - \alpha(u(t-t_1) - u(t-t_2))]\sin(\omega t)$	$0.1 \leq \alpha \leq 0.9$ $T \leq t_1 - t_2 \leq 9T$
Voltage Swell	$x(t) = [1 + \alpha(u(t-t_1) - u(t-t_2))]\sin(\omega t)$	$0.1 \leq \alpha \leq 0.8$ $T \leq t_1 - t_2 \leq 9T$
Interruption	$x(t) = [1 - \alpha(u(t-t_1) - u(t-t_2))]\sin(\omega t)$	$0.9 \leq \alpha \leq 1$ $T \leq t_1 - t_2 \leq 9T$
Flicker	$x(t) = [1 - \alpha(2\pi f t)]\sin(\omega t)$	$0.1 \leq \alpha \leq 0.2$ $5\text{ Hz} \leq f \leq 20\text{ Hz}$
Oscillatory Transient	$x(t) = \sin(\omega t) + \alpha \exp(-(t-t_1)\tau)(u(t-t_1) - u(t-t_2))\sin(2\pi f_d t)$	$0.1 \leq \alpha \leq 0.8$ $8\text{ ms} \leq t \leq 40\text{ ms}$ $0.5T \leq t_1 - t_2 \leq 3T$ $300\text{ Hz} \leq f_d \leq 900\text{ Hz}$
Harmonics	$x(t) = a_1 \sin(\omega t) + a_2 \sin(3\omega t) + a_3 \sin(5\omega t) + a_4 \sin(7\omega t)$	$0.05 \leq a_1, a_2, a_3, a_4 \leq 0.15$ $\sum a_i^2 = 1$
Notch	$x(t) = \sin(\omega t) - \sin(\omega t) \sum_{k=1}^N k[u(t-t_1+0.2n) - u(t-t_2+0.2n)]$	$0.1 \leq k \leq 0.4$ $0 \leq t_1, t_2 \leq 0.5T$ $0.01T \leq t_2 - t_1 \leq 0.05T$
Spike	$x(t) = \sin(\omega t) + \sin(\omega t) \sum_{k=1}^N k[u(t-t_1+0.2n) - u(t-t_2+0.2n)]$	$0.1 \leq k \leq 0.4$ $0 \leq t_1, t_2 \leq 0.5T$ $0.01T \leq t_2 - t_1 \leq 0.05T$

After generating the waveforms DWT is applied to all of them. For the mother wavelet db4 wavelet is chosen with 8 level decomposition [9]. The levels of decomposition are shown in Fig.3.

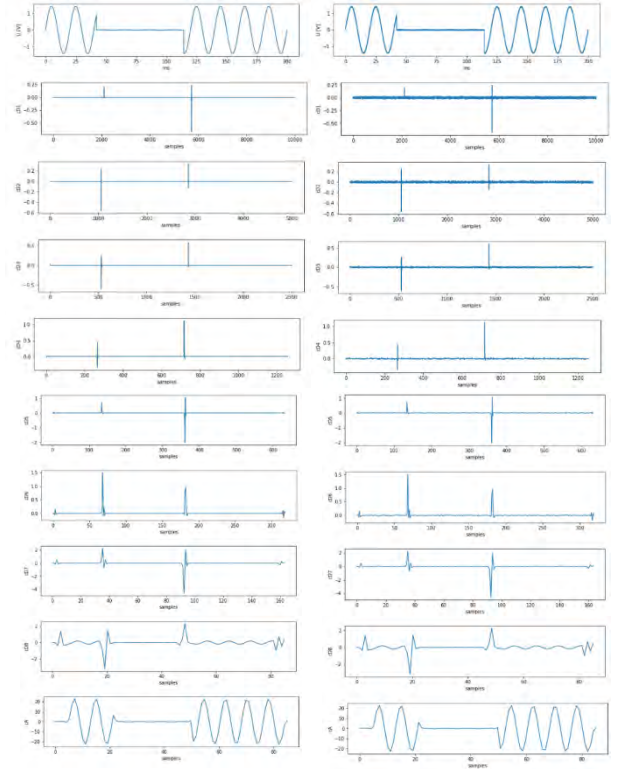


Fig. 3: Wavelet decomposition for interruption with and without noise

The decomposition coefficients that are obtained from the DWT are used to calculate the features. The features that can be extracted are shown in Table 2.

Table 2: DWT features

Energy	$E_l = \sum_{j=1}^N c_{ij} ^2$
Mean	$\mu_i = \frac{1}{N} \sum_{j=1}^N c_{ij}$
Standard Deviation	$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (c_{ij} - \mu_i)^2}$
Skewness	$SK_i = \sqrt{\frac{1}{6N} \sum_{j=1}^N \left(\frac{c_{ij} - \mu_i}{\sigma_i} \right)^3}$
Shannon Entropy	$SE_i = - \sum_{j=1}^N c_{ij}^2 \log(c_{ij}^2)$
RMS	$RMS = \sqrt{\frac{1}{N} \sum_{j=1}^N c_{ij}^2}$
Kurtosis	$KRT_i = \sqrt{\frac{N}{24} \left(\frac{1}{N} \sum_{j=1}^N \left(\frac{c_{ij} - \mu_i}{\sigma_i} \right)^4 - 3 \right)}$
Log-energy Entropy	$LOE_i = \sum_{j=1}^N \log(c_{ij}^2)$
Norm Entropy	$NE_i = \sum_{j=1}^N c_{ij}^p \quad 1 \leq p$

In the researches made in this field, different features are chosen. Some use the energy parameters, some the statistical and in some combination of energy and statistical parameters is used. In this paper Energy, Mean, Shannon Entropy and Log-energy Entropy are proposed. In the equations in Table 2, c_{ij} represent the detailed coefficient of a given liven and the approximation coefficient of the N^{th} level [3]-[7].

III. IMPLEMENTATION OF THE MACHINE LEARNING ALGORITHM

Two Machine Learning algorithms are tested in this paper, RF and DT. The two algorithms are tested with the both the noise and no noise dataset. The size of the train set is 80% and 20 % for the test set of the full dataset. An important part of building the Machine Learning Model is tuning of the hyper parameters. This is important so both high accuracy and fast execution time are achieved, because this classifier is meant to be used in real measurements. The RF is tuned with the number of estimators and the maximum number of features, meanwhile DT is tuned with the maximum depth, minimum number of leafs and maximum number of nodes. Accuracy is chosen for a condition for the hyper parameters tuning [10]. The results of the created models are:

Table 3: Achieved accuracy of the Machine Learning models

	Decision Tree	Random Forest
Without Gaussian Noise	97.47 %	99.195 %
With Gaussian Noise	87.67 %	92.91 %

Visually the accuracy of the models are represented via confusion matrix shown on Fig.3 and Fig.4.

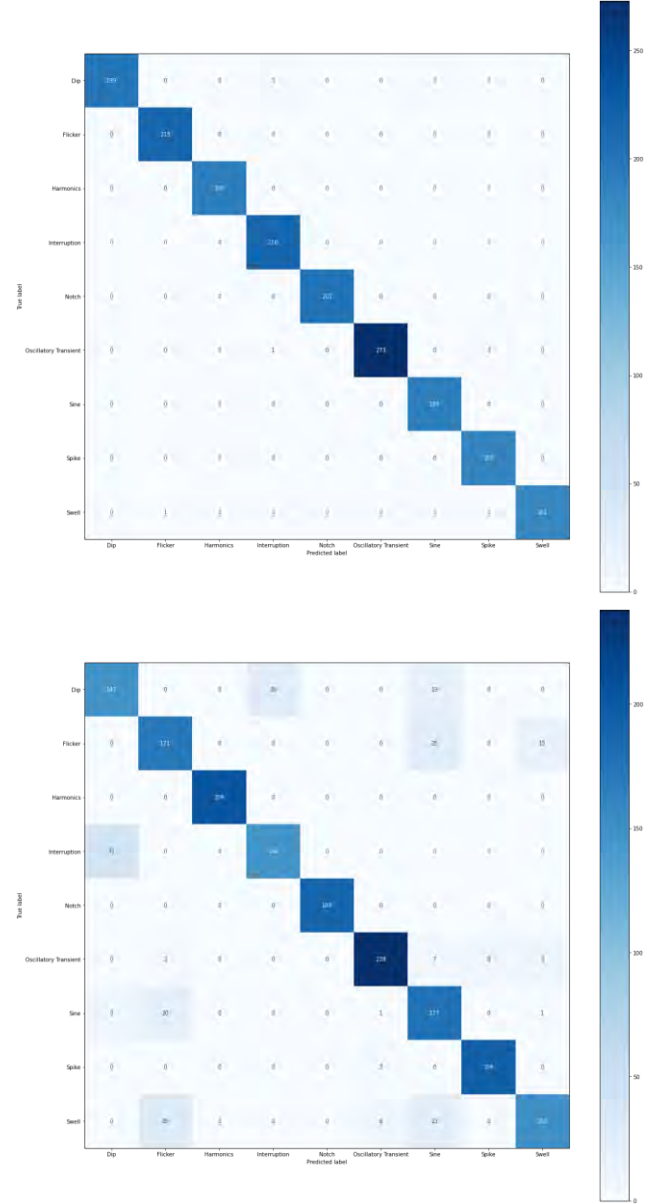


Fig. 4: DT model confusion matrix without and with Gaussian noise

V. TESTING OF THE REAL TIME CLASSIFIER

Important part of Real Time classifier is the subVI that finds the first positive zero crossing. This is crucial for proper operation of the Machine Learning algorithm, because the model is trained for 0° phase shift and 20000 samples. The problem with this is that enough data needs to be collected, so after the algorithm is performed there still need to be 20000 samples left. That is why in Fig. 7 and Fig. 8 22000 samples with sampling rate of 100 kHz are collected, that makes 11 50 Hz periods. After the position first positive zero crossing is found, the subVI takes from there 20000 samples of the scaled measurement data or 10 periods. This subVI is shown on Fig.10.

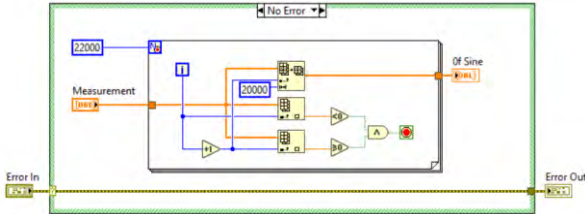


Fig. 10 subVI for finding first positive zero crossing

After the first positive zero crossing is found, all of the calculations and the classification is performed. The classification is performed by a subVI that calls the respective functions via Python Node, shown in Fig.11.

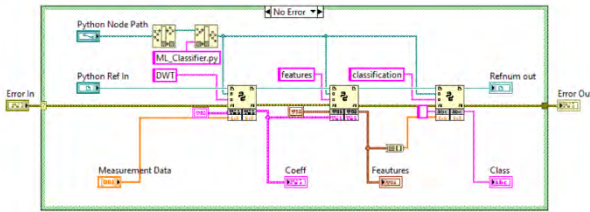


Fig. 11: subVI for calling Python Nodes

When the classification is finished the result updates on the front panel, and it is logged to a .csv file. The block diagram of the main VI is shown in Fig.12.

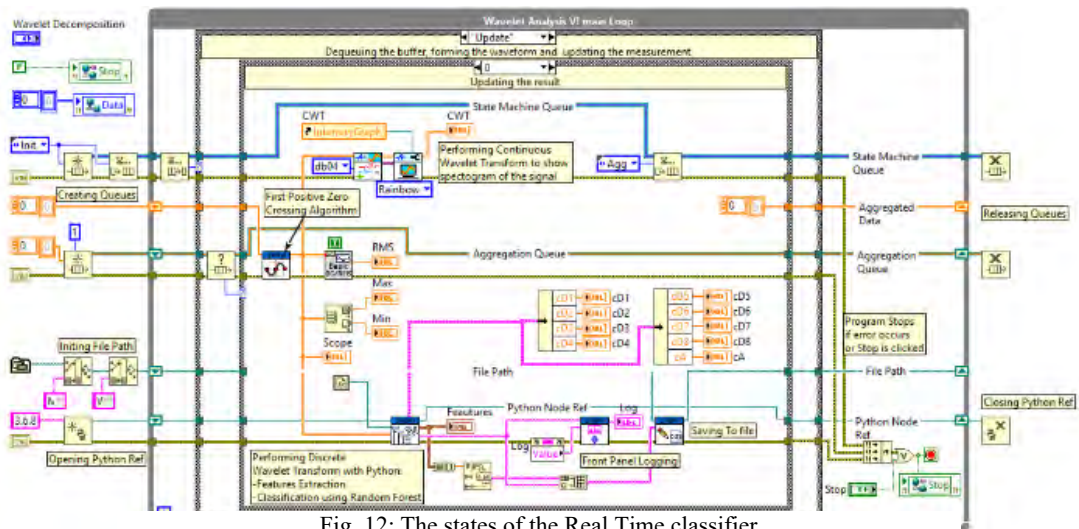


Fig. 12: The states of the Real Time classifier

For the testing of the classifier, PQ disturbance generator for a 200 ms window is created in LabVIEW (Fig.13, Fig.14). The disturbances are generated with the help of the mathematical models in Table 1, with a possibility to freely adjust the parameters. In this VI Python-LabVIEW integration is also used. Python is used for the generation of the disturbances, because it is easier to perform mathematical and array operations textually, also it helps in avoiding cluttered block diagram. The program architecture for this VI is producer-consumer with state machines. There are two consumer loops (so low coupling and high cohesion is achieved) [14]-[15]. One loop is used for generation of the signal, and one to write the signal to the DaQ card, in this case NI USB-6128. The analog output is done via Functional Global Variable (FGV). The analog output uses continuous generation, with 100 kHz and 22000 samples, so it can count for possible need for a zero crossing of the classifier. The generated signal is sent to the DaQ consumer via a queue. Because the program has 3 parallel loops, error handling algorithm is performed via User Event, so if error occurs in any of the consumer loops, the whole program is stopped.

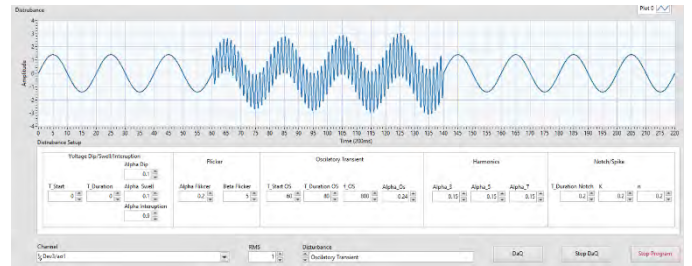


Fig. 13: Front panel of the Virtual Power Quality Disturbance generator

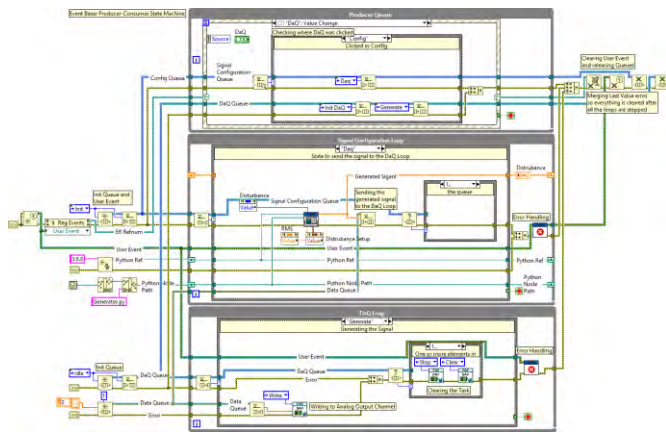


Fig. 14: Block diagram of the Power Quality Disturbance Generator

In the following figures couple of classification results are shown.

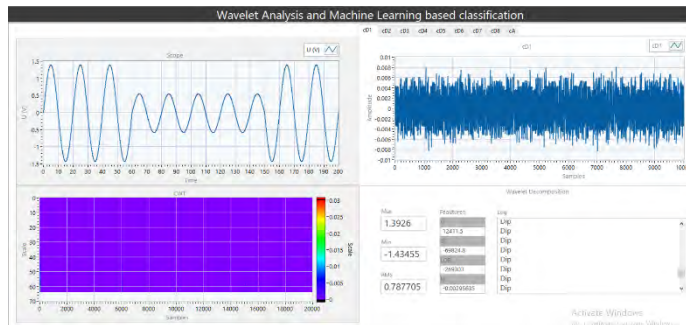


Fig. 15: Classification of a voltage dip

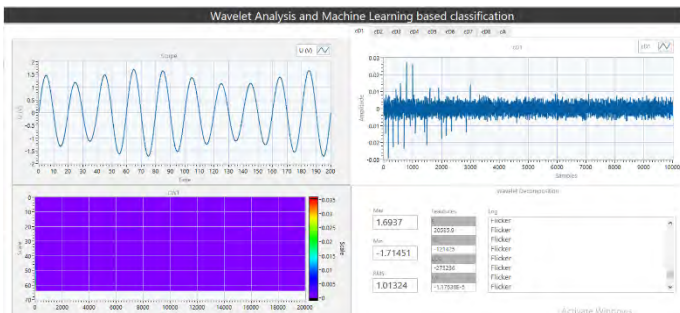


Fig. 16: Classification of a flicker

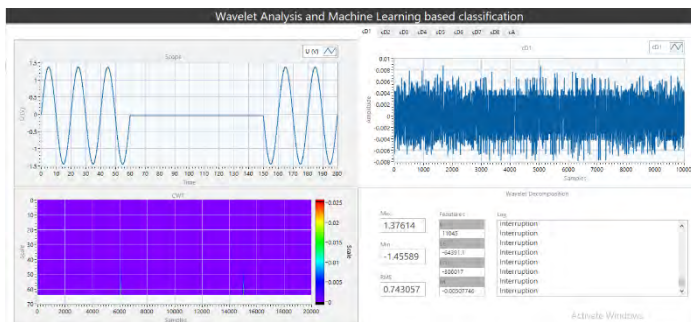


Fig. 17: Classification of an interruption

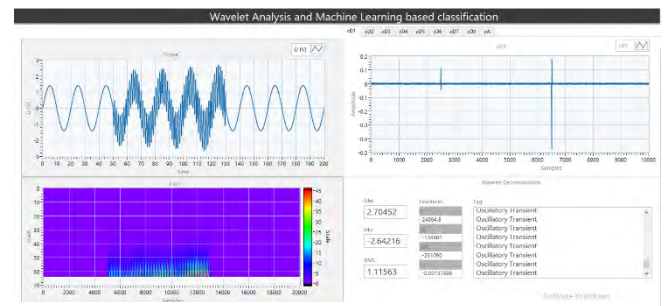


Fig.18: Classification of an Oscillatory Transient

VI. CONCLUSION

The great capabilities of LabVIEW for developing instruments and the easy implementation of Machine Learning algorithms in Python perform excellent together. The classifier shows satisfactory results. The future research will be in direction to increase the accuracy of the model, test other Machine Learning algorithms, and inspect the influence of the noise, but the main goal is to use the classifier in real measurements like in an industrial production line. Because of the great scalability of the virtual classifier can be easily expanded for 3 phase voltage disturbance classification.

REFERENCES

- [1] Bollen M., Gu I.: Signal Processing of Power Quality Disturbances, IEEE Press, Wiley-InterScience, 2006.
- [2] IEC 61000-4-30: Electromagnetic compatibility (EMC) part 4-30: Testing and measurement techniques-Power quality measurement methods, 2003.
- [3] Markovska M., Taškovski D.: Efficient feature extraction and classification of power quality disturbances, Journal of Electrical Engineering and Information Technologies, Vol. 3, No. 1-2, Year: 2018, Faculty of Electrical Engineering and Information Technologies, Skopje.
- [4] Markovska M., Taškovski D.: On the choice of optimal methods for feature extraction and classification of voltage disturbances, Vol. 4, No. 1-2, Year: 2019, Faculty of Electrical Engineering and Information Technologies, Skopje.
- [5] Chen S. Zhu H.Y.: Wavelet Transform for Processing Power Quality Disturbances, EURASIP Journal on Advances in Signal Processing, Volume 2007.
- [6] Deokar S.A., Waghmare L.A.: Integrated DWT-FFT approach for detection and classification of power quality disturbances, 2013, Elsevier M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
- [7] Poisson O., Rioual P., Meunier M.: Detection and Measurement of Power Quality Disturbances Using Wavelet Transform, IEEE TRANSACTIONS ON POWER DELIVERY, VOL. 15, NO. 3, JULY 2000.
- [8] Kutz J.N., Brunton L.S.: Data-Driven Science and Engineering Machine Learning, Dynamical Systems, and Control, Cambridge University Press, 2019.
- [9] Gregory R. Lee, Ralf Gommers, Filip Wasilewski, Kai Wohlfahrt, Aaron O'Leary (2019). PyWavelets: A Python package for wavelet analysis. Journal of Open Source Software, 4(36), 1237.
- [10] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
- [11] National Instruments, User Guide and Specifications NI myRIO-1900, National Instruments, 2013.
- [12] National Instruments, LabVIEW FPGA, National Instruments, 2012.
- [13] National Instruments, LabVIEW FPGA Exercises, National Instruments, 2012.
- [14] National Instruments, "LabVIEW Core 3", 2016
- [15] National Instruments, "Advanced Architectures in LabVIEW", 2012

Improving the Efficiency of Grounding System Analysis Using GPU Parallelization

Bodan Velkovski, Blagoja Markovski,
Vladimir Gjorgievski, Marija Markovska
Ss. Cyril and Methodius University in Skopje,
Faculty of Electrical Engineering and Information
Technologies

Skopje, Macedonia
bodan@feit.ukim.edu.mk, bmarkovski@feit.ukim.edu.mk,
vladjor@feit.ukim.edu.mk, marijam@feit.ukim.edu.mk

Leonid Grcev
Macedonian Academy of Sciences and Arts
Skopje, Macedonia
leonid.grcev@iee.org

Stefan Kalabakov, Elena Merdjanovska
Jožef Stefan Institute,
Ljubljana, Slovenia
stefan.kalabakov@ijs.si, elena.merdjanovska@ijs.si

Abstract— Safety analyses of the effectiveness of large grounding systems are often hampered by the lengthy computation times. Using even the simplest image models, the evaluation of touch and step voltages can require from several minutes to several hours of computations on modern CPUs. Our analysis shows that substantial reduction of computation times can be achieved by utilizing GPU parallelization. In this paper we provide basic steps in the implementation of GPU parallelization on the simplest equipotential model for grounding analysis in homogeneous earth, and we test the effectiveness of this approach in different scenarios.

Keywords—ArrayFire; grounding; parallelization; GPU; step voltages; touch voltages; ground potential rise

I. INTRODUCTION

Grounding systems are an important part of industrial and power plants [1]. They should provide proper equipment operation and personnel protection in normal and fault conditions, including lightning. The effectiveness of grounding systems is often evaluated by computer models [2]-[4], during the design of the plant. The image theory-based models are commonly used in practical engineering analysis of grounding grids due to the simplicity of their implementation and the fact that they can provide accurate results for frequencies of up to a few kilohertz [5]-[6]. For analysis of transients in grounding systems and connected equipment, accurate full-wave models are required [7].

Safety analyses of the effectiveness of large grounding grids are often hampered by the lengthy computation times. While image theory based-models are much simpler and computationally more efficient than full-wave models [8], their calculations can take up to several hours when analyzing touch and step voltages on modern CPUs. For analysis in multilayer soil or in case of transient analysis using the full-wave models, the computation times will be increased severalfold.

Computational efficiency can substantially be improved by utilizing parallelization of the models using a GPU. This is a complex procedure since it requires proper optimization and

This work was supported by the Ss. Cyril and Methodius University in Skopje, project: Electromagnetic Modeling of Transients in Large Systems, NIP.UKIM.20-21.10

organization of the code, and understanding some limitations imposed by the hardware, related to memory optimization, bottlenecks in the data transfer to and from the GPU etc.

In this paper we provide some basic steps in utilizing GPU parallelization of the equipotential image theory-based model, for analysis of grounding systems in homogeneous earth [8]. This approach can be considered as a simplest example for GPU parallelization of computer model for grounding analysis and also as a first step towards GPU parallelization of the full-wave electromagnetic model [7].

We analyze the effectiveness of this approach by comparing the computation times on CPU and GPU, as a function of the number of mathematical operations required for the analysis.

II. DESCRIPTION OF THE MATHEMATICAL MODEL FOR GROUNDING ANALYSIS

Here we follow the mathematical model described in [8]. We consider a grounding grid made from perfectly conducting wires with radius a , energized by fault current I_f . The grounding grid is divided into n segments, and each segment dissipates leakage current I_k to ground, where $k=1..n$. Leakage currents are sources of electric scalar potential in their surrounding, so the potential of the k -th segment can be calculated by superposition as:

$$\varphi_k = r_{k,1}I_1 + r_{k,2}I_2 + \dots + r_{k,n}I_n \quad (1)$$

where $r_{k,k}$ and $r_{i,k}$ are self and mutual resistances of segments, respectively. Each $r_{i,k}$ depends on the spatial position and orientation of the k -th and i -th segment, and the electrical characteristics of their environment. Their evaluation for homogeneous earth is described in section III. Following the assumption of equipotential grounding system, we can write the following expression:

$$\varphi_1 = \varphi_2 = \dots = \varphi_n = U_f \quad (2)$$

where U_f is the potential of the grounding grid with respect to remote neutral earth, when fault current I_f is dissipated.

Considering the entire grounding grid, the relations between the electric potential U_f and the leakage current from each segment can be written in matrix form:

$$[1]_n \cdot U_f = [r] \times [I] \quad (3)$$

and from Eq. (3), follows:

$$[r]^{-1} \times [1]_n \cdot U_f = [I] \quad (4)$$

where $[1]_n$ is n -element column matrix of ones, $[r]$ is $n \times n$ -element square matrix with self and mutual resistances between segments, and $[I]$ is an n -element column matrix with leakage currents from each segment.

The fault current I_f dissipated through grounding system equals the sum of leakage currents from all segments:

$$I_f = [1]_n^T \times [I] \quad (5)$$

and therefore Eq. (4) can be written as:

$$[1]_n^T \times [r]^{-1} \times [1]_n \cdot U_f = [1]_n^T \times [I] = I_f \quad (6)$$

Then the grounding resistance R_g of the grid can be evaluated as:

$$R_g = \frac{U_f}{I_f} = \frac{1}{[1]_n^T \times [r]^{-1} \times [1]_n} \quad (7)$$

while the grid potential U_f and the leakage currents from each segment $[I]$, for a given fault current I_f are calculated as:

$$U_f = R_g I_f = \frac{I_f}{[1]_n^T \times [r]^{-1} \times [1]_n} \quad (8)$$

$$[I] = [r]^{-1} \times [1]_n \cdot U_f = \frac{[r]^{-1} \times [1]_n}{[1]_n^T \times [r]^{-1} \times [1]_n} I_f \quad (9)$$

Once the leakage currents from each segment are evaluated, the potential at any point M can be evaluated as:

$$\varphi_M = r_{M,1} I_1 + r_{M,2} I_2 + \dots + r_{M,n} I_n = [r_M] \times [I] \quad (10)$$

where $r_{M,k}$ can be thought of as mutual resistances between the k -th segment and the point M , described in section III.

III. EVALUATION OF SELF AND MUTUAL RESISTANCES

Following the assumption that leakage current density is uniform over the entire length L_k of the k -th segment, the potential φ_M at any point M in a homogeneous environment with specific resistivity ρ can be calculated as [9]:

$$\varphi_M = \frac{I_k \rho}{L_k 4\pi} \ln \frac{r_1 + r_2 + L_k}{r_1 + r_2 - L_k} \quad (11)$$

The parameters r_1 and r_2 are the distances between segment endpoints and the point M , as illustrated in Fig. 1.

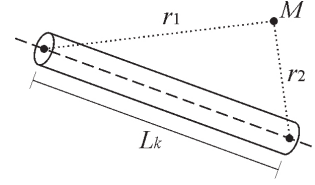


Fig. 1. Geometry parameters for calculating potential at point M

The simplest method of calculating mutual resistance $r_{i,k}$ between two segments in homogeneous medium is to calculate the potential $\varphi_{i,k}$ at the i -th segment's midpoint, due to leakage current I_k from k -th segment. The mutual resistance depends solely on the specific conductivity of the medium and the position between two segments:

$$r_{i,k} = \frac{\varphi_{i,k}}{I_k} = \frac{\rho}{4\pi L_k} \ln \frac{r_1 + r_2 + L_k}{r_1 + r_2 - L_k} \quad (12)$$

If both segments are in conductive half-space, which is a typical assumption for grounding analysis, then the influence of the air-earth interface can be considered by the contribution of reflected image of the source segment, from the interface between upper and lower half-space, (see Fig. 2):

$$r_{i,k} = \frac{\rho}{4\pi L_k} \left(\ln \frac{r_1 + r_2 + L_k}{r_1 + r_2 - L_k} + \ln \frac{r_{1i} + r_{2i} + L_k}{r_{1i} + r_{2i} - L_k} \right) \quad (13)$$

When the self-resistance $r_{k,k}$ is evaluated, the potential $\varphi_{k,k}$ is calculated at the surface the k -th segment, and near its midpoint.

IV. IMPLEMENTATION OF THE GPU PARALLELIZATION

A. Considerations

Over the last two decades GPUs have become widely available in consumer as well as developer computers. Despite this, GPU software development adoption has had a slow rise, which is mainly attributable to the difficulty in programming GPUs.

Currently, there are a number of programming platforms and languages for parallel programming, which cover a range of different approaches. Choosing the right approach for a certain application depends on the nature and complexity of the problem and requires analysis of the pros and cons of each approach.

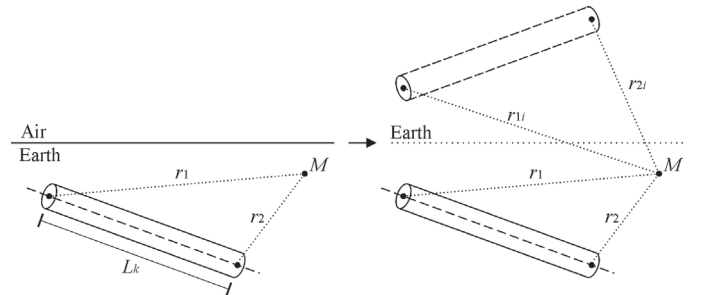


Fig. 2. Equivalence for modeling the influence of the air-earth interface in calculating potentials at point M within earth

At one end, there are platforms that require fairly detailed knowledge of the hardware architecture and which usually use a C-based programming code, that sets out the programming details in explicit parallel fashion. A market leader is CUDA (Computer Unified Device Architecture), a massively parallel computer platform and programming language. OpenCL is another widely adopted open-source framework, which provides wider support for different devices, but has slightly worse performance when compared to CUDA. At the other end, there are platforms which allow the user to make use of their own original non-parallel code, but to include instructions, typically pragmas, in the code which enable it to be compiled in parallel form using an appropriate compiler. Examples are OpenMP and OpenACC. The aim of these platforms is saving the user from spending a large amount of effort in understanding hardware details and corresponding programming code. In between are platforms like ArrayFire and Thrust which are C-based libraries of flexible constructs and subroutines which carry out commonly occurring parallel calculations. These platforms are not mutually exclusive and can be used as part of a CUDA application.

B. ArrayFire Library

For this application, the ArrayFire library approach was chosen, because it offers a back-end that manages memory optimization, cross-platform support for a wide range of devices and does not require writing separate kernels. ArrayFire is a GPU matrix library used for rapid development of general purpose GPU (GPGPU) computing applications within C, C++, Fortran, and Python. ArrayFire contains a simple API and provides full GPU computation capability on CUDA and OpenCL capable devices. It revolves around a single matrix object (array) which can contain floating point values (single- or double-precision), real or complex values, and boolean data. ArrayFire arrays are multidimensional and can be manipulated with arithmetic and functions. Additionally, ArrayFire provides a parallel FOR-loop implementation, “gfor”, which arbitrarily executes many instances of independent routines in a data parallel fashion.

The original code implementing the equipotential model for grounding system analysis described in the previous section was refactored using the ArrayFire library and its matrix-oriented approach. This was done with the goal of accelerating the calculations, which, for large systems, could take up to several hours to complete. It is important to note that the memory allocation and optimization, as well as the optimization of the parallel execution of the operations is handled entirely by the ArrayFire library’s backend.

C. Code Structure

The algorithm that implements the equipotential model for grounding system analysis is straightforward and is shown in Fig 3. The most compute-intensive step in the algorithm is the highlighted step “Calculate Ground Potential Rise”. This owes to the fact that this step consists of executing several mathematical statements a number of times equal to the product of number of segments and number of grid points. This, as can be seen in the following section, can amount to a

very large number of operations. The performance benchmark is addressed only to this section of the code.

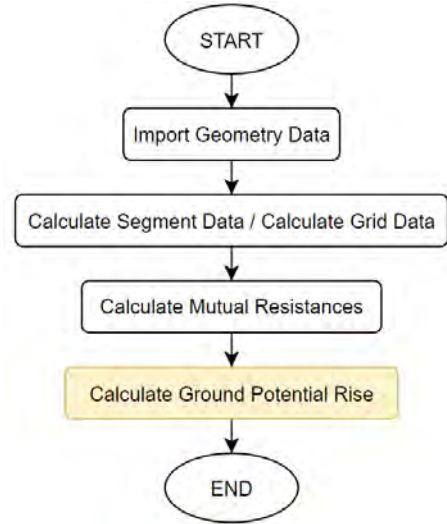


Fig. 3. Equipotential model calculation algorithm

V. PERFORMANCE BENCHMARK

In order to assess the performance of the parallelized GPU code, tests were carried out for grounding analysis of systems containing different numbers of segments as well as grids containing different numbers of points on a predefined geometry. Tests were run for every combination of segments and grid points presented in Table 1. The values in the column “Grid Points” correspond to the values in the column “Grid Step” for the analyzed geometry. The term “Grid Points” is for regularly distributed points on the earth surface, where surface potentials for estimating step and touch voltages have been evaluated.

A. Used Hardware

The benchmarking of the parallelized code was carried out by direct comparison between the execution times of the original code run on a commercial grade CPU and the parallelized code run on a GPU.

TABLE I. NUMBER OF SEGMENTS AND NUMBER OF GRID POINTS USED IN THE PERFORMANCE BENCHMARK

Segments	Grid	
	Grid Step	Grid Points
1,000	2 m	28,860
2,500	1 m	115,440
5,000	0.5 m	461,136
7,500	0.25 m	1,841,819
10,000		

The CPU that was used is an Intel Core i5-8265U CPU with a base frequency of 1.6 GHz. The used GPU is an

NVIDIA GeForce MX150, with a base clock speed of 1469 MHz and 384 CUDA cores used for parallel processing. Both the used CPU and GPU are commercially available for personal computers and are classified as mid-range units.

B. Obtained Results

For the presented combination of number of segments and grid points, the execution times of the CPU and the parallel GPU codes are shown graphically in Fig. 4 and Fig. 5 respectively. Table II shows the factors by which calculation times are decreased for each combination. Furthermore, Fig. 6 shows a comparison between execution times for the CPU and GPU codes on a single case of a grounding system containing 10,000 segments as a function of the grid points density.

It can easily be seen from the obtained benchmarking results that the GPU code outperforms the CPU code by a large margin. It can be concluded that the parallelized GPU implementation of the equipotential model for grounding system analysis generally provides for a 35-fold decrease in calculation times on the given hardware. Also, it can be observed that the improvement of calculation speeds increases for larger numbers of segments.

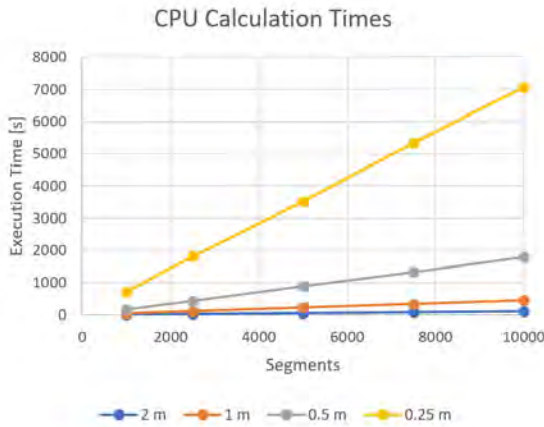


Fig. 4. Execution times of the CPU code

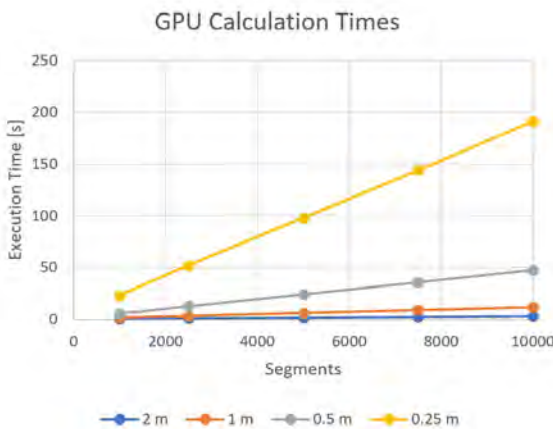


Fig. 5. Execution times of the parallelized GPU code

TABLE II. SPEED INCREASE FOR DIFFERENT NUMBER OF SEGMENTS AND GRID POINTS

Grid Step Segments	2 m	1 m	0.5 m	0.25 m
1,000	27.7	32.0	31.3	31.5
2,500	36.1	35.5	34.8	35.0
5,000	36.4	36.5	37.2	36.1
7,500	36.8	37.5	37.6	37.1
10,000	37.0	37.4	38.2	36.9

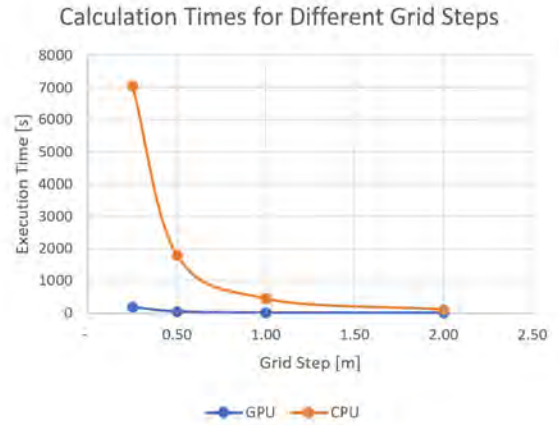


Fig. 6. Execution times of the CPU and GPU codes for 10,000 segments and different grid steps

VI. CONCLUSION

This paper presented the approach chosen for GPU parallelization of an equipotential model for grounding system analysis. The original code developed for CPU was refactored using the ArrayFire library and achieved a 35-fold decrease in computation time on GPU. In practice, calculations that could take up to several hours, now can be conducted in just several minutes using the parallelized code. It is important to state that the used GPU is a mid-range class unit, and using a more powerful GPU would yield substantially shorter calculation times.

This paper presented the first step in the parallelization of a simple computer model for grounding system analysis. Further work in this area will include testing the performance of the parallel code on more powerful GPUs, parallelization of the two-layered earth model for grounding system analysis and eventually parallelization of the full-wave electromagnetic model for grounding analysis in multilayered soil.

REFERENCES

- [1] IEEE Guide for Safety in AC Substation Grounding, IEEE Std. 80-2013, Dec. 2013.
- [2] L. Greev, F. Dawalibi, "An electromagnetic model for transients in grounding systems," IEEE Trans. Power Del., vol. 5, no. 2, pp. 1773-1781, Oct. 1990.
- [3] B. Nekhoul, C. Guerin, P. Labie, G. Meunier, R. Feuillet, X. Brunotte, "A finite element method for calculating the electromagnetic fields

- generated by substation grounding systems," *IEEE Trans. Magn.*, vol. 31, no. 3, pp. 2150–2153, May 1995.
- [4] M. Tsumura, Y. Baba, N. Nagaoka, A. Ametani, "FDTD simulation of a horizontal grounding electrode and modeling of its equivalent circuit," *IEEE Trans. Electromagn. Compat.*, vol. 48, no. 4, pp. 817–825, Nov. 2006.
 - [5] V. Arnautovski-Toseva and L. Grcev, "Image and exact models of a vertical wire penetrating a two-layered earth," *IEEE Trans. Electromagn. Compat.*, vol. 53, no. 4, pp. 968–976, Nov. 2011.
 - [6] V. Arnautovski-Toseva and L. Grcev, "On the image model of a buried horizontal wire," *IEEE Trans. Electromagn. Compat.*, vol. 58, no. 1, pp. 278–286, Feb. 2016.
 - [7] B. Markovski, L. Grcev, V. Arnautovski-Toseva, "Fast and Accurate Transient Analysis of Large Grounding Systems in Multilayer Soil," *IEEE Transactions on Power Delivery*, vol. 36, no. 2, pp. 598–606, Apr. 2021.
 - [8] R. K. Ackovski, *Groundings and Grounding Systems in Power Networks*, 2-nd ed., Ss. Cyril and Methodius University in Skopje, Skopje, 2008.
 - [9] J. Nahman, V. Mijailovic, *Odbrana poglavlja iz visokonaponskih postrojenja*, Akademska misao, Beograd, 2002.



ETAI 7: CLOUD AND IOT TECHNOLOGIES

Technological, Regulatory and Business Aspects of LPWAN Implementation in IoT

Atanas Godzoski

Toni Janevski

Aleksandar Risteski

Ss. Cyril and Methodius University, Faculty of Electrical Engineering and Information Technologies
Karpos 2 bb, 1000 Skopje, Republic of Macedonia
atanas.godzoski@gmail.com

Abstract — The Internet of Things (IoT) concept consists of a large number of connected devices placed on a large area and the most suitable communication for this type of implementations are the Low Power Wide Area Networks (LPWANs) which are characterized with low power consumption and high coverage range.

These features of LPWANs are obtained by using digital modulation techniques that produce a sufficiently strong signal to transmit long distances.

For efficient use of the radio spectrum, which from the aspect of telecommunications is not an infinite resource, several regulations have been adopted that define the work of LPWANs.

IoT has an increasing role in everyday life (business, medicine, education, entertainment, etc.) and it is necessary to adjust the characteristics of LPWANs (range, transmission speed, security, number of nodes) in order for IoT applications to be appropriate for the field in which they are implemented.

Keywords — ICT, Internet of Things, IoT, ISM, LoRa, LPWAN, NB-IoT, Sigfox.

I. INTRODUCTION

THE Information and Communication Technology (ICT) rapid development was followed by emerging the new technology concepts and applications in all aspects of human life - business, entertainment, medicine, education and more. One of the new concepts that is more and more used in all life areas is the IoT which in general can describe as a large number of interconnected devices (sensors, actuators, electronic devices etc.) which share data with each other and other devices, execute commands etc, [1-2].

A basic feature of IoT is the connection between the devices (and here when we say connection we usually mean wireless connection) and depending on the specific application, the frequency bands within which the IoT devices operate range from 13.56MHz for contactless payment devices up to 2.4GHz and more for apps that require video streaming and other large data exchanging.

To achieve these characteristics of LPWAN and depending on the specific application, several different types of modulations are used, between which the most commonly used are Narrow Band Modulation and Spread Spectrum Modulation, described in section II.

Although theoretically the radio spectrum is an unlimited resource, in terms of wireless telecommunications and wireless signal transmission, the spectrum has several limiting factors: the lowest frequency at which signals can

differ from noise, high frequencies that are dangerous to human health, the signal loss over distance (free space loss) etc. Therefore, from the aspect of telecommunications, the radio spectrum is a limited resource and that is why the countries and international regulators have adopted several rules and restrictions in order to use it with the maximum efficient, which are described in section III.

Modern technology, which includes the IoT concept too, has a crucial use in all aspects of life and its application facilitates and promotes activities in all areas where it is used in terms of efficiency, information processing, economy, etc. Section IV presents the main areas where IoT can be or is applied, with specific application implementations, but also the required features of LPWAN in relation to the area specifics in which the IoT application would be applied, possible shortcomings and necessary improvements and adjustments.

II. LPWANs MODULATIONS AND TECHNICAL OVERVIEW

A. LPWANs features overview

The shortest description of the IoT concept is that it is a network of a large number of connected smart devices (sensors, actuators, Systems On a Chip, etc.) that bi-directionally exchange information with higher-level software apps that apps return feedback commands, transmit information for further processing, perform statistical analysis, make decisions, and so on.

The frequency bands within which IoT devices operate range from 13.56MHz for contactless payment devices to 2.4GHz and more for applications that require video streaming and large data exchanging. In which frequency range the IoT communication will be placed depends on the specific application requirements and the devices location and density. An urban area (with many obstacles, concrete structures, etc.) has less effect on the signal operating at low frequency and on the other hand, communication in sub-GHz bands provides both signal security and stability.

The most suitable medium for connecting IoT devices are the LPWANs due to their low power consumption, large coverage and low costs operation and maintenance.

These networks operate in the licensed and unlicensed ISM (Industrial, Scientific and Medicine) SubGHz frequency bands (e.g., 868MHz in Europe, and 915MHz in the U.S.) and the communication range reaches up to 15km in rural areas and up to 5km in urban areas.

Such a large reach and coverage of LPWAN is possible with a new approach to the design of the physical level of communication that provides high sensitivity (e.g., -

130dBm) with the limitation - the low transmission speed (from a few hundred to a few thousand bps).

B. LPWAN modulation techniques

i) Modulation techniques in telecommunications

In telecommunications, the process in which the carrier frequency (carrier) is combined with another signal containing the data to be sent over the channel is called modulation. Each signal has three characteristics that can be modulated: frequency, amplitude and phase. The frequency of a wave defines how often it repeats, the amplitude gives us the strength or power of the waveform, and the phase gives us the state of the waveform with respect to time in a given cycle. The final waveform of the modulated signal is formed by the modulation of the carrier and the data signal.

The three basic types of modulation are:

Frequency modulation: Frequency modulation is used in FM radio, radar, telemetry, music synthesis, etc. Here the signal to be sent is integrated into the carrier by changing its frequency.

Amplitude modulation: AM radio is a common example of using amplitude modulation, by which the data signal is integrated by changing the amplitude of the signal carrier.

Digital modulation: The previous two types of modulation are used for analog signals and are not suitable for the modulation of digital signals represented by a series of 1's and 0's. Amplitude shift keying (ASK), Phase shift keying (PSK) and Frequency shift keying (FSK) are used to modulate digital signals.

ii) LPWAN modulation techniques

The LPWAN technologies have a range of tens of kilometers (rural areas) to a few kilometers (urban areas) and most of the them slow down their modulation rates in order to put more energy in each transmitted bit or symbol so that the receivers are able to decode even the weakest signals without any errors. In general, the receiver sensitivity for LPWAN technologies is around -150dBm.

The two most used modulation techniques in LPWAN are:

Narrow Band Modulation: The signals are encoded at a very low bandwidth (25kHz or less), so the overall spectrum can be used by multiple links. With this modulation there is minimum effect of the noise and it becomes easier for the receivers to decode the receiving signal. This kind of modulation techniques use WEIGHTLESS-P and NB-IoT.

Another technologies like SigFox use ultra narrow band (UNB) which more squeezes the signals (in bandwidths of 100Hz), which further reduces the noise effect but with disadvantage of the more reducing the signal data rate.

Spread Spectrum Modulation: The narrow band signal is spread over a much wider frequency, closer to the noise level, and since the power of transmitted signal is very close to the noise power level, it becomes less susceptible to external interference and much secure.

But, thus the signal and noise levels are very close, with this modulation technique there is a need for powerful receivers that can receive and decode the low energy, close to the noise signals.

The modulation technique is used in networks which require a higher degree of interference robustness and Chirp Spread Spectrum (CSS) and Direct sequence Spread Spectrum (DSSS) as derivatives of this modulation are used by LoRa, [3], and INGENU-RPMA protocols respectively.

The Fig. 1 shows the frequency bands and distances for the most widely used protocols in wireless communications, and it can be seen that LPWANs operate at low frequency bands and can cover a wide range of distances, [4].

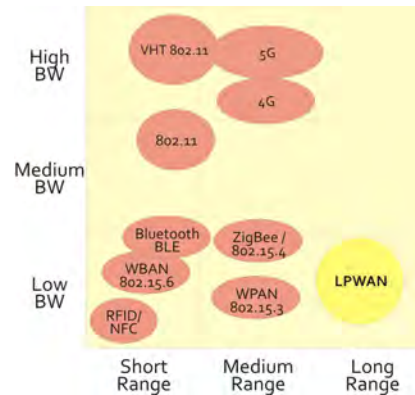


Figure 1. LPWANs frequency bands and range

C. Sigfox, LoRa and NB-IoT comparison

In the practical implementation of IoT systems, depending on the specific implementation (urban / rural environment, number of nodes, number and size of messages during the day, etc.), several types of LPWANs are used, [3-5].

Table 1 presents a comparison of the characteristics of the some conventional and some LPWAN networks.

TABLE 1
CONVENTIONAL & LP NETWORKS CHARACTERISTICS COMAPRISON

	SIGFOX	NB-IoT	LoRa WAN	WiFi	ZigBee
Standard	Sigfox	3GPP	LoRa Alliance	IEEE 802.11	IEEE 802.15.4
Modulat.	BPSK	QPSK	CSS	DSSS, OFDM	DSSS, QPSK
Freq.	433 MHz, 868 MHz, 915 MHz	Licensed (LTE)	433 MHz, 868 MHz, 915 MHz	2.4 GHz, 5 GHz	868MHz, 2.4 GHz
Coverage	10 - 40 km	2 - 20 km	1 - 10 km	10-100 m	10-100 m
BW	100 Hz	200 kHz	125 kHz, 250 kHz	20/40/80/160 MHz	2 MHz
TX Limit	140 packets / day	Unlim.	Duty Cycle Lim.	Unlim.	Unlim.
Max Data Rate	100 bps	200 kbps	50 kbps	250kbps @ 2.4 GHz	250kbps @ 2.4 GHz
Private Depl.	No	No	Yes	Yes	Yes
Energy Consum.	Low	Low	Low	High	Low
Security	Low	High	High	Low-High	High

i) Sigfox technical overview

Sigfox is an LPWAN network and end-to-end IoT connectivity solution based on its patented technologies. Sigfox network consists of base stations equipped with cognitive SDR connected to supporting servers via IP. The network nodes are connected to base stations using binary phase-shift keying (BPSK) modulation in an ultra-narrow band (100 Hz) subGHz ISM band. By using the UNB, Sigfox is characterized with very low power consumption, high receiver sensitivity, and low-cost antenna design at the expense of a throughput of only 100 bps and in practical implementations the number of uplink messages is limited to 140 msg/day and the max payload length is 12 bytes.

Without the acknowledgments support, the uplink communication reliability is ensured using time and frequency diversity and transmission duplication. Each end-device message is transmitted multiple times (three by default) over different frequency channels. As the base stations can receive messages simultaneously over all channels, the end device can randomly choose a frequency channel to transmit their messages. This simplifies the end device design and reduces its cost.

ii) LoRa technical overview

LoRa is a physical layer technology that modulates the signals in subGHz ISM band using a proprietary spread spectrum technique (CSS - Chirp Spread Spectrum).

LoRa and other technologies operating in a narrow frequency band must be extremely energy efficient, and to maintain this, LoRaWAN uses a simple channel management strategy, so that the end nodes and the network itself can be energy efficient. LoRaWAN uses pure Aloha MAC protocol with additional confirmation mechanism.

The LoRaWAN MAC level defines one mandatory and two optional classes of operation to cover all possible working scenarios, which are presented in Fig. 2, [6].



Figure 2. LoRaWAN classes

LoRaWAN has three main channel access strategies, presented as Class A, B and C.

Class A is an obligatory class for all network nodes and the incoming link (link to the node: downlink - DL) is controlled by the node itself. With this approach, each message sent by the node to the gateway activates two incoming windows through which the gateway can send data to the node, otherwise, the gateway has to wait for information from the node to send new data to it.

In Class B, the nodes open DL windows at a specified time interval and only then the node can receive incoming packets.

Class C devices are the least energy efficient because they are always open to receiving traffic.

Table 2 shows the LoRaWAN classes comparison.

TABLE 2
THE LoRaWAN CLASSES COMPARISON

CLASS	BATTERY CONSUMPTION	DESCRIPTION
A	Most energy efficient	Must be supported by all end-nodes. DL after TX
B	Efficient with controlled DL	Slotted communication sync. with beacon frames
C	Least efficient	Devices listen continuously. DL without latency.

iii) NB-IoT technical overview

NB-IoT is a Narrow Band IoT technology which can coexist with GSM and LTE under licensed frequency bands. NB-IoT occupies a frequency band width of 200 KHz, which is the same size with the one resource block in GSM and LTE transmission. Here, the following operation modes are possible:

- Stand-alone operation: a possible scenario is the utilization of GSM freq. bands currently used.
- Guard-band operation: utilizing the unused resource blocks within an LTE carrier's guard band.
- In-band operation: utilizing resource blocks within an LTE carrier.

The NB-IoT communication reduces LTE protocol functionalities to the minimum required for IoT applications. It is optimized to small and infrequent data messages and avoids the features not required for the IoT purpose, e.g., monitor the channel quality, carrier aggregation, and dual connectivity. Therefore, the network nodes are energy and cost efficient. NB-IoT supports up to 100K nodes per cell with the potential for scaling up the capacity by adding more NB-IoT carriers. NB-IoT uses the single-carrier frequency division multiple access (FDMA) in the uplink and orthogonal FDMA (OFDMA) in the downlink, and use the quadrature phase-shift keying modulation (QPSK). The data rates are up to 200 kbps (downlink) and up to 20 kbps (uplink). The maximum payload size for each message is 1600 bytes.

D. LPWANs comparison in terms of IoT

Many factors should be considered when choosing the appropriate LPWAN technology for an IoT application including quality of service, battery life, latency, scalability, payload length, coverage, range, deployment, and cost. In the Fig. 3, Sigfox, LoRa and NB-IoT are compared in terms of these factors and their technical differences, [4].

	Scalability	Range	Coverage	Cost Efficiency	Battery Life	QoS	Payload Length	Latency Performance
Sigfox	Medium	High	High	High	High	Low	Low	Low
LoRa	Medium	Medium	High	High	High	Low	Low	Low
NB-IoT	High	Low	Low	Low	Medium	High	High	High

Figure 3. Sigfox, LoRa and NB-IoT comparison

i) QoS - Quality of Service

Sigfox and LoRa are asynchronous communication protocols which use unlicensed spectrum. They cannot offer the same QoS provided by NB-IoT which use a licensed spectrum and an LTE-based synchronous protocol, which are optimal for QoS at the expense of cost. So, from QoS aspect, NB-IoT is preferred for applications that require guaranteed QoS and applications that do not have this constraint should choose LoRa or Sigfox.

ii) Battery Life & Latency

In Sigfox, LoRa, and NB-IoT, end devices (nodes) are in sleep mode most of the time, which made them highly energy efficient.

The NB-IoT nodes consumes additional energy because of synchronous communication and QoS handling, and its OFDM/FDMA access modes require more peak current, and it can be said that the Lora/Sigfox LPWANs is more energy efficient. LoRa has class C which also handle low bidirectional latency at the expense of increased energy consumption. As conclusion, for applications with small data exchange and that the latency is not a priority, Sigfox and class-A LoRa are the best options. For applications that require low latency, NB-IoT and class-C LoRa are the better choices.

iii) Network coverage & range

The Sigfox major utilization advantage is that an whole city can be covered with single base, LoRa has a lower range (i.e., range <20km) that requires, for example, three base stations to cover an entire city like Skopje. NB-IoT has the lowest range and coverage capabilities (i.e., range <10km). and the deployment of NB-IoT is limited to LTE base stations.

III. LPWANs REGULATIVE

A. Radio spectrum limitations

Although theoretically the radio spectrum is an unlimited resource, in terms of wireless telecommunications and wireless signal transmission there are two main limiting factors:

a) The limitation in the low frequencies, comes from the Shannon-Hartley theorem, shown in equation (1), which relates the amount of information potentially transmitted over a channel.

$$C = B \log_2 \left(1 + \frac{S}{N} \right) \quad (1)$$

where C is the channel traffic capacity (in bps), B is the channel bandwidth (in Hz) and SN is the channel signal to noise ratio. So, to transmit an amount of information in a time period, either we have to use enough bandwidth or we have enough S/N ratio. In this trade-off there is a limited ability to increase the signal to noise ratio and it is easier to select the carrier frequencies that will provide enough bandwidth to allow the required traffic capacity.

b) The frequencies above PHz are known as ionizing radiation and harmful to human life, so they are avoided.

There are also other limitations - free space path loss (FSPL), earth curvature, etc. and the above facts it can be

said that, in terms of telecommunications, the radio spectrum is a limited resource and therefore countries and international regulators have adopted several rules and restrictions in order to use it effectively.

B. Radio spectrum and IoT

IoT is based on the idea of a large number of connected devices and it is generally assumed that wired networks will be part of the infrastructure, but most of the connections will be wireless in order to reduce the cost of infrastructure and that the wireless networks allow mobility, [7].

In such wireless infrastructure, one of the main concerns is the power supply of wireless devices. There are four strategies to power IoT devices (network nodes):

- a) direct connection to the power grid
- b) rechargeable battery power
- c) energy scavenging
- d) life-long batteries.

The chosen energy strategy has a big impact on the communication capabilities of IoT devices, e.g. the more power is available, the more bandwidth the device can use. There are currently several available wireless technologies with different properties which will determine the network behavior: latency, mobility, cost, capacity, power consumption, complexity, reliability, interference immunity, symmetrical uplink and downlink channels, etc.

According to ETSI classification, there are four main groups:

Cellular based: cellular technologies optimized for IoT: LTE-CATM, NB-IoT and E-GSM. All this technologies take advantage of the licensed band.

Dedicated Star Networks: optimized for IoT technologies which network typology is a star and are built over shared spectrum: Sigfox, LoRaWAN, Weightless, Telensa, etc.

Dedicated Mesh Network: mesh networks covering wide area with multi-hops connectivity. These systems are also known as Network-Based SRDs.

Low power versions of LANs & PANs: WiFi, Bluetooth (5.0/4.2/4.1/4.0, Low Energy), WiGig, Ingenu, ZigBee, Thread, Z-wave, EnOcean, etc. They are also operate on ISM freq. bandwidth but the coverage range is much shorter than the dedicated star networks.

The first two subgroups (Cellular and dedicated star networks) are known as LPWAN. These two types of radio techniques share the common use of high sensitivity for increased radio coverage and the low power consumption. The term IoT-LTN (IoT - Low Throughput Network) refers to the Dedicated Star Networks category, which, in addition to the characteristics of LPWAN, adds the properties of shared spectrum, random channelization, star topology and half duplex communication.

C. LPWANs regulation technical constrains

With spectrum management regulators force spectrum maximum utilization and aim to foster economic activity around the use of the radio spectrum. The regulator selects the appropriate incentives that encourage the investors to invest their resources to create new business opportunities and paradigm. For unlicensed bands, [8], the main task of the regulations is, in addition to the spectrum efficient use,

to regulate its use in a way that will allow access to as many services as possible without mutual congestion and interference. Therefore, regulators select some technical parameters of radio transmission and decide some arbitrary thresholds following reasoned criteria. The first decision of the regulator is to select the frequency bands and corresponding applications. Higher frequencies are usually used for high throughput networks, lower frequencies are used for longer range communications (based on transmission power and distance correlation).

For a given band, the regulators usually establish a maximum Tx power limit, which results in determining a maximum coverage radius. Table 3 shows some of the important values affecting the regulations for the higher frequencies of the unlicensed bands in main world markets.

TABLE 3
REGULATION TECHNICAL CONSTRAINS

	US	Europe	China
Frequency Range (MHz)	902-928	863-875.6	779-787
Maximum TX Power (dBm)	30 (>50 ch.) 24 otherwise	27 (869.4-869.6) 14 otherwise	10
Minimum Number of Hopping Channels	50 (BW2 < 250 kHz) 25 otherwise	-	-
Maximum Bandwidth of Hopping Channels (kHz)	500	-	-
Maximum Spurious Emission Threshold. (dBuV/m@3m)	54	66	66
Parameters for Medium Access based on Duty Cycle			
Band Duty Cycle (%)	-	0.1 (863-868) 1 (865-868) 0.1 (868.7-869.2) 10 (869.4-869.6) 1 (870-875.6)	-
Band Duty Cycle Period (s)	-	3600	-
Channel Duty Cycle (%)	2 (BW2 < 250 kHz) 4 (250 kHz < BW2 < 500 kHz)	-	-
Channel Duty Cycle Period (s)	Period (s) 20 (BW2 < 250 kHz) 10 (250 Hz < BW2 < 500 kHz)	-	-

IV. IoT & LPWAN PRACTICAL APPLICATIONS

A. IoT applications overview

With the emergence of the Internet of Things (IoT), massive growth in the network node deployment is

expected and the estimations are that there will be more than 75 billion IoT device connections by 2025, [9]. (connected cars, medical devices, sensors, point-of-sale terminals, wearables, etc).

The exponential growth in IoT is impacting virtually all human life aspects – business, industry, medicine, entertainment, human science and redefining the ways of designing, managing, and maintaining connections, networks, data, cloud apps and all other inter-connected technologies like artificial intelligence, machine learning, data analytics, blockchain etc.

With the great progress in the fields of electronics, communication, computing, sensing, actuating, and battery technologies, now it is possible to design highly power efficient LPWANs with many years of battery life and tens of kilometers of coverage which can be deployed for a broad range of smart and intelligent applications, including environment monitoring, smart cities, smart utilities, agriculture, healthcare, industrial automation, asset tracking, logistics and transportation, and many more, [10].

Some examples of the possible applications of LPWAN and IoT are:

Smart Cities: Smart parking, air and water quality measurement, sound noise level measurement, traffic congestion and traffic light control, trash collection optimization and waste management, utility meters, fire detection, environment management etc

Smart Environment: Water quality, air pollution reduction, climate temperature rise reduction, forest fire, landslide, animal tracking, snow level monitoring, and earthquake early detection

Smart Metering: Smart electricity meters, gas meters, water flow meters, gas pipeline monitoring, and warehouse monitoring

Smart Grid and Energy: Network control, load balancing, remote monitoring and measurement, transformer health monitoring, and windmills/solar power installation monitoring

Retail: Supply chain control, intelligent shopping applications, smart shelves, and smart product management

Automotives and Logistics: Insurance, security and tracking, lease, rental, share car management, quality of shipment conditions, item location, storage incompatibility detection, fleet tracking, smart trains, and mobility as a service

Industrial Automation: M2M applications, robotics, indoor air quality, temperature monitoring, production line monitoring, ozone presence, indoor localization of assets, vehicle auto-diagnosis, machine health monitoring, preventive maintenance, energy management, machine/equipment as a service, and factory as a service

Smart Agriculture and Farming: Temperature, humidity, and alkalinity measurement, wine quality enhancement, smart greenhouses, agricultural automation and robotics, meteorological station network, compost, hydroponics, offspring care, livestock monitoring and tracking, and toxic gas levels

Smart Homes: Energy and water use, temperature, humidity, fire/smoke detection, remote control of appliances, intrusion detection systems, art, goods preservation, and space as a service

eHealth: Patient health and parameters, connected medical environments, healthcare wearable, patients surveillance, ultraviolet radiation monitoring, telemedicine, fall detection, assisted living, medical fridges, sportsmen care, tracking chronic diseases, and tracking mosquito and other such insects' population and growth

B. LPWANs technical characteristics for real life apps

The main features of LPWAN are the size of the area they cover, energy consumption and efficiency, the number of nodes they support, the cost of deployment, operation and maintenance, and depending on the specific application where the priority of each of these features should be applied, is different.

Also, depending on the type of applications, the ability to connect with other types of technology, support for communication protocols and data exchange, etc. is especially important.

Thus, there are specialized LPWANs that can be used for a small number of highly specialized applications, but there are also universal LPWANs that are designed to be applied to a wide range of applications in various fields.

IoT applications in which LPWANs are used can include measurement of environmental parameters with different accuracy and frequency of measurements (eg temperature, humidity, pollution with PM particles, CO, CO₂, NO gases, noise measuring, etc.) and at the same time measuring the health parameters of the population and requesting a mutual correlation of the measured parameters and assessing the impact of environmental pollution on the health of the population.

Another application would be waste management, detection of the waste containers fullness and creation of optimal routes for garbage collection.

With smart parking applications, users always will have information on free parking spaces near the places where they want to park, will be able to reserve these parking spaces and thus reduce the time required for parking and pollution due to longer driving.

If the intended application is to be used in urban, densely populated areas, it is clear that the primary requirement of the deployed LPWANs will be support for a large number of network nodes, at the expense of energy efficiency and coverage, due to the greater availability of energy sources in urban areas and the smaller area to be covered.

Unlike this type of implementation, applications that would measure the parameters of the environment in large areas needs high energy efficiency and coverage, at the expense of transmission capacity and number of nodes, so in this case suitable networks are energy efficient LPWANs with high coverage ability.

In practice, mesh network configurations are common, due to the requirements for a large area coverage, with a large number of nodes and a large transmission capacity.

Table 4 provides the relevance of the major characteristics - coverage, capacity, cost, and low power operation to the different applications. The relative scales for applicability of the characteristics to the application are high (H), medium (M), and low (L).

TABLE 4
RELEVANCE OF LPWANs CHARACTERISTICS TO APPLICATIONS

Applications	Cover.	Capacity	Cost	Low Power
Smart Cities	H	H	H	M
Smart Environment	M	H	H	H
Smart Metering	H	H	H	M
Smart Energy	H	H	M	M
Smart Agriculture and Farming	H	H	M	H
Smart Homes	H	M	L	L
eHealth	H	H	M	H

V. CONCLUSION

The paper describes the features and modulation techniques used in wireless communications, with a special description of the modulation used in LPWAN and on the basis of which these networks have the characteristics suitable for use in IoT - low energy consumption, support for a large nodes number, ability to cover large areas, integrated security, etc.

The frequency bands in which LPWANs operate and the possible problems and limitations in case of unregulated operation are described. Regulations and restrictions by states and state and international agencies and regulatory bodies for effective use of radio spectrum and uninterrupted communication are shown.

Based on the areas in which the IoT concept has a practical use, specific applications are presented, as well as the necessary features of LPWANs adapted to the specific implementations.

REFERENCES

- [1] Toni Janevski, "NGN Architectures, Protocols and Services", Wiley, UK, 2014.
- [2] "Technical and operational aspects of low-power wide-area networks for machine-type communication and the Internet of Things in frequency ranges harmonised for SRD operation", Rep. ITU-R SM.2423-0, ITU, Geneva 2018.
- [3] Brian Ray, "NB-IoT vs. LoRa vs. Sigfox", Link Labs (<https://www.link-labs.com/>), 2018.
- [4] Kais Mekki, Eddy Bajic, Frederic Chaxel, Fernand Meyer, "A comparative study of LPWAN technologies for large-scale IoT deployment", Research Centre for Auto. Control of Nancy, 2018.
- [5] U. Raza, P. Kulkarni, M. Sooriyabandara, "Low Power Wide Area Networks: An Overview", IEEE Commun. Surv. Tutor, 19, 855–873, 2017.
- [6] Ertürk M.A., Aydın M.A., Büyükkakka M.T and Evirgen H, "A Survey on LoRaWAN Architecture, Protocol and Technologies", Future Internet 11 (2019): 216.
- [7] Dhaval Patel, "Low Power Wide Area Networks (LPWAN): Technology Review And Experimental Study on Mobility Effect," South Dakota State University, 2018.
- [8] David Castells-Rufas, Adrià Galin-Pons, Jordi Carrabina, "The Regulation of Unlicensed Sub-GHz bands: Are Stronger Restrictions Required for LPWAN-based IoT Success?", Autonomous University of Barcelona, November 2018.
- [9] <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide>, accessed July 2021.
- [10] Philip J. Basford, Florentin M. J. Bulot, Mihaela Apetroaie-Cristea, Simon J. Cox, Steven J. Ossont, "LoRaWAN for Smart City IoT Deployments: A Long Term Evaluation", Sensors | An Open Access Journal from MDPI, Jan 2020.

Extended Performance Evaluation of the Tendermint Protocol

Jovan Karamachoski

Faculty of Electrical Engineering and Information
Technologies, UKIM
Skopje, Republic of Macedonia

Liljana Gavrilovska

Faculty of Electrical Engineering and Information
Technologies, UKIM
Skopje, Republic of Macedonia

Abstract— Blockchain technology emerges as a promising solution to a wide variety of problems in the financial, industrial, healthcare, and logistic sectors. The Bitcoin and the Ethereum networks are the most often used as Blockchain technologies. However, they are unable to scale with the increased demand and are highly energy inefficient. Besides these two technologies, there are alternatives that are promising the same functionality but with increased scalability. As one of the best alternative Blockchain technology is the Tendermint protocol. There is a lack of papers showing the performance of the Tendermint protocol. This paper presents a performance analysis of the Tendermint protocol in a controlled environment. The achieved transaction throughput makes the Tendermint protocol a promising consensus mechanism for large-scale implementation scenarios. The evaluation also shows the ability of the Tendermint protocol to serve applications that require exchanging larger transactions with acceptable transaction throughput.

Keywords— Tendermint, performance, throughput, transaction size

I. INTRODUCTION

The Blockchain technologies are emerging technologies that are promising high level of synchronization in a distributed network of users. The main enabler of this characteristic is the consensus mechanism that is implemented to coordinate the consensus achievement between unknown users. The consensus mechanism also allows the network of users to maintain a common distributed database with consistent records. The replication of the common database in every participant node makes the system highly redundant, and brings liveness to the records.

A key step in the process of consensus achievement is the hashing function [1]. The hashing function is one-way function that digests a block of input data and gives unique output data, which further propagates into the next steps of the Blockchain structure. The Blockchain structure (see. Fig 1) is new data structure build from blocks of transaction, in-between related with the hashing function output from the previous block. This way the records in the Blockchain structures are immutable, traceable and tamper-proof.

The development and implementation of the Blockchain are mainly related to two well-known Blockchain technologies: Bitcoin [2] and Ethereum [3]. They both are sharing the highest Blockchain community adoption rate. As ones of the

most promising Blockchain technologies, they proved the high level of reliability and security, provided by the Proof-of-Work (PoW) consensus mechanism [2]. The PoW consensus mechanism offers immutability of the records, protection against 51% attack, transactions' traceability, small scalability and a high price paid for the electricity consumption required for consensus maintenance. The drawbacks set a new goal for the researchers, to define a consensus mechanism that will offer high level of security, reliability, increased system scalability and use less energy to maintain the consensus among the users. One of the promising alternatives is the Tendermint protocol [4]. This protocol offers almost instantaneous transaction verification cycle and throughput of thousands transaction per second in an energy efficient consensus mechanism. However it can protect the correctness of the consensus in the system from up to 1/3 malicious nodes. The additional security layer (e.g. user authentication) can further enhance the overall system security.

This paper presents some performance analysis of the Tendermint protocol in a controlled environment. The evaluation is conducted on a Tendermint network built from nodes deployed on three distinct servers in a local area network. The assessment confirms and extends the results found in the literature, proving high transaction throughput associated with the Tendermint protocol and determining limitations of the transaction throughput in a Tendermint network.

II. RELATED WORK

The extensive literature search shows lack of papers that evaluate the real performances of the Tendermint protocol. One of the most insightful and detailed analysis provides [5].

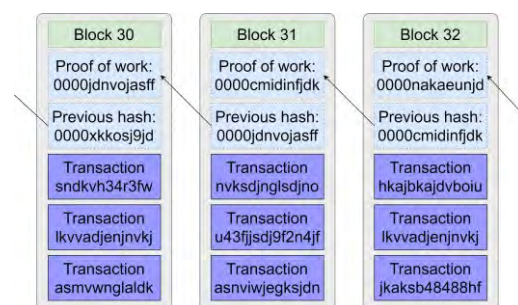


Fig. 1. Blockchain structure

The author of this thesis describes complete Tendermint protocol, and analyses the transaction throughput and latency in different network setups. Moreover, the presented performance results reflect the Tendermint network behavior under crash failures, random network delays and Byzantine failures. According [5], the Tendermint can offer transaction throughput of 30000 transactions per second promising high level of scalability and broader spectrum of services. In this thesis the author proves the ability of the network to cope with Byzantine failures and still deliver acceptable level of services with acceptable transaction latency.

The authors in [6] present the theoretical comparison between the different types of Blockchain technologies, their advantages and disadvantages. The paper presents the results from the performance evaluation of the Ethermint Blockchain technology. Ethermint is a Blockchain technology that creates synergy between the Tendermint consensus mechanism and Ethereum as an overlay. The evaluation in [6] covers performance analysis of the Tendermint consensus mechanism regarding transaction throughput, network topology influence on the transaction throughput and validation time.

The work in [7] presents the analysis conducted over several Blockchain technologies determining the bottlenecks and hotspots in consensus achievement. The analysis points out that the Tendermint consensus mechanism faces its main bottleneck in achievement of high transaction throughput in the computation process for serialization and deserialization of the messages and not in the network bandwidth.

The author in [8] shows the results from the analysis of the problems related to the extensive message exchange in the Tendermint protocol, during the process of gossiping transaction among the users and achieving consensus. The evaluation is directed only toward optimization of the gossiping procedure between the users. The author proposes solutions that can reduce the requirement of network bandwidth and the number of the exchanged messages between the users.

III. TENDERMINT FRAMEWORK AND TENDERMINT PROTOCOL

The Tendermint protocol is based on the Byzantine Fault Tolerance (BFT) algorithm [9]. It applies three-stage process (pre-vote, pre-commit and commit) for consensus achievement. The validation cycle is presented in Fig. 2. Dedicated nodes called validators conduct the process of validation. The transition from one stage to another is done by collection 2/3 of the votes from the validators. The single height or one validation block is finalized in 1 to 3 seconds. In this period the validators are conducting network heavy process of exchanging votes between them, making package storm of votes. Due to the package storm, the voting process can experience bottleneck problems in validators' network bandwidth.

Tendermint protocol has several improvements over the other BFT consensus mechanisms, among which the most important is the implementation of the gossiping protocol, nil voting and locking rules for pre-voting and pre-committing.

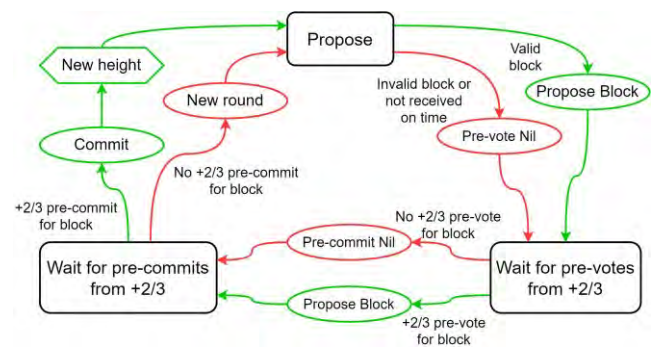


Fig. 2. Tendermint validation cycle

These mechanisms improve the deadlock problems in the validation process and significantly shorten the voting cycle, while still protecting the consensus mechanism.

The Tendermint consensus mechanism is part of the Tendermint framework (see Fig. 3). The main implementation of the Tendermint framework is in Cosmos network [10], which allows inter-operability between different types of Blockchain technologies and consensus mechanisms, thus creating Inter-Blockchain ecosystem. The modularity of the Tendermint framework allows layered development of application by placing distinct applications on top of the Application Blockchain Interface (ABCI) [10]. The ABCI acts as a translation layer between the Tendermint Core layer (or the Tendermint consensus mechanism) and the application. The ability to create Inter-Blockchain ecosystem gives the Tendermint protocol significant scaling potential.

The scalability of a Blockchain system is important parameter in the process of building an application. The degree of scaling can be in several dimensions: transaction throughput, storage capacity, address spacing and power consumption. The last decade of development and deploying Blockchain systems shows the limiting scalability potential for the Bitcoin and Ethereum networks, or more precisely, the Proof-of-Work based Blockchain technologies. In search of alternative consensus mechanisms, the Tendermint protocol gives promising scaling results.

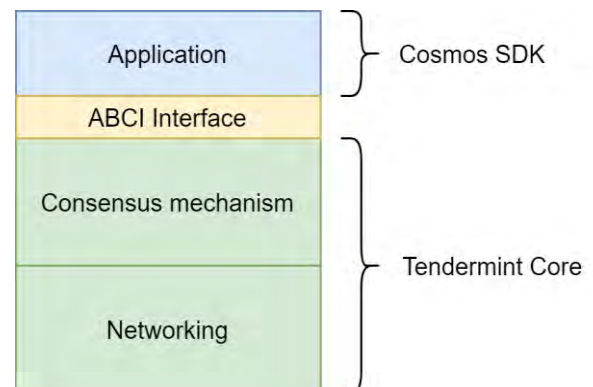


Fig. 3. Tendermint framework

IV. EXTENDED PERFORMANCE EVALUATION

This section presents the extended performance evaluation of the Tendermint consensus mechanism in emulation environment based on [5].

A. Emulation environment

The emulation environment consists of 64 virtual PCs in Hyper-V hypervisor and 2 virtual PCs for load testing and monitoring of the Tendermint network.

The Tendermint nodes have 1 GB of RAM and 10 GB of hard disk storage, implementing off-the-shelf Tendermint software [11] to set the node to participate in the Tendermint network. The Tendermint nodes are starting the *tendermint* service and participate in the Tendermint network depending on the pre-loaded default configuration. In order for the validators to participate in the same network, the configuration has to be synchronized over the parameters like the Blockchain ID and the validators' set. The Tendermint *testnet* configuration generation software generates particular configuration. Moreover, every virtual machine has software called *tcconfig* [12], capable to emulate network delay, packet loss and other network parameters.

The load-generating machine has 2 GB of RAM and 10 GB of hard disk storage, implementing benchmarking software off-the-shelf, *tm-load-test* [13]. The *tm-load-test* is software developed by the Tendermint community for performance evaluation of the Tendermint network. Some of the customizable parameters of the *tm-load-test* which are of our importance are: transaction size, number of transaction per batch of transactions, number of connection to every endpoint in the network, time period for load testing and period of transaction generation.

The monitoring machine has 2 GB of RAM and 20 GB of hard disk storage, and Ubuntu operating system with *InfluxDB* [14] for data storage, *Grafana* [15] for data visualization and *Prometheus* [16] software for data gathering. This set up is activated on every Tendermint node.

The network of virtual PCs, load generating machine and monitoring machine are orchestrated using the *Ansible* playbooks [17] and *Linux Bash* scripts.

B. Tendermint node setup

The conducted emulations are following the descriptions of [5]. The emulations of the Tendermint network are extending the set of evaluation parameters found in [5]. One extension is regarding the block size beyond 32768. The other extension includes the evaluation of broad range of parameters that might optimize the Tendermint network. Some optimization parameters of interest are: transmission and reception rate, number of inbound and outbound peers and transaction size. A list of the manageable parameters, their default values and their short descriptions, are present in Table 1. The default values are part of the off-the-shelf default configuration file that comes pre-configured with the Tendermint software.

The behavior of the Tendermint network and the performance analysis are investigated through monitoring of the transaction throughput for networks of 4, 8, 16, 32 and 64 Tendermint nodes. One complete emulation consists of sprint-sets where a sprint-set is an emulation of a Tendermint network with the predefined number of validators that run for 60 seconds. In a sprint-set the validators are inserting same number of transactions in the network. The cumulative number of transactions inserted by all validators per second in a sprint-set is 2^n where $n = [7, 8, \dots, 19]$.

C. Emulation results

This subsection presents the emulation results. Fig. 4 presents the results from the emulation with the default parameters found in Table 1. We can compare the obtained figures of the default configuration with the large-scale scenario of the [5]. The obtained results are showing maximum transaction throughput of around 70000 transactions/second, which is significantly higher than the results obtained in [5] for large-scale scenario. One reason is the extended range of evaluated transaction rates, but also possible reason is the inequality in the topology. Our emulation implements high degree connected set topology with latency smaller than 1ms, contrary to the topology in [5] where the validator nodes are deployed in 5 continents and 7 data centers, with appropriate network latency and jitter for inter-continental links. Obviously the obtained results are

TABLE I. TABLE I. SIMULATION PARAMETERS, THEIR DEFAULT VALUES AND SHORT DESCRIPTIONS

Parameter	Default value	Parameter description
Transmission rate	5120000 B/s	Rate at which the packets can be sent
Reception rate	5120000 B/s	Rate at which the packets can be received
Number of inbound peers	40	Number of peers to get transaction from
Number of outbound peers	10	Number of peers to send the transaction to
Message package payload	1024B	Maximum size of message package payload
Block proposal timeout	3s	Time interval to collect transactions for the next consensus round
Block proposal delta timeout	500ms	Current time interval overlap with the next round interval, to include transactions that arrived late in the current round
Prevote delta timeout	500ms	Interval overlap of the prevote stage with the precommit stage to include the votes that arrived late
Precommit delta timeout	500ms	Interval overlap of the precommit stage with the commit stage to include the votes that arrived late
Transaction size	250B	The size of the generated transaction
Period of batch transactions	1s	Period of generating batches of transactions in the node
Additional network delay	0ms	The additional network delay inserted artificially to emulate a realistic network conditions
Packet loss	0%	The package loss inserted artificially to emulate a realistic network conditions

comparable to the single datacenter scenario from the [5].

Fig. 5 shows slight improvement of the network performance with increase of the transaction throughput. The configuration of the network differs to the default configuration in the number of inbound and outbound peers. The number of inbound peers is decreased from 40 to 5 and the number of outbound peers is decreased from 10 to 5. Obviously the large number of connected peers will decrease the performance of the network due to network congestion. This in turn decreases the decentralization of the system as a whole.

Fig. 6 shows the emulation results for configuration with increased transmission and reception rate from the default 5120000B/s to 10240000B/s and decreased number of inbound peers from 40 to 5 and decreased number of outbound peers from 10 to 5. The results are showing no improvement compared to the default configuration limited by the network congestion, which in this case of doubled transmission and reception rate congests the network very fast due to emulated environment and shared network

interfaces.

The next emulation results are regarding the parameters of transmission and reception rate equal to 10240000B/s, and increased number of inbound and outbound peers to 65 in order to obtain fully connected set of peers. From Fig. 7 it is obvious that the transactions throughput decreases compared to the default configuration and to the results presented in Fig. 5 and Fig. 6, which additionally shows the limiting factor of the actual network bandwidth in a scenarios when the transmission and reception rate is increased and the number of inbound and outbound peers is increased.

Fig. 5, Fig. 6, Fig. 7 and Fig. 8 show that the transaction throughput gets to a point of saturation when the consensus mechanism obtains maximum throughput. After this point, the system performance declines and lot of transactions in the system are unverified and rejected because the consensus mechanism is overwhelmed with messages.

Next emulation set is related to the transaction size influence on the transaction throughput. The transaction size

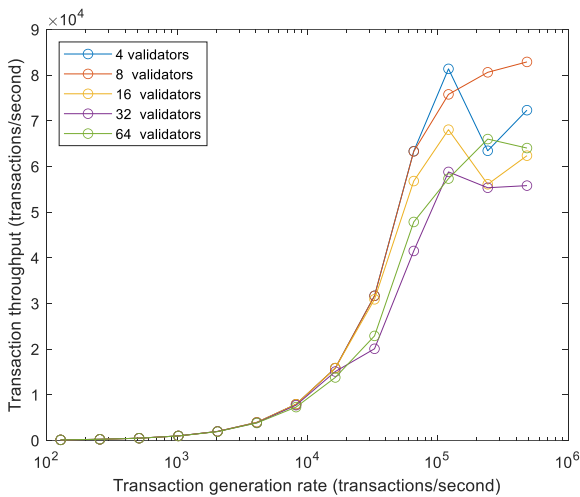


Fig. 4. Default configuration

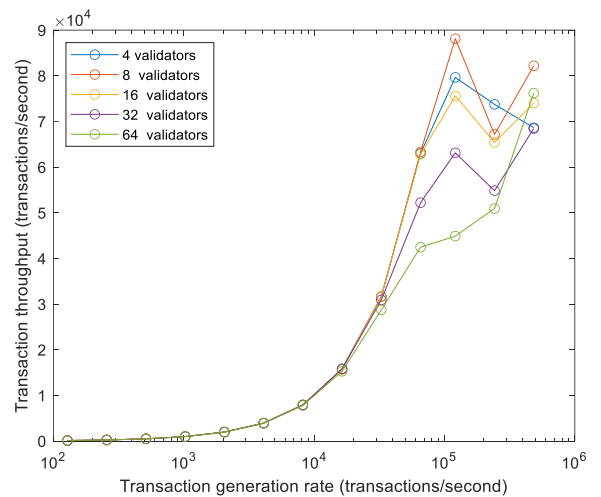


Fig. 6. Transaction throughput for inbound and outbound peers equal to 5, transmission and reception rate equal to 10240000B/s

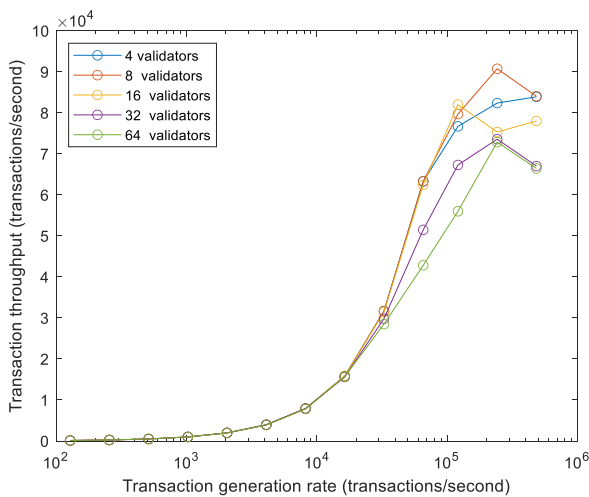


Fig. 5. Transaction throughput for inbound and outbound peers equal to 5

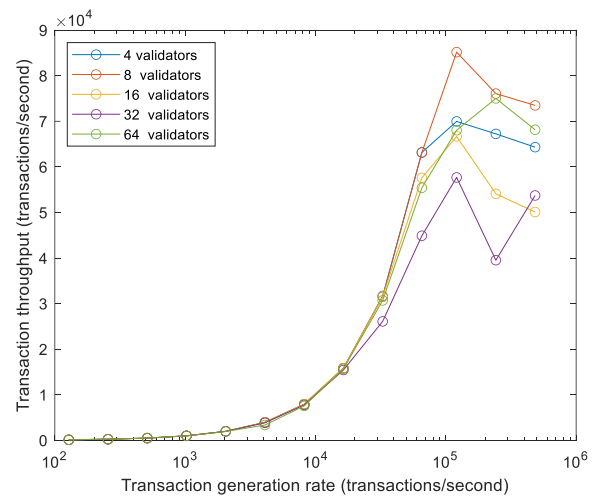


Fig. 7. Transaction throughput for inbound and outbound peers equal to 65, transmission and reception rate equal to 10240000B/s

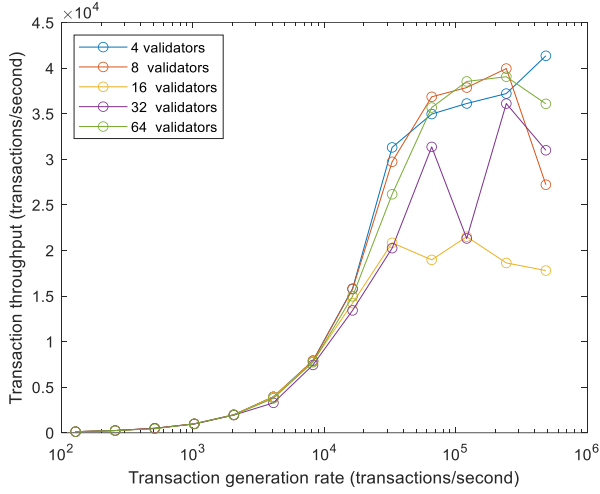


Fig. 8. Transaction throughput for transaction size of 1KB

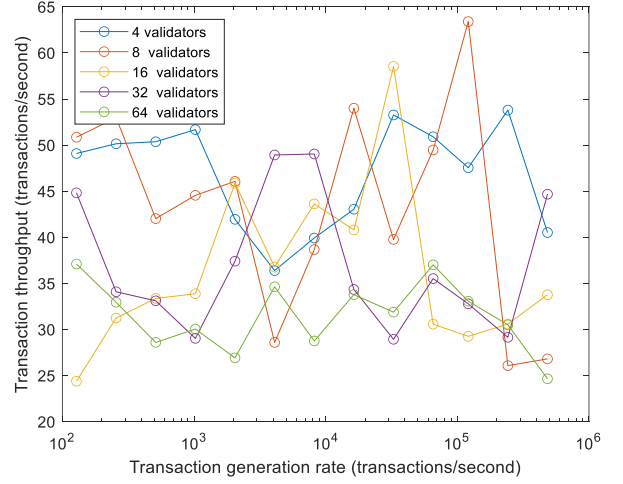


Fig. 10. Transaction throughput for transaction size of 1MB

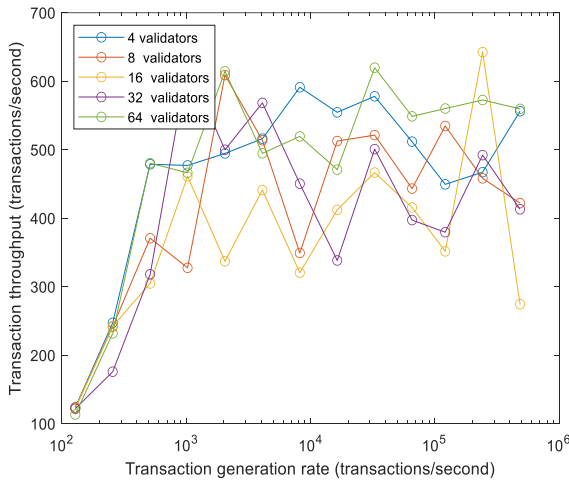


Fig. 9. Transaction throughput for transaction size of 100KB

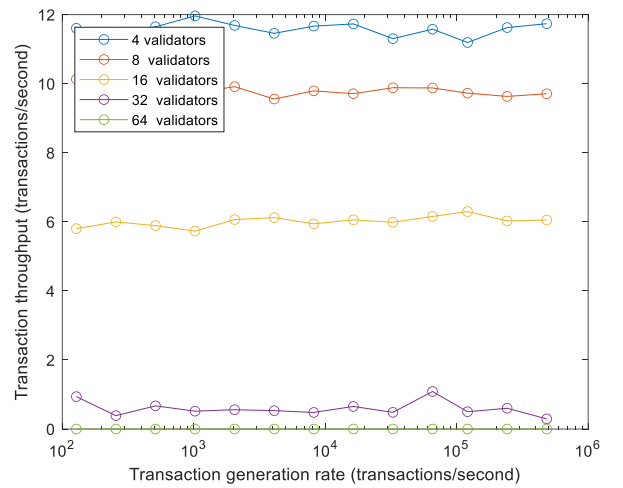


Fig. 11. Transaction throughput for transaction size of 5MB

was increased from the default 250B to 1KB, 100KB, 1MB and 5MB. Comparing the Fig. 8, Fig. 9, Fig. 10 and Fig. 11 it is obvious the decreasing trend of the transaction throughput as the transaction size increases. The big transactions have huge impact in the transaction throughput due to the requirement for the transaction to be disseminated to all the validators in the network and also it need more time for the transaction to be transmitted over the network. From Fig. 11 it is obvious that in a network with large validator set and transactions of 5MB the obtained transaction throughput is 0 transaction per second, because most of the transactions cannot even pass through the network to get to all the validators in the timeout proposal period.

V. CONCLUSION

The number of applications built by using Blockchain technologies is increasing, also the spectrum of implementation areas is widening. This require determination a Blockchain technology that will serve the increased scalability demand due to massive adoption of the applications. The extended evaluation of the Tendermint protocol shows potential for implementation in application

requiring high transaction throughput, which further will allow the application to scale accordingly. The conducted evaluation also proves the ability for the Tendermint protocol to serve applications that require transaction size bigger than the default transaction size. This will allow development of applications beyond the financial applications.

REFERENCES

- [1] US. NIST, "Descriptions of SHA-256, SHA-384 and SHA-512," 2001, Technical report, [Online]. Available: <http://www.iwar.org.uk/comsec/resources/cipher/sha256-384-512.pdf>
- [2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008 [Online]. Available: <https://bitcoin.org/bitcoin.pdf>
- [3] V. Buterin and others, "Ethereum white paper: A next-generation smart contract and decentralized application platform," *Ethereum* [Online]. Available: http://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf, 2014.
- [4] J. Kwon, "Tendermint: Consensus without mining," *Draft v. 0.6, fall*, vol. 1, no. 11, 2014.
- [5] E. Buchman, "Tendermint: Byzantine fault tolerance in the age of blockchains," University of Guelph, 2016 [Online]. Available:

- <https://allquantor.at/blockchainbib/pdf/buchman2016tendermint.pdf>
- [6] O. Dib, K. -L. Brousmiche, A. Durand, E. Thea, and E. B. Hamida, "Consortium blockchains: Overview, applications and challenges," *International Journal On Advances in Telecommunications*, vol. 11, no. 1&2, 2018.
 - [7] S. Alqahtani and M. Demirbas, "Bottlenecks in blockchain consensus protocols," *arXiv preprint arXiv: 2103.04234*, 2021.
 - [8] L. Miletic, "Formal and simulation analysis of data dissemination algorithms in a blockchain network," University of Belgrade, 2018.
 - [9] M. Castro, B. Liskov, and others, "Practical byzantine fault tolerance," in *OSDI*, 1999, vol. 99, no. 1999, pp. 173–186.
 - [10] J. Kwon and E. Buchman, "A Network of Distributed Ledgers," *Cosmos, dated*, pp. 1–41, 2018.
 - [11] "Tendermint GitHub repository." [Online]. Available: <https://github.com/tendermint/tendermint>
 - [12] "tcconfig GitHub repository." [Online]. Available: <https://github.com/thombashi/tcconfig>
 - [13] "tm-load-test GitHub repository." [Online]. Available: <https://github.com/informalsystems/tm-load-test>
 - [14] "InfluxDB website." [Online]. Available: <https://www.influxdata.com/>
 - [15] "Grafana website." [Online]. Available: <https://grafana.com/>
 - [16] "Prometheus website." [Online]. Available: <https://prometheus.io/>
 - [17] "Ansible website." [Online]. Available: <https://www.ansible.com/>

Analysis of Security Mechanisms of Containers in Cloud

Martina Janakieska

Ss. Cyril and Methodius University, Faculty of Electrical
Engineering and information Technologies,
Skopje, N. Macedonia
janakieska.martina@outlook.com

Aleksandar Risteski

Ss. Cyril and Methodius University, Faculty of Electrical
Engineering and information Technologies,
Skopje, N. Macedonia
acerist@feit.ukim.edu.mk

Abstract— As more and more organizations expand their use of tools for container orchestration, and Kubernetes is becoming a popular choice for many companies, it is crucial to understand its architecture and how the components are connected to protect the mutual communication and the applications that are installed on containers. Development speed and agility are key elements in the world of containerization. Increased use of containers leads to inability to use security mechanisms, only at the end of the software lifecycle. Instead of that, more frequent checks, and installation of security mechanisms in more stages of container lifecycle, will bring greater opportunities in detecting software vulnerabilities and shorter time to find a solution for them. This paper presents the security issues in Kubernetes and mechanisms that are used to protect the cluster.

Keywords—Cloud; Container; Docker; Kubernetes; Security

I. INTRODUCTION

A cloud-based services are constantly developing. More and more companies are migrating their services to a remote structure management. Kubernetes and Docker are the most used tools in the world of containerization. A recent study found that 84% of companies run containers in production, and 78% are using Kubernetes in some form [1]. Historically speaking, containers work on virtual machines. Kubernetes is an open-source platform for containers orchestration, automation, deployment, scaling, and management computer applications. Docker, on the other hand, is a software platform that simplifies the process of creating, maintaining, and distributing applications, and all this through virtualization of the operating system on which it is installed [2].

Security plays a very important role in the world of cloud servers and containerization. Most companies do not dare to move so fast to use containers as these are relatively new platforms. Furthermore, traditional security tools, such as firewalls, are not suitable for application in the domain of containers. Additionally, the operating system kernel is shared between containers which makes it difficult to create good insulation or prevention of resource abuse. Kubernetes is growing into a well-adopted container configuration solution, but, yet security for these environments is not always top-notch. A 2020 survey found that only 6% of Kubernetes users did not have a security incident in the last 12 months [1]. The purpose of this paper is to make analysis of the security

problems that containerization domain is faced with. For example, the ability to customize the Kubernetes in different environments, as an advantage, so it is weakness. Kubernetes is designed to be customizable in different scenarios, so it is the duty of users to include certain functionalities to protect their cluster. Accordingly, the engineers who are responsible for the implementation and maintenance of Kubernetes, should know about potential target points of attack and configuration vulnerabilities. Kubernetes security tools should provide digital signatures for a certain level of code trust, visibility, and transparency, not only in the code, but also in the configuration and of course to prevent input or output exchange of the information from/to unsecured services. However, in order to properly secure Kubernetes environment, it is recommended to implement the “shift left” technique [3], right from the first phase of development. This means that security is built into the process and designed into the container and application at an earlier stage of the development cycle. This process is different from previous development models where security mechanisms are implemented at the end of deployment. When you shift left, security is built into every stage of the container lifecycle:

- Develop
- Distribute
- Deploy
- Runtime

The paper is organized as follows: Section II briefly presents the Kubernetes architecture. Section III describes the security issues in Kubernetes and mechanisms that are used to protect the cluster. In Section IV is shown the installation of Kubernetes cluster using the security mechanisms. We conclude the paper in Section V.

II. KUBERNETES ARCHITECTURE

Kubernetes is an open-source system that automatically deploys containers based on the resources that each container requires [4]. Figure 1 shows the basic Kubernetes architecture. The smallest unit that can be deployed in Kubernetes is called a pod. A pod is a logical group of one or more containers that share the same IP address and port range. In Kubernetes, node represents physical or virtual machine on which Kubernetes is installed. These nodes are called worker nodes, where Kubernetes launches containers. Each node has a software that

allows them to work and manage the components. Furthermore, kubelet tool, which is responsible for starting, stopping, and managing individual containers, according to the requirements of the Kubernetes control plane. Kube-proxy is responsible for networking and load balancing. Besides the core elements, there are additional elements which extend the functionality and improve the user experience.

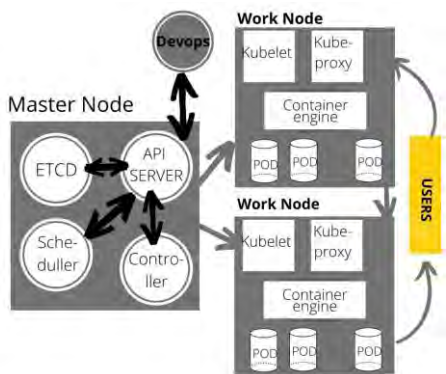


Fig. 1. Kubernetes Architecture

For the cluster to function properly, there must be someone who manages it and stores the information related to its members. That element is called master node and it is part of the Kubernetes control plane. Master node has few key elements:

- Application Programing Interface (API) Server
- Scheduler
- Controller Manager
- ETCD (/etc distributed)
- Cloud Controller

Namespace in Kubernetes plays an important role when it comes to security. It allows us to define virtual clusters by providing scope for the used names. Names of the resources within the namespace must be unique.

III. ANALYSIS OF KUBERNETES SECURITY

Kubernetes provides built-in security mechanisms. Software image for creating the application is not updated or no additional features and functionalities are added to it. Instead, the image is replaced by a completely new image, with a new version. With this we can have control over the version and, also it allows a quick rollback if any vulnerabilities are detected in the code. Besides, Kubernetes security tools should provide digital signatures for a certain level of code trust, visibility, and transparency, and to prevent input or output exchange of the information from/to unsecured services.

API server is the entry point in Kubernetes cluster. To protect and restrict access to the API server, three stages are included, authentication, authorization, and access control. The analysis and implementation of security mechanisms covered in the three phases, for protection of Kubernetes cluster, is carried out in a test environment where Kubernetes cluster is installed, on ubuntu server, version 20.0 LTS, that is booted into a virtual environment, VMWare. The cluster contains one

main node and two work nodes. The hardware configuration depends on needs of the application to be installed, but there are minimal requirements for hardware to be met. That is, at least two CPU units, 2 GB RAM and Internet access. For this case, every node has 4 GB RAM and 20 GB virtual disk space.

Node addresses are assigned automatically via the Dynamic Host Configuration Protocol - DHCP protocol during the installation of the virtual machine. Because for each type of cluster it is recommended to use static addresses, after the node is installed dynamically allocated addresses are replaced with static IPs. Static address are shown in Table 1.

TABLE I. NODES INFORMATION

Node	IP address	Node name
Master node	192.168.44.21	master-node1
Work node 1	192.168.44.31	work-node31
Work node 2	192.168.44.32	work-node32

Once all three nodes are installed and configured with a static IP address, via Secure Shell network protocol, SSH access is assigned to all nodes, which allows them to communicate in a secure way over an unsecured network. SSH server uses the Transmission Control Protocol protocol, on port 22. From the /home directory of the created username Kubernetes-user, we generate a pair of SSH keys. These are a set of cryptographic keys, one private and one public key. The pair of keys can be used to access a remote Linux console via SSH without the use of passwords. Figure 2 shows the process of generating a pair of cryptographic keys. As encryption protocol, we use SHA256, for generating the keys. It is very important to install the kubeadm and kubectl tools. Kubeadm performs the necessary actions to activate the minimum sustainable Kubernetes cluster. On the other hand, Kubernetes command line tool, kubectl, allows executing of the commands in Kubernetes. It can be used for deploying applications, view and manage cluster resources, and view events and logs.

```
kubernetes-user@master-node1:~$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/kubernetes-user/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/kubernetes-user/.ssh/id_rsa.
Your public key has been saved in /home/kubernetes-user/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:KJknw3o0MTzMGwsdLZQGponJkVi35gAjtd/2z2TUNCc kubernetes-user@master-node1
The key's randomart image is:
+---[RSA 2048]---+
|*+0 =+00|
|=* @ 0+|
|+ * % . E .|
| B @ . 0 +|
| # = S . .|
| o B . .|
| . . . 0|
| . =|
| . 0|
+---[SHA256]---+
```

Fig. 2. Process of generating a pair of cryptographic keys

A. Authentication in Kubernetes

The components that are part of Kubernetes, as well as the users who give commands through command line, kubectl, communicate with the API server. API server needs to process the requests and therefore to verify and authenticate those who sends the requests. Several authentication strategies are available in Kubernetes. Depending on deployment size, target users (human users vs. processes) and organizational policies, as a cluster administrator, we can choose one or more strategies in the authentication process like certificates for clients, tokens, proxy servers for authentication or Hypertext Transfer Protocol – HTTP authentication, to authenticate requests coming to the API through the so-called authentication plugins. For bigger security at least two methods for user authentication are recommended. For example, tokens for authentication of service user accounts and at least one of the methods for user authentication [5].

As described in the previous chapter, Kubernetes consists of few main components and therefore communication between them needs to be secure and elements should be able to authenticate each other. For that purpose, we use certificates which will be signed by a certification authority - CA. CA has the role of a central unit that all customers trust. As API server is entry point in the whole cluster, the port through which the API is accessed is recommended to be exposed via the Hypertext Transfer Protocol Secure. Based on this, a certificate is required with appropriate key, and the certificate will have to be signed by the CA. Certificates in Kubernetes can be generated manually or automatically. Kubernetes offers certificates.k8s.io API that allows us to send a request for generating a certificate. Typically, the client creates a certificate request and sends to the API server, and the certificate is issued immediately after the request is approved.

Manually, certificates can be created with open-Secure Socket Layer (SSL), easyRSA or CF-SSL tools. Easy-RSA (Rivest–Shamir–Adleman) acts as an authority for certificates (CA) and creates certificates also. When CA is generated, a number of new files are created with a combination of Easy-RSA and (indirectly) open-SSL. Ca.crt file which represents the CA certificate is included and also CA.key, which represents the private key of the certificate. Generated certificate is shown in Figure 3. Kubelet needs two types of certificates, one certificate for API access and other certificate for its own API server. Instead of the administrator constantly to create certificates for each kubelet, kubelet itself can request a certificate immediately as soon as it starts. This is based on the API Server for certificates and authenticator, so called Bootstrap tokens. This type of authenticator authenticates clients based on short tokens.

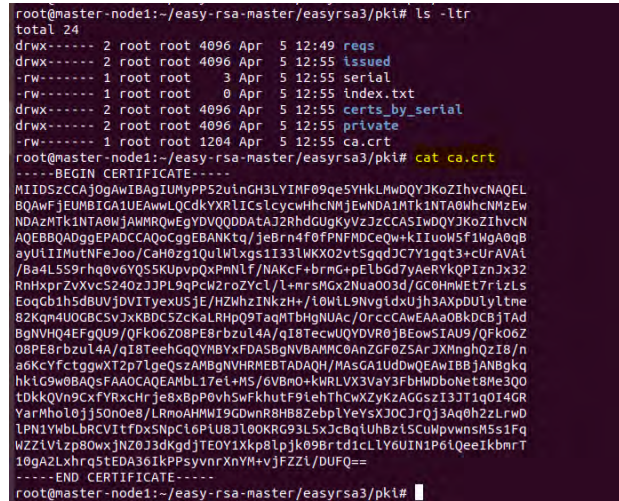


Fig. 3. Manually generated certificate with easy-RSA

Bootstrap tokens are stored as Secret in one of the standard namespaces, kube-system, where they can be dynamically created and managed. They occur in the following form, shown in [1].

[a-z0-9]{6}].[a-z0-9]{16} [1]

The first component is the token identifier and it is public information used when invokes the token without displaying the secret part used for authentication. The second component is a token secret and should only be shared with trusted parties. The token is specified in the HTTP request header as follows, [2].

Bearer 781292.db7bc3a58fc5f07e [2]

The object “Secret” for Bootstrap token is shown od Figure 4. The identifier of the token shows a Base64 string denoting a value encoded by the base64 algorithm. The type of secret must be bootstrap.kubernetes.io/token and the name is bootstraptoken-. The usage-bootstrap- * parameter indicates what this secret is used for. The value must be set to true for the token to be usable. Usage-bootstrap authentication field indicates that the token is used for API server authentication, while usage-bootstrap-signing indicates if use of the token is for signing ConfigMap. Expired tokens are rejected when they are used for authentication and they are ignored in signing up process. The expiration value is coded as absolute time UTC using RFC3339. It is recommended to activate the controller for tokens, for automatic deletion of expired tokens. Bootstrap tokens are used in the process of initially raising the cluster or connecting new nodes to an existing cluster. Initially, this type of token was built to support kubeadm, but also it can be used in another context, ie to be used by users who want to start a cluster without kubeadm.

```

root@master-node1:~# cat secretForToken.yaml
apiVersion: v1
kind: Secret
metadata:
  name: bootstrap-token-093oac
  #Imeto mora da bide vo sledniot format "bootstrap-token-<tokenId>"
  namespace: kube-system
#Type mora da bide vo sledniot format "bootstrap.kubernetes.io/token"
type: bootstrap.kubernetes.io/token
stringData:
  description: "Bootstrap token generiran so kubeadm init"

  token-id: base64(093oac)
  token-secret: base64(f395accd246ae52d)

  expiration: base64(2021-08-01T11:00:00Z)

  usage-bootstrap-authentication: base64("true")
  usage-bootstrap-signing: base64("true")

  #dopolnitelni grupi za avtenikacija na tokenot
  auth-extra-groups: system:bootstrappers:worker,system:bootstrappers:ingress
root@master-node1:~#

```

Fig. 4. Secret for Bootstrap Token

B. Authorization in Kubernetes

The authorization process in Kubernetes takes place immediately after the process of authentication. Kubernetes authorizes incoming requests using the API server. Authorization process consists of few steps:

- We assume that the client has been successfully authenticated (According to the procedure described in the previous chapter)
- The credentials of the client who sent the request are taken as input in the authorization module (username, the group user identificatory, ID)
- The second entry in the authorization module is a resource-containing vector, namespace and other secondary attributes
- If the client or application is allowed to perform a specific action on a particular resource, the request will be forwarded further to the admission controllers, otherwise the request is rejected with code 403, prohibited user.

Kubernetes offers several ways to implement authorization, through several modules: Node authorization, attribute-based admission control, admission control based on assigned roles (permissions) and webhook authorization. If multiple authorization modules are configured, each is checked by order. If any authorizer approves or rejects the request, the decision is automatic process and do not consult the following authorizer. If none of them have a decision in relation to the request, it is rejected and a response with code 403 is returned. The Role Based Access Control - RBAC authorization process uses the API group named `rbac.authorization.k8s.io` to be able to make authorization decisions. In this way admins are allowed, to dynamically create policies through the API. RBAC defines four types of objects depending on the resource. Within the cluster we distinguish two objects, `ClusterRole` and `ClusterRoleBinding`. In relation to namespace, `Role` and `RoleBinding`. The `Role` object specifies certain rules for the resources that are defined within a namespace, therefore, when creating a `Role`, the namespace must be specified. `ClusterRole` does not assign rules within a

namespace, but in a cluster or in multiple namespaces in one cluster [6].

The `RoleBinding` object gives permissions that are pre-defined in `Role`, for a specific user or group of users. This object has a list of resources and refers to the role to which it is assigned. `RoleBinding` gives permissions in a specific namespace while `ClusterRoleBinding` can grant permissions to access the entire cluster. Once a permission has been granted and linked to a certain role, for an entity, it is not possible to change it later. In this way, if someone has a permission to updating these objects will not be able to change the role assigned to entities. This prevents the cluster and resources from accessing of unwanted users. The API server for RBAC prevents users from escalating the privileges granted to them. The users are not able to change roles or tying new roles.

Within the RBAC process, there are two sets of roles that can be assigned. One group provides object-level security and is more secure, unlike the other group that provides security in a broader sense but is easier to manage. From this point of view, the safest approach is to assign a role to application-specific service account. This requires to specifies the name of the service user account in the pod specification, not to assign a role to the standard service account, which is created after creating a pod. The least secure approach is to assign a super-user role to all service user accounts within the entire cluster. This is how everyone will have full access to all resources and the system will be easier to be attacked.

IV. INSTALLING CONTAINER IN KUBERNETES FROM SYBER SECURITY PERSPECTIVE

The software in Kubernetes comes in the form of an image and it is very important to check that image, whether it can be used, whether it includes known vulnerabilities and whether meets certain image policy requirements. These images create containers in Kubernetes, so it is necessary that the image is secure, but not enough. The most basic procedure is to scan the image using a scanner that should inspect the packages included in the image and detect any known vulnerabilities. Scanning images should not be executed only once, instead of that it should be done more than once, because malicious code can be inserted into the image at almost any level [7]. A scan should be performed in the moment the image is built, when it is in the registry and at certain intervals, during its lifetime. In case vulnerabilities are detected, the container needs to be updated, use a fixed version of the package. Also, SSH access to the container is not recommended. Container's software can be stored in public and private registers. Organizations who are aware of the vulnerabilities of the images have their own private registers and it's allowed to create containers and applications from those registers only. In this way, they have bigger control over the software and who has the permission to read and use those images. Except that, they can restrict the registry from accessing the network by using firewall that would allow access only to known IP addresses. Software from the public sources can be risky because it is not scanned for security detection vulnerabilities. However, according to a 2020 study, almost 40% of images were extracted from public sources [8]. Also, whenever you need to create a container, it is important to specify the correct version of the image that will

be used and whenever there is a new revision of the image, yaml file must be updated with the newly version. After image scanning, next phase is admission control procedure. Admission controller is code that intercepts requests coming to the Kubernetes API server, but only after requests are authenticated and authorized. Controllers can be mutational and validation. Admission controllers restrict requests to create, delete, modify, or link requests with a proxy but do not support read requests. Recommendation is to use controllers who will inspect the software and requests and will not accept all the requests without checking.

A. Security Context for a Pod and Container

Security context for a container defines the privileges and settings to control access to a pod or container. To be able to set the security context for a given Pod, the spec for that pod should include the *securityContext* field. This field is also called the *PodSecurityContext* object. The security settings, which are specified for the given pod are applied to every container included in that pod. Figure 5 shows a Pod with security context.

By default, Kubernetes doesn't apply network policies and security contexts to a pod. This means that all the pods can talk to each other in a Kubernetes environment, potentially enabling lateral movement during a security breach. If we want to create a security context for a particular container, *securityContext* field must be included in the file from which we are creating the pod [9]. This type of context applies only to the specific container for which it is specified.

By using the Linux Capabilities certain privileges can be assigned to a process, without having the same privileges as a root user. This is also configured within the file with which we create the container. To use Linux Capabilities, field *capabilities* must be included in that file.

```
root@master-node1:~# cat security-context2.yaml
apiVersion: v1
kind: Pod
metadata:
  name: security-context
spec:
  securityContext:
    runAsUser: 10
    runAsGroup: 30
    fsGroup: 20
  volumes:
  - name: sec-ctx
    emptyDir: {}
  containers:
  - name: sec-ctx
    image: busybox
    command: [ "sh", "-c", "sleep 1h" ]
    volumeMounts:
    - name: sec-ctx
      mountPath: /data/sec-context-new
    securityContext:
      allowPrivilegeEscalation: false
root@master-node1:~# kubectl get pod
NAME                READY   STATUS    RESTARTS   AGE
security-context    1/1     Running   0           12s
security-context-demo 1/1     Running   0           6m36s
static-web          1/1     Running   2           35h
root@master-node1:~#
```

Fig. 5. Creating Pod with security context

Figure 6 shows the file for creating container, which includes *capabilities*.

For additional security, seccomp parameter needs to be used. Seccomp is set with the *seccompProfile* parameter in the pod security context or container. It consists of two parameters, *type* and *localhostProfile*. Valid type options are *RuntimeDefault*, *Unconfined* and *localhost*. The localhostProfile parameter can be enabled only if type is set as localhost. Seccomp is a powerful tool that should be used whenever it's needed. At least a Kubernetes auditing account should be enabled to track all system calls, and then, based on this to create an appropriate profile of seccomp.

```
root@master-node1:~# cat without_capabilities3
apiVersion: v1
kind: Pod
metadata:
  name: security-context-demo-3
spec:
  containers:
  - name: sec-ctx-3
    image: gcr.io/google-samples/node-hello:1.0
root@master-node1:~# v1_withCapabilitiesAdded4
root@master-node1:~# kubectl apply -f withCapabilitiesAdded4
pod/security-context-demo-4 created
root@master-node1:~# kubectl get pod
NAME                READY   STATUS    RESTARTS   AGE
security-context    1/1     Running   50          3d5h
security-context-demo 1/1     Running   50          3d5h
security-context-demo-2 1/1     Running   0           47h
security-context-demo-3 1/1     Running   0           17m
security-context-demo-4 1/1     Running   0           6s
static-web          1/1     Running   2           4d16h
root@master-node1:~# cat withCapabilitiesAdded4
apiVersion: v1
kind: Pod
metadata:
  name: security-context-demo-4
spec:
  containers:
  - name: sec-ctx-4
    image: gcr.io/google-samples/node-hello:1.0
    securityContext:
      capabilities:
        add: ["NET_ADMIN", "SYS_TIME"]
root@master-node1:~#
```

Fig. 6. Creating container with Linux Capabilities

B. Security Policies for a Pod and Container

In order to have more secure cluster, at the cluster level, a security policy is implemented. Security policy controls sensitive aspects of pod specification and defines a set of rules and conditions for the pod to be accepted in the system. Security policy is defined by the *PodSecurityPolicy* parameter [10].

Security policies for pod are implemented by activating the admission controller. However, all of this without authorizing any policies would prevent creation of the pod in the cluster. If only the *PodSecurityPolicy* resource is added to the pod specification, it will not mean anything. For service account to be able to use it, at first place it must be authorized to use that policy, by adding the verb *use* to the security policy. With granting controller access to the policy, access is obtained for all pods created with that controller, hence the preferred policy authorization method for security is to grant access only to the service account of the corresponding pod. For this purpose, we use the RBAC authorization method, which is standard authorization mode in Kubernetes and is also used for authorization security policies. First, the Role (or ClusterRole) parameter should assign access to the service account so that the desired policies can be used. If RoleBinding is used (not

ClusterRoleBinding) it will only allow the use of pods belonging to the same namespace [11]. In addition to restrict the creation and updating of the pod, security policies can be used to provide standard values for most fields that are in their control. When multiple security policies are available, the controller for security policies selects the policy based on the following criteria, first policies that allow the pod to remain as it is, without any change, are preferred. The order of these non-mutating security policies does not matter and the second, if the pod must be mutated, the first security policy is selected (sorted by name). To create a security policy, at first we create a namespace, service account for that namespace to which we grant privileges. In this example, the service account has the privilege to modify the pod, Figure 7. Then we create a security policy as an object in the file. This security policy will prevent the creation of privileged pods, shown in Figure 8. It is advisable to use more restrictive security policies to prevent unnecessary escalation of privileges, thus preventing further attacks.

```
root@master-node1:~# kubectl create namespace mynamespace
namespace/mynamespace created
root@master-node1:~# kubectl get namespace
NAME                STATUS AGE
default              Active 20d
kube-node-lease      Active 20d
kube-public           Active 20d
kube-system           Active 20d
mynamespace          Active 6s
root@master-node1:~# kubectl create serviceaccount -n mynamespace myaccount
serviceaccount/myaccount created
root@master-node1:~# kubectl create rolebinding -n mynamespace myrole --cluster-role=edit --serviceaccount=mynamespace:myaccount
Error: unknown flag: --cluster-role
See 'kubectl create rolebinding --help' for usage.
root@master-node1:~# kubectl create rolebinding -n mynamespace myrole --clusterrole=edit --serviceaccount=mynamespace:myaccount
rolebinding.rbac.authorization.k8s.io/myrole created
```

Fig. 7. Adding privileges to service account for specific namespace

```
root@master-node1:~# vi mysecpodpolicy.yaml
root@master-node1:~# cat mysecpodpolicy.yaml
apiVersion: policy/v1beta1
kind: PodSecurityPolicy
metadata:
  name: myexample
spec:
  privileged: false #We dozvoluwa krerianje na prillivigrani Pods.
  selinux:
    rule: RunAsAny
  supplementalGroups:
    rule: RunAsAny
  runAsUser:
    rule: RunAsAny
  fsGroup:
    rule: RunAsAny
  volumes:
    - '*'
root@master-node1:~#
root@master-node1:~# kubectl-admin create -f mysecpodpolicy.yaml
error: error validating "mysecpodpolicy.yaml": error validating data: ValidationError(PodSecurityPolicy.spec.volumes): invalid type for io.k8s.api.policy.v1beta1.PodSecurityPolicySpec.volumes: got "string", expected "array"; if you choose to ignore these errors, turn validation off with --v
allidate=false
root@master-node1:~# vi mysecpodpolicy.yaml
root@master-node1:~# kubectl-admin create -f mysecpodpolicy.yaml
error: unable to recognize "mysecpodpolicy.yaml": no matches for kind "PodSecurityPolicy" in vers
ion "policy/v1"
root@master-node1:~# vi mysecpodpolicy.yaml
root@master-node1:~# kubectl-admin create -f mysecpodpolicy.yaml
podsecuritypolicy.policy/myexample created
```

Fig. 8. Security Policy that prevents creating privileged pods

V. CONCLUSION

The purpose of this paper was to analyze the security mechanisms that can be implemented in Kubernetes. From the analysis so far, several key points are identified that need to be taken for the protection of Kubernetes:

- Scanning images in the initial stage of deployment
- Updated, robust tools for scanning and verifying that image has no vulnerabilities
- Proper configuration of all Kubernetes parameters, at each stage, which will not leave "doors open" for the attackers
- Selection of appropriate runtime software

Some of the news about Kubernetes security future are that PodSecurityPolicy will be replaced with PSP Replacement Policy which is designed to be as simple as practically possible while providing enough flexibility to really be useful in production.

Kubernetes is the future. According to research, it is expected that by 2022, more than 75% of global organizations will be running containerized applications in production.

REFERENCES

- [1] Kim McMahon, "2019 CNCF Survey results", March 4, 2020
- [2] Isam Mashhour al Jawarneh, Filippo Bosi, Paolo Bellavista, Luca Foschini, "Container Orchestration Engines: A Thorough Functional and Performance Comparison", 2019
- [3] Mike Ensor, Drew Stevens, "Shifting Left on Security-Securing software supply chains", Google Cloud, Feb 25, 2021
- [4] Eddy Truyen, Nane Kratzke, Dimitri Van Landuyt, Bert Lagaisse, Wouter Joosen, "Managing Feature Compatibility in Kubernetes: Vendor Comparison and Analysis", 2020
- [5] Kubernetes Documentation, "Authenticating with Bootstrap Tokens", 2021
- [6] Kubernetes Documentation, "Certificate Management with kubeadm", 2021
- [7] Liz Rice, Michael Hausenblas, "Kubernetes Security – Operating Kubernetes Cluster and Application Safety", O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA, 2018
- [8] Shay Berkovich, Jeffrey Kam, Glenn Wurster, "UBCIS: Ultimate Benchmark for Container Image Scanning", 2020
- [9] Tanner Luxner, "Cloud Computing Trends: 2021 State of the Cloud Report", Cloud Management, 2021
- [10] Paulo Gomes, "Seccomp in Kubernetes — Part 3: The new syntax plus some Advanced topics", 2020
- [11] Gaurav Chaware, "Kubernetes Pod Security Policies with Open Policy Agent", 2020
- [12] Aneta Poniszewska-Mara, Ewa Czechowska, "Kubernetes Cluster for Automating Software Production Environment", Sensors 2021
- [13] Pawan Shankar, "Sysdig 2020 Container Security Snapshot: Key image scanning and configuration insights", 2020

The Application of the Internet of Things in Everyday Equipment Affects to Have a More Efficient and Quality Life

Prof.Dr. Aleksandar Ristevski
Faculty of Electrical Engineering and Information
Technology
University of „Ss. Cyril and Methodius“ ,Skopje,
North Macedonia

Mr.sc.Avni Rustemi
Faculty of Electrical Engineering and Information
Technology
University of „Ss. Cyril and Methodius“ ,Skopje,
North Macedonia

Abstract- With the advancement of Automation technology, life is becoming simpler and easier in all aspects. In today's world, automated systems are more preferred than manual systems. With the rapid increase in the number of Internet users over the past decade it has made the Internet a very important part of life, and IoT is the latest technology in under development. IoT is a network of facilities that are constantly evolving from industrial machinery to consumer goods that can share information and perform tasks while you are busy with other activities. Wireless Home Automation system (WHAS) using IoT is a system that uses computers or mobile devices to control basic home functions and features automatically via the Internet from anywhere in the world, an automated home is sometimes called a smart home. It aims to save electricity and human energy. The home automation system differs from other systems by allowing the user to operate the system from anywhere in the world via the internet connection. Through this paper, we will try to briefly describe the importance of IoT in people's daily lives, and the possibilities of applying IoT to devices used in people's daily lives. The IoT is undoubted of particular importance, as through intelligent devices every day more and more our lives are being facilitated and opportunities for efficiency and a better quality of life.

Keywords: internet of things, smart house, smartphones, security, technology, models.

I. INTRODUCTION

The Internet of Things is definitely a technology that is simplifying people's daily lives and is making it more efficient in utilizing multiple devices in daily life. This is because a large number of devices that are used today in everyday life, thanks to technological inventions, can be controlled by digital devices, smartphones, or even tablets. What is much more important is how the devices can be controlled from a certain distance and how the devices communicate to control each other.

Many of the devices are controlled and monitored to help human beings. Moreover, various wireless technologies help connect from remote locations to improve the intelligence of the home environment. IoT technology has been used to come up with innovative and growing ideas for smart homes to improve living standards. Various wireless technologies that are able to support a type of transfer, knowledge, and remote knowledge management such as Bluetooth, Wi-Fi, and mobile networks are used to introduce ample levels inside the home. Smart homes for the elderly and disabled limited provide increased quality of life for these persons[1]. It can provide an interface for home appliances or the automation system itself, through the telephone line or the Internet, to supply management and implementation through a mobile phone or personal computer. Even over long distances the user can monitor and manage the gate of his home, various devices and turn on / off the TV without any human intervention. Despite these advantages, home automation has nevertheless received widespread approval and attention due to its importance and high complexity. Internet of Things security is a growing concern because our networked devices are data collectors. Personal information collected and stored on these devices - such as our name, age, health information, location, etc. - can assist cyber criminals in stealing our identity. But the more functionality we add to our smartphone, the more information we store on the device. This can make smartphones and everything related to them vulnerable to a number of attacks of various kinds.

II. BRIEF DESCRIPTION OF COMMUNICATION MODELS BETWEEN IOT DEVICES

What is very important is to clarify how communication is achieved between these IoT devices. There are several methods of communication, where we will briefly stop to explain in brief the essence of communication for each method in particular.

Device-to-Device Communications - This type of communication used by IoT devices, enables remote devices to communicate directly with each other, without utilizing any intermediary services in between. These devices communicate through many types of networks, including IP networks or the Internet. Often these devices use protocols like Bluetooth, Z-Wave, or ZigBee to establish direct communication from the device in equipment. The following is an example of how electrical equipment (eg. electric light) communicates with an electrical switch.

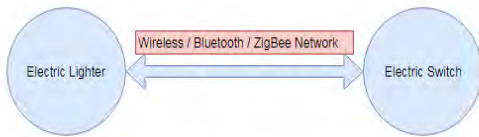


Figure1. *Device-to-Device Communications model*

Device-to-Cloud Communications. In a device-to-cloud communication model, the IoT device connects directly to an Internet cloud service as a service provider to exchange data and control messaging traffic. This approach often takes advantage of existing communication mechanisms, such as traditional Wired Ethernet or Wi-Fi connections to establish a connection between the device and the IP network, which eventually connects to the cloud service. An example of using this method is the case with Samsung Smart TV technology, where the television uses the Internet connection to transmit viewing information to Samsung for analysis and to enable interactive features of TV voice recognition.

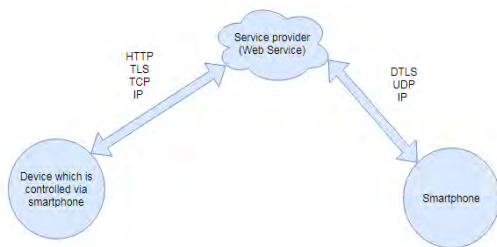


Figure2. *Cloud Device Communication Model*

Device-to-Gateway Communication. In the Device-to-Gateway model, or more commonly, device-to-application-layer gateway (ALG), the IoT device connects via an ALG service as a channel to reach the cloud service. In simpler terms, this means that there is application software running on a local gateway device, which acts as an intermediary between the device and the cloud service and provides security and other functions such as translating data and protocols.

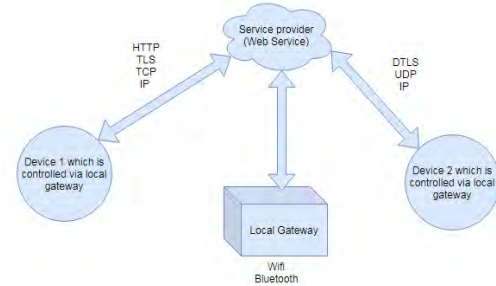


Figure3. *Device-to-Gateway Communication Model*

Back-End Data-Sharing Communication. The Back-End Data-Sharing model refers to a communication architecture that enables users to export and analyze smart object data from a cloud service in combination with data from other sources. This architecture supports "the user's desire to grant access to the uploaded sensor data to a third party". This approach is an extension of the device-to-cloud communication model, which can lead to isolated data where "IoT devices upload data to a single application provider". A Back-End Data Sharing architecture enables the collection and analysis of data collected from the data streams of an IoT device. [2]

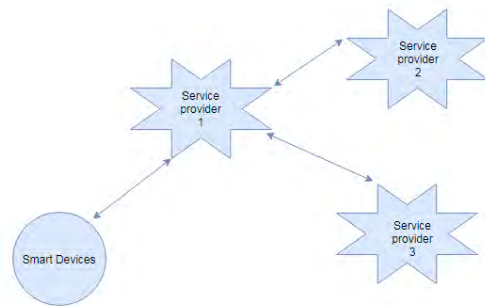


Figure4. *Back-End Data-Sharing Communication Model*

III. SECURITY PROBLEMS RELATED TO IOT DEVICES AND IOT APPLICATION

The rapid development of Information Technology (IT) and the Internet have found applications in almost all walks of life. This development is making society increasingly dependent on technology. In addition to the rise of technology and numerous services, cyber-attacks have also increased to a large extent. Therefore, defending against these attacks is becoming increasingly challenging. These attacks are advancing day by day and attackers are managing to access sensitive data; in governmental and non-governmental organizations, but also in the personal data of many technology users. Such unauthorized approaches are causing great damage, such as theft of confidential data, damage to the image of companies, large financial losses, etc.[3]. Data security includes a set of measures that must be taken to ensure that a system will be able to meet its intended purpose, while on the other hand reducing intentional or unintentional negative consequences. When new functionalities are added to a system, then security measures should be applied to ensure that these functionalities do not compromise the intended functionality or introduce new attack vectors [4]. The world is increasingly descending into the ancient vision, known as "good" versus "evil". The focus of great talent is resulting in new products and applications, offering new opportunities to a large number of users. From smartphones to smart cars, we now have the Internet of Things (Internet of Things-IoT), which enables devices to communicate with each other. Large enterprises continue to experiment and introduce technologies that meet human needs. But, at the same time, "evil" is growing very fast, using these technologies for malicious purposes[3]. The key to maintaining a high level of performance in relation to implementation measurements is to first establish an initial security base which will be continuously improved until approximately 100% security is achieved. Once we have established such a basis, we must have a constant flow of information to respond to vulnerabilities, as well as have constant system updates[4].

The Internet of Things has so many applications in everyday life. Some of the IoT applications are:

- Home Automation: Smart home or home automation can have so many features like power monitoring, intelligent lights, temperature monitoring, humidity monitoring, smoke detector, security and surveillance, smart door, baby monitoring, intelligent devices, etc.
- Health: Patient monitoring, staff health monitoring, remote patient monitoring, etc.
- Retail: The retail sector also has so many applications like inventory monitoring and real-cameras, and alarm systems.

These smart devices are connected to the local server

time sales.

- Industry: Supply chain management, logistics management, etc.
- Transport: Parking management, traffic management, transport management, monitoring and tracking of vehicles, smart cars, etc.
- Security and oversight: Home security, building security, environmental security.
- Intelligent infrastructure: Intelligent parking, street lighting, pollution monitoring, environmental monitoring, waste management, disaster management, etc.

IV. SMART HOUSES

IoT-based homes consist of several sensors, which are connected to the wireless network to develop distributed support networks. Each node of these sensors includes three subsystems: (1) the subsystem for environmental sensors such as temperature, humidity, and light intensity; (2) a processing subsystem, consisting of a microcontroller and integrated circuit to process the sensor data for calculation; and (3) a communication subsystem for the exchange of data collected between different sensors [5-7].



Figure 5. Schematic diagram of sensors in a house

Wireless Sensor Networks (WSNs) offer specific applications which are primarily designed as a closed system; however, IoT-based applications are independent of specific applications [8] and are more focused on developing a large-scale WSN infrastructure that can support open standards. In most IoT-based home systems, intelligent actuators and smart sensors are installed inside the home environment to monitor and monitor its operation. It also includes intelligent devices to run the entire house intelligently without misdirection. Some of the applications are home appliances, lights, security assumed to be more isolated and powerful.

through a wireless medium for data collection and analysis. A house to be called smart must meet at least some criteria, respectively it must have some IoT equipment installed in that house, and they are:

- Fire/smoke detector: The fire or smoke detector is one of the essential sensors to build a smart home and protect the house from fire.
- Humidity Detector: The flow sensor is used to leak water into a supply unit. The sensor can be placed around water heaters, dishwashers, refrigerators, sinks, water pumps, and wherever there may be a risk of water leakage.
- Smart thermostats: The thermostat provides control over home heating and cooling - from anywhere. They are always useful to save money by monitoring the temperature and humidity inside and outside the house.
- Motion sensors: This sensor detects movements in an area. These sensors can alert you immediately if there is any movement inside the house, or if the doors or windows are open. They can even turn the lights on and off when the doors open and close.
- Home automation is improved by considering a Wireless sensor node. A modern home integrates various home electrical appliances and automates them without minimal user intervention. The modern home keeps track of the various variables of the present environment and guides the equipment to work according to the needs of the user. Not only the automation of household appliances for daily use but also the notification of the user of the price of his electric bill at regular intervals and the automatic reservation of the gas cylinder, if the gas level reaches below the threshold. We have achieved the development of a House modern using IoT technologies.

V. SECURITY THREATS TO THE IOT

In the IoT, security threats are generally organized into two main classes. Threats in the first class are attached to the CIA which is the confidentiality, integrity, and availability associated with the conventional network ecosystem. Although, the confusion and ability of such threats are significantly more bitter, due to the large size and variety of objects. Since IoT facilities are being placed around us viz. capturing accurate estimates from heartbeats to measure room temperature, and also being used for several other purposes in different areas, therefore, the data residing in the IoT ecosystem are

While some IoT devices collect a large amount of confidential information about IoT users such as their behavior, account passwords, geographical location, daily habits, etc., data that possess such information is considered private property and any harm may expose the user's privacy information which an offender may obtain and disclose the privacy of the IoT user. One of the most critical problems in securing information has been accepted as Privacy Policy [9]. To protect user privacy, cryptographic tools were used. A certificate primarily certifies that simply certified personnel have access to the user's personal information, while cryptographic tools ensure the security of sensitive information during the transmission, storage, and alteration of such information. Recently proposed research regarding privacy protection primarily in the IoT is either at a higher level of research based on physical layer transmission security. Scaling is the only major concern that separates IoT from the legendary Internet. There are about trillions of objects associated with the IoT network. Network-connected devices on this massive scale are quite difficult to handle by the general naming policy. Therefore, to give the convention of new names, it is required that the current naming policy be revived or regenerated. Unlike naming policy, methods related to identification and authentication are also required to be revived respectively. The other two basic challenges, i.e. transparency and reliability, further make the design even more difficult in terms of identification and authentication methods. Potential security threats to IoT are:

- Unauthorized access: If we take as an example a door that is locked by means of a "smart lock", we assume that it is attacked by an unauthorized party, the attacker can break into the wise house without destroying the door. The result of this scenario can be loss of life or property. To overcome such attacks, passwords need to be changed frequently and must be long because it is much harder for attackers to crack the long password. Also, similarly, more sophisticated authentication and access control mechanisms can be applied.
- Monitoring: Security is one of the important goals of a smart home. Therefore there are many sensors used for fire monitoring, baby monitoring, housebreaking, etc. If these sensors have been hacked by an attacker, he can monitor the house and access personal information. To avoid this attack, data encryption between the central system and the sensors must be implemented or user authentication for unauthorized parties can be applied.

- DoS / DDoS: Attackers can enter the home network and send multiple messages to smart devices. They can also attack the target device using malicious code in order to carry out DoS attacks on other devices that are connected to a smart home. As a result, intelligent devices are unable to perform their proper functions due to the depletion of resources that respond to such attacks.
- Forgery: When smart home devices communicate with the application server, an attacker could change the flow of packets by changing the packet routing table. In this way, the attacker can change the content of the data or the confidentiality of the data can be revealed.
- Some of the challenges that need to be considered to make the IoT as secure as possible are:
- Data privacy: Some smart TV manufacturers collect data about their customers to analyze their viewing habits, so the data collected by smart TVs pose a challenge to data privacy over time. transmission.
- Data Security: Data security is also a major challenge. While transmitting data smoothly, it is important to protect yourself from being monitored by cybercriminals.
- Insurance Concerns: Insurance companies that install IoT devices on vehicles collect health and driving status data in order to make insurance decisions.
- Lack of Common Standard: Since there are many standards for IoT equipment and the IoT manufacturing industries, distinguishing between permitted and unauthorized Internet-connected devices is a challenge in itself.
- Technical Concerns: Due to the increased use of IoT devices, the traffic generated by these devices is also increasing. There is therefore a need to increase network capacity .

VI. CONCLUSION

By the time we have arrived, technology has developed quite a lot and we are witnessing that it is not stopping. The Internet has made it possible to connect people and cars on land, in the air, and at sea. The IoT offers some very interesting applications to make our lives easier in both health, transportation, and business. However, various factors such as security, privacy, data storage should also be considered. device in our homes, our cars, our buildings, our workplaces, and everything else.

As much as technology is designed to make our lives easier, to make us think about the things we need to think about, it also has its downsides.

Nowadays avoiding technology is impossible, but the least we can do is be careful about privacy.

VII. REFERENCES

- [1] Atzori, Luigi, Antonio Iera, and Giacomo Morabito. "The internet of things: A survey." *Computer networks* 54.15 (2010): 2787-2805.
- [2] Swamy s, Narasimha AU - Nayak, Shantharam AU - M.N, Vijayalakshmi PY - 2016/05/01 SP - 75 EP - 78 T1 - Analysis on IoT Challenges, Opportunities, Applications and Communication Models VL - 2 JO - International Journal of Advanced Engineering, Management and Science (IJAEMS) ER .
- [3] Ab Razak, M.F., Anuar, N.B., Salleh, R., Firdaus, A., (2016); The rise of "malware": bibliometric analysis of malware study. *J. New. Comput. Appl.* 75, 58–76.
- [4] Ahmad, F., Sarkar, A., (2016); Analysis of dynamic web services: Towards efficient Discovery in cloud. *Malays. J. Comput. Sci.* 29 (3).
- [5] He D, Chan S, Guizani M, Yang H, Zhou B. Secure and distributed datadiscovery and dissemination in wireless sensor networks. *IEEE Transactionson Parallel and Distributed Systems*. 2015;26(4):1129–1139. [accessed Jan 02 2021].
- [6] Ghormare S, Sahare V, Editors. Implementation of data confidentiality forproviding high security in wireless sensor network. 2015 InternationalConference on Innovations in Information, Embedded and CommunicationSystems (ICIIECS), Coimbatore, India. IEEE, 2015.[accessed Jan 03 2021].
- [7] Ghayvat H, Liu J, Mukhopadhyay SC, Gui X. Wellness sensor networks:A proposal and implementation for smart home for assisted living. *IEEE Sensors Journal*. 2015;15(12):7341–7348.
- [8] Li J, Zhang Y, Chen Y-F, Nagaraja K, Li S, Raychaudhuri D, Editors.A mobile phone based WSN infrastructure for IoT over future internetarchitecture. 2013 IEEE International Conference on Cyber, Physical andSocial Computing, Green Computing and Communications (GreenCom),and Internet of Things (iThings/CPSCoM), Beijing, China. IEEE, 2013.[accessed Jan 03 2021].
- [9] Vikas Hassija, Vinay Chamola, Vikas Saxena, Divyansh Jain, A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures, *IEEE Access* (Volume: 7).

User-to-Cloud Latency Performance Characteristics in an European Cloud infrastructure

Teodora Kochovska, Marija Kalendar
Faculty of Electrical Engineering and IT, Skopje
University "Ss. Cyril and Methodius" in Skopje
Skopje, N. Macedonia
{teodora.kochovska, marijaka}@feit.ukim.edu.mk

Simon Bojadzievski
A1 Macedonia,
A1 Telekom Austria Group
Skopje, N. Macedonia
simon.bojadzievski@gmail.com

Abstract— The cloud infrastructures of today become a de-facto standard in everyday life being used for work, study, as well as private home needs. As a result, the performance of cloud infrastructures and cloud services availability is quite crucial. The academic community is continuing its efforts to characterize the cloud infrastructure performance, since the cloud providers usually provide partial and only qualitative information about network performance. This paper will provide practical measurements of network latencies and the user experience of cloud services, focusing on European cloud infrastructures and user's experience in East European geographical regions. The work also encompasses in depth analysis of access segment latencies: user to exit from ISP provider, ISP provider to cloud provider and intra cloud segment. For this analysis we collected 10-day measurements from an Eastern European user location to services deployed in several distinct geographical locations using the infrastructure of one of the most renown European public-cloud infrastructure providers-Exoscale. The experiments enable a detailed latency performance characterization and fine-grained view of cloud data paths perceived by East European users. Results show that most network latency measurements are within the boundaries of the expected human perceivable latency of 50-100 ms, thus enabling even real-time cloud applications.

Keywords—cloud infrastructures, cloud performance measurements, network latency CDF, European cloud infrastructure, European GDPR

I. INTRODUCTION

Measuring and estimating network latency is significant in evaluating the cloud service performance from the perspective of the users and Internet providers. Taking accurate measurements of the network latency is pivotal in providing highly available and reliable cloud services. Cloud providers start to host latency-critical applications, such as cloud-based gaming ([1], [2]) that are already available in the market. Consequently, it is essential to analyze latency and variations on the cloud paths. As a result, measuring network latency becomes a critical, but not an easy task, due to constant network dynamics, fluctuations and failures.

In this paper we decided to focus on measuring and characterizing the user-to-cloud latency experience from an East European user location to several distinct European cloud datacenter locations using the cloud infrastructure of one of the most renown European public-cloud infrastructure providers-Exoscale. The measurements will present an extensive study of the user experience for reaching the cloud servi-

ces in different network segments and different cloud datacenter locations. The setup is even more interesting, since it comprises an East European user location which is not an integral part of the European Union.

The network latency is measured by using two diagnostic tools: traceroute and ping. We gathered and processed the results from both diagnostic tools every hour in a continuous period of ten days. Our goal is to make a complete analysis of data obtained from the diagnostic tools by calculating average round trip times (RTT) among different segments of the network. The rest of the paper is organized as follows: Section II summarizes the related work, Section III explains the experimental setup, Section IV presents and elaborates the results, while Section V concludes the paper.

II. RELATED WORK

An extensive amount of research work is done for characterizing the Internet network, especially taking into account the recent massive use of the cloud platforms and the various cloud services, ranging from data storage, virtual machines, application services, up to online cloud gaming and provision for real-time IoT services. The versatility of services, sparks the need for different network and cloud infrastructure characterization studies including user experienced latencies for service access, network traffic paths, loss of packets, unavailability of services, loss of connectivity.

A number of cloud reachability studies have been carried out, considering vastly spread geographical regions. The authors in [3], [4] work on a global scale involving hundreds of cloud datacenters and thousands of probes (users). Two key points are analyzed: user-to-cloud latency and path lengths between end-users and the datacenters. The results show that the current cloud can keep up with many latency-critical applications and a larger part of the users around the world can approach a cloud within 100 ms. This is the threshold for many future latency-critical applications ([5], [6]). The geographical distance from the end-user to the cloud was also investigated showing that large-scale data center deployment is essential to shorten the network latencies and to be in line with the critical-latency applications.

The research in [4] focuses on edge-computing latency-critical applications compared to the cloud. The results imply that cloud infrastructures in well-connected areas (North America and Europe) can meet the requirements of latency-

critical applications to operate on the edge, but the benefits of edge computing remain small from the network performance point of view of. Greater benefits of edge computing have been measured in Latin America, Africa, and parts of Asia.

The authors in [7] implemented a tool, BlameIt, that allows cloud operators to localize the cause of network latency. The tool gives a comprehensive view of the WAN latency of the cloud provider Azure hosting many services over hundreds of billions of TCP connections.

The authors in [8] and [9] focus on cloud-to-user latency, performing a 14-day measurement employing 25 vantage points from the Planetlab infrastructure by using services in two popular cloud infrastructures: Amazon Web Services and Microsoft Azure. The results show the presence of both spatial and temporal latency trends. Additionally, it was not possible to choose a best-performing provider, since the results changed with the considered specific cloud regions, but that an ideal multi-cloud deployment leads to non-negligible latency gains in 7% of the cases.

Other authors, as in [10], investigate the delay, but also the throughput among node pairs within cloud networks. Latency in cloud networks is targeted in works as [11], [12], [13], [14]. Some characterize the latency in inter-datacenter networks ([11]), while others ([12]) also measure intra-datacenter latency, focusing on the provider's view of the network. Works like [13] and [14] measure user-end experience when interacting with the cloud, giving guidelines for geographically spread deployment of cloud services in general [13]. The specific cloud gaming applications [14] additionally show the need to bring the edge infrastructure closer to the end-users in order to satisfy the stricter requirements.

Even though the research work in this area is vast, we have seen few examples targeting specifically European cloud-infrastructures and the experience of the users accessing such infrastructures. Moreover, latency measurement studies for the user experience from the East European geographical regions when accessing European cloud infrastructures has rarely been addressed. Finally, we decided to measure the user experience with this specific setup taking into account the European GDPR regulations that require access to cloud data specifically deployed in European geographical regions.

III. EXPERIMENTAL SETUP

This section will present details of the investigational approach taken in order to measure the network latency from the end-user to a certain cloud location. Here, we describe the methodology for measuring and evaluating network latency. Here we present the experimental set up: comprehensive overview of the European cloud service provider - Exoscale; configuration and selection of instances in targeted data centers of the cloud provider; and the detailed methodology for measuring the network latencies.

A. About Exoscale

Exoscale, a part of A1 Telekom Austria Group, is a cloud service provider offering scalable infrastructure and platform hosting. Exoscale offers comprehensive building blocks for

cloud applications. Currently the provider enables six(6) data centers located only across Europe, with several in Germany. Each data center is connected with a redundant 10 Gbps Internet link, securing customers to access their own VPSs (Virtual Private Server) with an inbuilt Firewall.

With datacenters based on European locations, Exoscale specifically emphasizes its fulfillment and dedication for the European General Data Protection Regulation (GDPR) compliance. The GDPR is an European Union (EU) law on data protection and privacy that protects the fundamental rights and freedoms, by protecting users' personal data. It protects citizens of the EU and the European Economic Area (EEA) and persons whose data is processed by EU or EEA businesses [15]. The GDPR imposes special requirements for the transfer of personal data outside the EU and EEA thus cloud datacenters deployed in Europe are a preferred option. Consequently, we believe that a European cloud provider will be a preferred choice for European companies dedicated to data privacy.

B. Cloud measurements

In this section, the experimental setup for making network latency measurements for the European cloud infrastructure Exoscale is described. As it can be seen in Fig. 1, Exoscale is exclusively a European cloud service provider with data centers set up in Germany, Austria, Bulgaria and Switzerland. One VPS in each datacenter has been deployed (in six different locations), in order to include geographical region diversification (Fig. 2). On the client side, due to limited resources, we have performed the measurements from one geographical location, located in Skopje, R. N. Macedonia. This setup outlines our intention to measure and characterize an East European location connecting to a West European implemented cloud infrastructure.



Fig. 1. European cloud provider Exoscale data-center locations

The setup regarding the measurements includes diagnostic tools that are implemented on the user-side and directed to each geographically distributed datacenter cloud location where a specified VPS is configured and deployed (Fig. 2). Consequently, the measurements and estimations of the latency have been conducted from the end-user to all cloud instances. The latency measurements were scheduled to execute once every hour for a period of ten days from the end-user point to each cloud location. We addressed the latency measurements by using the network diagnostic tools traceroute and ping. The ping network tool is used to test the reachability of each instance deployed in the data centers. The traceroute

command is used to identify each hop's IP address as well as the latency to the hop until the packet reaches the last destination. The round trip time (RTT) measurements from both tools are used as an indicator of the latency between the end-user and the six VPSs. Each latency measurement probe is analyzed to investigate the network performance connectivity to the cloud. Our goal is to discover and localize the moments and parts of the network where increased latency and distance (hop count) between the end-user and cloud datacenters occur.

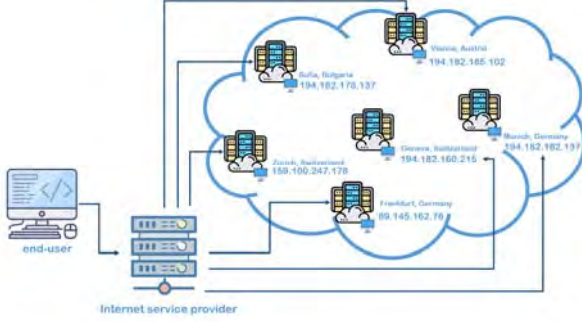


Fig. 2. Experimental setup with end-user and six Exoscale VPS locations

We analyzed results from traceroute and ping diagnostic tools to present characterization of the network latency. We have presented and analyzed the RTT distribution of the measured data from the traceroute and ping diagnostic tools considering each cloud datacenter location. Finally, a comparison of the measured data from both tools will be presented and discussed.

In order to estimate the delay characteristic of a connection, the Cumulative Distribution Function (CDF) as described in [16] and presented on Fig. 3 will be used.

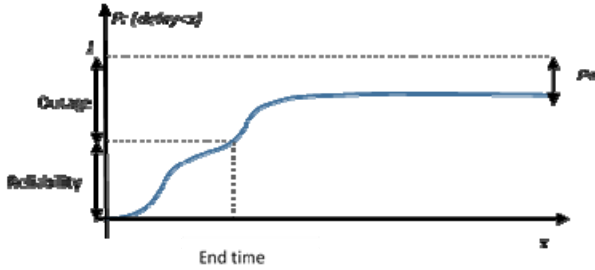


Fig. 3. Ratio between reliability, delay, outage and end time [16]

Fig. 3 depicts the generic needs from aspect of reliability and reliability of a data communication links, where specific values depend on particular environments. The asymptote of the CDF is equal to $1 - P_e$, where P_e is the probability for packet loss due to link outage. The blue curve presents the probability that the packet is delivered to the destination within the expected delay time. Based on the CDF function, we can estimate the system end-to-end delay and the expected performance for the applications that will run on top of this connection. If the system performs with low delay, most of the results will be on the left side and will reach higher value very fast, since Y presents the relative number of events for $X < x$.

IV. RESULTS

This section will present the measurements results from the experimental setup described in Section III.

A. Evaluating results from traceroute diagnostic tool

Within the platform, measurements were taken every hour for a continuous period of ten-days from the end-user point to each cloud VPS on a different geographical location. The network path from the end-user to the cloud VPSs has been considered in three different segments: user Internet Service Provider (ISP) segment (**client segment**); network traversal segment (**middle segment**); and final intra-platform segment (**cloud segment**). The gathered data were analyzed from two perspectives: the time the packet spent in each segment, and cumulative latency behaviour for reaching each segment.

The average RTT values are calculated every hour from the three packets obtained from traceroute. The traceroute diagnostic tool measures the packet latency from the end-user to each hop until it reaches the final destination; in this paper - from the end-user to the VPS on each cloud location. After analyzing the data we consider three vantage points from the results: 1. the hop IP address when the packet leaves the user ISP; 2. the hop IP address when the packet enters the cloud, 3. the final cloud destination.

In the first perspective, we started evaluating the traceroute results for the three different segments: **client**, **middle**, and **cloud** segment. We observed and analyzed the time that each packet spent in each segment by processing the obtained RTT values. The average RTT values in each of the segments describe the following:

- In the **first (client) segment**, the average value of RTT is the time it takes for the packet to leave the ISP, e.g. the time the packet spends in the ISP network;
- In the **second (middle) segment**, the average value of RTT is the time it takes for the packet to travel from the moment of leaving the user ISP, through the Internet network, until it enters the cloud, e.g. **network traversal**;
- In the **third (cloud) segment**, the average value of RTT is the time it takes for the packet to reach the final destination from within the Exoscale cloud; From the moment it enters the Exoscale cloud platform, up to the moment of its arrival at the destination.

As a result we visualize the distribution of all latencies from the end-user to the cloud instances, separately for each segment, throughout our measurement duration of 10 days.

In the second perspective, we evaluated the traceroute results by considering three points of view:

- After leaving the network of the end-user ISP, the packet is always forwarded to the next hop gateway in Bulgaria. This happens for all six final cloud destination locations. Thus, we define the first average RTT value as the total time it takes for the packet to travel from the end-user ISP to the next hop IP address in Bulgaria.
- The second average RTT value is defined as the total time it takes for the packet to travel from the end-user through

the Internet network up to the moment the packet enters the cloud infrastructure.

- The third average RTT value is the total time it takes for the packet to travel from the end-user through the Internet network until the packet reaches the final destination.

B. Evaluating results from ping diagnostic tool

Similarly, we measured and analyzed the results obtained from the ping diagnostic tool. The measurements were taken at the same time with traceroute, for each cloud location, every hour, for a period of ten days. The results are summarized as the minimum, maximum, and average RTT from the end-user to the final destination. The average RTT from all four packets sent from the end-user to the final cloud destination are considered. This average RTT is the total time that packet travels from the end-user through the network until it reaches the final deployed VPS in each cloud location.

C. Network latency measurement results

Here we present and comment on the obtained results by depicting the measured values and corresponding distributions. The results are organized and discussed in three sections:

- distribution of the network latency from traceroute diagnostic tool in the three identified segments.
- distribution of the network latency from ping diagnostic tool.
- distribution of the network latency in comparison (both diagnostic tools).

Regarding the **first segment - the client segment** - the distribution of all network latencies for connections to all six instances (in geographically distributed cloud data centers) from the traceroute tool is presented in Fig. 4. Here it can be perceived that the time it takes for the packet to leave the user ISP's network is similar for all packets, regardless of the location of the final destination. This is an expected conclusion since the ISP's network is the same. Nevertheless, we can also conclude that the RTT measurements in the **first - client or exit - segment** are normally distributed centered around 10-15 msec, which is a good result if compared to the measurements in [3].

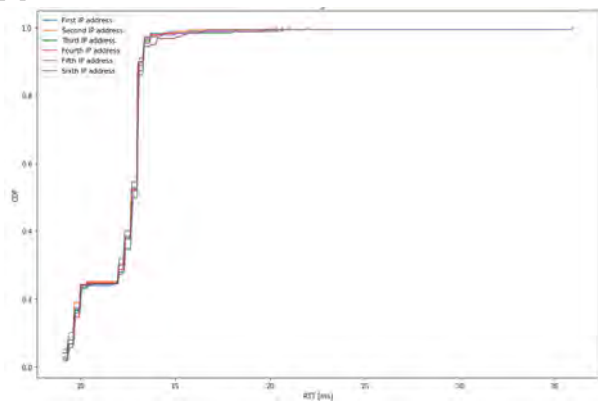


Fig. 4. Distribution of all RTTs in the first segment for all destinations (traceroute)

Noticeably, 25% of the packets stay in the first segment for a period of 10ms, and majority of the rest stay up to 15 ms,

following the same pattern for all six datacenters. Of course, we can observe some greater latencies, which result from the ISP's network instability at short intervals of time.

The distribution of all network latencies in the second (middle, transport) segment is presented in Fig. 5. Again, the graph depicts measurements taken with traceroute, for the packets that travel to all six deployed VPSs. Fig. 5 shows that there are variations in network latency. In other words, the time it takes for the packet to travel from the moment of exiting the user ISP's network through the Internet, until it enters the Exoscale cloud infrastructure is significantly different for the deployed VPSs in the different data center geographical locations.

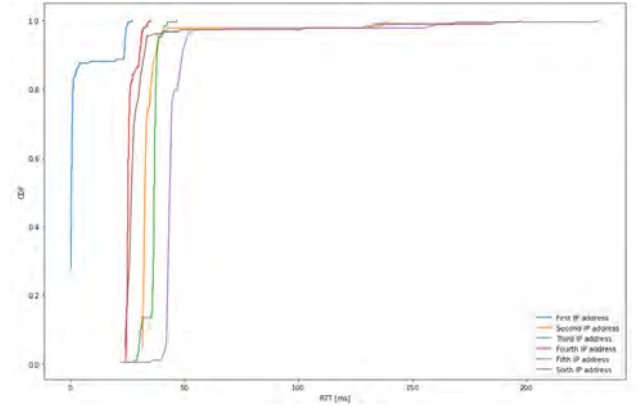


Fig. 5. Distribution of all RTTs in the second segment for all destinations (traceroute)

The shortest network latency is measured to the first IP address. Similar network latency is presented by the sixth and fourth IP addresses, but with slightly longer network latencies. The second and third IP addresses present visibly greater network latencies, but similar among each other. The highest network latency is measured when the packet travels to the fifth IP address (Geneva, Switzerland - 213ms). However, we can note that most of the measured RTTs for the second segment are under 50 msec which is way below the 100 msec human perceivable latency (HPL) threshold as elaborated in [3], and is acceptable for large number of applications. Some high values (spikes) are reported when the packet travels to the second, fifth, or sixth IP address, possibly due to the load of some data center or the congestion on the network.

Lastly, in Fig. 6 we illustrate the distribution of all network latencies for all destinations for the **third (cloud) segment**. The time it takes for a packet to travel "in" the cloud, from the moment of entry until it reaches its final destination in the cloud infrastructure, is similar for each packet traveling to each VPS in the six data centers. As expected, this usually presents as a very short latency (several msec) due to the good network configuration inside the cloud infrastructure in each datacenter location.

High values (spikes) that report bad events when the packet travels to the second, fifth, and sixth IP address can also be observed, possibly due to the load of a data center, congestion on the intra-network, or due to redundant network-

equipment in the cloud data centers, configured for reliability and speed. The traceroute can not distinguish between redundant hops in the network, thus reporting RTTs for all of them.

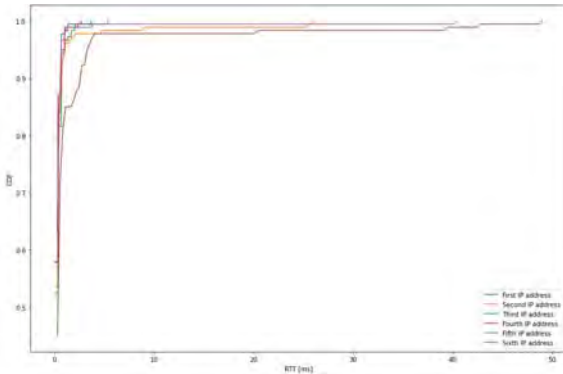


Fig. 6. Distribution of all RTTs in the third segment for all destinations (traceroute)

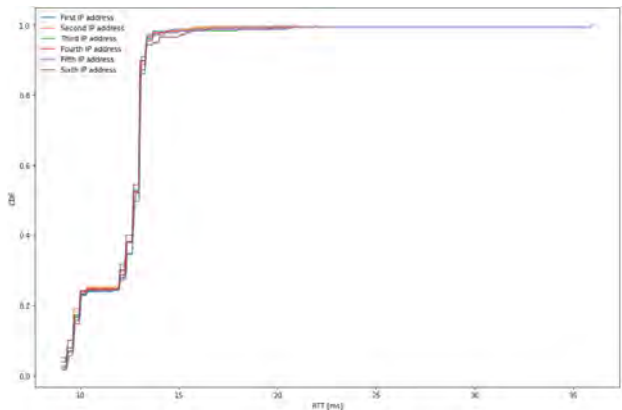


Fig. 7. Distribution of the RTTs from the ISP's exit point to the next hop (in Bulgaria) (traceroute)

The second part of the observations is focused on cumulative RTTs (not by segments). We also characterized the RTT needed for the packets to travel from the exit point (hop) of the end-user's ISP to the next hop (in all our measurements the next hop after ISP exit is located in Bulgaria). From Fig. 7 we can observe that this RTT is very similar for all six final destinations. The observed differences and spikes can be attributed to the load of some network device and congestion on the network. This is especially observed in the sixth IP address (Munich, Germany - 241ms). The CDFs presented in Fig. 4 and Fig. 7 have similar distributions since the time that the packet needs to leave the first segment is similar to the time that it needs to travel to the next-hop IP address in Bulgaria.

In Fig. 8 we present the distribution of the RTTs needed for the packet to travel from the end-user through the network until the packet enters the cloud infrastructure (first and second segment). The packets traveling to the first IP address present the shortest network latency. The packets that travel to the other five geographical locations have similar network latency. Moreover, it can be observed that there are high values (spikes) for the second, fifth, and sixth IP addresses.

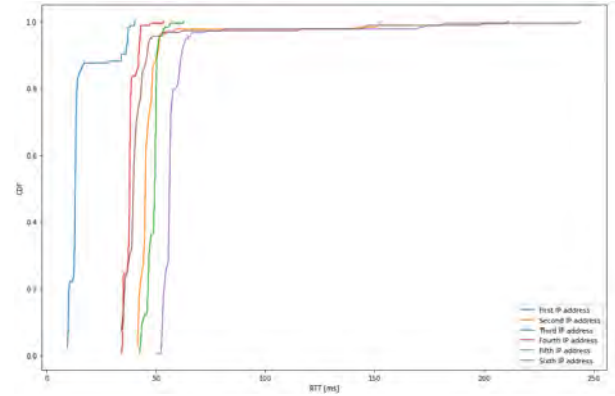


Fig. 8. Distribution of the RTTs from the end-user to the entry point of the cloud infrastructure (traceroute)

Finally, in Fig. 9 we evaluate the RTTs for the entire trip of the packet from the end-user, through the network, until the final destination VPS in the Exoscale cloud. It can be observed that the Fig. 8 and figure Fig. 9 have a very similar visual footprint, so we can conclude that the time that packet spent in the cloud is significantly shorter than the time that the packet spent to reach the cloud. This was also true for the RTTs of the third (cloud) segment. Thus, this last segment does not add much variation on the overall trip of the packet.

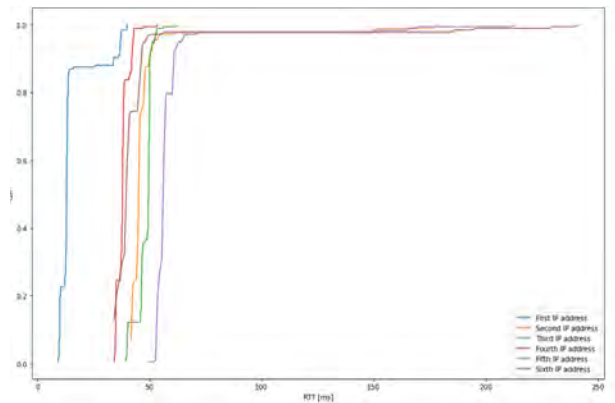


Fig. 9. Distribution of the RTTs for the full packet trip to the final destination (traceroute)

Finally, we present the results obtained from the ping diagnostic tool. It is interesting to compare the results from both diagnostic tools, especially since both have been used at the same time intervals and started just seconds apart. Fig. 10 shows the distribution of all RTTs recorded by the ping diagnostic tool sent from the end-user and directed to the six VPSs as final destinations in the cloud infrastructure. It can be observed that this figure and Fig. 9 have similar distribution as expected. Both distributions have been plotted side by side in Fig. 11 in order to visualize the differences. Fig. 11 depicts the RTT comparison for the full packet trip from the traceroute and ping diagnostic tools. As it can be seen the distributions of the total trip time of a packet from the end-user, through the network, until the final VPS destination in the cloud obtained from traceroute and ping is similar, but with small differences that probably result from the different techniques that traceroute and ping use to measure the RTTs.

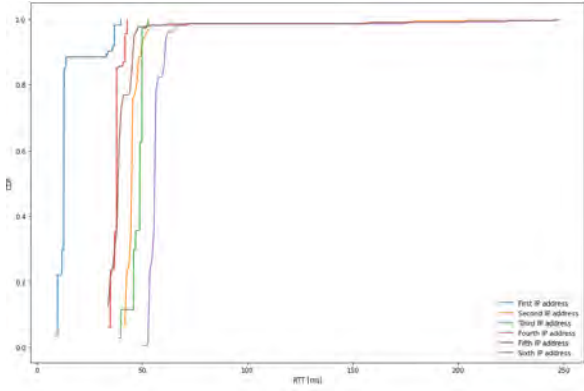


Fig. 10. Distribution of the RTTs for the full packet trip to the final destination (ping)

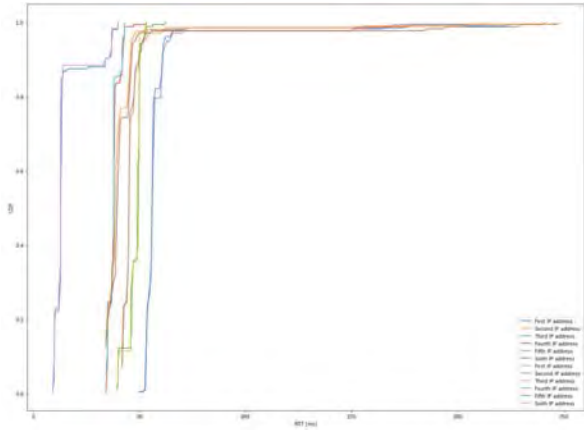


Fig. 11. Comparison of RTTs for the full packet trip to the final destination (ping v.s. traceroute)

V. CONCLUSION

The main goal in the paper was to characterize the network latency to the cloud and to conclude if the cloud services offered fall within the range of Human Perceivable Latency (HPL) for real-time cloud services. The latency measurements of the three identified segments (user, middle, cloud) have shown that they satisfy the boundaries of HPL [3] (between 50-100 msec) and can support real-time applications, even for East European end-users. Most of the VPS locations, even though geographically apart, present similar behavior. The geographically closest location in Bulgaria always presents best results. On one side this location is geographically closest, but on the other, probably the end-user ISP has a direct physical network link to this geographical location. The geographical locations Geneva, Switzerland - 213ms [fifth IP address] and Munich, Germany - 241ms [sixth IP address] usually present the worst metrics, most probably due to the greater number of hops. The results show that even though the ISPs on one side, and cloud providers on the other, enable good configuration of their own networks, the free Internet path of the packets has great influence on the overall network latency perceived by the user when accessing cloud services.

Some potential directions for future exploration is expanding the number of end-users and cloud platforms, while

keeping the European focus due to the geographical vicinity of the users and the European legal GDPR requirements. One further approach would be to apply machine learning techniques by exploiting the measurement latency data in order to predict the network latency and availability of cloud services. Lastly, we can further investigate the edge computing paradigm and give an assessment of network latency variations for users that will direct their requests to edge servers and cloud data centers estimating the network latency benefits of edge servers opposite to cloud data centers.

REFERENCES

- [1] Google Stadia, <https://stadia.dev/>. (2019).
- [2] Microsoft Xbox Project xCloud. <https://www.techradar.com/news/project-xcloud-everything-we-knowabout-microsofts-cloud-streaming-service>, (2019)
- [3] L. Corneo, M. Eder, N. Mohan, A. Zavodovski, S. Bayhan, W. Wong, P. Gunningberg, J. Kangasharju, and J. Ott, "Surrounded by the Clouds: A Comprehensive Cloud Reachability Study", in Proceedings of the Web Conference 2021, Ljubljana, Slovenia. April 19–23, 2021.
- [4] N. Mohan, L. Corneo, A. Zavodovski, S. Bayhan, W. Wong, and J. Kangasharju, "Pruning Edge Research with Latency Shears", in Proceedings of the 19th ACM Workshop on Hot Topics in Networks, Association for Computing Machinery, NY, USA, pp.182–189, 2020.
- [5] K. Raaen, R. Eg, and C. Griwodz, "Can gamers detect cloud delay?". In Proceedings of 13th Annual Workshop on Network and Systems Support for Games. IEEE, pp. 1–3, 2013.
- [6] D. L. Woods, J. M. Wyma, E. W. Yund, T. J. Herron, and B. Reed. "Factors influencing the latency of simple reaction time", *Frontiers in human neuroscience* 9, 131, 2015.
- [7] Y. Jin, S. Renganathan, G. Ananthanarayanan, J. Jiang, V. N. Padmanabhan, M. Schroder, M. Calder, and A. Krishnamurthy, "Zooming in on Wide-area Latencies at a Global Cloud Provider", in Proceedings of ACM SIGCOMM 2019 Conference, Beijing, China, August 19–23, 2019
- [8] F. Palumbo, G. Aceto, A. Botta, D. Ciuonzo, V. Persico, and A. Pescapé, "Characterizing Cloud-to-user Latency as perceived by AWS and Azure Users spread over the Globe", in proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1-6, 2019.
- [9] F. Palumbo, G. Aceto, A. Botta, D. Ciuonzo, V. Persico, A. Pescapé, "Characterization and analysis of cloud-to-user latency: The case of Azure and AWS", *Computer Networks*, vol. 184, Elsevier, Jan 2021.
- [10] B. Karacali, J. M. Tracey, P. G. Crumley, and C. Basso, "Assessing cloud network performance." in IEEE ICC'18, 2018.
- [11] C. Guo, L. Yuan, D. Xiang, Y. Dang, R. Huang, D. Maltz, Z. Liu, V. Wang, B. Pang, H. Chen, Z. Lin and V. Kurien, "Pingmesh: A Large-Scale System for Data Center Network Latency Measurement and Analysis", *ACM SIGCOMM Computer Communication Review*, pp. 139–152, 2015.
- [12] Y. A. Wang, C. Huang, J. Li, and K. W. Ross, "Estimating the performance of hypothetical cloud service deployments: A measurement-based approach", *IEEE INFOCOM'11*, 2011.
- [13] V. Persico, A. Botta, P. Marchetta, A. Montieri, A. Pescapé, "On the performance of the wide-area networks interconnecting public-cloud datacenters around the globe", *Computer Networks*, vol. 112, pp. 67–83, 2017.
- [14] S. Choy, B. Wong, G. Simon, and C. Rosenberg, "The brewing storm in cloud gaming: A measurement study on cloud to end-user latency," in *IEEE/ACM NetGames*, 2012.
- [15] Dennis G. Jansen, Whitepaper: Cloud Act vs. GDPR <https://www.a1.digital/en-de/blog/whitepaper-cloud-act-vs-gdp>
- [16] E. G. Strom, P. Popovski, and J. Sachs, "5g ultra-reliable vehicular communication", *arXiv:1510.01288 [cs.IT]*, 2015.



ETAI 8: ARTIFICIAL INTELIGENCE IN AUTOMATION

Forecasting Dynamic Tourism Demand using Artificial Neural Networks

Cvetko Andreeski

University “St. Kliment Ohridski” – Bitola, Faculty of
Tourism and Hospitality
Ohrid, North Macedonia
cvetko.andreeski@uklo.edu.mk

Biljana Petrevska

University “Goce Delčev” – Štip, Faculty of Tourism and
Business Logistics
Štip, North Macedonia
biljana.petrevska@ugd.edu.mk

Abstract—Planning tourism development means preparing the destination for coping with uncertainties as tourism is sensitive to many changes. This study tested two types of artificial neural networks in modeling international tourist arrivals recorded in Ohrid (North Macedonia) during 2010-2019. It argues that the MultiLayer Perceptron (MLP) network is more accurate than the Nonlinear AutoRegressive eXogenous (NARX) model when forecasting tourism demand. The research reveals that the bigger the number of neurons may not necessarily lead to further performance improvement of the model. The MLP network for its better performance in modelling series with unexpected challenges is highly recommended for forecasting dynamic tourism demand.

Keywords— *Time series, Tourism demand, Tourism planning, Modeling, COVID-19.*

I. INTRODUCTION

Planning tourism development particularly in turbulent times during and after the COVID-19 (declared as a pandemic by the WHO, 12 March 2020), becomes not an easy task. Tourism as one of the most important contributors to the world's economy was found to be extremely fragile and vulnerable, facing enormous losses leading to a worldwide recession and depression. A severe drop in international tourist arrivals (estimations to -78%) and an enormous loss of US\$ 1.2 trillion in export revenues from tourism, is the largest decline ever [1]. It may take a while before tourism will start again to generate a large financial portion in exports and job creation since COVID-19 provoked many transformations to global economic, socio-cultural, and political systems.

Tourism planners and policy-makers are already eager to continue the forecasting process as a way to furnish information for recovering exhausted economies. Creating solid tourism development plans based on accurate forecasted values envisages success and quick recovery. It is often a case, tourism development to be interrupted for various crises (e.g. terrorism, SARS, natural disasters, earthquakes, political conflicts, Ebola, regional instability, etc.), thus, provoking a structural change in the tourism time series. This disables smooth prediction of tourism values and modeling the series and makes it difficult to analyze expected tourism development. Currently, due to the many measures and strategies related to the COVID-19 (e.g. social distancing,

national lockdowns, quarantine, mobility bans etc.), tourism has never experienced such a global collapse. Despite studies that argue the importance of managing the pandemic and finding another context for reimagining and transforming tourism to go a step beyond [2], [3], the inability to create a valid tourism forecasting model will continue long after the pandemic is gone. Structural changes interrupt the series, and the new trend rapidly differs from the previous one.

Many studies explore forecasting models, generally to assist in mitigating the potential negative impacts for the planning process. Although the classical linear models for the identification of time series, such as the ARIMA model [4], can be used in such cases, their application becomes quite complex due to the need to identify all individual structural changes and their influence on the series. Often, modeled series have poor performance in forecasting values [4],[5]. Classical models are linear and therefore unable to model the built-in nonlinear nature of certain time series [6]. On the other hand, models based on artificial neural networks (ANN) can be applied to both, linear and nonlinear time series.

In general, scholars apply the ANN and argue their suitability for forecasting in various fields, but with no focus on an in-depth identification of the cause that makes the model simple and more accurate. This study fills this gap by determining whether the greater number of neurons contributes to better results in modeling and forecasting. To this end, the research tests two types of ANN – the MultiLayer Perceptron network (MLP) and the Nonlinear AutoRegressive eXogenous model (NARX). The main research aim is to identify which model better describes and forecasts international tourism demand. The case of Ohrid is elaborated, as the most popular tourist destination in North Macedonia. Besides adding to the literature on forecasting methods, this study contributes to the scarce empirical academic work in North Macedonia, with some exceptions [7],[8],[9].

The paper is structured as follows: after the introduction, Section 2 provides a brief overview of the literature on forecasting models. Section 3 presents background material on the case study selected for the analysis, i.e., Ohrid as a top tourist destination in North Macedonia. The description of the applied methodology in terms of data and models is presented in Section 4. Section 5 covers the modeling, main results, and

discussion, while the conclusion and some future issues to be discussed are drawn in the final section

II. LITERATURE REVIEW

Forecasting tourism demand is vastly explored in academia. The forecasting methodology varies as scholars employ both the time series and econometric approaches in predicting tourism demand [21]. Often, a combined forecast is advocated for obtaining more accurate models [10],[11],[12].

On the other side, any change in the level or variance of the series is considered a structural change, and the analyzed series is not stationary in the entire analyzed period [5]. Nonlinear models can identify series that have a change in the level or variance of the series and are therefore suitable for modeling complex time series with structural changes [5],[12]. Neural network models are not limited to some specific type of series or some specific field of research. Yet, numerous studies use different types of neural networks to model tourism time series [13],[15],[5]. In [15] three types of neural networks are tested: multi-layer perceptron network, a radial basis function network and an Elman neural network to determine which one gives the best results in predicting future values of the series. The authors in [16] analyze the series on rural tourism by using the multi-layer perceptron network. [17] propose a Bayesian estimation and prediction procedure and assume that even in the period of forecasting future values, the possibility of structural changes should be considered.

Although indicators for describing tourism demand differ in academia, the most applicable one is the tourist arrivals. This is further decomposed into in-depth variables as holiday tourist arrivals, business tourist arrivals, as well as tourist arrivals for visiting friends and relatives [18],[19].

III. DATASET

Ohrid is the most famous tourist destination in North Macedonia. Due to favorable natural attractors (sun and lake) along with many additional factors (usage of vacations and ferries, personal preferences for summer season, etc.), Ohrid generally develops summer tourism simultaneously with other tourism forms (cultural, congress, etc.). The peak points for the international tourism demand are visible in the third quarter (summer months July-September) (Figure 2). So Ohrid is characterized by an unequal seasonal distribution of tourist arrivals and the presence of strong and powerful seasonality [20],[21].

For its exceptional natural values, first in 1979, and then in 1980 for its cultural and historical area, the Lake Ohrid region was inscribed as a transboundary mixed UNESCO property [22]. This adds value to this site in attracting tourists. In 2019, before the COVID-19, Ohrid accounted for a quarter of all tourist arrivals (322,573) and for almost one-third of all registered overnights in the country (1,101,563) [23]. 59.5% of all registered tourists in 2019 were foreigners, while in 2020 due to the total COVID-19 lockdown, this rapidly decreased to only 9.6% [24]. As such, Ohrid was experiencing a complete fiasco in its tourism development.

Figure 1 depicts the COVID-19 dramatic reshape in the international tourism demand as of March 2020 when a huge decrease of 66% was noted. The decrease was even more profound in April and May 2020 with less than 0.001%, and June with less than 0.1% of foreign tourists being registered compared

to the same months in 2019. During July-December 2020, only 3-6% of foreigners were registered compared to 2019 [24]. This practically meant that Ohrid had no foreigners that season and no tourism development at all. So, COVID-19 has been so far the most significant crisis provoking unforeseen trajectories. This requires a redesign of tourism policy and building a new model since the 'old' exploratory models may be outdated.

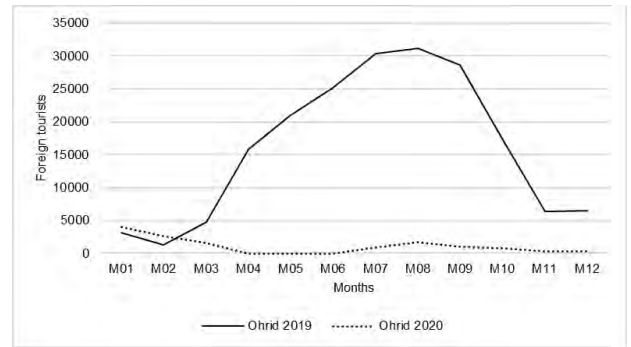


Fig 1. International Tourism Demand in Ohrid, 2019 vs. 2020

A good model of the series before 2020 and forecasting for 2020 and 2021 can give us information about loss of income in tourism sector if we compare forecasted and real data.

The research is based on available official statistical data further processed by the software E-views and Matlab. The original time series is the number of foreign tourists per month being registered in Ohrid in the period 2010-2019 (Figure 2). Data of 2020 are disregarded due to the long-standing structural change in the series provoked by the COVID-19. It is a common standpoint to omit structural breaks which do not allow good modeling of the series based on its previous values [13],[25].

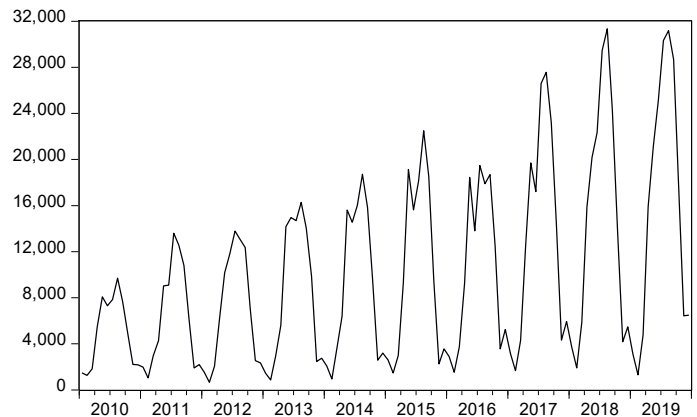


Fig 2. Monthly International Tourist Arrivals in Ohrid, 2010-2019

Based on Figure 2, several features of the series can be noticed: (i) The series is growing, i.e. there is a positive trend in almost the entire analyzed period; (ii) The series is heteroskedastic, the variance increases over time; (iii) The series has a seasonal character, i.e. every year the seasonality is expressed; and (iv) There is a change in the level in 2016 which indicates a possible structural change.

The first three features are visually evident from Figure 2, but the fourth assertion is tested by performing a Breakpoint Unit Root Test (Table 1). This test detects change of levels and trends that differ across a single break date. In combination with Dickey Fuller t-test we can detect significant change in the level or trend of the series at a certain point.

Table 1. Breakpoint Unit Root Test

Null Hypothesis: FOREIGN has a unit root		
Trend Specification: Intercept only		
Break Specification: Intercept only		
Break Type: Innovational outlier		
Break Date: 2011M02		
Break Selection: Minimize Dickey-Fuller t-statistic		
Lag Length: 0 (Automatic - based on Schwarz information criterion, maxlag=12)		
Augmented Dickey-Fuller test statistic	t-Statistic	Prob.
	-3.429158	0.4250
Test critical values:		
1% level	-4.949133	
5% level	-4.443649	
10% level	-4.193627	

The analysis of the structural change indicates a presence of a robust structural change in 2011 (Figure 3). After the World economic crisis in 2010, the government introduced a set of financial measures to support tourism development. The national Agency for Promotion and Support of Tourism introduced a new Rulebook to subsidize incoming tourism. So as of 2011, all tourism arrangements agreed between national incoming agencies and foreign tour operators were substantially subsidized, thus supporting tourism development in the country. This explains the structural change that occurred in 2011.

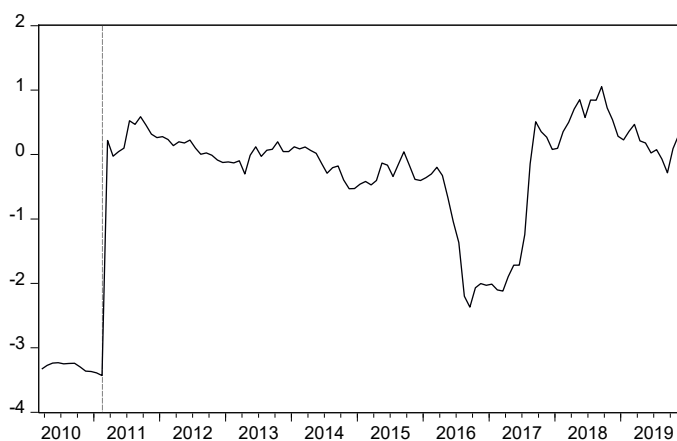


Fig 3. Dickey-Fuller t-statistics

A closer look at the period 2016-2017 (Figure 3), puts a shed-light for a second potential structural change. To check the presence of such, the series was shortened to 2012-2019 and the Breakpoint Unit Root Test was re-performed only to this segment of the series (Table 2).

Results in Table 2, and the visual presentation in Figure 4, point to a conclusion for the presence of another structural

change, this time in the first quarter of 2017. There isn't any known causal event that we can mention for this structural break. There can be several different events that should cause such a break like: canceled flights, bad weather conditions, reduced number of airlines, change in the interest of tourists from important countries, etc.

Table 2. Breakpoint Unit Root Test, Cropped Time Series

Null Hypothesis: FOREIGN has a unit root		
Trend Specification: Intercept only		
Break Specification: Intercept only		
Break Type: Innovational outlier		
Break Date: 2017M03		
Break Selection: Minimize Dickey-Fuller t-statistic		
Lag Length: 11 (Automatic - based on Schwarz information criterion, maxlag=11)		
Augmented Dickey-Fuller test statistic	t-Statistic	Prob.
	-2.492220	0.9049
Test critical values:		
1% level	-4.949133	
5% level	-4.443649	
10% level	-4.193627	

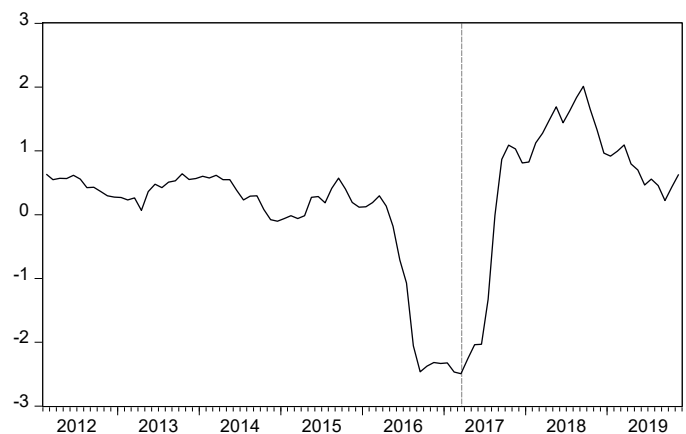






































Fig 4. Dickey-Fuller t-statistics

Concluding that the analyzed time series has a presence of seasonality and two structural changes, the built-in character makes the time series unsuitable for linear analysis with the ARIMA model [26][27][28]. The complex nature of the series itself indicates to model with nonlinear models that can detect all confirmed features of the series without having to do preprocessing of batch data.

In order to detect valid inputs, we made an correlogram of the lags. Results are given in table 3. From the values given in the table, we can conclude that there is a serial correlation pattern in the lags of the correlogram, and the 12th lag is significant, and it should be part of the inputs. These results are expected according to the emphasized seasonality of the analyzed series.

Table 3. Correlogram of the analyzed series

Sample: 2010M01 2019M12
Included observations: 119

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.253	0.253	7.8245	0.005
		2	0.004	-0.064	7.8268	0.020
		3	0.089	0.111	8.7993	0.032
		4	-0.168	-0.239	12.320	0.015
		5	-0.355	-0.270	28.253	0.000
		6	-0.574	-0.553	70.291	0.000
		7	-0.361	-0.303	87.079	0.000
		8	-0.183	-0.376	91.413	0.000
		9	0.094	0.090	92.561	0.000
		10	0.006	-0.508	92.565	0.000
		11	0.265	-0.160	101.94	0.000
		12	0.857	0.546	200.77	0.000
		13	0.237	-0.109	208.39	0.000
		14	0.022	-0.104	208.45	0.000
		15	0.085	-0.062	209.44	0.000
		16	-0.151	-0.122	212.65	0.000
		17	-0.328	0.005	227.84	0.000
		18	-0.511	0.057	265.03	0.000

IV. ANN MODELS

So the research applied two types of ANN models, the MLP and the NARX. For both networks, the input data, and the series to be modeled are identical.

The first network model is the MLP (figure 5) that uses a sigmoid function in the hidden level, linear at the output, two inputs, one output (target values) without feedback, and the way to set the network parameters is by gradient descent training process. The input parameters in the model are the first and 12th delays of the values in the series. Their selection is made based on previous analysis of autocorrelation and partial autocorrelation analysis of delays (Table 3). The series has a serial autocorrelation, for which the first delay is used, and for the seasonal component of the series, the 12th lag is used. Batch heteroskedasticity can be removed by preprocessing batch values using a logarithmic function [29], but nonlinear models can adapt inner values to the variance change without pre-processing of input data [15],[30][16]. No series stationing has been done, as nonlinear models can model non-stationary series. The only change to the original time series is to normalize the series using the maximum value method. The series was divided on three parts: training part 7 years, testing part 18 months and forecasting part 18 months.

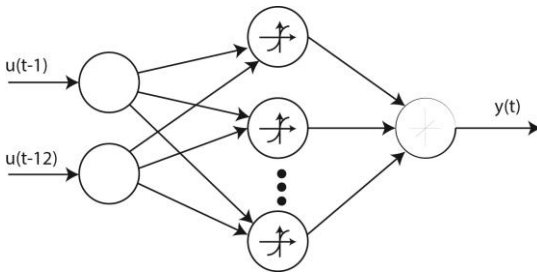


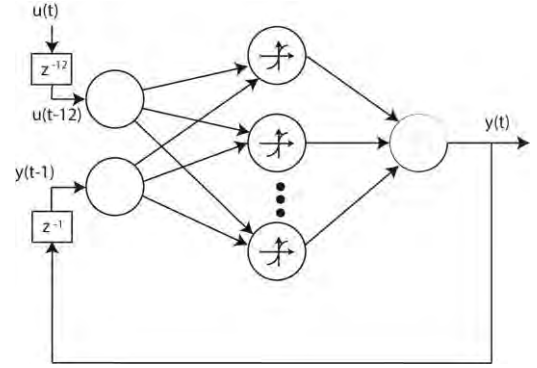
Figure 5. MLP network for time series modelling

The second model is the NARX neural network (figure 6), which is a recurrent neural network that uses exogenous values at the input. Concerning linear ARMA models, this

network provides the possibility to use autoregressive parameters in time series modeling. These networks are intended for modeling dynamic nonlinear systems and are widely applied [31][32]. Yet, this network does not have the Moving Average (MA) component of the linear model but can model the non-linear behavior of the series. The basic formula for determining the output values from the network is given by (1).

$$y(t)=f(y(t-1),y(t-2),\dots,y(t-n_y),u(t-1),u(t-2),\dots,u(t-n_u)) \quad (1)$$

where $y(t)$ is the value of the output at moment t , and $u(t)$ is the value of the exogenous input at moment t .

Figure 6. NARX network for time series modelling with one delayed input of 12th lag and one delayed output

For our NARX network, as inputs we use the 12th delay of the input series, and the first delay of the output $y(t-1)$. The recurrent input is intended for elimination of serial correlation, and the input is another valid lag for time series modelling according to the results of autocorrelation table.

Both networks were trained using the Levenberg Marquardt - LM optimization algorithm, which enables faster adjustment of the network weights, using larger memory. As the series is not large, this method is optimal for faster modeling results. Networks with 3, 4, 5, and 10 neurons in the hidden level were used for modeling, testing, and forecasting. The Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) calculations were used to measure the modeling results, test the model, and predict future values for the model output error, relative to the original series. In the process of model testing, the so-called 'In-sample' forecasting is done. Opposing, in the forecasting process, the network output is closed at the input, and out of sample forecasting is done to calculate the real error of the model in predicting data. Those data were not part of the series used for adjustment of internal parameters.

In the NARX neural network, one network delay is used to eliminate the serial correlation and different initial values are used to determine whether they will lead to better results in modeling and forecasting. Only the best values are presented and elaborated. Due to the detected serial correlation in the series, a dynamic one-step-ahead prediction was used. Both series have one output at the output level of the network, which is sufficient for forecasting values with one-step ahead.

V. MODELING RESULTS AND DISCUSSION

Table 4 presents the modeling results of the series, with the MLP model, with different numbers of neurons (3, 4, 5, and 10) in the hidden level to determine whether a larger number of neurons affects the model performance. The values of the parameter R^2 are also presented to identify the degree of variance modeling of the original series.

Table 4. Parameters of the MLP network

Neurons	Process	R	R^2	RMSE	MAPE
3	Training	9.72E-01	9.44E-01	1914.467	7479.918
	Valid.	9.94E-01	9.88E-01	1196.718	5489.201
	Forec.	9.69E-01	9.39E-01	1743.005	7096.724
4	Training	9.81E-01	9.62E-01	1666.191	5564.072
	Valid.	9.81E-01	9.62E-01	1692.736	5197.05
	Forec.	9.82E-01	9.65E-01	1116.864	4592.031
5	Training	9.80E-01	9.60E-01	1683.177	4502.398
	Valid.	9.93E-01	9.87E-01	1040.613	5166.994
	Forec.	9.68E-01	9.36E-01	1617.207	3574.327
10	Training	9.83E-01	9.65E-01	1580.946	5717.576
	Valid.	9.72E-01	9.46E-01	1231.411	6004.149
	Forec.	9.89E-01	9.78E-01	1648.797	3847.769

Figure 7 visually presents the errors (RMSE and MAPE) for the forecasted values by the MLP model. According to the presented values of the errors, the network with four neurons in the hidden level has better results compared to all others, because the value of RMSE error is lowest compared to other networks, MAPE error is close to the lowest value, and the R^2 parameter has higher value than the model with three or five neurons. So, increasing the number of neurons in the hidden layer to some extent improves the performance of the network, in terms of better prediction.

Table 5 presents the corresponding parameters for the NARX network.

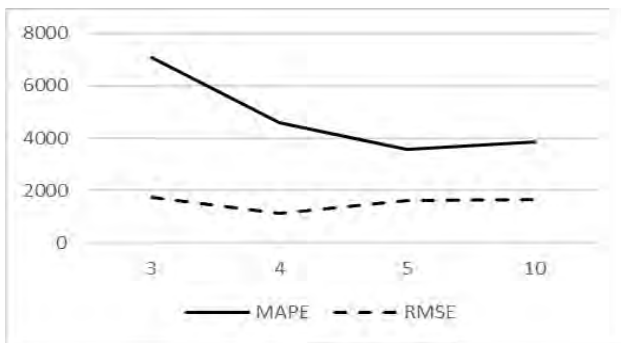


Fig 7. RMSE and MAPE errors of the time series with the MLP network

Yet, Figure 8 gives a glance that a bigger number of neurons than five does not necessarily lead to further performance improvement. The same conclusion derives when screening the degree of follow-up of the variance of the predictions (Table 4, MLP values). Namely, the R^2 does not increase.

Table 5. Parameters of the NARX network

Neurons	Process	R	R^2	RMSE	MAPE
3	Training	8.26E-01	6.83E-01	4758.3522	12502.686
	Valid.	8.03E-01	6.45E-01	4829.5567	6201.3085
	Forec.	9.32E-01	8.68E-01	2845.4526	7393.1373
4	Training	8.31E-01	6.90E-01	4663.392	1411.7103
	Valid.	8.64E-01	7.47E-01	4556.742	2416.408
	Forec.	9.19E-01	8.44E-01	3235.7296	2109.4327
5	Training	8.25E-01	6.81E-01	4730.9297	14187.085
	Valid.	8.42E-01	7.10E-01	4201.033	9118.467
	Forec.	8.58E-01	7.35E-01	14151.295	10143.034
10	Training	8.15E-01	6.64E-01	5031.637	5587.0758
	Valid.	8.98E-01	8.07E-01	4738.0125	2553.4265
	Forec.	8.61E-01	7.42E-01	4338.8036	3064.3625

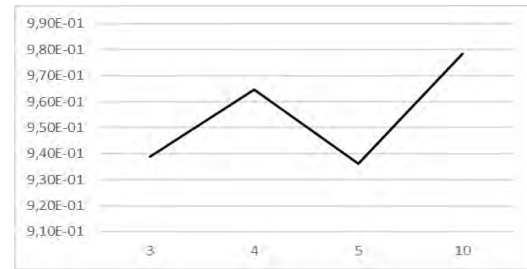


Fig 8. R^2 Parameter for the forecasting with the MLP network

Figure 9 visually presents the errors for forecasted values with the NARX network, where the network with four neurons in the hidden level has better-comparing results. In the NARX networks, there is no defined tendency for the error to decrease or increase with different number of neurons in the hidden layer. So, the network with four neurons in the hidden level shows the best results. These values are not followed by the parameter R^2 presented in figure 10. This parameter decreases its values as the number of neurons in hidden layer increase. The values of R^2 parameter is much lower for the model of NARX network, compared with the same parameter of MLP network.

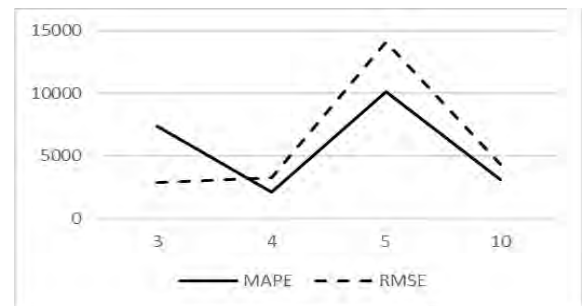


Fig 9. RMSE and MAPE errors of the time series with the NARX network

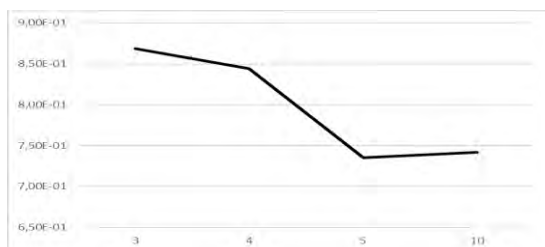


Fig 10. R^2 Parameter for the forecasting with the NARX network

Finally, when comparing the results of the modeling and prediction of the series made with two different types of neural networks, it may be concluded that the MLP network offers significantly better forecasting results than the NARX network. Values of RMSE error are lower for the MLP network in comparison with NARX error. According to the values of MAPE error, NARX networks give better results, but the MLP network gives us information about the optimal number of neurons in the hidden layer.

Despite many scholars who recommend that the NARX networks are suitable for modeling dynamic systems or time series with sufficiently rich input [22],[31], this research revealed the opposite, but only for time series with previously discussed features. The time series that we analyze in this paper has 120 input data. On the other hand, the most complex network that we use has $2 \times 10 + 10 = 30$ internal variables (weights). We have four time more input data than the number of weights that ensures sufficiently rich input. In cases of dynamic tourism time series with structural breaks and uncertain trends, the MLP network provides better results in forecasting tourism demand.

VI. CONCLUSION

Planning tourism development, particularly in times of uncertainties like the COVID-19 pandemics, must be relied on consistent forecasting values. Due to fact that tourism trend is often interrupted by structural changes, linear models are disabled to successfully model the original time series, particularly if missing sufficient data after the occurrence of the structural change. However, lasting changes in structure of the series prevent any known model on identification and forecasting of different behavior of the same series. Periods of crisis, such as the current COVID-19 pandemics, require models that after completing the change in a relatively short time will be able to make valid modeling of the series and predict future values. Neural networks, due to the nonlinear functions used in creating the model, are suitable for modeling complex time series that have short time built-in structural changes, an evident trend in the series, and the occurrence of heteroskedasticity.

This study employed two artificial neural networks (MLP and NARX) to investigate their accuracy when forecasting international tourism demand for the city of Ohrid, the most popular tourist destination in North Macedonia. By employing monthly data for the international tourist arrivals for the period 2010-2019, the study elaborated and found that generally,

does not mean that more neurons will result in better model performance. According to the number of neurons in the hidden level, it is necessary to determine the optimal number of neurons to obtain the optimal solution. The bigger the number of neurons may not lead to further performance improvement of the model.

Moreover, the study argued that the MLP network is more accurate compared to the NARX network and suggests applying this model more intensively when forecasting tourism demand. Further, it practically raises the need for using the ANN for predicting tourism values, particularly the MLP network for its better performance in modeling series when unexpected short-term challenges occur. Totally different behavior of the series are more challenging, and in the period of lasting different behavior impossible to identify and predict. However, in these challenging periods, we can compare actual and forecasted data to be able to detect the losses and to make decisions about support of tourism.

Some further refining in forecasting may be additionally added if employing the Convolutional neural networks for batch modeling. The research may be upgraded with a larger number of time series with similar characteristics to obtain more information on the benefits of different series modeling networks with several structural changes.

REFERENCES

- [1]. UNWTO. *UNWTO World Tourism Barometer* (Vol. 18, Issue 2, May 2020). Madrid, Spain: UNWTO, 2020
- [2]. Gössling, S., Scott, D., and Hall, C.M., „Pandemics, tourism and global change: a rapid assessment of COVID-19”. *Journal of Sustainable Tourism*, <https://doi.org/10.1080/09669582.2020.1758708>, 2020
- [3]. Hall, C.M., Scott, D., and Gössling, S., „Pandemics, transformations and tourism: be careful what you wish for”. *Tourism Geographies*. <https://doi.org/10.1080/14616688.2020.1759131>, 2020
- [4]. Casini, A. and Pierre, P., „Structural Breaks in Time Series. In *Oxford Research Encyclopedia of Economics and Finance*”. Boston: Palgrave Macmillan., 2018
- [5]. Hang Xie, Hao Tang and Yu-He Liao, "Time series prediction based on NARX neural networks: An advanced approach," *International Conference on Machine Learning and Cybernetics*, Baoding, China, pp. 1275-1279, 2009
- [6]. Zhang, P., *Neural Networks for Time-Series Forecasting*. Berlin: Springer. 2012
- [7]. Petrevska, B., „Forecasting International Tourism Demand: the Evidence of Macedonia”. *UTMS Journal of Economics*, 3(1), pp. 45-55. 2012
- [8]. Petrevska, B., „Estimating tourism demand: the case of FYROM. *Tourismos*” *An International Multidisciplinary Journal of Tourism*, 8(1), pp. 199-212. 2013
- [9]. Petrevska, B., „Predicting tourism demand by ARIMA models”. *Economic research-Ekonomska istraživanja*, 30(1), pp. 939-950., 2017
- [10]. Song, H., Witt, S.F., Wong, K.K.F. and Wu, D.C., „An empirical study of forecast combination in tourism”. *Journal of Hospitality & Tourism Research*, 33(1), pp. 3-29., 2008
- [11]. Song, H. and Li, G., „Tourism demand modelling and forecasting” – A review of Recent research. *Tourism Management*, 29(2), pp. 203-220., 2008

- [12]. Wong, K., Song, H., Witt, S.F. and Wu, D.C., 2007. Tourism forecasting: to combine or not to combine? *Tourism Management*, 28(4), pp. 1068-1078.
- [13]. Asghar, Y. and Amena, U., „Structural Breaks, Automatic Model Selection and Forecasting Wheat and Rice Prices for Pakistan”. *Pakistan Journal of Statistics and Operation Research*, pp. 1-20., 2012
- [14]. Falat, L., Stanikova, Z., Durisova, M., Holkova, B. and Potkanova, T., „Application of Neural Network Models in Modelling Economic Time Series with Non-constant Volatility”. *Procedia Economics and Finance* 34, pp. 600-607., 2015
- [15]. Claveria O., Monte, E., and Torra, S. "A multivariate neural network approach to tourism demand forecasting". Barcelona: University of Barcelona, Regional Quantitative Analysis Group., 2014
- [16]. Shi, X. „Tourism culture and demand forecasting based on BP neural network mining algorithms”. *Personal and Ubiquitous Computing*, pp. 1-10., 2019
- [17]. Timmermann, A., Pettenuzzo, D. and Pesaran, M.H., „Forecasting time series subject to multiple structural breaks” (No. 1237). CESifo Working Paper. 2004
- [18]. Kulendran, N. and Wong K.K.F. „Modeling Seasonality in Tourism Forecasting”. *Journal of Travel Research*, 44, pp. 163-170., 2005
- [19]. Turner, L.W. and Witt, S.F., „Factors influencing demand for international tourism: Tourism demand analysis using structural equation modelling”, Revisited. *Tourism Economics*, 7, pp. 21-38. 2001a
- [20]. Petrevska, B. and Nikolovski, B., „Level of seasonality in Macedonian tourism and strategies and policies for coping with it”. 3rd International Thematic Monograph: “Modern Management Tools and Economy of Tourism Sector in Present Era”, Association of Economists and Managers of the Balkans (Belgrade, Serbia) & Faculty of Tourism and Hospitality – Ohrid, Macedonia, pp. 17-29., 2018
- [21]. Petrevska, B. „Effects of tourism seasonality at local level.” Scientific Annals of the “Alexandru Ioan Cuza” University of Iasi, Economic Sciences Series, 62(2), pp. 241-250., 2015
- [22]. UNESCO. [online]. World Heritage List. Available at: <<https://whc.unesco.org/en/list/>>.
- [23]. Statistical Yearbook for [online]. Available at: <<http://www.stat.gov.mk/Publikacii/SG2020/SG2020-Pdf/14-TransTurVnatr-TransTourTrade.pdf>>., 2019
- [24]. Statistical data for 2020. [online]. Available at: <http://makstat.stat.gov.mk/PXWeb/pxweb/mk/MakStat/MakStat_TirizamUgostitel_Turizam_TuristiNokevanja/125_Turizam_Op_BrTurNok_ml.px/table/tableViewLayout2/?rxid=46ee0f64-2992-4b45-a2d9-cb4e5f7ec5ef>.
- [25]. Kaushik, R., Jain, S., Jain, S. and Dash, T., „Performance evaluation of deep neural networks for forecasting time-series with multiple structural breaks and high volatility”. *Computer Science, Mathematics* (arXiv preprint arXiv:1911.06704). 2019
- [26]. Box, G.E.P. and Jenkins, G.M., „Time Series Analysis: Forecasting and Control”. San Francisco: Holden-Day Inc., 1976
- [27]. Junttila, J., „Structural breaks, ARIMA model and Finnish inflation forecasts”. *International Journal of Forecasting*, 17(2), pp. 203-230., 2001
- [28]. Turner, L.W. and Witt, S.F. „Forecasting tourism using univariate and multivariate structural time series models”. *Tourism Economics*, 7, pp. 135-147. 2001b
- [29]. Andreeski, C. and Mechkaroska, D., „Modelling, Forecasting and Testing Decisions for Seasonal Time Series in Tourism”. *Acta Polytechnica Hungarica*, 17(10), pp. 149-171, 2020
- [30]. Mamuda, M. and Sathasivam, S., „On the fusion of neural network models in the case of heteroscedasticity”. In *AIP Conference Proceedings* (Vol. 1974, No. 1, p. 020010). AIP Publishing LLC., June 2018
- [31]. Boussaada, Z., Curea, O., Remaci, A., Camblong, H. and Mrabet Bellaaj, N., „A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation”. *Energies*, 11(3), p. 620., 2018
- [32]. Xie, H., Tang, H., & Liao, Y.-H. "Time series prediction based on NARX neural networks: An advanced approach", *International Conference on Machine Learning and Cybernetics*. doi: 10.1109/ICMLC.2009.5212326, (pp. 1275-1279). Baoding. 2009

Forecasting Power Consumption for Residential Sector

Aleksandra Zlatkova, Aneta Buchkovska and Dimitar Tashkovski

Faculty of Electrical Engineering and Information Technologies

Ss Cyril and Methodius University-Skopje

Skopje, North Macedonia

aleksandraz@feit.ukim.edu.mk

Abstract— The fast increase of power consumption as a result of fast development of the technology and the growth of population, has a big impact of stability of the electric grid. These two main reasons arise the need of forecasting the power consumption. In this paper statistical methods: exponential smoothing method and ARIMA are used for forecasting power consumption on daily and monthly basis. They are autonomous methods because they used historical data to predict future values. The data is gathered from household in France, in a period of four years. The proposed models give good experimental results in one-year forecasting regarding to RMSE and MAE.

Keywords—ARIMA; exponential smoothing method; forecasting; power consumption; statistical methods;

I. INTRODUCTION

Nowadays, forecasting power consumption is a significant information for the power plant. With the fast growth of population and the development of the technology, power plants are burdened and the stability of the electric grid becomes critical [1]. According to [2], it is expected the power consumption to increase by 1% in the period to 2040. A big impact in increasing the power consumption has the residential sector and it is estimated that this sector uses 27% from overall power consumption [3]. This fact exposes the need of forecasting the power demand in residential sector.

In last years, there are a lot of studies about forecasting power demand, but the most of them are about how climate changes will impact the power demand. They research how power demand is correlated with weather parameters as temperature, humidity and speed of wind [4–8]. In [6], the authors focus to build multiple linear regression for monthly forecasting power demand. They use climatic parameters as independent variables and they put attention on multicollinearity to eliminate the variables that are strong correlated to other independent variables. They use humidity, rainy days, cooling degree-days and heating degree-days as predictors. Braun et al. also use regression analysis to forecast power and gas consumption of supermarket in the UK [7]. They use temperature and humidity as independent variables. In the paper [8], the researchers use simple and multiple linear regression to forecast power consumption using outdoor temperature and solar radiation as predictors. As we can see there are a lot of researchers that focus on forecasting power consumption using linear regression. Markovska et al. predict power consumption for public institution, the building of Faculty of Electrical

Engineering and Information Technologies using only historical data of power consumed in the period of January 2014 to May 2016 [9]. They use Auto Regressive Integrated Moving Average (ARIMA) and Holt-Winters model for forecasting power consumption on daily, weekly and monthly basis. Tae-Young and Sung-Bae use machine learning technique, more precisely combination of convolutional neural network (CNN) and long short-term memory (LSTM) to predict power consumption of French household [1]. They used neural network to extract spatial and temporal features to forecast the power consumption.

The prediction models can be autonomous or conditional models [10]. Autonomous models use past data of the power consumption to build the model while conditional models use other variables that are correlated to power consumption as temperature, humidity, speed of wind etc.

In this paper, autonomous models are used for prediction power consumption for residential sector in a period of one year. The used models are based on exponential smoothing and ARIMA methods. From the exponential smoothing methods was chosen Holt-Winters model for both prediction, exponential method with level and seasonal component for prediction on daily basis and exponential smoothing only with the level of time series for monthly prediction. The models will be compared using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

The paper is organized as follows. In Section II the theoretical framework is presented, Section III gives details about the used data, Section IV explains the experiment and results and Section V concludes the paper.

II. THEORETICAL FRAMEWORK

Forecasting the power consumption for a residential sector on a daily and monthly basis is done by using linear regression. In forecasting on daily and monthly basis three models are used and a comparison between them is made. These models are based on two methods which are: exponential smoothing and ARIMA method. Linear regression is commonly used in forecasting power consumption.

The exponential smoothing methods predict the future values using the past measured values. They give determined weight of the past observations and how the observation gets older the appropriate weight gets smaller. In other words, the recent

observation has bigger contribution in forecasting the future values.

A. ETS model

ETS (Error, Trends and Seasonality) models are often used for time series forecasting. These models use error, trend and seasonal component to predict some future points [12]. The error can be additive or multiplicative. In all cases, abbreviations are used for type error denotation, trend and seasonality component as “A” is for additive, “M” is for multiplicative, and “N” is for none. In our case $ETS(A, N, A)$ model is used where the first letter denotes that model is with additive errors, the second letter denotes that the trend component is not used and the third letter denotes additive seasonality. The function $ets()$ is used for automatic determination of specific ETS model. For forecasting, the function chose to use $ETS(A, N, N)$ on daily basis and $ETS(A, N, A)$ for forecasting od monthly basis. The appropriate equation for $ETS(A, N, A)$ model is:

$$\hat{y}_{t+h|t} = l_t + s_{t+h-m(k+1)} \quad (1)$$

Level component:

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)l_{t-1} \quad (2)$$

Seasonal component:

$$s_t = \gamma(y_t - l_{t-1}) + (1 - \gamma)s_{t-m} \quad (3)$$

And the appropriate equation for $ETS(A, N, N)$ model is:

$$\hat{y}_{t+h|t} = l_t \quad (4)$$

Level component:

$$l_t = \alpha y_t + (1 - \alpha)l_{t-1} \quad (5)$$

Where,

- $\hat{y}_{t+h|t}$ is forecasted value for h periods ahead
- l_t denotes the series level at time t
- s_t denotes the seasonal component at time t
- y_t denotes the observed value at time t
- α and γ are smoothing parameters
- m is the number of seasons in a year
- k is the integer part of $(h - 1)/m$

B. Holt-Winters' model

Holt-Winters' model was also used from exponential smoothing methods. This model captures trend and seasonality of the data. The model uses three smoothing equations – one for the level l_t , for the trend b_t and for seasonal component s_t . This model has two variations: additive and multiplicative. The additive model is used when the seasonal variations are similar through the series and multiplicative when variations slowly proportionally change [12]. In our paper, additive model is used because the data is characterized with constant seasonal pattern.

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)} \quad (6)$$

Level component:

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (7)$$

Trend component:

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (8)$$

And the seasonal component:

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (9)$$

where

- $\hat{y}_{t+h|t}$ is predicted value,

- l_t is level component,
- b_t is trend component
- s_t is seasonal component,
- m is frequency of seasonality,
- h is the period for forecasting
- k is the integer part of $(h - 1)/m$.
- The coefficients α, β and γ are smoothing parameters.

C. ARIMA model

ARIMA is statistical model that uses past values and its own lags to predict future values. ARIMA model is combination of two models: Auto Regressive (AR) and Moving Average (MA) model. AR model uses only its own lags. If the data contains the trend component, that means that the data is not stationary and contain seasonal component. That means that data should be differenced to reduce the seasonality. The MA model uses errors of the previous time values to forecast values.

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (10)$$

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (11)$$

Where

- y'_t is differenced time series
- c is the averaged of the changes between consecutive observations
- ε_t is the noise
- ϕ is the smoothing parameter
- θ is the MA parameter
- p is order of AR part
- d is degree of first differencing involved
- q is order of MA part
- B is backward shift operator

Definition of ARIMA model includes defined parameters p , d , and q . The function $auto.arima()$ was used to determine the values of parameters p , d and q . The function finds best model according to Hyndman-Khandakar algorithm which integrate unit root tests, minimization of corrected Akaike's Information Criterion (AICc) and maximum likelihood estimation (MLE) [13].

To determine which models, give better results, two criteria will be used: RMSE and MAE. They are calculated using following equation:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}} \quad (12)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t| \quad (13)$$

Where,

- \hat{y}_t is predicted value
- y_t is observed value
- n is number of predictions

III. DATASET

In this study, the used dataset was provided by UCI learning repository [11]. The dataset contains household electric power consumption over a period of almost four years, from December 2006 to November 2010. The measurements of electric power consumption are with a one-minute sampling rate. The measurements are collected from a house located in Sceaux, Paris in France. The dataset contains 9 variables: date, time, Global Active Power (GAP), global reactive power, voltage, global intensity, energy sub-metering in the kitchen, laundry room and for water heater and air conditioner. The dataset contains 2075259 measurements but around 1.25% of the values are missing. On the Fig. 1 the histogram of the daily power consumption is shown over four years, from December, 2006 to November 2010. From the histogram we can notice that the power consumption in winter is higher than in the hot months.

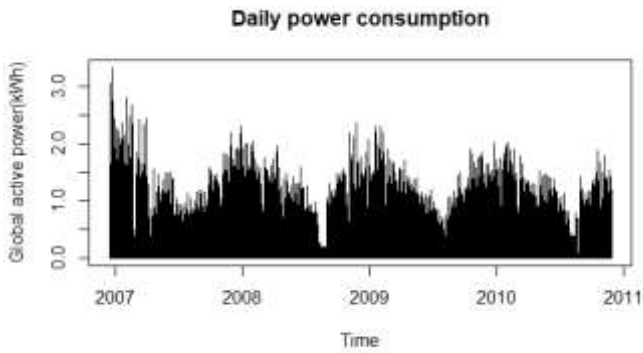


Fig. 1 Daily power consumption

On the Fig. 2 the power consumption is shown on monthly basis. We can notice that the data has seasonal pattern because over cold months the power consumption increase and for hot months the power consumption goes down.

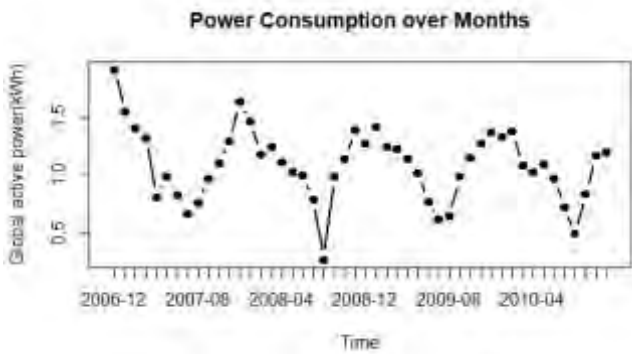


Fig. 2 Monthly power consumption

IV. EXPERIMENTAL RESULTS

A. Experiment

The subject of this research is to find model that will predict power consumption for one year precisely. The experiment is done using R programming language. Dataset has a big importance in the process of building the model. The data was

loaded and split in two subsets, one for training the model and the second for test. We tend to predict the power consumption using date, time and GAP on daily and monthly basis, so we need to resample the data. The data is sampled with high sampling rate, so we use *ts()* function to change the sample rate. In this function, the input parameter *frequency* is used to set the number of observations per unit of time. In our experiment, frequency of 365 days was used for average household power consumption per day and frequency of 12 for average power consumption per month.

For better understanding of the used data decomposition is made. Data is decomposed on four components: seasonal, trend, random and observed. On the Fig. 3 and Fig. 4 decomposition of time series is shown.

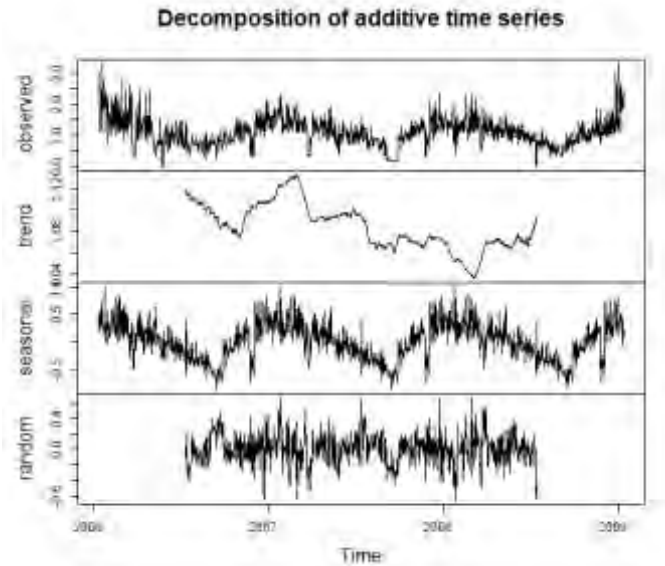


Fig. 3 Decomposition of daily power consumption

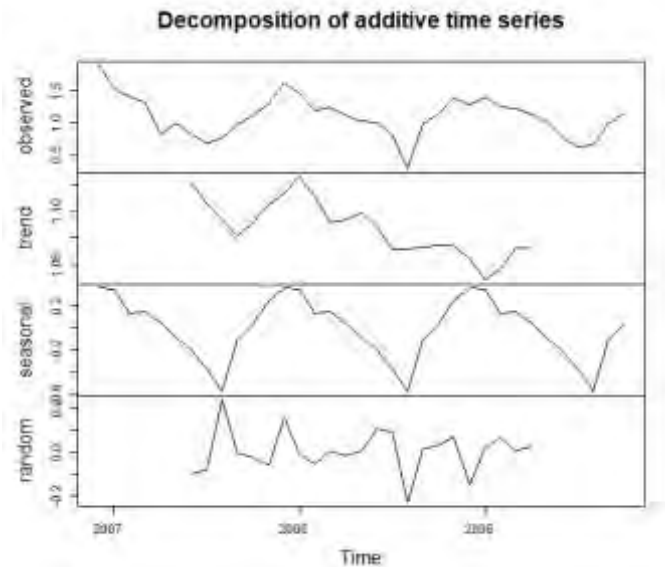


Fig. 4 Decomposition of monthly power consumption

From the presented graphics is notable that the data has a seasonal pattern and irregular trend. The seasonal and trend components have a big significance in building the model because they contain seasonal pattern of the data and the trend component catches long term increase or decrease of the data.

$ETS(A, N, A)$ is exponential smoothing method with additive errors and additive seasonality. This model was used only for power consumption on daily basis. Function $ets()$ was used where as input argument was past values and date for train the model. The function estimates the values of smoothing parameters: α, β and γ . The values of smoothing parameters for our models are:

- $\alpha = 0.0336$, $\beta = 0$ and $\gamma = 0.0011$ for daily sampled data

$ETS(A, N, N)$ is simple exponential smoothing method with additive error. This model does not use seasonal and trend component. This model is known as “error correction”. The values of smoothing parameters are:

- $\alpha = 0.121$ for monthly sampled data

$HoltWinters()$ function is also used to build exponential smoothing model, but there is a difference. $HoltWinters$ function use different optimization because this model use heuristic values for the initial states and the parameters are determined by optimizing the MSE [14]. The estimated parameters are:

- $\alpha = 0.180$, $\beta = 0$ and $\gamma = 0.6700$ for daily sampled data
- $\alpha = 0$, $\beta = 0$ and $\gamma = 0.145$ for monthly sampled data

Function $ARIMA(p, q, d)$ was used for ARIMA model where p stands for order of the AR model, q is order of the MA model and d is order of differencing. The estimated values are:

- $p=1$, $q=0$ and $d=5$ for daily sampled data
- $p=0$, $q=1$ and $d=0$ for monthly sampled data

B. Results

In this study several models were proposed for forecasting power consumption for residential sector using dataset that is provided by UCI machine learning repository [11]. To build the models past values of the data were used. The data was aggregated by daily and monthly units and prediction of the power consumption was made for 1 year on daily and monthly basis. For power consumption forecast on daily basis, the data set was split on 1077 days for train and 365 days for test. The trainset is from 16.12.2006 to 26.11.2009 and test set from 27.11.2009 to 26.11.2010.

In this research, the comparison between the used models is made. From the presented figures, is notable that all three models predict the power consumption with high accuracy. Also, on the Fig. 5 to Fig. 7 forecasting with $ETS(A, N, A)$, $HoltWinters'$ and $ARIMA$ model are shown for the same test set. With the red line forecasted data is shown, and with green is the real, measured data from the test set.

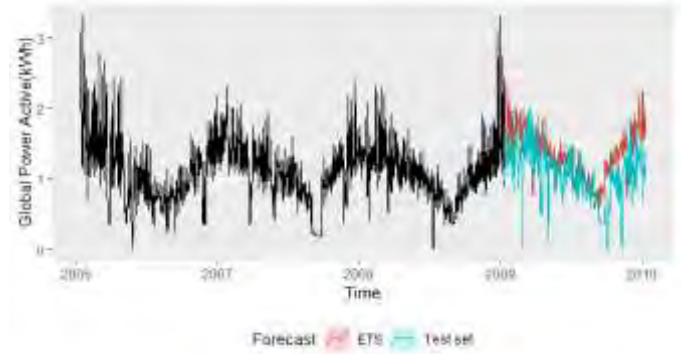


Fig. 5 Forecasting with $ETS(A, N, A)$ model on daily basis

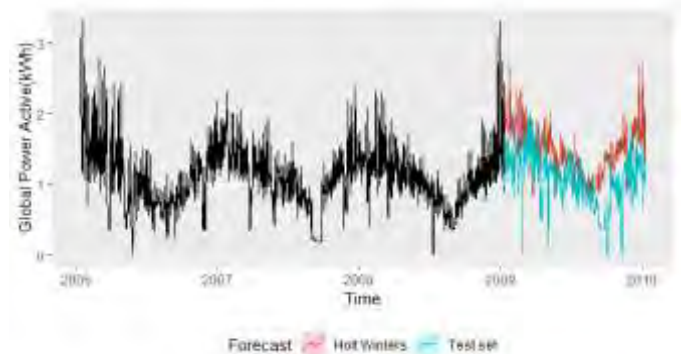


Fig. 6 Forecasting with Holt-Winters' model on daily basis

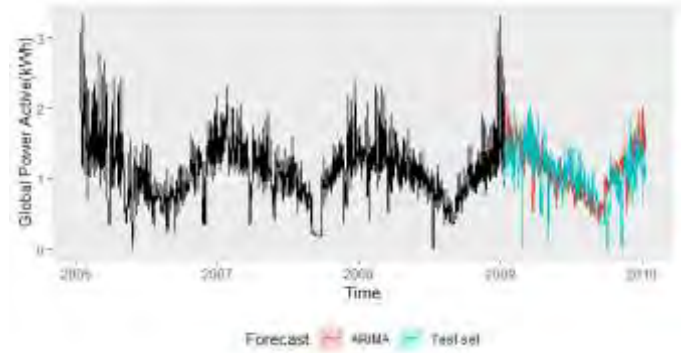


Fig. 7 Forecasting with ARIMA model on daily basis

Visually, all three models give significant good results, but in TABLE I comparison of RMSE and MAE for different models is presented. From the results in TABLE I, we can conclude that on daily basis active power forecast with ARIMA model is better fitted, than the other two models.

TABLE I Comparison of $ETS(A, N, A)$, $HoltWinters'$ and $ARIMA$ model on daily basis

Model	RMSE	MAE
$ETS(A, N, A)$	0.48	0.37
$HoltWinters'$	0.58	0.465
$ARIMA$	0.12	0.11

On the Fig. 8 – Fig. 10 the forecasting with ETS, Holt-Winters' and ARIMA model on monthly basis is showed. From the presented results we can note that all three models predict the power consumption with high accuracy.

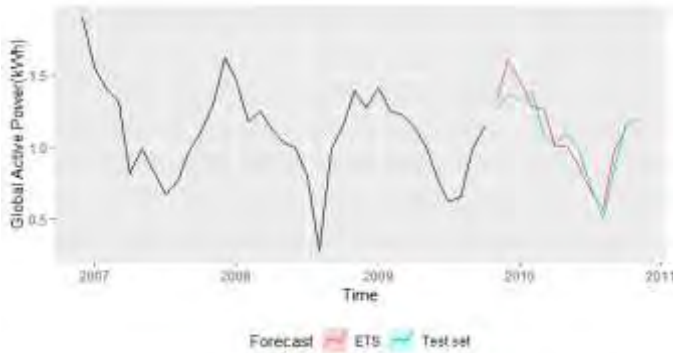


Fig. 8 Forecasting with $ETS(A, N, N)$ model on monthly basis

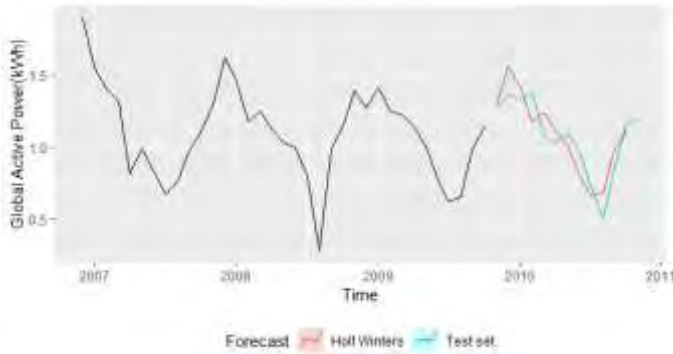


Fig. 9 Forecasting with Holt-Winters' model on monthly basis

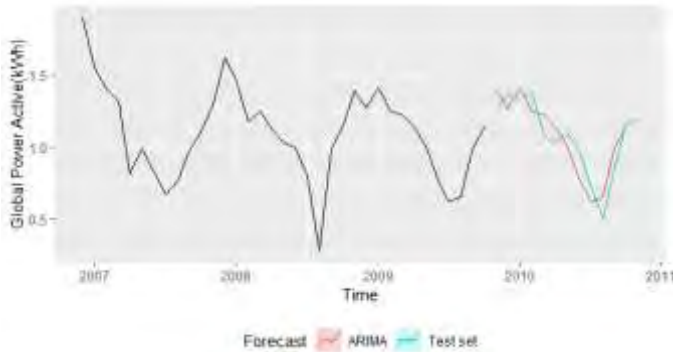


Fig. 10 Forecasting with ARIMA model on monthly basis

In the TABLE II compassion of the $ETS(A, N, N)$, Holt-Winters' and ARIMA model is shown regarding to RMSE and MAE. We can note that all three models are characterized with high and similar accuracy. The chosen ARIMA model has the same accuracy for power forecasting on daily and monthly basis.

TABLE II Comparison of $ETS(A, N, N)$, Holt-Winters' and ARIMA model on monthly basis

Model	RMSE	MAE
$ETS(A, N, N)$	0.12	0.1
Holt-Winters'	0.132	0.12
ARIMA	0.12	0.11

V. CONCLUSION

In this research exponential smoothing and ARIMA methods were proposed for forecasting power consumption for residential sector. The models were trained and tested on dataset that contains power consumption in household in France from December, 2006 to November, 2010. They use only historical data and according to past values, predict the future values. The presented results show high accuracy in prediction power consumption regarding to RMSE and MAE. ARIMA model show highest accuracy in prediction on daily basis and all three models have the same accuracy on monthly basis.

REFERENCES

- [1] Tae-Young K., Sung-Bae C., Predicting residential energy consumption using CNN-LSTM neural networks, Energy, Volume 182, 2019, Pages 72-81, ISSN 0360-5442.
- [2] Sieminski A. International energy outlook. Energy Information Administration(EIA); 2017. p. 5e30.
- [3] Nejat P, Jomehzadeh F, Taheri MM, Gohari M, Majid MZA. A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries). Renew Sustain Energy Rev 2015;43:843e62.
- [4] Sailor D. Relating residential and commercial sector electricity loads to climate evaluating state level sensitivities and vulnerabilities. Energy 2001;26:645–57. T. Ahmed, K.M. Muttaqi, A.P. Agalgaonkar,
- [5] Tariq A., Muttaqi K.M., Agalgaonkar A.P., Climate change impacts on electricity demand in the State of New South Wales, Australia, Applied Energy, Volume 98, 2012, Pages 376-383, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2012.03.059>.
- [6] Vu D.H., Muttaqi K.M., Agalgaonkar A.P., A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables, Applied Energy, Volume 140, 2015, Pages 385-394, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2014.12.011>.
- [7] Braun M.R., Altan H., Beck S.B.M., Using regression analysis to predict the future energy consumption of a supermarket in the UK, Applied Energy, Volume 130, 2014, Pages 305-313, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2014.05.062>.
- [8] Fumo N., Biswas M.A. R., Regression analysis for prediction of residential energy consumption, Renewable and Sustainable Energy Reviews, Volume 47, 2015, Pages 332-343, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2015.03.035>.
- [9] Markovska M, Buckovska A. and Taskovski D., Comparative study of ARIMA and Holt-Winters statistical models for prediction of energy consumption, ETAI, 2016
- [10] Al-Alawi SM, Islaw SM. Principles of electricity demand forecasting. I.Methodologies. Power Eng J 1996;10:139–43.
- [11] H_ebrail G, B_erard A. "Individual household electric power consumption dataset," UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science; 2012. Retrieved from <http://archive.ics.uci.edu/ml>.
- [12] Hyndman R.J. and Athanasopoulos G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 21.05.2021

- [13] Selva Prabhakaran, Complete Guide to Time Series Forecasting in Python, Accessed on 26.05.2021
<https://www.machinelearningplus.com/time-series/arma-model-time-seriesforecastingpython/#:~:text=ARIMA%2C%20short%20for%20'Auto%20Regressive,used%20to%20forecast%20future%20values.>
- [14] Hyndman R.J., Akram M. and Archibald B.C. The admissible parameter space for exponential smoothing models. *AIIM* **60**, 407–426 (2008).
<https://doi.org/10.1007/s10463-006-0109-x>

Modulation Classification with Deep Learning

Comparison of deep learning models

Selçuk Balsüzen

Electronics and Communication Engineering
Istanbul Technical University
Istanbul, Turkey
balsuzen18@itu.edu.tr

Mesut Kartal

Electronics and Communication Engineering
Istanbul Technical University
Istanbul, Turkey
kartalme@itu.edu.tr

Abstract—Automatic modulation classification (AMC), which defines the modulation type of the received signal, is an important part of non-cooperative communication systems. Automatic modulation classification plays an important role in many civil and military applications such as cognitive radio (CR), adaptive communication and electronic reconnaissance. Effective modulation classification is required for CR systems to describe the modulation techniques applied for data transmission. Classical signal identification methods used in the past are based on complex feature extraction methods such as cyclic stationarity, high order cumulants and complex hierarchical decision trees. It should also be noted that conventional methods cannot be generalized over all types of signals and are not readily adaptable when a new wireless communication technology emerges. On the other hand, deep learning (DL) has been suggested as a useful method for such classification problems, and has recently been intensively researched in the field of communications. In this paper, a convolutional neural network model, a long/short term memory model and a fully connected neural network models were designed and applied to datasets for implementing radio signal identification tasks, which is an important facet of constructing the spectrum-sensing capability required by software defined radio. Simulation results and training times are given, the advantages of the models over each other are stated and innovations that can be added in the future are proposed.

Keywords—deep learning; cognitive radio; convolutional neural network; long/short term memory; fully connected network; automatic modulation classification

I. INTRODUCTION

In modern wireless communication, automatic modulation classification (AMC) plays a very important role [1]. Signal processing as well as signal analysis can only be performed when the modulation type of the signal is known [2]. Signal analysis and processing finds application in a variety of commercial and military platforms. Various methods have been developed for AMC. Currently, classical AMC methods can be divided into two categories: probability-based (LB) and feature-based (FB) [3]. The LB modulation classifier defines the modulation of the signal by comparing the probability function value of the received signal within the known modulation pool [4]. It has been used for high accuracy modulation classification in multi-channel environment [5]. While some parameters such as carrier frequency, code rate

and channel parameters need to be known beforehand, they become very complex when unknown parameters are added. Therefore, it is difficult to design a system for signal acquisition. Some researchers have investigated the consequences of lack of information and the way to simplify the probability function, which would lead to false results [6]. The LB method cannot be applied in many practical communication scenarios because it is sensitive to parameter estimation deviations or model mismatches.

In the FB modulation classification method, first the features of the received signal are extracted, and then the modulation of the signal can be identified by comparing the features with threshold values or by giving them to the pattern recognizer [7][8]. Many traditional pattern recognition methods require manual extraction of signal features such as snapshot statistics, high-order statistics, time-frequency features, asynchronous delay sampling features [9]. These features are then used as inputs to a classifier such as a support vector machine and a decision tree. Although the method is simpler because it requires less mathematical operations, it performs poorly for nonlinear problems. Other than that, manually extracted features may not reflect the characteristics of differently modulated signals, and improper feature selection will reduce the accuracy of the classifier. Therefore, it is difficult to generalize.

In recent years, great progress has been made in the field of artificial intelligence, and the computing power of computers has increased greatly. These developments encourage the widespread use of deep learning algorithms in modulation classification [10][11]. AI solves the fundamental problem of how to automatically select and extract features from data. It also uses the combination of simple features to obtain more efficient and complex features to achieve superior classification performance [12]. In addition, deep neural networks have a multi-layered structure that can better extract the characteristics of the signal by avoiding the manual selection of data features [13].

In the modulation classification area, the basic models in DL can be evaluated in four categories: deep belief network (DBN) [14], convolutional neural network (CNN) [15], recurrent neural network (RNN) [16] and some hybrid models [17], respectively. These basic models or improved models

have been used by researchers in recent years to recognize and classify modulation types.

In this paper, three different deep learning models have been proposed, namely CNN (Convolutional Neural Network), LSTM (Long-short Term Memory) and FCNN (Fully Connected Neural Network). These models were trained on RadioML2016.10b dataset, and the results and the superiority of the models to each other were examined.

II. DATASET AND PROPOSED MODELS

In this paper, a CNN, LSTM and FCNN models are built by using Keras which is an open-source machine learning library.

A. RadioML2016.10b dataset

RadioML dataset is heavily used in modulation classification studies and it is a well-accepted dataset by the literature. In this paper, RadioML2016.10b dataset is employed. It consists of synthetic signals with 10 modulation types. The modulation types covered by the dataset are listed as: AM-DSB, WBFM, GFSK, CPFSK, 4-PAM, BPSK, QPSK, 8-PSK, 16-QAM, and 64-QAM. 1,200,000 signals for 20 different SNRs (-20 dB, -18 dB, -16 dB, -14 dB, -12 dB, -10 dB, -8 dB, -6 dB, -4 dB, -2 dB, 0 dB, 2 dB, 4 dB, 6 dB, 8 dB, 10 dB, 12 dB, 14 dB, 16 dB, 18 dB) were generated and divided as 6000 signals per modulation per SNR. Details for the generation and packaging of the dataset can be found in [18].

Here, the dataset is split into two parts (training and test) with the ratio of 7:3. 70% of the dataset is used for training. After training procedure, the models are tested with the rest of the signals.

B. CNN Model

CNN model is designed for modulation classification with deep learning. The proposed CNN model layer structure is given in TABLE I. Dropout value is chosen as 0.6.

In this model, the convolution layers narrow in terms of the number of filters. Similarly, fully connected layers also decrease in number of neurons. Our experience with many different configurations has shown that models with narrower convolutional layers give better results in terms of classification success and reduce training time.

TABLE I. CNN MODEL TABLE

Model layer architecture		
Layer	Output Dimensions	Filter size
Convolution	(2, 126, 256)	1x3
Batch normalization	(2, 126, 256)	
Dropout	(2, 126, 256)	
Convolution	(1, 124, 128)	2x3
Batch normalization	(1, 124, 128)	
Dropout	(1, 124, 128)	
Convolution	(1, 122, 64)	1x3
Batch normalization	(1, 122, 64)	
Dropout	(1, 122, 64)	
Convolution	(1, 120, 32)	1x3
Batch normalization	(1, 120, 32)	
Dropout	(1, 120, 32)	
Flatten	3840	

Model layer architecture		
Layer	Output Dimensions	Filter size
Dense	256	
Dense	128	
Softmax	10	

C. LSTM Model

LSTM model is designed for modulation classification with deep learning. The proposed LSTM model layer structure is given in TABLE II. Dropout value is chosen as 0.3. In the LSTM model, 1 LSTM layer and 3 dense layers are used.

TABLE II. LSTM MODEL TABLE

Model layer architecture	
Layer	Output Dimensions
LSTM	256
Dropout	256
Dense	256
Dropout	256
Dense	128
Dropout	128
Dense	64
Softmax	10

D. FCNN Model

FCNN model is designed for modulation classification with deep learning. The proposed FCNN model layer structure is given in TABLE III. Dropout value is 0.6. In this model, 4 dense layers that narrow in terms of the number of neurons are used. The first dense layer has 1024 neurons.

TABLE III. FCNN MODEL TABLE

Model layer architecture	
Layer	Output Dimensions
Dense	1024
Batch normalization	1024
Dropout	1024
Dense	1024
Batch normalization	1024
Dropout	1024
Dense	512
Batch normalization	512
Dropout	512
Dense	128
Batch normalization	128
Dropout	128
Softmax	10

The epoch number specifies the number of times an entire data set is fed to the model while the model is being trained. A small epoch value will shorten the training time, but the performance of the model may not be at the optimum level. The larger this value, the higher the training time and the performance of the model. But it can cause a problem such as overfitting. As a result of the trials of this study, it was seen that the trainings of 100 epochs gave good results in terms of model performance, so the number of epochs was chosen as 100. The batch size indicates how many data will be input to the model at the same time. After all batches have been processed, the next epoch is passed. Small batch size will increase the training time as more optimization calculations will be required. A large number will shorten the training time, but may adversely affect the performance of the model. The batch size is usually chosen as an exponential multiple of two.

In this study, 64,128, 512, 1024 values were chosen as batch size in order to see the effects of batch size on training time and performance. The learning rate is adaptive, 'adam' is chosen as the function. This function learns the learning speed itself and has a dynamic structure. The activation function 'ReLU' is selected. It is the most used activation function in the field of deep learning. Since multiple classifications were made in this study, the loss function was determined as categorical cross-entropy. Hyperparameters of the models are given in TABLE IV.

TABLE IV. HYPERPARAMETER TABLE

Hyperparameter	Value
Epoch	100
Batch size	64, 128, 512, 1024
Activation function	ReLU
Optimization	Adam
Loss function	Categorical cross-entropy

III. SIMULATION RESULTS

The proposed models are tested in RadioML2016.10b dataset. The test results are provided below.

A. Classification with CNN

Proposed CNN model confusion matrices are given in Fig. 1 for -12 dB, -6 dB, 0 dB, 12 dB and batch size 512.

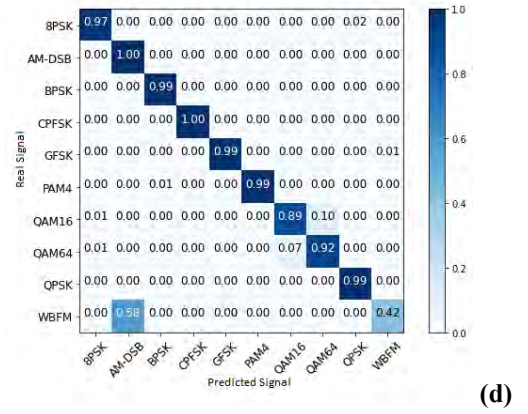
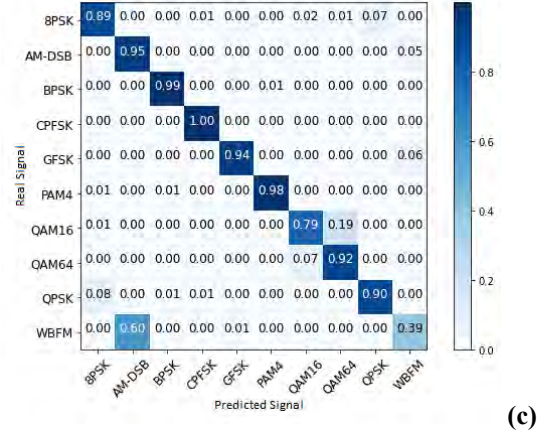
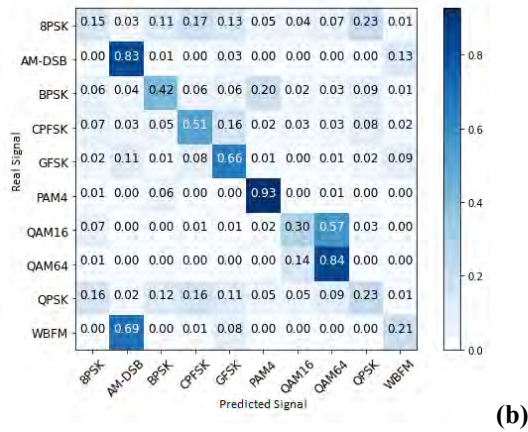
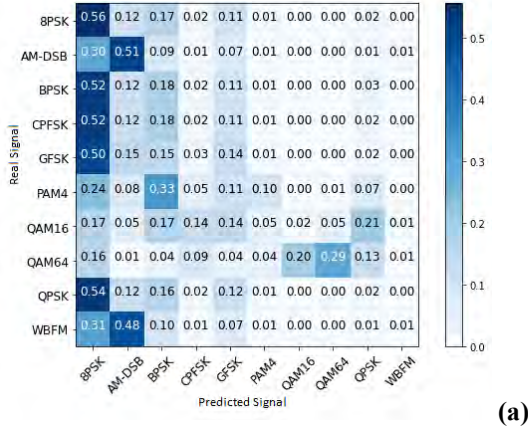


Fig. 1. CNN model confusion matrix: (a) -12 dB, (b) -6 dB, (c) 0 dB, (d) 12 dB

CNN model accuracy graph according to SNR values from -20 dB to 18 dB can be found in Fig. 2.

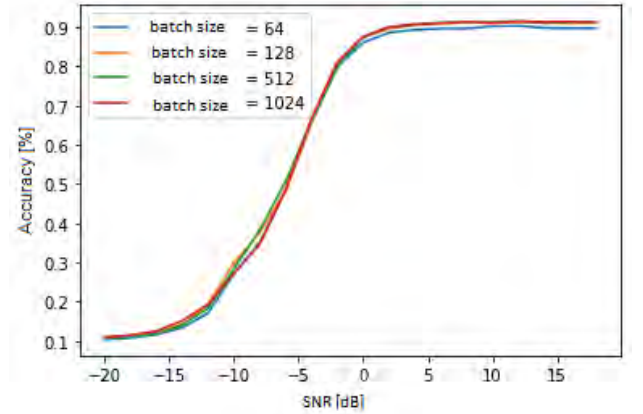
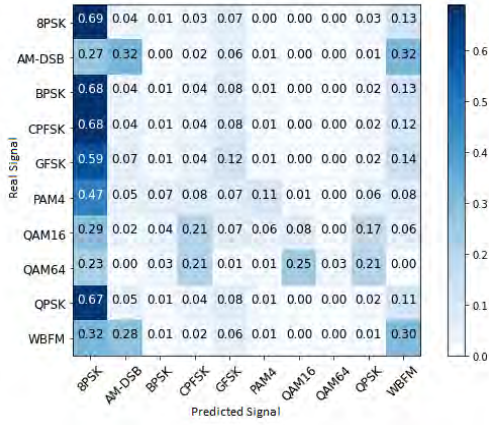


Fig. 2. CNN model accuracy graph

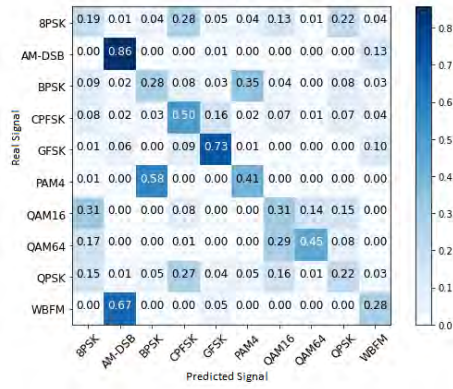
This model showed 18% classification performance at -12 dB, 50% at -6 dB, 87% at 0 dB, and 91% at 12 dB.

B. Classification with LSTM

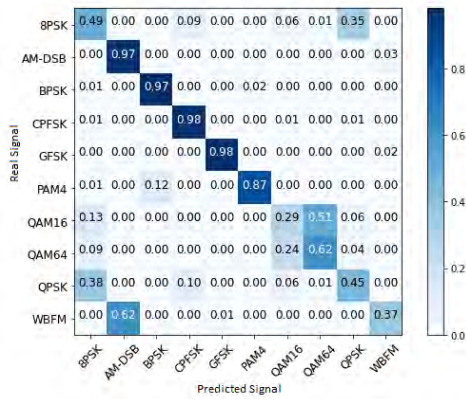
Proposed LSTM model confusion matrices are given in Fig. 3 for -12 dB, -6 dB, 0 dB, 12 dB and batch size 512.



(a)



(b)



(c)

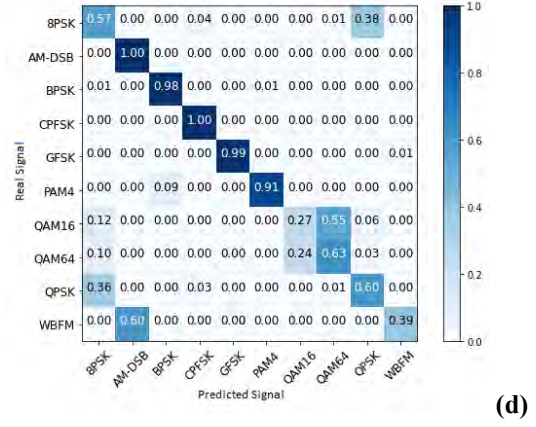


Fig. 3. LSTM model confusion matrix: (a) -12 dB, (b) -6 dB, (c) 0 dB, (d) 12 dB

LSTM model accuracy graph according to SNR values from -20 dB to 18 dB can be found in Fig. 4.

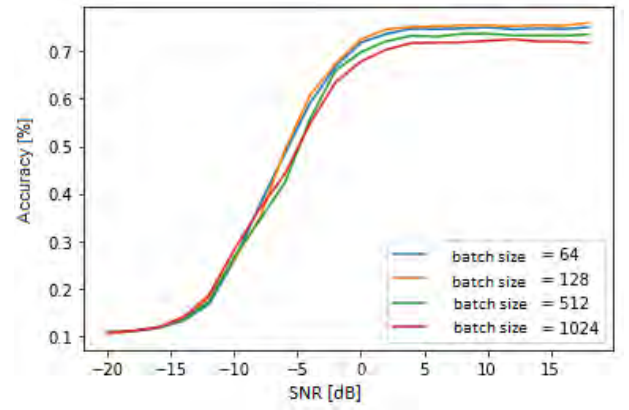


Fig. 4. LSTM model accuracy graph

This model showed 17% classification performance at -12 dB, 47% at -6 dB, 69% at 0 dB, and 73% at 12 dB.

C. Classification with FCNN

Proposed FCNN model confusion matrices are given in Fig. 5 for -12 dB, -6 dB, 0 dB, 12 dB and batch size 512.

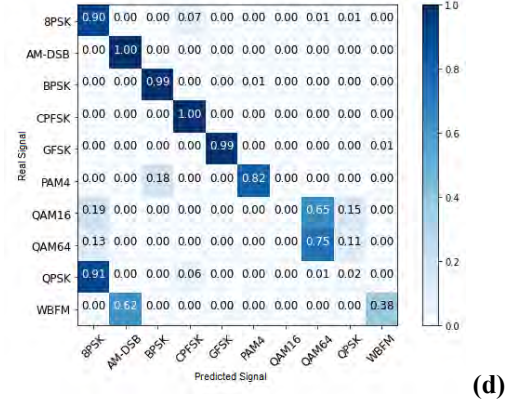
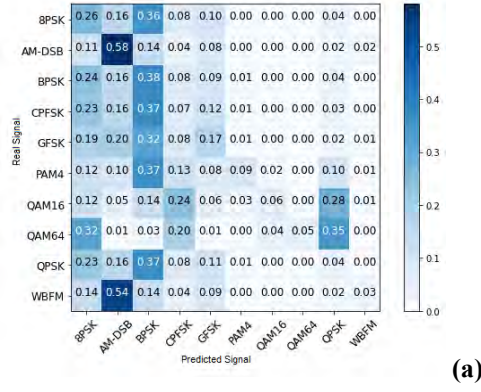


Fig. 5. FCNN model confusion matrix: (a) -12 dB, (b) -6 dB, (c) 0 dB, (d) 12 dB

FCNN model accuracy graph according to SNR values from -20 dB to 18 dB can be found in Fig. 6.

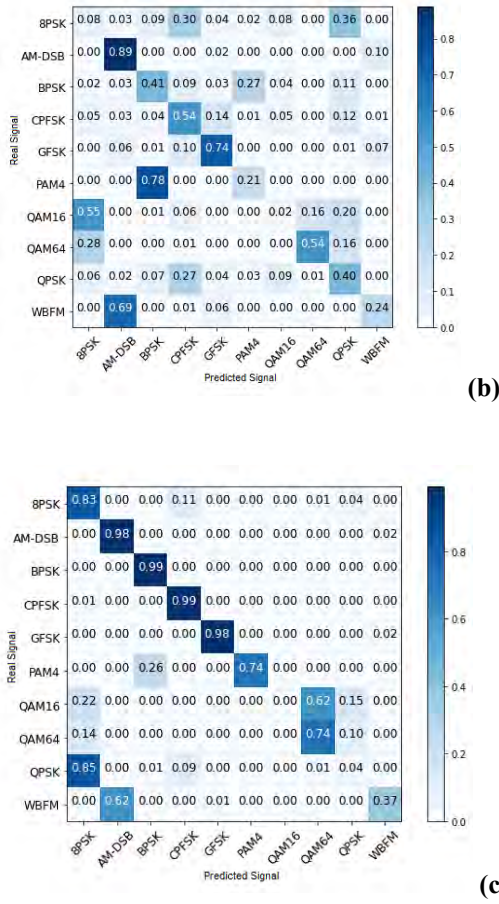


Fig. 6. FCNN model accuracy graph

This model showed 17% classification performance at -12 dB, 40% at -6 dB, 66% at 0 dB, and 68% at 12 dB.

As the batch size decreases, the time taken to train increases. As can be seen in Table V, the CNN model has the highest training time and the LSTM model the lowest training time for this study. In the training and test stages, we employ NVIDIA Tesla T4 graphics processing units (GPUs).

TABLE V. TRAINING TIMES TABLE

Proposed models training times			
Batch size	CNN	LSTM	FCNN
64	110.8[sec/epoch]	42.3[sec/epoch]	56.4[sec/epoch]
128	88.1[sec/epoch]	24.4[sec/epoch]	38.6[sec/epoch]
512	68.4[sec/epoch]	7.4[sec/epoch]	11.7[sec/epoch]
1024	64.7[sec/epoch]	4.5[sec/epoch]	7.9[sec/epoch]

In table Table VI, classification accuracies of all SNR and all batch size for designed three models can be seen.

TABLE VI. CLASSIFICATION ACCURACIES

SNR [dB]	CNN				LSTM				FCNN			
	64	128	512	1024	64	128	512	1024	64	128	512	1024
-20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
-18	0.10	0.11	0.11	0.11	0.10	0.11	0.10	0.11	0.10	0.10	0.10	0.10
-16	0.11	0.12	0.11	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
-14	0.13	0.14	0.14	0.15	0.13	0.13	0.13	0.13	0.12	0.13	0.13	0.13
-12	0.17	0.19	0.18	0.19	0.16	0.18	0.17	0.18	0.16	0.15	0.17	0.16
-10	0.27	0.29	0.28	0.27	0.25	0.26	0.26	0.28	0.25	0.23	0.24	0.24
-8	0.35	0.37	0.38	0.34	0.37	0.34	0.34	0.36	0.31	0.31	0.33	0.33
-6	0.48	0.48	0.50	0.48	0.48	0.49	0.42	0.44	0.38	0.39	0.40	0.41
-4	0.66	0.65	0.65	0.66	0.59	0.60	0.55	0.54	0.52	0.53	0.52	0.53
-2	0.80	0.80	0.79	0.81	0.66	0.67	0.65	0.63	0.62	0.62	0.62	0.63
0	0.86	0.87	0.87	0.87	0.71	0.72	0.69	0.67	0.66	0.65	0.66	0.67
2	0.88	0.89	0.89	0.90	0.73	0.74	0.72	0.70	0.67	0.66	0.67	0.67
4	0.89	0.90	0.90	0.90	0.74	0.75	0.73	0.71	0.67	0.67	0.68	0.68
6	0.89	0.91	0.91	0.90	0.74	0.75	0.73	0.71	0.67	0.67	0.68	0.68
8	0.89	0.91	0.91	0.91	0.74	0.75	0.73	0.71	0.67	0.67	0.68	0.68
10	0.90	0.91	0.91	0.91	0.75	0.75	0.73	0.72	0.67	0.67	0.68	0.68
12	0.90	0.91	0.91	0.91	0.74	0.75	0.73	0.72	0.67	0.67	0.68	0.68
14	0.89	0.91	0.91	0.91	0.74	0.75	0.73	0.72	0.67	0.67	0.68	0.68
16	0.89	0.91	0.91	0.91	0.74	0.75	0.73	0.72	0.67	0.67	0.68	0.68
18	0.89	0.91	0.91	0.91	0.75	0.76	0.73	0.71	0.67	0.67	0.68	0.68

In RadioML2016.10b dataset, when the performances of the 3 models are examined according to their SNR values, it is seen that they show similar classification performance for low SNR (< -8 dB). For high SNR values, the CNN model gives approximately 17% better results than the LSTM model and 23% better than the FCNN model.

As the batch size decreases in the models, the training time increases. The CNN model has the longest training time. The fastest trained model is the LSTM model.

In summary, this paper examined the modulation type classification performances of CNN, LSTM and FCNN models under different parameters. In this study, it has been tried to bring different perspectives by examining the effects of the hyperparameter changes of the dataset on the model performance. The obtained results are promising in terms of modulation classification on real systems in the future.

We are excited that work on DL and ML for communications has high potential and as the field develops, it can contribute to future wireless communications systems. For now, there are many open problems to be solved and practical gains to be made.

Optimization can be done by changing the hyperparameters of the models to improve the results in this study. For example, choosing different sizes of filters, increasing the number of layers, using pooling for CNN models, using different activation functions and creating different architectures can be given. Filter selection for CNN models can significantly affect performance. Increasing the number of layers of the models can increase the number of features that can be learned, but it will increase the probability of encountering the vanishing gradient problem. Pooling can be used to solve the problem of overfitting and reduce the number of mathematical operations, but it can lead to reduced model performances. Deep learning model optimization brings with it many tradeoffs. A large number of tests should be carried out to find the most optimum parameters.

It can be said that the content and number of datasets have as much effect on the performance results as the model architecture and hyperparameters used. The performance of deep learning models is directly proportional to the amount of training data. Increasing the number of data can improve model results and especially the accuracy of certain modulations that are misclassified.

IV. CONCLUSION

Wireless spectrum monitoring and signal classification over frequency, time and space dimensions is still an active research topic. In this study, several promising applications of deep learning to the modulation classification problem are introduced.

In this paper, the RadioML2016.10b data set was used. However, diversity can be achieved by creating more data sets in this area. Data diversity can be achieved by analyzing the amplitude/phase and frequency domain of the signals. Finally, performance can be increased by creating hybrid models. Considering the high feature extraction capacity of CNN and the success of LSTM in time series, more complex models can be obtained by combining CNN and LSTM models.

ACKNOWLEDGMENT

This research work is supported by Istanbul Technical University.

REFERENCES

- [1] F. Wang, S. Huang, H. Wang, and C. Yang, 'Automatic Modulation Classification Exploiting Hybrid Machine Learning Network', *Math. Probl. Eng.*, vol. 2018, pp. 1–14, Dec. 2018.
- [2] Y. Wang, H. Zhang, Z. Sang, L. Xu, C. Cao, and T. A. Gulliver, 'Modulation Classification of Underwater Communication with Deep Learning Network', *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–12, Apr. 2019.
- [3] Y. Kumar, M. Sheoran, G. Jajoo, and S. K. Yadav, 'Automatic modulation classification based on constellation density using deep

- learning', *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1275–1278, Jun. 2020.
- [4] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, 'Survey of automatic modulation classification techniques: Classical approaches and new trends', *IET Communications*, vol. 1, no. 2, pp. 137–156, 2007.
 - [5] J. Zhang, D. Cabric, F. Wang, and Z. Zhong, 'Cooperative Modulation Classification for Multipath Fading Channels via Expectation-Maximization', *IEEE Trans. Wirel. Commun.*, vol. 16, no. 10, pp. 6698–6711, Oct. 2017.
 - [6] E. Nachmani, Y. Bachar, E. Marciano, D. Burshtein, and Y. Be'ery, 'Near Maximum Likelihood Decoding with Deep Learning', Jan. 2018.
 - [7] A. Ali and F. Yangyu, 'Unsupervised feature learning and automatic modulation classification using deep learning model', *Phys. Commun.*, vol. 25, pp. 75–84, Dec. 2017.
 - [8] T. J. O'Shea, N. West, M. Vondal, and T. C. Clancy, 'Semi-supervised radio signal identification', in *2017 19th International Conference on Advanced Communication Technology (ICACT)*, Nov. 2017, pp. 33–38.
 - [9] M. Abu-Romoh, A. Aboutaleb, and Z. Rezki, 'Automatic modulation classification using moments and likelihood maximization', *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 938–941, May 2018.
 - [10] Z. Zheng, T. Huang, H. Zhang, S. Sun, J. Wen, and P. Wang, 'Towards a resource migration method in cloud computing based on node failure rule', *J. Intell. Fuzzy Syst.*, vol. 31, no. 5, pp. 2611–2618, Oct. 2016.
 - [11] X. Shi, Z. Zheng, Y. Zhou, L. He, B. Liu, and Q. S. Hua, 'Graph processing on GPUs: A survey', *ACM Computing Surveys*, vol. 50, no. 6, Association for Computing Machinery, pp. 1–35, Jan. 01, 2018.
 - [12] T. J. O'Shea, T. Roy, and T. C. Clancy, 'Over-the-Air Deep Learning Based Radio Signal Classification', *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.
 - [13] T. Wang, C. K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, 'Deep learning for wireless physical layer: Opportunities and challenges', *China Commun.*, vol. 14, no. 11, pp. 92–111, Nov. 2017.
 - [14] G. J. Mendis, J. Wei, and A. Madanayake, 'Deep learning-based automated modulation classification for cognitive radio', in *2016 IEEE International Conference on Communication Systems (ICCS)*, Dec. 2016, pp. 1–6.
 - [15] C. Wang, J. Wang, and X. Zhang, 'Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network', in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Jun. 2017, pp. 2437–2441.
 - [16] D. Hong, Z. Zhang, and X. Xu, 'Automatic modulation classification using recurrent neural networks', in *2017 3rd IEEE International Conference on Computer and Communications, ICC3 2017*, Mar. 2018, vol. 2018-January, pp. 695–700.
 - [17] X. Liu, D. Yang, and A. El Gamal, 'Deep neural network architectures for modulation classification', in *Conference Record of 51st Asilomar Conference on Signals, Systems and Computers, ACSSC 2017*, Apr. 2018, vol. 2017-October, pp. 915–919.
 - [18] T. O'shea, 'Radio Machine Learning Dataset Generation with GNU Radio', Sep. 2016.

Machine Learning Approach for Autonomous Control of Vertical Cement Roller Mills

Othon Manis¹, Gorjan Nadzinski¹, Mile Stankovski¹

¹Ss. Cyril and Methodius University, Faculty of Electrical Engineering and Information Technologies, Skopje, North Macedonia
manis.othon@gmail.com, gorjan@feit.ukim.edu.mk, milestk@feit.ukim.edu.mk

Abstract—The production of cement requires fine grinding of raw materials and thus consumes a lot of energy. In the search for savings, the development of the vertical cement roller mills (VCRM) significantly reduced the energy consumption. The control of any cement roller mill is a complex process and requires the monitoring of many variables by the operator. This work proposes a novel Real Time Optimizer (RTO) for autonomous control of VCRM, focusing on a machine learning algorithm which uses historical data to predict the values of essential process variables. These support vector machine (SVM) prediction models are trained on real plant data and are an integral part of the optimizer designed for the process. The final optimization goal is the improvement of the mill operation with increase of the production, decrease of energy consumption, and reduction in the variability of the cement quality.

Keywords—Real-time optimization; Machine learning; Vertical cement roller mill.

I. INTRODUCTION

Big Data presents many promising opportunities and challenges for manufacturers, especially for industrial manufacturers who have operational and business data about their finances, inventories, products, human resources, distributors, and partners. A big number of different sensors are installed in production systems and assist operators in the supervision, monitoring and control of production, so industrial manufacturers are poised to use Big Data technologies to capitalize on these and other sources of data to optimize manufacturing and field operations. The use of distributed control systems in industrial plants has generated large amounts of historical process data, especially in large-scale processes [1].

Therefore, in order to ensure the reliability and safety of modern large-scale industrial processes, data-driven methods have been receiving considerably increased attention for the purpose of process monitoring. Among them, and under the complex real operating conditions, machine learning has been used specially to estimate and anticipate events of interest regarding industrial assets and production processes [2].

This paper outlines the methodology of building a tool for real time optimization and control of a vertical cement roller mill (VCRM). It is a complex production process where many process variables such as vibrations, temperatures, pressures, weight feeders of raw material, quality cement's characteristics must be dealt with simultaneously and continuously. The rise of commercially available computing power has significantly reduced the costs of complex on-line computations, so an

approach utilizing machine learning could prove more efficient and suitable for large-scale industrial applications than before.

An optimizing tool for such a process should monitor and predict process parameters, analyze the propagation of process faults, and optimize the cement quality characteristics in real time, adjusting and controlling the production process in the most efficient way via a ranking algorithm that recommends the best set of process manipulated variables. An important segment of such an optimizer are the prediction models for the essential process variables, and this paper will especially focus on them. These models have been built upon data from a real cement plant. The implementation of the entire real time optimizer for the plant is underway but the results of the optimal online control will be presented in a future publication.

The paper is organized as follows: Section 2 describes the vertical cement roller mill, Section 3 outlines the proposed optimization approach, Section 4 presents a case study of machine learning based prediction of essential variables at a real cement plant, before a conclusion is given in Section 5.

II. VERTICAL CEMENT ROLLER MILL

Cement production requires extraction of a large amount of raw materials from the environment. Most of these materials have to be crushed and then ground to a very fine size. This intermediate material then feeds a rotary kiln which is a type of oven with temperatures high enough to complete the chemical changes. The output of this process is a material called "clinker" which is used as the main constituent for the cement production. There are two different technologies for cement grinding: the ball mills (Fig. 1) and the vertical roller mills (Fig. 2).

The grinding process differs fundamentally in these two technologies; reference [3] presents the grinding processes as based on ball mills in which the comminution takes place by impact and attrition from the grinding balls tumbling inside the mill. Efficiency and output of the ball mills primary depend on optimum utilization of the ball charge energy for coarse and fine grinding.

The comminution in the vertical roller mill takes place by exposing a bed of material to a pressure sufficiently high to cause fracture of the individual particles in the bed, although the majority of the particles in the bed are considerably smaller than the thickness of the bed. Many studies have been performed on comparison of conventional grinding systems (ball mills) and VCRM [4], many of them indicating that

VCRM grinding efficiency is more than 30% higher [5], which has made VCRM overtake old technologies [6].

The main grinding parts of a VCRM are given in Fig. 2 [5]. The rotating table (1) with a horizontal grinding track and rollers (2), are pressed onto the table by lever arms and a hydro pneumatic spring system. The particle bed comminution takes place between the working surfaces of the track and rollers. A dynamic air separator (3) is located above the grinding chamber, which classifies the ground particles. The transport of the particles from the grinding table to the air separator is done pneumatically.

A. Mill Functions

The VCRM has four main functions: grinding, drying, separating, and transporting [7]. The VCRM is widely used in the grinding of cement raw meal, slag, cement clinker, raw coal and other raw materials.

The working principle of VCRM is the following: the motor drives the grinding table to rotate through the reducer, and raw material falls from the feeding port to the center of the grinding table as hot air enters the grinding chamber from the air inlet.

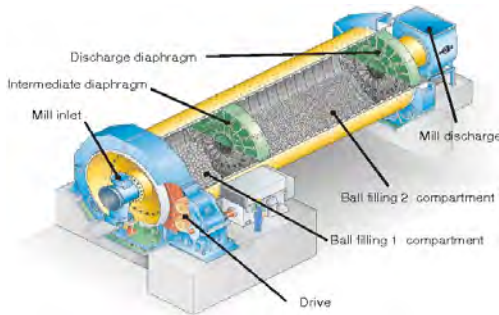


Fig. 1. Example of a cement ball mill.

Influenced by centrifugal force, the material moves to the edge of grinding table and is there crushed by the grinding roller through the annular groove on the grinding table. The pulverized material is then taken up from the edge of the grinding table by the high-speed airflow of the wind ring. When raw material in the airflow passes through the separator, the coarse particles fall under the action of the rotating rotor. The fine particles move together with the airflow to the dust-collecting device from which the final product is collected. The air flow through the mill is also necessary for transporting the heat for the drying process. The moisture-containing material is dried during the contact with the hot air, coming from a hot gas generator which uses petroleum coke as a fuel. The hot gases are necessary in order to achieve the mill exit temperature target.

B. Main Mill Controls

A VCRM makes intensive use of a lot of sensors and actuators, and industries are obliged to monitor it intensively in order to increase the run factor of the different units. However, in order to control the VCRM, operators usually control just some key variables, the main ones being:

- The amount of material inserted into the mill (material feed). The feed rate has to be carefully controlled in order to avoid vibrations and high power consumption but it is usually a variable that is supposed to be maximized during the optimization of the mill.
- The pressure of the rollers on the table (grinding pressure). This pressure is the main factor affecting the quality of the product and the operation of the mill. The grinding pressure is adjusted according to the amount of the mill feed, the material size and its grindability. The pressure must be controlled in order to maintain a layer of material with a certain thickness on the grinding disc, reduce the vibration of the mill and ensure stable mill operation. When the grinding pressure is high, the vibration speed of the mill increases and the component damage is accelerated. Therefore, maintaining a proper grinding pressure is a critical operation.

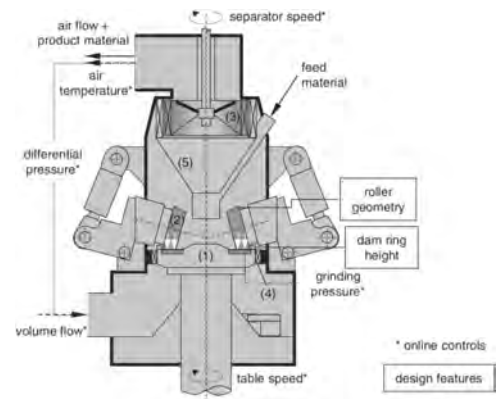


Fig. 2. Example of a vertical cement roller mill.

- Fan speed and dampers. The air flow through the mill transports the material inside of the mill and the product out of the mill to the bag filter and provides heat exchange between the hot gases and the material in the mill for the drying process. Furthermore, the air fluidizes and stabilizes the grinding bed and is the critical factor for the efficiency of the separator. By controlling the fan power and some related dampers it is possible to control the force of the air that passes from the mill circuit. When the system air volume is too high, the internal differential pressure and the main motor current both decrease, the thickness of the material layer is too low, the vibrations are high, and the sieve residue increases. When the system air volume is too low, the thickness of the material layer increases, the internal differential pressure and the current of the main motor both increase.
- Separation speed. The speed of the separator is important to achieve the fineness target, and helps to achieve the proper size distribution of the final product. Through the separator speed, the amount of the material which would go back to the disc for more grinding (and thus the quality of the final product) can be controlled.
- Water injection. The grinding water spray system plays an important role in the stabilization of the material bed,

especially in the case of more powdery materials or raw materials with low moisture content. It helps the material to stay on the disc, to avoid vibration, and to increase the grinding efficiency of the mill.

- **Mill Outlet Gas Temperature.** The temperature at the exit of the mill is critical for the cement quality and must be controlled very carefully. If the temperature is high then gypsum dehydration occurs, and if it is low then there will be problems in the cement storage silo.
- **Vibration.** Excessive vibration of the vertical mill will not only directly cause mechanical damage, but also affect production and quality. The factors that cause vibration are the key points of a cement mill operation, such as the grinding pressure, the thickness of the material layer, the air volume and the air temperature, and the wear of the roller surface or the grinding disc.

It is apparent that the VCRM is a complex process shaped by many important variables which must be properly monitored and/or controlled.

III. PROPOSED OPTIMIZATION APPROACH

As stated previously, the cement industry (as most others) uses many sensors which give information about the status of every relevant unit in fixed time periods, generating enormous amounts of data. It is therefore possible to use this data and create models in order to develop a real time optimizer (RTO) which will improve the operation of the plant. The methodology to building a real time optimizer is given in Fig. 3 and will be outlined in this Section. This paper will then focus on the modeling process using machine learning, while the rest of the steps towards real time optimization are left for future work.

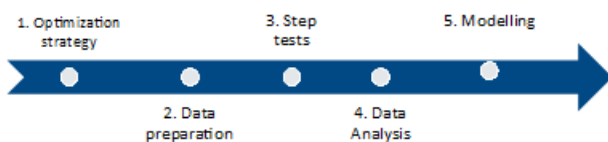


Fig. 3. Methodology to building a real time optimizer.

A. Optimization strategy

The definition of the optimization strategy includes the determination of an objective function, the variables that the RTO will control, the necessary constraints, and the models which will be used.

The selection of the variables that will be controlled/manipulated by the RTO must be carefully done and based on expert knowledge of the process itself. The vertical cement mill's process also has to be deeply analyzed in order to determine which are the existing constraints, which are those process variables that will prevent the RTO from taking further actions to improve the objective function because of some process limitations. Some of the basic constraints are the limits of the production unit's equipment and machinery defined by the manufacturers, as well as limits regarding cement quality. It is important to note that the definition of any variable as a constraint is directly related with how that variable affects the

RTO's manipulated variables. Process constraints which are not related to the RTO manipulated variables cannot be considered part of the optimization strategy as the system will not be able to take any corrective action to improve these variables.

The final step is the development and building of separate models as part of the final application. One model for each constraint must be created in order to be able to predict their value and ensure that their limits are not overridden during the RTO operation. Similarly, one model per quality variable is also required, in order to ensure that the quality specs will be fulfilled at any time based on the predictions. The definition of input variables for each model is done based on process knowledge and deep investigation of any correlation between each of the model's manipulated variables.

B. Data preparation

The process of data preparation for a RTO product consists of data definition, to determine the information that will be required during the product development, and of data integration, which includes the data collection from various sources and ensuring consistency among all gathered inputs.

The availability of the historical data must be considered in order to ensure that enough data exist on the current operation process which can be used in modeling – this should ideally be at least a year's worth of historical data of a minute average sampling time. Detailed understanding of process variables is the key factor for successful data preparation.

The variables identified can be classified into four groups:

- manipulated variables (for control over the process),
- constraints (representing the machinery and equipment limits),
- quality characteristics of the cement,
- auxiliary variables.

The sampling time for the data must also be carefully chosen, as a key point for further analysis and in the modeling phase of RTO development. Long sampling times could cause information losses, whereas too short sampling times may capture negligible information or noise. A smaller sampling time is always preferable since no information is lost.

Additionally, step tests are often required in order to create enough data variability. During a step test, only one of the manipulated variables is modified and its effect on the rest of the process variables is evaluated.

C. Data Analysis

Data analysis encompasses all the data manipulation activities that need to be done before the modeling. It is done in order to obtain relevant information about the process behavior and dynamics. During this phase it is paramount to prioritize the engineering point of view over the purely mathematical one, or risk ending up with an overtrained model that does not align with process logic. To do this, full understanding of the

process is required. Some of the most important data analysis tasks are:

- Defining which are the normal working conditions of the process unit – They are defined by process experts based on the operation of the unit. This eliminates the startup and shut down periods and defines the range of the variables in which the operation of the unit is steady and smooth.
- Detection and exclusion of outliers that fall outside of normal working conditions – Abnormal or inaccurate values need to be removed since they do not represent the steady state of the system. Outliers have to be detected based on process expert knowledge or through statistical analysis.
- Preliminary data exploration and visualization to assess correlations and time patterns, and to ensure correct input-output correlations within the process.
- Isolation of data periods showing stable conditions – This can also help to find periods in which only one of the input variables changes, making it possible to segregate the effect of that input variable from the rest.

With proper data analysis, some key information can be retrieved, such as the ranges of operation for the different variables in the process (with particular interest on the operative and objective variables as well as their variability), the operating limits for constraints, detection of events such as shut downs, breakages, maintenance, etc.

D. Modeling

The modeling effort for the RTO consists of two main steps, namely of building separate predictive process models for the variables of interest, and of building a composite model which integrates all of the separate models, determining the control structure to be used by the RTO and setting up the optimization algorithm.

Regarding the separate predictive models, the effective inputs (operative and informative variables) for each output variable of interest must be determined, along with the lag or time delay for each input-output relation. After the validation of all these models, an all-connecting composite model, which also integrates the process constraints and objective function, is created. Once the RTO has been completely deployed, it must be tested offline extensively before its unsupervised closed loop operation.

Finally, the process objective function OF determines the target variables and the aim of the optimization (increasing cement production, decreasing thermal energy or electrical or water consumption, etc.). This function is a weighted sum of normalized process variables PV_i with coefficients α_i (1).

$$OF = \sum \alpha_i PV_i / PV_{i,max}, \sum \alpha = 1. \quad (1)$$

The RTO optimizes the process by maximizing the objective function. This is done by periodically analyzing the separate model predictions in order to select better settings for

the manipulated variables, which will in turn achieve feasible operational conditions with the best possible objective function values.

IV. PREDICTIVE MODELS FOR MANIPULATED AND CONSTRAINED VARIABLES

The real time optimization strategy was defined in collaboration with the plant team and the main purposes are to increase mill productivity, reduce energy consumption (combination of mill motor power and fan speed) and water consumption, reduce the number of mill stops (by keeping the production constant), and maintain the cement quality (blaine and fineness), all by using process and equipment constraints. This is implemented by using machine learning algorithms in order to predict the values of all necessary variables and suggest parameters adjustments where required.

All values from the different sensors along the process line of the VCRM were recorded by a local historian server connected to the distributed control system of the plant. At this stage more than two years worth of data sampled at a two-second-period was available at the server.

All variables of interest were grouped as follows:

- The manipulated variables are: mill total feed, separator speed, bag filter pressure, mill inlet temperature, mill exit temperature, mill inlet pressure, grinding pressure.
- The constrained variables are: bucket elevator power, mill motor power, fan motor power, separator motor power, mill vibration.
- The quality characteristics are the cement blaine and the cement fineness.
- The auxiliary variable is the mill differential pressure.

The sampling time was selected after a deep analysis of the historical data and was set on 30 seconds or 1 minute for the majority of the process data, except for some specific cases (e.g. vibrations) which have to be analyzed on shorter time period of 2 seconds.

It was observed that the plant mostly operates within the same working conditions on a regular basis, so it was not possible to capture different operating regimes. To solve this issue and to diversify the information contained in each variable, step tests were executed to create enough data variability. During a step test only one manipulated variable was modified and the effects on the rest of the process variables were evaluated.

Data analysis was performed on both the historical and the step test generated data, during which the variable signals were filtered and smoothed, and highly correlated variables were removed. Normal working conditions were defined, the startup and shut down periods were overlooked, and any data knowingly caused by sensor malfunctions or human error were detected and removed.

After the data was prepared, fourteen variables in total were predicted with separate machine learning based predictive

algorithms, seven manipulated and seven constrained. The manipulated variables were:

- the bag filter pressure,
- the mill inlet temperature,
- the mill exit temperature,
- the mill inlet pressure,
- the grinding pressure,
- the mill total feed,
- the separator speed.

The constrained variables, whose prediction was necessary in order to avoid operational problems caused by exceeding process or equipment limits, were:

- the bucket elevator power,
- the mill motor power,
- the fan motor power,
- the separator motor power,
- the mill vibration,
- the blaine,
- the sieve residue.

Four different approaches were initially used, support vector machines (SVM), multi-layer perceptron (MLP), k-nearest neighbors (kNN), and long short-term memory neural network (LSTM), all trained on 70% of the data. All models were validated on a portion of the data unseen during the training phase, which consisted of 30% of all available data. In order to clarify the results achieved to validate each block, the output's prediction for all models was compared with the real values captured by the sensors. The models predictions were referring to 400 samples ahead, which with a sampling time of 30 seconds equals to a time period of 200 minutes ahead.

Evaluation of the modeling was done with three different commonly used metrics: mean absolute error (MAE), mean absolute percentage error (MAPE), and mean absolute scaled error (MASE).

The results showed that the SVM algorithm was superior to the rest of the approaches, as shown in Table I for an example of the bag filter pressure model. For brevity, this paper will only present the prediction results from this algorithm.

TABLE I. PREDICTION ERRORS FOR BAG FILTER PRESSURE FOR ALL FOUR USED ALGORITHMS

Algorithm	Prediction Errors According to 3 Metrics		
	MAPE	MASE	MAE
SVM	1.3384	43.9349	0.115
MLP	4.3194	144.0288	0.3769
kNN	1.4996	49.399	0.1293
LSTM	2.3704	14837.5914	0.1724

In continuation, Fig. 4 – Fig. 12 show a time-series comparison between the real data for four out of the seven manipulated variables and five out of the seven constrained variables, and the respective SVM predictions over the same time window. The axes are unlabeled to preserve the real values and measurements of these variables, due to the sensitivity of the process. Also, Table II presents the prediction errors for all nine models, according to all three metrics.

While most of the errors are low, it is clear that not all models are very effective and accurate, but in general they provide satisfying results and are able to qualitatively predict the variables behavior. The final step of the RTO design is the connection of all these models into a composite model which uses them to predict the trends within the process and suggest in-time changes of the manipulated variables in order to improve the quality characteristics and optimize the production. Early work on the RTO has shown the results to be satisfactory in their precision.

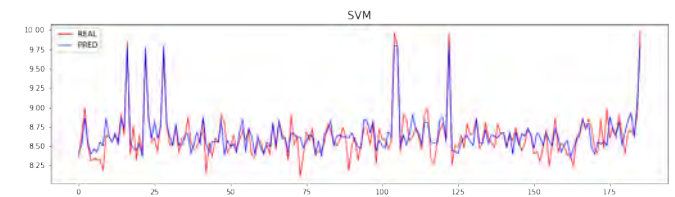


Fig. 4. Real data vs. SVM prediction for bag filter pressure.

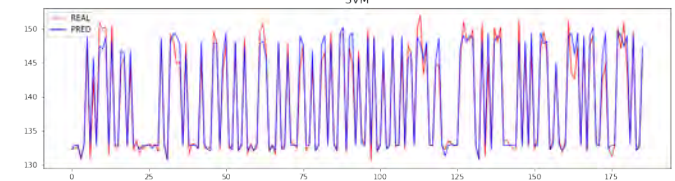


Fig. 5. Real data vs. SVM prediction for mill inlet temperature.

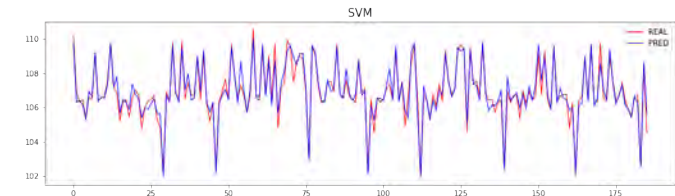


Fig. 6. Real data vs. SVM prediction for mill exit temperature.

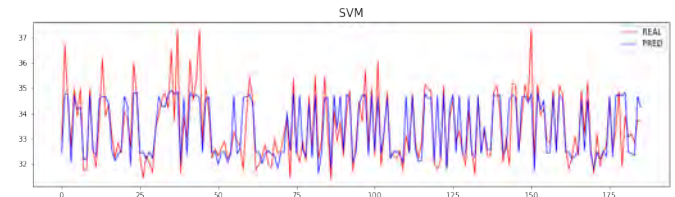


Fig. 7. Real data vs. SVM prediction for mill inlet pressure.

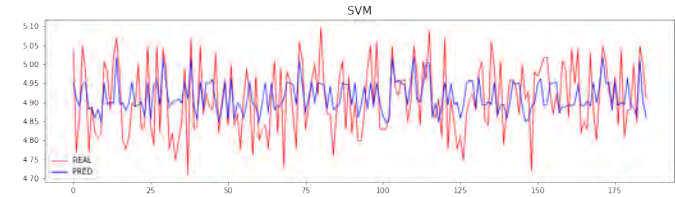


Fig. 8. Real data vs. SVM prediction for bucket elevator power.

V. CONCLUSION

This paper outlined the necessary steps for creation of a real time optimizer for the autonomous control of a vertical cement roller mill. It analyzed the process, listed the most important variables and the challenges their analysis pose, and presented the initial results from using a machine learning approach to build predictive models for the essential signals. The results are encouraging and work is being done on the next stages of the optimizer, which consist of building a composite model out of the developed predictors, defining an objective function, and practical implementation in line with the actual VCRM process. The success of the machine learning prediction models also owes to the access to a large quantity of historical process data. Encouraged by some initial results from the testing of a developed real time optimizer on the plant which uses the prediction models outlined here, the authors hope to be able to present those achievements in a future publication.

REFERENCES

- [1] V. Uraikul, C.W. Chan, and P. Tontiwachwuthikul, "Artificial intelligence for monitoring and supervisory control of process systems," *Engineering Applications of Artificial Intelligence*, vol. 20, pp. 115-131, 2007.
- [2] A. Diez-Oliván, J. Del Ser, D. Galar, and B. Sierra, "Data fusion and machine learning for industrial prognosis: Trends and perspectives towards industry 4.0," *Information Fusion*, vol. 50, pp. 92-111, 2019.
- [3] O. Labahn, *Cement engineers' handbook*, Bauverlag, 1971.
- [4] D. Altun, H. Benzer, N. Aydogan, and C. Gerold, "Operational parameters affecting the vertical roller mill performance," *Minerals engineering*, vol. 103-104, pp. 67-71, 2017.
- [5] M. Reichert, C. Gerold, A. Fredriksson, G. Adolfsson, and H. Lieberwirth, "Research of iron ore grinding in a vertical-roller-mill," *Minerals engineering*, vol. 73, pp. 109-115, 2015.
- [6] J.I. Bhatti, F.M. Miller, S.H. Kosmatka, and R. Bohan, *Innovations in Portland cement manufacturing*, Portland Cement Association Washington DC, 2004.
- [7] J. Koch, *Vertical mills for raw and cement grinding – inspection and evaluation*, Heidelberg Cement, 2010.

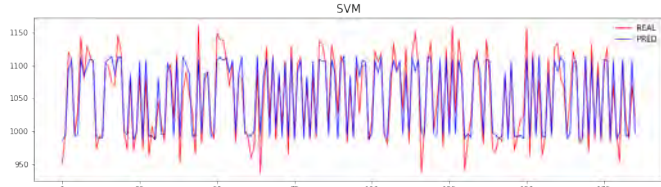


Fig. 9. Real data vs. SVM prediction for mill motor power.

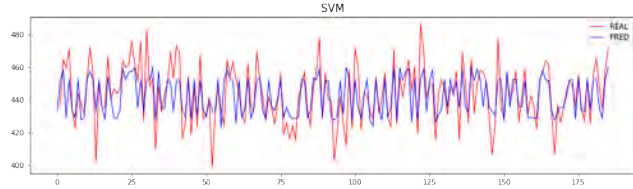


Fig. 10. Real data vs. SVM prediction for fan motor power.

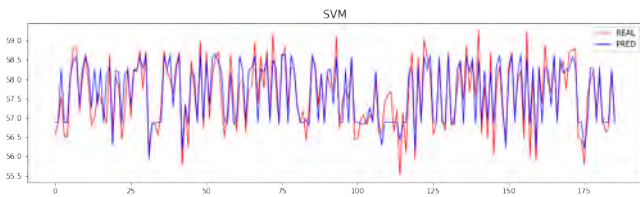


Fig. 11. Real data vs. SVM prediction for separator motor power.

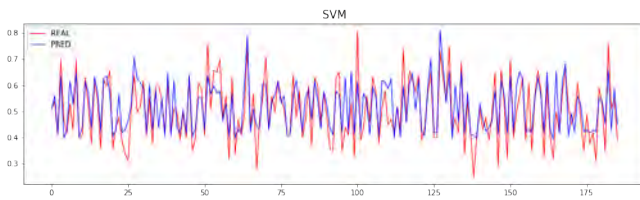


Fig. 12. Real data vs. SVM prediction for mill vibration.

TABLE II. SVM PREDICTION ERRORS FOR ALL VARIABLE MODELS

Variable	SVM Prediction Errors According to 3 Metrics		
	<i>MAPE</i>	<i>MASE</i>	<i>MAE</i>
Bag filter pressure	1.3384	43.9349	0.115
Mill inlet temperature	0.7491	12.8687	1.0764
Mill exit temperature	0.2754	17.3162	0.2944
Mill inlet pressure	1.6731	37.5416	0.5631
Bucket elevator power	1.3328	61.245	0.0654
Mill motor power	1.7992	24.6989	18.898
Fan motor power	2.1030	47.5448	9.3161
Separator motor power	0.5769	32.1765	0.3317
Mill vibration	11.8166	37.1169	0.0552

Selecting an Optimisation Algorithm for Optimal Energy Management in Grid-Connected Hybrid Microgrid with Stochastic Load

Natasha Dimishkovska, Atanas Iliev, Borce Postolov

Faculty of Electrical Engineering and Information Technologies

Ss. Cyril and Methodius University in Skopje

dimishkovskan@feit.ukim.edu.mk, ailiev@feit.ukim.edu.mk, borce.postolov@hotmail.com

Abstract— Decentralisation of the power system and the implementation of microgrids into the standard power system, leads to a complex system which requires a reliable operation and a proper energy management. Finding the right set of optimisation algorithms is the base for solving the optimisation problem. This paper overviews the usage of optimisation algorithms for microgrid energy management, with an accent on a classical optimisation algorithm (Dynamic Programming), and heuristic algorithms, such as Genetic Algorithm and Particle Swarm Optimisation. The paper also proposes a methodology for optimal energy management in a hybrid grid-connected distribution microgrid, with a storage system and stochastic load. The algorithm analyses the optimal scheduling of the installed generators considering the state of charge of the battery and electricity price for power trading with the utility grid. The optimal solution is the most economically justified solution from which the microgrid can benefit, and the one with the least impact on the nodal voltages.

Keywords—*Microgrid; Optimisation Methodology; Energy Management System; Unit Commitment*

I. INTRODUCTION

The battle for nature salvation, against the fossil fuel using power plants, triggers the alarm for enhancing the operation of Renewable Energy Sources (RES) within the power system. Using clean energy is both economic and environmental friendly. Implementing the renewables for local power generation enables the consumers to become producers of electrical energy, who are independent and self-sustained. Driven by the will for providing electricity for every household and at the same time respecting nature, the number of microgrids implemented into the standard power systems increases. As far, there are some basic rules accepted by the microgrids community, regarding the proper operation and maintenance [1]. However, under the lack of regulations and technical guidelines, there are still some obstacles to the seamless operation of grid-connected microgrids.

A microgrid represents a small power system, connected to the utility grid, consisting of RES and local consumers, and, optionally, a diesel generator and storage systems for storing the unused electrical energy, for its further usage or trading

with the main power grid. These components are interconnected, and they operate as a single controllable unit. That way, microgrids enable a clean and self-maintained way of power generation, meaning that they can work as a separate entity, isolated from the power system. In addition to the environmental benefits they provide, the latter is another reason microgrids are accepted worldwide [2]. Since the main sources of power in the microgrids depend on the weather conditions, their switching off and on can cause disturbance in the power system. Therefore, the microgrid has to be secured with a stable voltage and frequency.

Grid-connected microgrids can also trade electric power with the utility grid they are connected to. The price of the electricity is previously determined. For that purpose, there is a smart energy management system in the microgrids, which determines whether it is more economically justified to store the power or to sell it [3]. This is beneficial to the power system because it enhances the economy, the reliability of supply and lowers the burden that the spinning reserve carries.

The optimisation of microgrids is similar to the one for standard power systems, except that it has to take into consideration the weather conditions. The change of the weather conditions has a huge impact on the nodal voltages in the microgrid. Therefore, proper optimisation regarding the operation of the distributed generators is required.

The literature consists of many different optimisation algorithms, but the choice for a certain algorithm is based on the number of constraints and the complexity of the problem.

This paper analyses the optimisation algorithms used in the literature and proposes a methodology for enhancing the algorithms for improving the results.

II. LITERATURE REVIEW

Along with the increased microgrid implementation, the need for its optimisation increased in the last decade. The numerous research on finding the best optimisation algorithm testify to the importance of the optimisation of the microgrids. The optimal work of the microgrids means a proper and safe power system operation. That includes choosing the best

location, size, and configuration of the microgrid, management, and control of the distributed generators and loads.

Depending on the aspect of view, whether the costs for generation are optimised, nodal voltage disruptions, or microgrids' impact on the power system, there are many different algorithms and mathematical methods applied for solving the microgrid optimisation problem. Researchers are constantly working on proposing novel optimisation algorithms or improving the classical algorithms which can solve the unit commitment problem of microgrids by simplifying it [4] [5] [6] [7].

For instance, paper [8] proposes a solution to the unit commitment problem in a microgrid supported with a battery system, by implementing the Most Valuable Player Algorithm (MVPA). This algorithm is a new metaheuristic optimisation algorithm inspired by actual sports events. The optimisation is subject to the operation costs. The results using the proposed method are satisfactory for the analysed microgrid configurations and operation scenarios, neglecting the power demand and power generation variations.

But, generally, there are three most applied algorithms for this problem: the Dynamic Programming Method (DP), Genetic Algorithm (GA), and Particle Swarm Optimisation (PSO) [9] [10]. In [11] these algorithms are overviewed and compared. The paper provides a clear picture of each method's usage and application.

Reference [12] presents an overview of six metaheuristic algorithms for cost minimisation of microgrids. The paper, through a comparative analysis, using different performance indicators for a microgrid, provides directions for choosing the most suitable optimisation technique for a grid-connected hybrid microgrid cost minimisation.

The unit commitment problem in microgrids is a complex problem requiring an algorithm that gathers all of the constraints. Dynamic Programming method (DP) is a classical optimisation method that can be used for solving unit commitment problems in microgrids, as presented in [13] and [14]. However, adding the nodal voltages' and distribution lines' limits, the optimisation requires a more evolutionary algorithm.

A comparison between a classical optimisation algorithm and a metaheuristic algorithm is presented in [15]. The algorithms are used for minimising the fuel costs and CO₂ emissions for a micro gas turbine in a microgrid. The results show that the PSO algorithm is applicable for solving the unit commitment problem in the microgrids, and it is more effective compared to DP.

In [16] a hierarchical GA is implemented for maximising the profit from energy exchange of a microgrid with the utility grid, assuming a Time of Use (TOU) energy policy.

The [1] presents an improved GA which minimises the costs of an islanded microgrid and maximises the benefits when it is connected to the grid. The algorithm uses a simulated annealing technique to accelerate the convergence, leaving the bad individuals in the GA in the earlier stages.

Paper [17] presents a day-ahead energy storage system scheduling in a microgrid, by using the GA and PSO. The paper gives a contribution to optimal microgrid scheduling by minimising the costs of microgrid operation, which are defined by dynamic pricing. The goal is to optimise the operation of the distributed generators and battery so that in times of high prices, the stored energy would be used. The paper compares the applicability of the two optimisation algorithms, which results in a better performance of the PSO.

The [18] proposes an improved PSO algorithm for unit commitment in microgrids. Additionally, cost functions for determining the state of charging and discharging of the battery and a dynamic penalty function are introduced. The results show improvement in cost reduction by 12 %.

In [19] wind power-based microgrid supported with fuel cells, a diesel generator, and an electrolyser is analysed. The fuel cell is used in times of energy demand which is not satisfied by the wind turbine. The paper proposes a PSO-based algorithm to minimise the operation costs. The results show nearly 70% cost reduction and economic operation of the microgrid.

The PSO algorithm is used for cost optimisation in a grid-connected microgrid, with a capability of islanded work in [20]. The proposed algorithm considers the variations of the distributed generators and power demand proposing a day-ahead forecast for overcoming this issue.

In [21] voltage disruptions caused by connected distributed generation are analysed. This is the starting point to finding the optimal placement of the distributed generation. Using PSO, the objective function of line losses, voltage stability index, and node voltage deviation of the system is optimised to determine the capacity and location of distributed generation.

Another methodology for optimisation of distributed generation considering the costs and voltage stability was introduced in [22]. The multi-objective optimisation uses two techniques: the sum-weighted Pareto front and an adapted goal programming methodology. In this paper, the voltage stability is "measured" by the load index value (L -index).

III. PROBLEM DEFINITION

The implementation of the microgrids represents a big step into a future clean energy power system and it is a big change that has come along to a very positive reaction from the people. However, it is still challenging for people to adjust their behaviour and their habits to the microgrid operation. For instance, cooking or showering during a certain part of the day. If people's habits follow the weather conditions and power generation practise, the implementation of the microgrids will be very easy and there would not be a need for a smart energy system that follows the consumption habits. However, since that practice is not very likely applicable, and power demand is a stochastic process, the microgrids' operation has to be adjusted to the consumption while respecting certain constraints.

Grid-connected microgrids can rely on the utility network, as a backup power source in times of need. However, its

operation should not impact the normal operation mode of the utility grid, especially not the consumers. Therefore, it is necessary to determine the optimal schedule of distributed generators and battery systems.

The constraints usually refer to the technical limits of the installed equipment and system's balance. But, besides the technical limitations of the installed generator and battery, the microgrid operation should consider the nodal voltages, limitations of the power bought from the utility on occasions when the generators do not produce any power and the battery is empty, and the power demand. These parameters, have to be in a perfect balance, in which they can overcome the variety of uncertainties regarding the weather conditions and power demand.

In a grid-connected microgrid, additionally, the electricity prices have to be considered in order to find an optimal operation plan. This adds to the complexity of the unit commitment problem in the grid-connected microgrids, which is different from the unit commitment problem in standard power systems [1].

IV. PROPOSED METHODOLOGY

The optimisation of a microgrid is a complex problem consisting of multiple constraints, from the technical limits of the equipment to the balance between the production and consumption of power and the stable power supply. Grid-connected microgrids have a great advantage of being connected to the utility grid, which represents a backup in emergencies when there is an interruption in the power supply. However, being connected to the utility grid brings a big responsibility to voltage stability.

This paper proposes a methodology for creating an algorithm that considers the probability of power supply from the installed distributed generators, the uncertainty of power demand, state of charge of the battery system, and the probability of voltage sags and proposes an optimal solution by minimizing the operation costs.

Most of the microgrid optimisation research focus on the operation costs and technical constraints of the microgrid's components. However, the nodal voltages should be inspected too, in order to define one solution as the optimal one. This invokes the penalty costs for not satisfying the defined conditions for a proper microgrid's solution.

The objective function is subject to the total costs for microgrid operation:

$$F(C) = \max \left\{ \sum_{i=1}^T (B_{DER,i} - C_{grid,i}) - C_{penalty} \right\} \quad (1)$$

where, $B_{DER,i}$ refers to the profit for selling the excess power to the utility grid in the i -th hour, and $C_{grid,i}$ refers to the costs for buying power from the grid in the i -th hour. Additionally, the penalty costs $C_{penalty}$ for not supplying quality electrical power, with stable voltage, are considered.

Adding penalty costs levels up the reliability of the power supply of the microgrids and the standards for electrical power quality and proper operation.

The constraints consider the installed power capacity of the distributed generators:

$$P_{\min,DER} \leq P_{DER} \leq P_{\max,DER} \quad (2)$$

Power limits of the battery:

$$0 \leq P_{bat} \leq P_{\max,bat} \quad (3)$$

Buying power from the grid:

$$0 \leq P_{grid} \leq P_{\max,grid} \quad (4)$$

Nodal voltage levels:

$$0.95 \cdot U_{r,node} \leq U_{r,node} \leq 1.05 \cdot U_{r,node} \quad (5)$$

The power bought from the utility grid should be enough to supply the load in the microgrid. However, if there is a legal frame that defines some of the load as a priority, then the maximum quantity of power bought from the grid can be enough to supply the priority load.

Since the weather conditions are not predictable and the forecast is not a hundred percent accurate, the algorithm should be able to follow the power production from the distributed generators and take information about the battery state of charge constantly, as often as possible. Only then it can maintain the voltage levels and optimise the microgrid operation.

Additionally, the algorithm should take information for the electrical power prices, and then decide whether the excess power from the microgrids is going to be sold to the utility grid or stored for further use. Also, this decision applies in times of power production shortage, i.e. whether the needed power should be bought from the grid or taken from the storage system. A solution to this problem was proposed in [23] and [24] by using the convex optimisation technique.

The selection for the optimal solution is based on satisfying the before mentioned constraints. That means that the optimisation algorithm should optimise the generators' operation to optimise the power losses and maximise the profit from power trading. The optimisation is described by the following steps:

- First, the data for maximum power generation for the installed generators and power demand is entered.
- Then, the values for power generated from the generators at the analysed moment are compared to the constraints.
- Additionally, each solution is applied to the network and nodal voltages are inspected. This step is very important since the microgrids are small-scale systems, whose stability depends highly on the generators' performance.

- The process continues for a limited number of iterations. The algorithm memorises the optimal solution in a way that compares it to the global best.
- In the final step, total profit from trading power with the utility grid is calculated and the penalty costs are evaluated based on the time of an outage. The solution with the highest profit is considered to be the optimal one.

V. DESCRIPTION OF THE OPTIMISATION ALGORITHMS

In this section, the Dynamic Programming (DP) method, GA and PSO basic settings will be discussed, and the comparison between the three methods will be presented.

A. Dynamic programming

DP method is a classical optimisation method set by Bellman in the 1950s. The method provides an optimal solution to a certain issue by dividing the main problem into many smaller sub-problems. The DP optimisation method uses a set of algorithms for finding an optimal solution to a wide range of input data. The optimisation is done by maximising or minimising an objective function.

The solution to each of the sub-problems eventually gives the optimal solution to the main problem. Although the method can be classified as a “divide and conquer” group of methods, it works opposite of them [25]. The optimisation is done by analysing the sub-problems first, which are simpler expressions. The solution of each sub-problem is memorised. The set of all conditionally optimal solutions leads to the solution of the main problem.

Many nonlinear problems, from any field, can be solved using the DP method. Its application is widely known for power system planning, optimal unit commitment in complex power systems, which cannot be solved by standard methods of nonlinear programming and energy management optimisation. In power systems optimisation, usually, the method is used for minimising the costs or maximising the profit. In energy management optimisation, the method is mostly used for the optimisation of emissions from the power plants [26].

The simple microgrid optimisation and unit commitment problem can be solved using classical optimisation techniques, such as the Dynamic Programming (DP) method, as presented in [27]. However, adding the voltage stability constraint makes the problem more complex, and therefore a different optimisation technique is required.

B. Genetic Algorithm

The most commonly used heuristic optimisation algorithms are the Genetic Algorithm (GA) and Particle Swarm Optimisation (PSO). Each of these algorithms is based on natural running processes.

GA is a heuristic optimisation technique, inspired by the Darwinian principle of evolution through genetic mutation and selection. The method is an abstract version of the evolutionary process in which for a certain number of populations of chromosomes, a mutation and selection are made [28]. The

chromosomes are encoded strings (usually in a binary system) that carry the information of one generation [29]. The chromosomes are tested for fitness in a certain function, which grades the solution to the analysed problem. If the solution is satisfactory, then the next generation is created.

Each generation of chromosomes has parents. The genes of the child chromosomes are created by two parent chromosomes. For that purpose, a crossover should be defined. The crossover is a point to which recombination of genes of the parent chromosomes is made. The new set of genes is the child chromosome. In the next step, a mutation of a certain gene is done.

The iterations continue for a certain population. As the number of iterations increases, eventually, the chromosomes' fitness increases and the solution improves. The process runs until the stopping criteria are reached. In this way, the optimal solution to a problem is determined [30].

C. Particle Swarm Optimisation

PSO is an optimisation technique inspired by the motion of bird flocks and schooling fish [31]. Similar to the GA, in PSO, the system is initialized with a population of random solutions, and the search for the optimal solution is performed by updating generations. However, this method does not have crossover and mutation steps and it requires a lower number of iterations [32].

PSO method is based on the movement of particles in space, which in the algorithm, represent the potential solutions [32]. At each point the algorithm memorises the best performance of the particle (the best solution), creating the optimal movement path. Although this might seem like an advantage, it decreases the method's accuracy.

D. GA and PSO combination

The comparison between GA and PSO performance on optimising a hybrid RES system presented in [33] shows that both, GA and PSO, are efficient for optimising complex problems. However, each of the optimisation methods achieves better results under well-defined objective functions and constraints. PSO is computationally more efficient than GA in terms of both speed and memory requirements. And although it is less practical, it is found to be quite applicable for unit commitment problems in microgrids [33].

Many research combine the GA and PSO, creating an even better and more efficient optimisation algorithm. In [35] a combined GA and PSO algorithm is proposed for optimisation and sizing of distributed generation. The proposed algorithm should minimise network power losses, improve voltage regulation, and improve voltage stability. The PSO algorithm is used for finding the optimal sizing of the distributed generation, and GA is used for calculating the optimal sitting of the distributed generation. The results show that the combination of these two algorithms provides a better solution than their separate usage.

This shows that for some problems, the best solution is provided by combining two optimisation algorithms. For the

presented microgrid optimisation problem, we propose a combination of GA and PSO.

VI. SELECTION OF AN OPTIMISATION ALGORITHM

The summary of the advantages and disadvantages of some optimisation algorithms in Table 1, shows that the DP method cannot be used for the optimisation of complex systems as the one presented in sections III and IV. However, the combination of GA and PSO can and will be used for that purpose.

TABLE I. COMPARISON OF OPTIMISATION ALGORITHMS

Optimisation Algorithms	Advantages	Disadvantages
<i>Dynamic Programming</i>	<ul style="list-style-type: none"> - Splitting a problem into simpler sub-problems. - Solving much simpler problems gives the optimal solution. - Easy to implement to any kind of problem. 	Complex problems which require multi-objective optimisation cannot be solved.
<i>Genetic Algorithm</i>	<ul style="list-style-type: none"> - The optimal solution is calculated through a selection of multiple iterations, each one better than the other. - Fast convergence. - Can be used in many fields. 	Selecting true selection criteria, crossover, and mutation parameters are essential for better optimisation.
<i>Particle Swarm Optimisation</i>	<ul style="list-style-type: none"> - Does not require crossover and mutation parameters. - Memorises the previous conditional optimal solution. - Fast convergence. - Applicable for many different types of optimisation problems. 	Requires complex computations.

Since GA requires a larger number of iterations and has a simpler computation algorithm, it will be used for the unit commitment problem of the microgrid. The output of the GA (the optimal solution) will be a parameter to which the voltage stability of the nodes will be computed. For that purpose, the PSO can be used.

The optimisation algorithm consisting of both, GA and PSO, will give the unit commitment and economic dispatch of the microgrid.

VII. CONCLUSION

Each microgrid is unique. There are many microgrid test-beds around the world and each of them differs from the others. Starting from the location and to its performance.

Therefore, there is not an empirical solution to the microgrid optimisation problem. And that is what makes it challenging.

This paper presented the most used algorithms for the optimisation of grid-connected microgrids. Since it is a complex problem that requires a detailed analysis of every entity of the microgrids, at each moment, it cannot be said that the solution is unified. Each research proposes a unique solution to this problem. This means that in the planning and operation of microgrids, one can choose which optimisation algorithm suits the best for one's microgrid, according to its unique constraints.

The paper proposed a methodology for solving a complex optimisation problem, considering the uncertainties of weather conditions and power demand. The algorithm considers the penalty costs, as much stronger criteria for obtaining an optimal solution.

Dividing the problem into two different problems (unit commitment and voltage stability), which will be computed separately by GA and PSO, will simplify this problem and provide the optimal solution of the microgrid's operation.

In future work, the implementation of the proposed methodology on a test example and the results of that case study will be presented and discussed.

REFERENCES

- [1] Liang, H. Z., & Gooi, H. B. (2010). Unit commitment in microgrids by improved genetic algorithm. *Conference Proceedings IPEC* (pp. 842-847). DOI: 10.1109/IPEC2010.5697083.
- [2] Delfino, F., Procopio, R., Rossi, M., Brignone, M., Robba, M., & Bracco, S. (2018, ISBN: 9781630811518). *Microgrid Design and Operation: Toward Smart Energy in Cities*. Norwood: Artech House.
- [3] Li, R. (2019). Chapter 6 - Grid-connected operation and engineering application of distributed resources. In R. Li, *Distributed Power Resources* (pp. 177-232, <https://doi.org/10.1016/B978-0-12-817447-0.00006-7>). Academic Press.
- [4] Dey, B., & Bhattacharyya, B. (2018). Dynamic Cost Analysis of a Grid-Connected Microgrid Using Neighbourhood Based Differential Evolution Technique. *Int Trans Electr Energ Syst*. 2019; 29:e2665, <https://doi.org/10.1002/etep.2665>.
- [5] Paliwal, N. K., Singh, N. K., & Singh, A. K. (2016). Optimal power flow in grid-connected microgrid using Artificial Bee Colony Algorithm. *2016 IEEE Region 10 Conference (TENCON)* (pp. 671-675, DOI: 10.1109/TENCON.2016.7848087). Singapore: IEEE.
- [6] Antonyraj, S., & Samuel, G. G. (2019). Optimal Energy Scheduling of Renewable Energy Sources in Smart Grid using Cuckoo Optimization Algorithm with Enhanced Local Search. *National Conference on Recent Advances in Fuel Cells and Solar Energy* (DOI: 10.088/1755-1315/312/012014). Karaikal, U.T of Puducherry, India: IOP Conference Series: Earth and Environmental Science 312 (2019) 012014.
- [7] El-Hendawi, M., A.Gabbar, H., El-Saady, G., & Ibrahim, E.-N. A. (2018). Optimal operation and battery management in a grid-connected microgrid. *Journal of International Council on Electrical Engineering*, 8:1, pp. 195-206, DOI: 10.1080/22348972.2018.1528662.

- [8] Ramil, M. A., Boucekara, H. R., & Alghamdi, A. S. (2019). Efficient Energy Management in a Microgrid with Intermittent Renewable Energy and Storage Sources. *Sustainability*, 11, 3839, <https://doi.org/10.3390/su11143839>.
- [9] Rendroyoko, I., Sinisuka, N. I., & Koesrindartoto, D. P. (2019). A Literature Survey of Optimisation Techniques of Unit Commitment Implementation in Microgrid Electricity System With Renewable Energy Sources. *2019 2nd International Conference on High Voltage Engineering and Power Systems (ICHVEPS)* (pp. 344-349, DOI: 10.1109/ICHVEPS47643.2019.9011080). Denpasar, Indonesia: IEEE.
- [10] Hassan, A. S., Sun, Y., & Wang, Z. (2020). Optimization techniques applied for optimal planning and integration of renewable energy sources based on distributed generation: Recent trends. *Cogent Engineering*, vol. 7, no. 1, DOI: 10.1080/23311916.2020.1766394.
- [11] Sanchez-Huertas, W., Gómez, V., & Hernández, C. (2018). Optimization Algorithms for Solving Microgrid and Smart grid Integration Problems. *International Journal of Applied Engineering Research*, Vol. 13, No. 21, pp. 14886-14892.
- [12] Khan, B., & Singh, P. (2017). Selecting a Meta-Heuristic Technique for Smart Micro-Grid Optimization Problem: A Comprehensive Analysis. *IEEE Access*, vol. 5, pp. 13951-13977, DOI: 10.1109/ACCESS.2017.2728683.
- [13] Dimishkovska, N., & Iliev, A. (2021). Application of Dynamic Programming for Optimal Unit Commitment and Economic Dispatch of Distribution Networks. *International Journal on Information Technologies and Security*, No.1 (vol. 13), pp. 17-26.
- [14] Park, K., Lee, W., & Won, D. (2019). Optimal Energy Management of DC Microgrid System using Dynamic Programming. *IFAC-PapersOnLine*, vol. 52, no. 4, pp. 194-199; <https://doi.org/10.1016/j.ifacol.2019.08.178>.
- [15] Borra, V. S., & Debnath, K. (2019). Comparison Between the Dynamic Programming and Particle Swarm Optimization for Solving Unit Commitment Problems. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 395-400, DOI: 10.1109/JEEIT.2019.8717481). Amman, Jordan: IEEE.
- [16] Leonori, S., Pascherno, M., Mascioli, F. M., & Rizzi, A. (2019). Optimization Strategies for Microgrid Energy Management Systems by Genetic Algorithms. *Applied Soft Computing Journal*, Vol. 86, 105903, <https://doi.org/10.1016/j.asoc.2019.105903>.
- [17] Raghavan, A., Maan, P., & Shenoy, A. K. (2020). Optimization of Day-Ahead Energy Storage System Scheduling in Microgrid Using Genetic Algorithm and Particle Swarm Optimization. *IEEE Access*, vol. 8, pp. 173068-173078, DOI: 10.1109/ACCESS.2020.3025673.
- [18] Hossain, M. A., Pota, H. R., Squartini, S., & Abdou, A. F. (2019). Modified PSO Algorithms for Real-Time Energy Management in Grid-Connected Microgrids. *Renewable Energy*, Vol. 136, pp. 746-757, <https://doi.org/10.1016/j.renene.2019.01.005>.
- [19] Nivedha, R. R., Singh, J. G., & Ongsakul, W. (2018). PSO-based economic dispatch of a hybrid microgrid system. *2018 International Conference on Power, Signals, Control and Computation (EPSCICON)* (pp. 1-5, DOI: 10.1109/EPSCICON.2018.8379595). Thrissur, India: IEEE.
- [20] Li, H., Eseye, A. T., Zhang, J., & Zheng, D. (2017). Optimal energy management for industrial microgrids with high-penetration renewables. *Prot. Control Mod. Power Syst.* vol. 2, article no. 12, <https://doi.org/10.1186/s41601-017-0040-6>.
- [21] Dejun, E., Pengcheng, L., Zhiwei, P., & Jiaxiang, O. (2014). Research of voltage caused by distributed generation and optimal allocation of distributed generation. *2014 International Conference on Power System Technology* (pp. 3098-3102, DOI: 10.1109/POWERCON.2014.6993599). Chengdu, China: IEEE.
- [22] Morais, H., Sousa, T., Perez, A., Jóhannsson, H., & Vale, Z. (2016). Energy Optimization for Distributed Energy Resources Scheduling with Enhancements in Voltage Stability Margin. *Mathematical Problems in Engineering*, vol. 2016, Article ID 6379253, 20 pages, <https://doi.org/10.1155/2016/6379253>.
- [23] Ramachandran, E. M., & Chandrakala, K. R. (2019). Dynamic Pricing Based Optimal Power Mix of Grid Connected Micro Grid Using Energy Management System. *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1-5, DOI: 10.1109/i-PACT44901.2019.8960088). Vellore, India: IEEE.
- [24] Silani, A., & Yazdanpanah, M. J. (2019). Distributed Optimal Microgrid Energy Management With Considering Stochastic Load. *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 729-737, DOI: 10.1109/TSTE.2018.2846279.
- [25] Réveillac, J.-M. (2016). Dynamic Programming. In J.-M. Réveillac, *Optimization Tools for Logistics* (pp. 55-75). ISTE Press - Elsevier.
- [26] Verbič, G., Mhanna, S., & Chapman, A. C. (2019). Chapter 5 - Energizing Demand Side Participation. In A. Taşçıkaraoğlu, & O. Erdiñ, *Pathways to a Smarter Power System* (pp. 115-181, <https://doi.org/10.1016/C2017-0-03015-X>). Academic Press.
- [27] Yuan, J., Chen, Z., Wang, X., Zeng, X., & Zhang, Y. (2020). An Energy Management System Based on Adaptive Dynamic Programming for Microgrid Economic Operation. *2020 Chinese Automation Congress (CAC)* (pp. 1459-1464, DOI: 10.1109/CAC51589.2020.9327528). Shanghai, China: IEEE.
- [28] Siau, K. (2003). E-Creativity and E-Innovation. *The International Handbook on Innovation*, pp. 258-264, <https://doi.org/10.1016/B978-008044198-6/50017-6>.
- [29] Mohamed, F. A., & Koivo, H. N. (2012). Online management genetic algorithms of microgrid for residential application. *Energy Conversion and Management*, vol. 64, pp. 562-568, <https://doi.org/10.1016/j.enconman.2012.06.010>.
- [30] McCall, J. (2005). Genetic algorithms for modeling and optimisation. *Journal of Computational and Applied Mathematics* 184 (2005), pp. 205-222.
- [31] Lu, H., Chen, J., & Guo, L. (2018). 5.7 Energy Quality Management. In I. Dincer, *Comprehensive Energy Systems*, ISBN 9780128149256 (pp. 258-314, <https://doi.org/10.1016/B978-0-12-809597-3.00521-6>). Elsevier,
- [32] Sahab, M. G., Toropov, V. V., & Gandomi, A. H. (2013). 2 - A Review on Traditional and Modern Structural Optimization: Problems and Techniques. In A. H. Gandomi, X.-S. Yang, S. Talatahari, & A. H. Alavi, *Metaheuristic Applications in Structures and Infrastructures* (pp. 25-47, <https://doi.org/10.1016/B978-0-12-398364-0.00002-4>). Elsevier,
- [33] Yang, X.-S. (2021). Chapter 8 - Particle Swarm Optimization. In X.-S. Yang, *Nature-Inspired Optimization Algorithms (Second Edition)* (pp. 111-121, <https://doi.org/10.1016/B978-0-12-821986-7.00015-9>). Academic Press.
- [34] Torres-Madroño, J. L., Nieto-Londoño, C., & Sierra-Pérez, J. (2020). Hybrid Energy Systems Sizing for the Colombian Context: A Genetic Algorithm and Particle Swarm Optimization Approach. *Energies*, vol. 13, no. 21:5648, <https://doi.org/10.3390/en13215648>.
- [35] Moradi, M., & Abedini, M. (2012). A combination of genetic algorithm and particle swarm optimization for optimal DG location and sizing in distribution systems, *International Journal of Electrical Power & Energy Systems*, vol. 3, no. 1, pp. 66-74, <https://doi.org/10.1016/j.ijepes.2011.08.023>.
- [36] Pancholi, R. K., & Swarup, K. S. (2003). Particle swarm optimization for economic dispatch with line flow and voltage constraints [power generation scheduling]. *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region* (pp. 450-455 Vol.1, DOI: 10.1109/TENCON.2003.1273363). Bangalore, India: IEEE.

Comparative Analysis of Different Heliostat Field Control Algorithms

Ivan Andonov¹, Vesna Ojleska Latkoska¹, Mile Stankovski¹

¹Faculty of Electrical Engineering and Information Technologies,

“Ss. Cyril and Methodius” University in Skopje,

Rugjer Boskovic bb, P.O. box 574, 1000 Skopje, Republic of Macedonia

ivan_andonov_94@hotmail.com, vojleska@feit.ukim.edu.mk, milestk@feit.ukim.edu.mk

Abstract - This study presents the use of various algorithms for control of a field of heliostats, through which a thermal power plant with concentrated solar energy is controlled. The design of the control algorithms consists of several steps. First, in order to obtain the mathematical model of the system, the real system is identified according to the gray box and the least-square method. The data used to identify the system is generated by step excitation on the real system, for a specific sampling period. The resulting mathematical model is used to design and simulate a continuous and discrete PID controller, Mamdani and Sugeno fuzzy logic controllers, as well as ANFIS based fuzzy logic controller. The results of the applied controllers are analyzed and compared, based on the output overshoot, the rise and settling time. It can be concluded that we got best results (least settling time and the least overshoot) when fuzzy logic controller with ANFIS was used, while in terms of speed and rise time, the best results were obtained when discrete PID control algorithm was used.

Keywords: *Heliostat, least-square, PID, fuzzy logic control, Adaptive Neuro-Fuzzy Inference System*

I. INTRODUCTION

Heliostat is a motorized mirror which is used to reflect the solar radiation into a receiver mounted on a tower. Heliostats are a basic component of a thermal power plant with a tower and even 40-50% of the cost of the entire plant. Recently, more and more investments are being made in research and analysis to reduce the cost of heliostats and thus the plant itself [1],[2].

The aim of heliostat is to reflect the sunlight on predefined target and therefore, for heliostat control, the orientation of the mirror needs to be known in order to determine deviations and precisely control the actuators to minimize the error of the overall system. The conventional approach uses open-loop calibration and control, while modern solutions use feedback sensors and closed-loop control [19]. A prerequisite for open loop control is very small statistical error or backlash and stable, observable system behavior. Calibration effort can be high with error systems. In addition, the heliostat geometry model must be appropriate to describe the real imperfections that can be changed [18].

Since the axis of each heliostat is driven by an electric motor, much of the heliostat control challenge comes down to control of the motor. An electric motor is a device that converts electrical energy into mechanical energy. The principle of working is the interaction between the magnetic fields generated by the stator or motor rotor magnets and the magnetic field created by the electric current in the windings, which generates a force in the form of torque on the motor axis. In applications such as the heliostat, DC motors are most commonly used for several reasons: higher starting power and torque required for the minimum heliostat rotation time for a certain angle, faster start-up response time, stopping or acceleration, which makes them more accurate and easier to control, simpler to install and cheaper.

The DC motor is controlled by changing the voltage with a PWM signal from the controller. The most commonly controlled variables are speed and position, and 95% of industry applications use a PID controller [7]. However, in processes where the dynamics change due to nonlinearity and interference, traditional PID controllers can not cope and system oscillations may occur due to precisely (crisp) adjusted controller parameters. The fuzzy logic controller is a good alternative to the PID controller, as it can handle nonlinear systems and can be designed using human operator knowledge without knowing the mathematical model of the system. Although the fuzzy logic controller does not have a better response in the time domain than the PID controller, it can still be applied to systems that have rapid changes, unlike PID which will need to adjust the values of the control parameters [20].

The main limitations of fuzzy logic controllers are the lack of a systematic design methodology and the difficulty in predicting the stability and robustness of a controlled system. Therefore, in many applications, fuzzy logic controllers are improved by fine tuning with the help of neural networks, i.e. the so-called hybrid fuzzy - neural controllers, which use a neural network to determine the rules and to make a conclusion [7].

This paper presents the open-loop control of heliostats as a systems, (Sun tracking algorithm and encoders) based on closed-loop position control of the DC motor explained in [19].

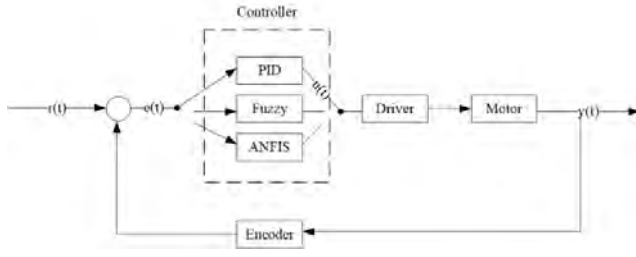


Figure 1: Heliostat control algorithms

In order to make a comparative analysis of the above-mentioned different control algorithms in DC motors, i.e. indirectly in those used in the heliostat system, the paper performs design and selection of control algorithm and strategy for control, described in detail. Several types of control algorithms are used to control the heliostat position: PID, fuzzy-logic, and ANFIS (combination of fuzzy-logic control and neural networks), after which a comparative analysis is made.

In order to achieve the aforementioned control goal, several steps are used to design and select the control algorithm, i.e. (Figure 1):

- 1) Identification of the real system and obtaining a mathematical model;
- 2) Control algorithm design;
- 3) Implementation on the mathematical model;
- 4) Analysis of the obtained results and selection according to the overshoot, settling time and rise time.

II. IDENTIFICATION

Mathematical models are commonly used to describe system behavior, and thus to simulate and design a controller. Depending on the knowledge (a priori information) about the system, obtaining the mathematical model can be done in the following ways [4]:

- Using white-box method - the mathematical model is obtained by applying the physical principles of modeling the system, while it remains to determine the most commonly given parameters.
- Using gray-box method - modeling occurs with the development of state space models with a known structure. For a given input/output system, there are infinite realizations with spatial variables that give the same connection for a given input/output. However, a particular structure may be desirable for identification. The limited optimization of the model parameters provides the necessary framework for the identification of this method.
- Using black-box method - the modeling uses input/output data without prior knowledge of system behavior. The mathematical model is obtained using neural networks and algorithms for their optimization.

Figure 2 shows the schematic and block diagram of a permanent magnet DC motor. By applying the laws of physics, the mathematical models is obtained, i.e. the transfer function of the system. The detailed parameters of the PMDC motor used are not known, so the gray box method is applied for their

identification (although it can be said that there is not much difference with the black box method except the limitation).

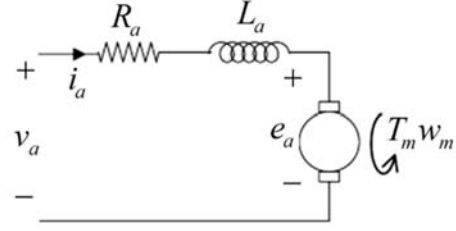


Figure 2: Schematic drawing of PMDC motor

The Laplace transformation is applied to the obtained mathematical model, which gives the transfer function that represents the relationship between the voltage and the rotational speed of the PMDC motor [13]:

$$G_1(s) = \frac{\omega_m(s)}{v_a(s)} = \frac{1/k_b}{t_m t_e s^2 + t_m s + 1} \quad (1)$$

where $t_e = L_a/R_a$ is an electrical time constant, $t_m = RJ/k_t k_b$ is the mechanical time constant, k_t and k_b are torque constants and the back voltage EMF. Mathematically, velocity is a derivative of position with respect to time, which means that position is obtained by integrating velocity with respect to time [15]:

$$\theta_m(t) = \int \omega_m(t) dt \quad (2)$$

which in s domain means multiplying the transfer function by s^{-1} :

$$G(s) = \frac{\theta_m(s)}{v_a(s)} = \frac{1}{s} \times G_1(s) = \frac{1/k_b}{t_m t_e s^3 + t_m s^2 + s} \quad (3)$$

The same principle is used to obtain the mathematical model and the transfer function of the linear actuator. In [14] the obtaining of the second order portable f for the velocity/voltage ratio is shown, while in [16] the third order transfer function is obtained which later due to a simpler analysis is approximated to the first order function.

The identification of the parameters of the transfer function is performed by applying the least-squares method on the output data that are generated with step excitation. The purpose is to give N output data of the variables $y = [y[0], y[1] \dots, y[N-1]]^T$ to get the best prediction (or approximation) of y using p descriptive variables (or regressors) $\varphi_i[k]$, for $i = 1, \dots, p$, so that the predictions $\hat{y} = [\hat{y}[0], \hat{y}[1] \dots, \hat{y}[N-1]]^T$ are collectively at minimum (vector) distance from y [4]. Assume that the approximation of $y[k]$ is through a linear model:

$$\hat{y} = \sum_{i=0}^p \theta_i \varphi_i[k] = \varphi^T[k] \theta \quad (4)$$

where θ is the unknown set of free parameters that need to be optimized to achieve the goal of the smallest squares. We introduce:

$$\Phi = [\varphi[0], \varphi[1] \dots, \varphi[N-1]]^T \quad (5)$$

$$Z = y U \Phi \quad (6)$$

since each $\varphi[k]$ is a $p \times 1$ vector, F is an $N \times p$ matrix. The Z matrix consists of known data. Then, the optimization problem can be written as:

$$\min_{\theta} J_N(Z, \theta) = \|y - \hat{y}\|_2^2 = (y - \hat{y})^T (y - \hat{y}) \quad (7)$$

where $\hat{y} = \Phi\theta$. Finding the minimum can be achieved by the method of descending gradient $dJ/d\theta = 0$:

$$\frac{\partial J}{\partial \theta} = -2\Phi^T (y - \Phi\theta) = 0 \quad (8)$$

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T y \quad (9)$$

where $\hat{\theta}$ represents the minimum obtained from the optimization.

Step excitation of 15 % (3.6 V) of the rated voltage (24 V) is used to obtain the output data y at input due to nonlinear reduction, while displacement and velocity are measured with the microcontroller ESP32 and encoder with a sampling period of 50 ms (20 Hz) [3]. The heliostat is mounted without mirrors, so that the instantaneous mass that affects the movement of the DC motor is 85 kg (horizontal pipe and construction), which is almost half of the total mass of 180 kg (Figure 3).



Figure 3: Real system used for identification

The parameters of the tested PMDC motor and gears are given in Table 1.

Table 1: Parameters of PMDC motor

PMDC motor	
Voltage [V]	24
Output power [W]	40
No-load speed [rpm]	3000
No-load current [A]	1 max
Load [mNm]	136
Load speed [rpm]	2800
Load current [A]	2.6
Gear ratio	1:180
Reduction speed [rpm]	16
Max. load [Nm]	10

By using MATLAB, the curves from the measured data and the identification of the system are calculated, and shown in Figure 4. The accuracy of the obtained transfer function of the identification system is 87.4 % and it is given by:

$$G_1(s) = \frac{\omega_m(s)}{v_a(s)} = \frac{0.4055}{0.0265s^2 + 0.4741s + 1} \quad (10)$$

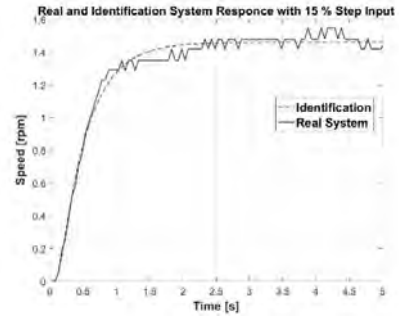


Figure 4: Real and simulated system output with 15 % step input

III. PID CONTROL ALGORITHM

The PID controller is the most widely used control algorithm. Most feedback loops are controlled by this algorithm or small variations of it. It can be implemented in various forms, as a standalone controller, as part of a DDC (direct digital control) or hierarchically distributed process control system. The mathematical representation of the PID controller is [5]:

$$u(t) = K_p \left(e(t) + \frac{1}{T_i} \int_0^t e(\tau) d\tau + T_d \frac{de(t)}{dt} \right) \quad (11)$$

where u is the control signal and e represents the control error ($e = r - y$). The control signal is a set of three terms: P-term is proportional to the error, I is proportional to the error integral, and D-term is proportional to the derivation (change) of the error. The control parameters are the proportional coefficient K_p , the integration time T_i , and the differentiation time T_d .

In cases where only proportional control is used, the control algorithm is represented only by $u(t) = K_p e(t)$, which means that the control signal is proportional to the error. The change of the coefficient K_p affects the change of the system error in the steady state and the occurrence of oscillations and overshoot. Thus, increasing K_p reduces the error in the steady state, but increases the response oscillations [5].

The main function of the integration term I is to ensure that the response of the system matches the reference value in the steady state. With proportional control, there is a steady error, while with the I-term, a small positive error will always lead to an increase in the control signal, and a negative error will give a decreasing control signal. In cases where the integration time $T_i = \infty$, the PI control combination switches to P control only. The steady state error is removed when T_i has finite values. For large values of integration time, the rise time is large, while for small values of T_i the rise time of the response is shorter, but oscillations and overshoot occur (increase of settling time). In the I term, the problem arises from the limitations of the physical systems (actuator length, limited speed, high latency) which leads to a situation where the controlled system reaches the limit, and the term continues to integrate the error and increases (windup). Then the error needs to have the opposite sign for a longer period of time for the control signal to return to normal. The consequence is that any controller with integrated action can cause major changes when the system is saturated (reaches the limit). This problem can be overcome in several ways:

- limiting changes of the reference value;
- back-calculation - when the output is in saturation state, the control term is also recalculated so that its new value gives an output at the saturation limit. And the term is reset dynamically with time constant T_i ;
- tracking - another input is added to the controller which is a tracking signal and is followed by the control signal;
- conditional integration - the term is also excluded when the control is far from steady state and thus the term is used under certain conditions, otherwise it is constant.

The term D is used to improve the stability of a closed loop. Usually due to the dynamics of the process, it takes time to notice the change of the control variable in the response. This will cause the control system to be delayed in correcting the error. The action of the PD control can be described so that the control is proportional to the predicted response of the system, where the prediction is made by extrapolating the error from the tangent to the error curve. It can be said that the term D is used to predict the error in the future. The disadvantage of using the term D is that the ideal output has a very high coefficient for high frequency signals. This means that the high frequency measuring noise will generate large variations of the control signal. This problem is overcome by implementing a first-order filter with a time constant T_d/N . Thus, for small s the transfer function is approximately sK_pT , and for large s it is equal to K_pN . The approximation acts as a derivative for the low frequency components of the signal, and the high frequency

coefficient is limited to K_pN . Thus, high frequency measurement noise is amplified mostly by the K_pN factor. The obtained transfer function for the PID controller is:

$$C(s) = K_p \left(1 + \frac{1}{sT_i} + \frac{sT_d}{1 + \frac{sT_d}{N}} \right) \quad (12)$$

In some cases, instead of filtering the D term, it is possible to filter the measured signal, which guarantees that the high frequency noise will not produce large control signals (high frequency roll-off).

The adjustment of the PID parameters is done with the frequency response method of Ziegler-Nichols. This method is the second of the two classical methods for determining the parameters of PID controllers presented by Ziegler and Nichols in 1942 and is used to adjust the parameters in a closed loop. These methods are still widely used, in their original form or with some modification. They are often the basis of adjustment procedures used by controller manufacturers and the processing industry. The methods are based on determining some characteristics of the process dynamics. The control parameters are then expressed in terms of features with simple formulas. These methods have a major impact on the practical adjustment of the PID controller even if they do not result in a good setup. Additional extensions of the method are presented in [5]. It is often necessary to supplement the design method with manual adjustment to obtain the desired behavior of the closed loop. The method consists of calculating the critical (limit) values of the parameters K_c and T_u with which the system is on the edge of stability, ie oscillates. K_c represents the critical amplification, while T_u represents the period of one oscillation. K_c is calculated using the Ruth-Hurwitz criterion for closed system stability at $K_i, K_d = 0$ ($T_i = \infty, T_d = 0$) [17].

$$G(s) = \frac{322.7}{s} G_1(s) \quad (13)$$

$$T(s) = \frac{K_p G(s)}{1 + K_p G(s)} \quad (14)$$

The critical value is $K_c < 0.1369$. Substitution of K_c gives $\omega_{cr} = 6,167$ rad/s.

$$T_u = \frac{2\pi}{\omega_{cr}} = 1.0183 \text{ s} \quad (15)$$

Then, by applying the obtained values according to Table 2, $K_p = 0.08214$, $T_i = 0.5092$ and $T_d = 0.1273$ are calculated, while $K_i = K_p/T_i$ and $K_d = K_p * T_d$.

Table 2: PID tuning with Ziegler-Nichols closed-loop method

	K_p	T_i	T_d
P controller	$0.5K_c$	∞	0

PI controller	$0.45K_c$	$\frac{T_u}{1.2}$	0
PID controller	$0.6K_c$	$\frac{T_u}{2}$	$\frac{T_u}{8}$

The response of the system with the calculated PID controller and step excitation is shown in Figure 5. It can be noticed that there is a significant overshoot of 74.2% and a settling time of 4.5 s. For this purpose, additional manual tuning of the parameters of the PID controller is used, so that by increasing K_p the overshoot is reduced and the response is faster, while by increasing the T_d the oscillations and the settling time are reduced. The newly obtained PID controller has values for $K_p = 0.1232$ and $T_d = 0.2546$, so the overshoot is 36.1 % and the settling time is 2.5 s.

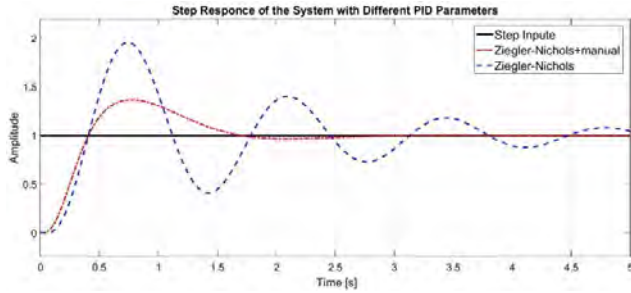


Figure 5: Ziegler-Nichols and additional manual tuning

Due to the implementation of the microcontroller, the PID controller is transformed from a continuous to a discrete form, and then a differential equation is obtained [3]. When converting the PID controller, it is important to select the sampling period which should be at least 10 times the system bandwidth [12]. The bandwidth of the closed system with PID controller is 14.5 rad/s, which means that the sampling period is $T_0 < 0.043$ s. A shorter sampling period will adjust the system response faster based on the changes that have occurred. Therefore 10 ms ($T_0 = 0.01$ s) is used for the sampling period. The conversion from s-domain to z-domain is performed using the method of backward Euler calculation, so that for $s = \frac{z-1}{T_0 z}$ we get a controller of the form:

$$C(z) = K_p + \frac{K_i T_0 z}{z-1} + \frac{K_d(z-1)}{T_0 z} = \frac{3.263z^2 - 6.397z + 3.137}{z^2 - z} \quad (16)$$

From Figure 6 can be seen that the obtained discrete PID controller gives better results than the continuous controller, with an overshoot of 32.7 % and a settling time of 1 s.

Using a shorter sampling period (1 ms) gives better results, but reduces the performance of the microcontroller needed to perform calculations for other processes.

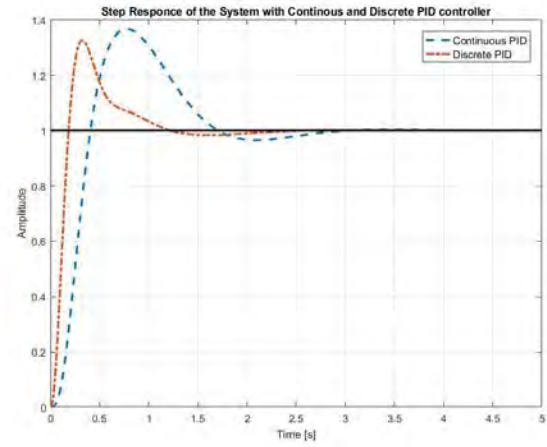


Figure 6: System response with continuous and discrete PID controller

The discrete PID controller can be presented in the form:

$$C(z) = \frac{U(z)}{E(z)} = \frac{3.263 - 6.397z^{-1} + 3.137z^{-2}}{1 - z^{-1}} \quad (17)$$

from which is obtained the differential equation:

$$u(n) = u(n-1) + 3.263 e(n) - 6.397 e(n-1) + 3.137 e(n-2) \quad (18)$$

IV. FUZZY LOGIC CONTROL ALGORITHM

In processes where the dynamics change as a result of nonlinearity and interference, conventional PID controllers can not cope and system oscillations may occur due to precisely adjusted control parameters. The fuzzy-logic controller is a good alternative to the PID controller, as it can handle nonlinear systems and can be designed using human operator knowledge without knowing the mathematical model of the system. Although the fuzzy logic controller usually does not have a better response in the time domain than the PID controller, it can still be applied to systems that have rapid changes, unlike PIDs that will need to adjust the values of the control parameters [6].

Fuzzy logic control is a control algorithm based on linguistic control, which derives from the expert knowledge applied in an automatic system control algorithm [7],[8]. The components of the fuzzy-logic controller are: fuzzification, rule base, inference system and defuzzification.

- Fuzzification - converts all inputs to a membership function so that there is a degree of membership for each linguistic term referring to the input variable.
- Rule base - is a collection of rules that are usually in the format "If-then" and formally the side "If" is called premise and the side "Then" is called a consequent. The computer is able to execute the rules and calculate the control signal depending on the inputs.

- Defuzzification - is the combination and conversion into a single output signal that is not fuzzy but crisp, which is the control signal of the system. The output signal depends on the rules of the system [9].
- Inference system - assesses which control rules should be ignited at a given moment and then decides what the control output signal will be. The most commonly used are Mamdani and Sugeno (Takagi-Sugeno) inference systems.
 - The Mamdani inference system is based on Lotfi Zade's 1973 work on fuzzy algorithms for complex systems and decision processes that expects all output functions to be fuzzy sets. This inference system is intuitive, and widely accepted, better suited to human input, but the main limitation is that the calculation for the defuzzification process takes longer;
 - Sugeno inference system is based on the Takagi-Sugeno-Kang fuzzy inference method, in their joint effort to formalize a systematic approach to generating fuzzy rules from a set of input-output data, which expects all affiliation functions to be singleton. This inference system is computer efficient, works well with linear techniques (PID control, etc.), works well with optimization and adaptation techniques, guarantees output surface continuity, and is more suitable for mathematical analysis. The results are very similar to the consequents from Mamdani's style.

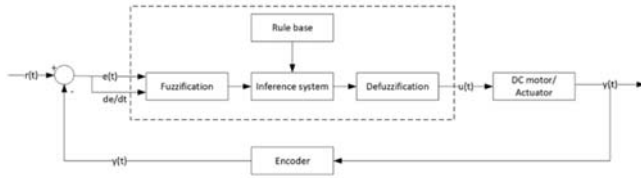


Figure 7: Fuzzy logic controller

In this paper, the Mamdani inference system is used to design the fuzzy logic controller. The fuzzification of the two inputs is performed with 5 triangular and two Γ (trapezoidal) membership functions for each input respectively, which creates a base of 49 rules. The output is formed by five triangular and two Γ membership functions. The rule base is presented in Table 3, where N – negative, P – positive, B – big, S – small.

The membership functions are shown in Figure 8, where the ranges of the linguistic variables are graphically represented. The range [-2000 2000] is used for the error, while for the change of the error [-27 27], taking into account that at a maximum speed for 10 ms a maximum of 27 units can be passed, and the voltage [-23.7 23.7]. The inference system is based on composition so that the fuzzy relations that represent the meaning of each individual rule are merged into one fuzzy-relation that describes the meaning of the whole set of rules.

Table 3: Rule base

Control voltage u(t)		Error e(t)						
		BN	N	SN	Z	SP	P	BP
Change of error Δe	BN	BN	BN	BN	BN	N	SN	Z
	N	BN	BN	BN	N	SN	Z	SP
	SN	BN	BN	N	SN	Z	SP	P
	Z	BN	N	SN	Z	SP	P	BP
	SP	N	SN	Z	SP	P	BP	BP
	P	SN	Z	SP	P	BP	BP	BP
	BP	Z	SP	P	BP	BP	BP	BP

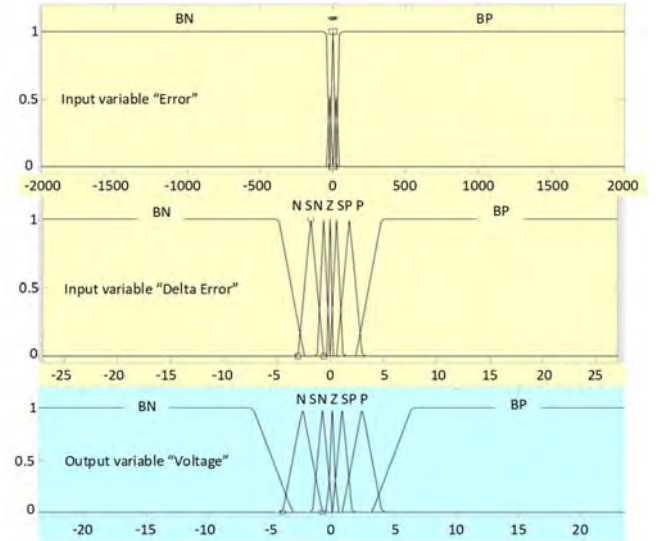


Figure 8: Membership functions

The inference is performed through an operation of the composition of the fuzzy input and the fuzzy relation which represents the meaning of the whole set of rules. The result is a fuzzy set that describes the fuzzy value of the total control output [3],[10]:

$$\mu_u(u) = \max_x \min_{e, \Delta e} (\mu_{ant}(e, \Delta e), \mu_R(e, \Delta e, u)) \quad (19)$$

The centroid method (center of gravity) is used for defuzzification. In continuous case the crisp value of the control signal is obtained with the following relation:

$$u = \frac{\int u * \mu(u) du}{\int \mu(u) du} \quad (20)$$

While in discrete case with the relation:

$$u = \frac{\sum_{i=1}^n u_i * \mu(u_i)}{\sum_{i=1}^n \mu(u_i)} \quad (21)$$

The resulting fuzzy logic controller with Mamdani inference system is used to generate fuzzy logic controller with Sugeno inference system which is more suitable for microcontroller implementation. Generation is performed using MATLAB (mam2sugeno or convertToSugeno) so that the newly obtained Sugeno inference system has constants as output membership functions. These constants are determined by the centroids of the output (consequential) membership functions of the original Mamdani mechanism while the input membership functions and the rules remain the same. The weighted-average method is used for defuzzification, for product - implication and for aggregation - sum. From Figure 9 can be seen that using the Mamdani inference system the system has a overshoot of 9.5%, and oscillates with an amplitude ± 0.05 around the reference value, while the oscillations decrease over time. This value is acceptable considering that the distance between two displacement units is 0.0176 mm, which in case of physical realization can occur oscillations due to the backlash of the reduction itself (gears) of the motor. Using the Sugeno inference system there are no oscillations and the response has a settling time of 1.1 s and a overshoot of 15.9%.

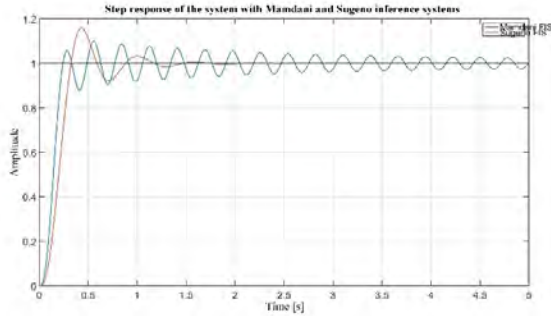


Figure 9: System response with Mamdani and Sugeno inference systems

V. ADAPTIVE NEURAL FUZZY INFERENCE SYSTEM – ANFIS

The Adaptive Neural Fuzzy Inference System (ANFIS) is a combination of two computational methods, neural networks and fuzzy logic. Fuzzy logic has the ability to change the qualitative aspects of human knowledge and insights in the process of precise quantitative analysis. However, there is no defined method that can be used as a guide in the process of transforming human thought into a fuzzy inference system and also takes a long time to adjust to membership functions. Unlike

fuzzy logic, neural networks have great capabilities in the process of adapting to their environment. Therefore, neural networks can be used to automatically adjust the membership functions and reduce the error rate in determining the rules in fuzzy logic [11].

A simple fuzzy inference system has limited learning (or adaptation) opportunities. If learning skills are required, it is convenient to place a fuzzy model within the supervised neural networks that can systematically calculate gradient vectors. Sugeno system is used for consequent and the typical fuzzy rule is:

$$\text{IF } x \text{ is } A \text{ and } y \text{ is } B \text{ THEN } z = f(x, y),$$

where A and B are fuzzy sets in the premise part and $z = f(x, y)$ is a sharp function in the consequent part. Typically, the z function is a first-order (moving single) or zero-order (constant single) fuzzy Sugeno model. An example of modeling a Sugeno first-order inference system is shown in Figure 10, which contains the following two rules:

$$\text{Rule 1: IF } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ THEN } f_1 = p_1x + q_1y + r_1,$$

$$\text{Rule 2: IF } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \text{ THEN } f_2 = p_2x + q_2y + r_2,$$

the functionally equivalent supervised neural network in Figure 10 that follows the general design algorithm has one input layer, three hidden layers and one output layer, the meaning of which is:

Layer 1 - each adaptive node in this layer generates values for the membership of the input vectors A_i , for $i = 1, 2$. For example, the membership function of the i -node can be a generalized bell membership function:

$$O_i^1 = \mu_{A_i} = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}} \quad (22)$$

where O_i^j denotes the output of the i -node in the j -layer, x is the input of node i , A_i are the input vectors connected to the i -node and $\{a_i, b_i, c_i\}$ are the parameter set that changes the form of the membership function. The parameters in this layer are listed as the parameters of the premise part.

Layer 2: Each node in this layer is fixed and calculates the ignition power of a particular product rule. The output of each node represents the ignition power of the rule:

$$O_i^2 = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y), i = 1, 2 \quad (23)$$

In fact, any other T-norm operator that performs fuzzy and operation can be used as a node function in this layer.

Layer 3: The fixed node i in this layer calculates the ratio between the ignition power of the i rule and the sum of the ignition powers of all the rules:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1, 2 \quad (24)$$

For simplicity, the outputs of this layer are also called normalized firing powers.

Layer 4: Adaptive node i in this layer calculates the contribution of i -rule to total output, with the following node function:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (25)$$

Where is the output of layer 3, and $\{p_i, q_i, r_i\}$ are the parametric set. The parameters in this layer are listed as parameters of the consequent part.

Layer 5: The only fixed node in this layer calculates the total output as the sum of the values of each rule:

$$O_i^5 = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i f_i} \quad (26)$$

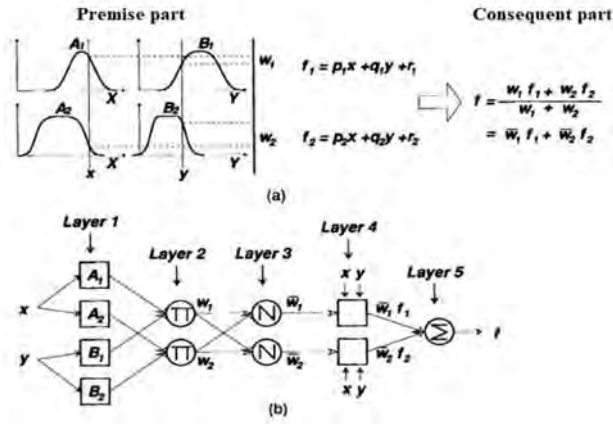


Figure 10: a) Sugeno fuzzy logic inference system b) ANFIS

The basic learning rule is a descending backpropagation gradient, which calculates the error signals (the change in the square error relative to the output of each node) recursively from the output layer back to the input nodes. This learning rule is exactly the same as the reverse propagation learning rule used in feedforward neural networks. The total output f can be expressed as a linear combination of the following parameters:

$$f = \bar{w}_1 f_1 + \bar{w}_2 f_2 = (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 + (\bar{w}_2) r_2 \quad (27)$$

Based on Equation (27), the hybrid learning algorithm combines descending gradient methods and the least squares for optimal parameter search.

The steps used to obtain ANFIS are:

1. Draw the Simulink model with the phase logic controller and simulate it with the given rule base.
2. The first step in designing ANIFS is to collect training and testing data while simulating the fuzzy logic controller.
3. The two inputs, i.e. $e(t)$ and Δe and the output signal $u(t)$ provide the data for training and testing.

4. Use the `anfisedit` command in MATLAB to create the ANFIS .fis file or use the `genfis1` and `anfis` commands.
5. The training data collected in step 2 is loaded and an inference mechanism is generated with selected membership functions (in this case triangular).
6. The collected data are trained with the generated inference system up to a certain number of epochs (iterations) and then tested [7].

In this paper, an ANFIS controller is designed based on the data obtained from the error signals, the error change and the control signal from a modified Sugeno inference system (the error change signal is multiplied by a coefficient of 1.5) from the previous chapter. Data were collected with a sampling time of 0.01 s over a period of 5 s and divided into 80% for training and 20% for testing. 7 triangular membership functions are selected for both inputs, a hybrid learning algorithm and 100 epochs. Trial and error have shown that the best results for ANFIS are obtained by collecting data with excitement greater than about 25% of the single. By using a single step excitation the obtained ANFIS destabilizes the system with small changes in the data (errors).

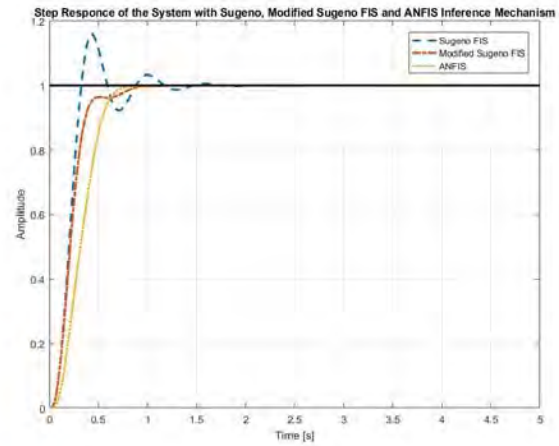


Figure 11: System response with Sugeno, modified Sugeno and ANFIS inference systems

The data obtained from the system response simulations with the applied controllers on the DC motor are numerically presented in Table 4. It can be noticed that the least settling time and the least overshoot has the ANFIS fuzzy logic controller. In terms of speed and rise time, the discrete PID controller has the best results.

Table 4: Results of the applied controllers

	Rise time [s]	Settling time [s]	Overshoot [%]
Continuous PID + ZN	0.2	4.5	74.2
Continuous PID + ZN + manual tuning	0.3	2.5	36.1

Discrete PID	0.1	1.1	32.7
Fuzzy logic with Mamdani	0.1	5	9.5
Fuzzy logic with Sugeno	0.2	1.1	15.9
Fuzzy logic with modified Sugeno	0.3	0.8	0
Fuzzy logic with ANFIS	0.4	0.7	0

VI. CONCLUSION

This paper presented the design of a various control algorithms for the control of actuator as part of a heliostat. Real system was used for identification of the data and obtaining mathematical model. Furthermore, the mathematical model was used for design and simulation of continuous PID controller from which was obtained discrete PID controller. In addition to the PID, fuzzy logic controller was designed both with Mamdani and Sugeno inference systems. Fuzzy logic controller with Sugeno inference system was used for generating data for the ANFIS. Finally, the results from all controllers were summarized and analyzed. It can be concluded that with the application of modified fuzzy logic controller and ANFIS, satisfactory results have been achieved with minimal rise and settling time, without overshoot and steady state error.

REFERENCES

- [1] Téllez, F., Burisch, M., Villasente, Sánchez, M., Sansom, C., Kirby, P., Turner, P., Caliot, C., Ferriere, A., Bonanos, C. A., Papanicolas, C., Montonen, A., Monterreal, R., Fernández, J., *State of the Art in Heliostats and Definition of Specifications: Survey for a low cost heliostat development*, STAGE-STE Project, 2014.
- [2] Wang, Z., *Design of Solar Thermal Power Plants*, Academic Press, 2019.
- [3] Станковски, М., Колемишевска-Гутуловска, Т., *Компјутерско водење на процеси*, Електротехнички факултет, Скопје, 2006.
- [4] Tangirala, A. K., *Principles of system identification – Theory and Practice*, CRC Press, 2015.
- [5] Åström, K. J., Hägglund, T., *Advanced PID Control*, ISA, 2006.
- [6] Salgado-Plasencia, E., Carrillo-Serrano, R. V., Toledano-Ayala, M., *Development of a DSP Microcontroller-Based Fuzzy Logic Controller for Heliostat Orientation Control*, Applied Sciences 10(5):1598, February 2020.
- [7] Khuntia, S. R., Mohanty, K. B., Panda, S., Ardil, C., *A Comparative Study of P-I, I-P, Fuzzy and Neuro-Fuzzy Controllers for Speed Control of DC Motor Drive*, International Journal of Electrical Systems Science and Engineering 1:1 2009.
- [8] Mustafa, G. Y., Ali, A. T., Bashier, E., Elrahman, M. F., *Neuro-Fuzzy Controller Design for a Dc Motor Drive*, UofKEJ Vol. 3 Issue 1 pp. 7-11, February 2013.
- [9] Zeghoudi, A., Chermiti, A., *Speed Control of a DC Motor for the Orientation of a Heliostat in a Solar Tower Power Plant using Artificial Intelligence Systems (FLC and NC)*, Research Journal of Applied Sciences, Engineering and Technology 10(5): 570-580, 2015.
- [10] Grgore, O., Florescu, A., Vasile, A., Stoichescu, D. A., *Part A: Fuzzy Design For DC Motor Speed Control*, International Workshop on Trends and Recent Achievements in Information Technology, Cluj-Napoca, Romania, May 2002.
- [11] Grgore, O., Florescu, A., Vasile, A., Stoichescu, D. A., *Part B: Neuro-Fuzzy Design For DC Motor Speed Control*, UPB Scientific Bulletin, Series C: Electrical Engineering 61(3-4):225-235, May 2002.
- [12] El-sharif, I. A., Hareb, F. O., Zerek, A. R., *Design of discrete-time PID controller*, International Conference on Control, Engineering & Information Technology (CEIT'14), 2014.
- [13] Wu, W., *DC Motor Parameter Identification Using Speed Step Responses*, Modelling and Simulation in Engineering, Article No.: 30, January 2012.
- [14] Ruiz-Rojas, E. D., Vazquez-Gonzalez, J. L., Alejos-Palomares, R., Escudero-Urbe, A. Z., Mendoza-Vázquez, J. R., *Mathematical Model of a Linear Electric Actuator with Prosthesis Applications*, 18th International Conference on Electronics, Communications and Computers, April 2008.
- [15] Mohamed, M. E. A., Guo, Y., *Separately Excited DC Motor Speed Tracking Control Using Adaptive Neuro-Fuzzy Inference System Based on Genetic Algorithm Particle Swarm Optimization and Fuzzy Auto-Tuning PID*, IOP Conference Series Earth and Environmental Science 300:042114, August 2019.
- [16] Mao, W.-L. Suprpto, S., Hung, C.-W., *Adaptive neural network-based synchronized control of dual-axis linear actuators*, Advances in Mechanical Engineering, Vol. 8(7) 1–17, 2016.
- [17] Tymerski, R., *ECE317: Feedback and Control, Lecture: Routh-Hurwitz stability criterion Examples*, Dept. of Electrical and Computer Engineering, Portland State University.
- [18] Pfahl, A., Coventry, J., Röger, M., Wolfertstetter, F., Vázquez-Arango, J. F., Gross, F., Arjomandi, M., Schwarzbözl, P., Geiger, M., Liedke, P., *Progress in heliostat development*, Solar Energy, Volume 152, August 2017, pp. 3-37
- [19] Prinsloo, G. J., Dobson, R. T., *Solar Tracking: High precision solar position algorithms, programs, software and source-code for computing the solar vector, solar coordinates & sun angles in Microprocessor, PLC, Arduino, PIC and PC-based sun tracking devices or dynamic sun following hardware*, Stellenbosch: SolarBooks, 2015
- [20] Salgado-Plasencia, E., Carrillo-Serrano, R. V., Toledano-Ayala, M., *Development of a DSP Microcontroller-Based Fuzzy Logic Controller for Heliostat Orientation Control*, Applied Sciences 10(5):1598, February 2020



ETAI 9: COMMUNICATION TECHNOLOGIES

Uncertain AQM/TCP Computer and Communication Networks: Fixed-time Congestion Tracking Control Using Gaussian Fuzzy-logic Emulator

Jindong Shen, Yuanwei Jing

Northeastern University of Shenyang
School of Information Science &
Engineering, Liaoning, P.R. China
Email: ywjing@mail.neu.edu.cn

Janusz Kacprzyk

Polish Academy of Sciences
Institute of Systems Research
Warsaw, Poland
Email: kacprzyk@ibspan.waw.pl

Georgi M. Dimirovski*

SSs Cyril and Methodius University
School FEIT, Karpos 2, Skopje
R. N. Macedonia
Email: dimir@feit.ukim.edu.mk

Abstract— A novel control synthesis for fixed-time congestion control problem of uncertainty AQM/TCP computer and communication networks with UDP flows and un-modeled uncertainties is revisited by thorough analysis. The fluid-flow based nonlinear dynamics model of AQM/TCP networks is analyzed, and a fixed-time adaptive congestion control algorithm along with fuzzy-logic emulator of unknown UDP flows and un-modeled uncertainties is proposed. The new control algorithm has been derived using combined requirements for the fixed-time control and the prescribed performance control by employing the back-stepping technique and the fuzzy approximation of uncertain quantities. Using the back-stepping methodology, a fixed-time AQM/TCP controller is designed which does not depend on the initial condition so as to ensure that the fixed time of the closed-loop response and system signals remain bounded. The closed-loop simulation results demonstrate the extent of effectiveness and superiority of the proposed design relative to known previous control designs.

Keywords— AQM/TCP communication/computer networks; nonlinear fluid-flow dynamics; back-stepping control synthesis; fixed-time practical stabilization; Gaussian fuzzy-logic emulator; congestion minimization.

I. INTRODUCTION

Network congestion is a common phenomenon of data transmission in computer and communication networks, which can lead to packet loss, increased network delay, and even collapse of network functionality. It is of great significance therefore to study deeply and thoroughly the congestion control mechanism of Transmission Control Protocol (TCP) network in order to prevent congestion phenomenon in TCP networks and to ensure quality of service communications of networks [1].

During the past three decades, more and more scholars have paid attention to the network congestion control problem, in general. Among those research endeavors, the Active Queue Management (AQM) in [1] based on router as the crucial node has become one of the most widely used solutions in the proposed TCP network congestion controls. The first proposed active queue management algorithm was based on RED (Random Early Detection) in [2]. Then some improvements

were proposed for this method in subsequent works [3-4], but the adjustment of the above method parameters was too sensitive to the overall network system. In order to find better solution to the problem of network congestion, Misra and co-authors in [5] have established for the TCP networks a fairly adequate nonlinear dynamic model by using stochastic theory. Then they derived a simplified version of this model in article [6]. Subsequently, Wang et al. [7] have improved the fluid-flow based model and extended it from the single-bottleneck link network to the multi-bottleneck link networks. They successfully verified that the appropriate compound combination of control theory and definition of congestion control was an adequate way to solving congestion control problem in computer and communication networks.

In time, some scholars adopted the feedback control theory to analyze and design new AQM control schemes based on the fluid-flow TCP model. This papers presents a thorough improvement analysis of the 2021 IJCAS article by Jidong Shen and co-authors [9,a1] from the view point of both the underlying functioning physics of AQM/TCP network and computational emulation power of Gaussian fuzzy-logic approximating emulator [10b1, 11b2, 20].

A retrospective insight to closely-related previous works reveals the background as surveyed in the sequel. In article [8], the author studied the robustness of network parameters uncertainty and the controller coefficient perturbation, and proposed a good non-fragile proportional-integral (PI) AQM control algorithm. A proportional-differential (PD) AQM feedback controller was proposed in [9], which could adjust the queue length with smaller oscillation and by means of a faster response. However, whenever the UDP (User Datagram Protocol) flows have occurred or sustained when external disturbances are present, then the properly inadequate or unresponsive robustness of the linear control strategies would appear considerably reduced. Thus, some scholars adopted the nonlinear control approach and designed the AQM control algorithm with considerably efficient affective interference suppression capacity.

In work [13], researchers have studied the issue of adaptive practical finite-time congestion control design for AQM/TCP

traffic flows in the presence of both unknown hysteresis and external disturbance. In [14], taking the advantages of the ‘*minimax*’ control and the *integral back-stepping* control synthesis, an AQM based design approach was proposed to ensure the overall system possess a effective anti-interference performance. In article [15], by considering the non-responsive flow interference such as caused by the UDP flows, the back-stepping control synthesis methodology was adopted to design an AQM controller with prescribed performance. In the real-world [16] communication network environment however, in addition to the interference caused by non-responsive flows in AQM/TCP communication network in operation, the possibly un-modeled uncertainty within AQM/TCP system model also reduces the accuracy of the overall network system performance [16-17]. Thus, authors have also applied control methods to solve the same problem in the presence of uncertainties in AQM/TCP networks.

In [18], the author designed a finite time congestion control algorithm based on sliding mode control, funnel control and using neural network. The AQM/TCP networks with external disturbances and un-modeled uncertainties were studied in article [19] and an integral back-stepping congestion controller was designed under the assumption of existent uncertainties. The assumptions considered however appear conservative to an extent that devaluates the effectiveness of derived results. As shown in article [20], the universal approximation property of fuzzy-logic derived system models can be effective to deal with uncertainty or nonlinearity of the system.

An adaptive fuzzy output feedback control was proposed in [21] to ensure stability and tracking control of a class of uncertain non-linear systems with arbitrary switching signals and un-modeled dynamics. In article [22], an adaptive fuzzy controller with given constraints was designed based on the idea of fuzzy approximation for the constrained control problem of strictly feedback nonlinear systems. This same method also had been applied in the AQM/TCP network system with a rather good result as shown in [23]. However study [23] (hence the previous two works too) does not consider the influence of the un-modeled uncertainty within AQM/TCP networks on the queue change rate of the routing node. This issue too precisely is taken into consideration as well in our current study presented in here. This study adopted fuzzy logic systems to deal with external disturbance and uncertainty in the AQM/TCP network, thus conservatism admitted in previous works [19 - 23] was greatly reduced.

In the actual TCP network systems, the request of QoS (Quality-of-Service) is known it can be improved by making the sender to respond as soon as possible before the routing node queue becomes full. Therefore, it is expected that the control objective can be achieved within the finite time as pointed out in [24]. The theory of finite time control was proposed at the right moment in [25] and developed rapidly ever since. In this direction certain authors have achieved fruitful results in [26, 27]. In [26], an adaptive state feedback controller was designed for the finite-time stabilization problem of a class of high-order uncertain nonlinear systems. Work [27] studied the problem of finite time adaptive neural tracking control for nonlinear systems with non strict feedback form, and the finite time control method also achieved good results in the network congestion control problem in works [28, 29]. However, the convergence time of the finite time

control method depends on the initial state value of the system. Therefore, because of this defect, this proposed method may be a fixed-time control with a deficiency, which gave rise to the definition of fixed-time stability in [30].

The definition of fixed-time stability and practical fixed-time stability does not only guarantee the rapid convergence of the system in closed loop, but also determines the upper bound of the existence of convergence time independent of the initial state of the system. Along this line of research, work [31] proposed a fixed-time controller design for the second-order, multi-agent system in order also to achieve the consistency tracking too in the overall system. In [32], a class of uncertain high-order strict feedback nonlinear system control problem was solved by using fixed time control, and a new fixed time tracking control algorithm based on back-stepping method was proposed. Further in [33], the author studied the fixed-time prescribed performance control problem for nonlinear systems with uncertain dead zones and strict feedback [34, 35]. The fixed time control method was also widely used in other application areas such as spacecraft, surface vessels, robots [36-41].

To the best of our awareness, in so far the approach and methodology combining the adaptive fuzzy control, the back-stepping synthesis methodology, and the fixed-time control theory has not been applied as yet to solve the traffic flow congestion problem in AQM/TCP networks. The current study is focused on this un-solved problem precisely by proposing a new design for system tracking control when both external interference and un-modeled uncertainty are present by employing the fixed-time control theory.

This paper thus brings the following contributions:

(1) According to the model used in [5, 6, 13, 14, 15, 17], this paper considers the impact of external interference and the influence of un-modeled uncertainty on the queue change rate. These innovations imply network operating circumstances are closer to the real-world network.

(2) To the best of our awareness, the fixed-time congestion control problem in a class of AQM/TCP network systems is considered for the first time in the literature. A new congestion control method is proposed, which employs fixed-time and prescribed-performance control by using adaptive back-stepping control synthesis and fuzzy-logic emulator system to solve the congestion tracking control in finite fixed time.

This new method possesses the following advantages.

(i) The upper bound on the stability time of the system does not depend on the initial conditions of the network system, but only on the design parameters.

(ii) By introducing the default performance measure, the tracking error $e_1(t) = x_1(t) - q_{ref}$ of the closed loop system can meet the design requirements.

(iii) The fuzzy-logic derived system is used to deal with the uncertain items and external interference of the AQM/TCP network, which ensures a good compensation effect.

The Section 2 introduces the system model of AQM/TCP network dynamics, certain definitions and as well as necessary lemmas of previous results. The new results are discussed in Section 3. An AQM/TCP network simulation example is given in Section 4 to illustrate the effectiveness of the proposed new methodology for fixed-time congestion control. Section 5 presents the conclusions and outlines some open issues for future research while references follow thereafter.

II. PROBLEM FORMULATION AND PRELIMINARIES

In this section, we present the precise problem formulation for the subsequently derived novel results and proceeding to the application example.

A. On the AQM/TCP Network Nonlinear Dynamics

In this paper, the dynamics of TCP/AQM network system is assumed to be described by the nonlinear model proposed in study [15]. In the existing literature, this particular model is generally considered as the most appropriate one:

$$\begin{cases} \frac{dW(t)}{dt} = \frac{1}{R(t)}(1-p(t)) - \frac{W(t)}{2} \frac{W(t)}{R(t)} p(t), \\ \frac{dq(t)}{dt} = \frac{N(t) \cdot W(t)}{R(t)} - C(t) + \omega(t), \quad q(t) > 0, \\ R(t) = T_p + \frac{q(t)}{C(t)}. \end{cases} \quad (1)$$

The quantities in (1) denote: $W(t)$ is the average congestion window size; $q(t)$ is the router's instantaneous queue length; $R(t)$ is the round trip transmission delay; $C(t)$ is the available link capacity; T_p is the path transmission delay; $N(t)$ is the number of TCP sessions; $p \in [0,1]$ is the probability of a packet being dropped or marked; $\omega(t)$ is the external interference caused by non-response flows such as the UDP flow. According to the study [19], we assume that $N(t)$, $C(t)$ and $R(t)$ are constants for a given network system hence denoted as N , C and R , respectively. The above TCP/AQM dynamic model can be transformed into the flow rate model below via considering that the number N of TCP sessions do share a bottleneck router as in the study [14]:

$$\begin{cases} \dot{r}(t) = \frac{N}{R^2} - \left(\frac{N}{R^2} + \frac{r^2(t)}{2N} \right) p(t), \\ \dot{q}(t) = r(t) - C + \omega(t). \end{cases} \quad (2)$$

In here, $r(t) = \frac{N}{R} W(t)$ and furthermore the system variables are denoted as $x_1(t) = q(t)$, $x_2(t) = r(t) - C$, $u(t) = p(t)$, the queue error $e_1(t) = x_1(t) - q_{ref}$ is defined as in work [10]; here q_{ref} represents the expected queue length. Then it follows:

$$\begin{cases} \dot{x}_1(t) = x_2(t) + \omega(t), \\ \dot{x}_2(t) = f(x) + g(x) \cdot u(t), \end{cases} \quad (3)$$

In order to improve the accuracy of the model, the above TCP network dynamics nonlinear model (3), after variable substitution, it can be transformed

$$\begin{cases} \dot{x}_1(t) = x_2(t) + \omega(t) + \Delta(x_1, x_2), \\ \dot{x}_2(t) = f(x) + g(x) \cdot u(t), \end{cases} \quad (4)$$

where $f(x) = \frac{N}{R^2}$, $g(x) = -\left(\frac{N}{R^2} + \frac{(x_2 + C)^2}{2N} \right)$, $\Delta(x_1, x_2)$

represents the un-modeled uncertainty in the timeout and slow

start phases.

A.1. Preliminary knowledge

Definition 1. Consider a nonlinear dynamic system as follows:

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0, \quad x \in R^n. \quad (5)$$

The origin of the state space of system (5) is semi-global practical, fixed-time stable if, for any initial condition $x(0) \in \Omega$, there is a finite convergence time $T(x_0)$, for all $t \geq T(x_0)$ such that the solution of system (5) satisfies $\|x(t, x_0)\| \leq \varepsilon$ and $\varepsilon > 0$, where the convergence time $T(x_0)$ is bounded, that is where $T(x_0) \leq T_{\max}$.

Lemma 1 [38-39]. If there exist design constants $\mu_1, \mu_2 > 0$, $p > 1$, $0 < q < 1$, $0 < \delta < \infty$, $0 < \phi < 1$, satisfy

$$\dot{V}(x) \leq -\mu_1 V^p(x) - \mu_2 V^q(x) + \delta, \quad (6)$$

where $V(x)$ is the selected Lyapunov function, then the system (5) is practical fixed-time stable, and the stabilization time T_{\max} can be estimated as:

$$T(x_0) \leq T_{\max} := \frac{1}{\mu_1 \phi (p-1)} + \frac{1}{\mu_2 \phi (1-q)}. \quad (7)$$

The solution of system (5) eventually converges to the following set of residuals:

$$x \in \left\{ V(x) \leq \min \left\{ \left(\frac{\eta}{(1-\phi)\mu_1} \right)^{\frac{1}{p}}, \left(\frac{\eta}{(1-\phi)\mu_2} \right)^{\frac{1}{q}} \right\} \right\}. \quad (8)$$

Definition 2 [40]. A smooth function $\rho: \mathcal{R}_+ \rightarrow \mathcal{R}_+$ is called a performance function, if $\rho(t)$ is decreasing and $\lim_{t \rightarrow \infty} \rho(t) = \rho_\infty > 0$.

In this article, we consider the following function as the prescribed performance function which limits the tracking error:

$$\rho(t) = (\rho_0 - \rho_\infty) e^{-at} + \rho_\infty, \quad \forall t > 0, \quad (9)$$

where $0 < \rho_0 < \rho_\infty$, a is a positive number and can be selected according to design requirements. Furthermore, the following inequality can be used to limit the tracking error within the set range:

$$-\delta_{\min} \rho(t) < e_1(t) < \delta_{\max} \rho(t), \quad (10)$$

where δ_{\min} , δ_{\max} are positive constants that can be subject to design in the course of this study.

Lemma 2 [41]. For $\forall \eta_i \in R^+$, $i = 1, \dots, n$ and $0 < b < 1$, the following inequality holds:

$$(\eta_1 + \dots + \eta_n)^b \leq \eta_1^b + \dots + \eta_n^b, \quad (11)$$

Lemma 3 [33]. For $\forall \eta_i \in R^+$, $i = 1, \dots, n$ and $s > 1$, the following inequality holds:

$$\eta_1^s + \dots + \eta_n^s \geq n^{1-s} (\eta_1 + \dots + \eta_n)^s, \quad (12)$$

Lemma 4 [27]. For any real m and n the following inequality holds:

$$|m|^l |n|^l \leq \frac{l_1}{l_1 + l_2} l_3 |m|^{l_1 + l_2} + \frac{l_2}{l_1 + l_2} l_3 \frac{l_1}{l_2} |n|^{l_1 + l_2}, \quad (13)$$

where l_1, l_2 and l_3 are any given positive constants.

Because in the real-world AQM/TCP network system dynamic model (4) does contain un-modeled uncertainties and undergoes sudden external interference, it is difficult to design the controller. To overcome this real-world caused trouble, this paper makes use of the universal approximation features of Gaussian fuzzy-logic systems [20] to approximate (and in simulations to emulate) uncertainties and nonlinear functions in the system representation.

Lemma 5 [11, 20, 22]. For any given continuous function $h(Z)$ defined on the compact set Ω_Z , under the desired accuracy $\varepsilon > 0$, there is a fuzzy logic system $W^T \varphi(Z)$, which can approximate $h(Z)$, which is

$$\sup_{x \in \Omega_Z} |h(Z) - W^T \varphi(Z)| < \varepsilon, \quad (14)$$

where $W = [\omega_1, \omega_2, \dots, \omega_n]^T$ is the adjustable parameter vector, $\varphi(Z)$ is the fuzzy basis function, defined as:

$$\varphi(Z) = \frac{[s_1(Z), s_2(Z), \dots, s_N(Z)]}{\sum_{i=1}^N s_i(Z)}, \quad (15)$$

where N is the number of fuzzy rules, $s_i(Z)$ is the selected Gaussian function:

$$s_i(Z) = \exp \left[\frac{-(Z - \xi_i)^T (Z - \xi_i)}{\eta_i^2} \right], i = 1, 2, \dots, n. \quad (16)$$

where $\xi_i = [\xi_{i1}, \xi_{i2}, \dots, \xi_{in}]^T$ is the center vector and η_i is the width of the Gaussian function.

A.2. Error transformation

In order to implement equation (10), the following equation is introduced to convert the tracking error $e_1(t)$ with inequality performance constraints to the equivalent unconstrained error $\zeta(t)$, making (10) an equation form.

$$e_1(t) = \rho(t)S(\zeta(t)) \quad (17)$$

where $S(\zeta) = (\delta_{\max} e^{\zeta} - \delta_{\min} e^{-\zeta}) / (e^{\zeta} + e^{-\zeta})$ is a smooth and strictly increasing function. It is easy to know that $\partial S(\zeta) / \partial \zeta = 2(\delta_{\max} + \delta_{\min}) / (e^{\zeta} + e^{-\zeta})^2$. And also

$$\begin{aligned} \dot{\zeta}(t) &= \frac{\dot{e}_1(t) - \dot{\rho}(t)S(\zeta)}{\rho(t) \frac{\partial S(\zeta)}{\partial \zeta}} \\ &= \frac{x_2(t) + \omega(t) + \Delta(x_1, x_2) - \dot{\rho}(t)S(\zeta)}{\rho(t) \frac{\partial S(\zeta)}{\partial \zeta}} \quad (18) \\ &= \nu_1(\zeta)(x_2(t) + \omega(t) + \Delta(x_1, x_2)) + \nu_2(\zeta), \end{aligned}$$

where $\nu_1(\zeta) = \frac{1}{\rho(t) \frac{\partial S(\zeta)}{\partial \zeta}}$, $\nu_2(\zeta) = -\nu_1(\zeta) \dot{\rho}(t)S(\zeta)$,

Thus we can get the converted TCP/AQM system model:

$$\begin{cases} \dot{\zeta}(t) = \nu_1(\zeta)(x_2(t) + \omega(t) + \Delta(x_1, x_2)) + \nu_2(\zeta), \\ \dot{x}_2(t) = f(x) + g(x)u(t). \end{cases} \quad (19)$$

Next, we carry out TCP/AQM controller design and its stability analysis on the model described by equation (19).

Objectives of the Control Synthesis: In this paper, a novel adaptive fuzzy fixed-time AQM controller is designed to achieve the actual fixed time stability of nonlinear TCP network dynamic system with external disturbance and uncertainty, and the following problems are solved:

(a) The tracking error $e_1(t) = x_1(t) - q_{ref}$ meets the prescribed performance, and the system output $q(t)$ tracks the expected queue length q_{ref} .

(b) All signals of AQM/TCP closed-loop system remain guaranteed bounded.

III. THE MAIN NEW RESULTS REVISITED

Firstly, an appropriate analysis of controlled AQM/TCP nonlinear dynamics is presented from the viewpoint of systems and control science.

3.1. Design of adaptive fuzzy fixed time controller

In this section, the design process of adaptive fuzzy fixed time controller for TCP/AQM system is proposed.

In order to use back-stepping technique for controller design, the following coordinate transformation is introduced:

$$z_1 = \zeta, \quad (20)$$

$$z_2 = x_2 - \alpha, \quad (21)$$

where α is the virtual control law.

Remark 1. Equations (20) and (21) are similar to those in literature [35], when z_1 is bounded, then $e_1(t)$ is to satisfy the prescribed performance in equation (10). The whole design process is divided into two steps. The first step is to design the virtual control law, and the second step is to design the actual control law.

The specific derivation design process is described below as follows.

Step 1. According to the first subsystem state equations of (20) and (19), there is

$$\dot{z}_1 = \dot{\zeta} = \nu_1(z_2 + \omega + \Delta) + \nu_1 \alpha_1 + \nu_2, \quad (22)$$

Choose the following Lyapunov function:

$$V_1 = \frac{1}{2} z_1^2 + \frac{1}{2\gamma_1} \tilde{\theta}_1^2, \quad (23)$$

where γ_1 is a designable positive constant, $\theta_1 = \|W_1\|^2$, $\tilde{\theta}_1 = \theta_1 - \hat{\theta}_1$, $\hat{\theta}_1$ is the estimated value of θ_1 and $\tilde{\theta}_1$ is the estimated error.

Derivation of V_1 with respect to time t ,

$$\begin{aligned} \dot{V}_1 &= z_1 \dot{z}_1 - \frac{1}{\gamma_1} \tilde{\theta}_1 \dot{\hat{\theta}}_1 \\ &= z_1 (\nu_1(z_2 + \omega + \Delta) + \nu_1 \alpha_1 + \nu_2) - \frac{1}{\gamma_1} \tilde{\theta}_1 \dot{\hat{\theta}}_1 \quad (24) \\ &= z_1 (-k_{2,1} z_1^{2q-1} + \nu_1 \alpha + h_1(Z_1)) - \frac{1}{\gamma_1} \tilde{\theta}_1 \dot{\hat{\theta}}_1 - \frac{1}{2} z_1^2, \end{aligned}$$

where $h_1(Z_1) = \nu_1 \omega + \nu_1 \Delta + \frac{1}{2} z_1 + k_{2,1} z_1^{2q-1}$, $k_{2,1}$ is a positive parameter that can be designed.

Then the fuzzy logic derived system $W_1^T \varphi_1(Z_1)$ is being used to approximate the external disturbances and uncertainties in the TCP/AQM system as follows

$$h_1(Z_1) = W_1^T \phi_1(Z_1) + \delta(Z_1), \quad \|\delta(Z_1)\| \leq \varepsilon_1, \quad (25)$$

Furthermore, by using Young's inequality to scale $z_1 h_1(Z_1)$, the following inequality can be obtained:

$$\begin{aligned} z_1 h_1(Z_1) &= z_1 (W_1^T \phi_1(Z_1) + \delta(Z_1)) \\ &\leq \frac{1}{2a_1^2} z_1^2 \|W_1\|^2 \phi_1^T(Z_1) \phi_1(Z_1) + \frac{a_1^2}{2} + \frac{z_1^2}{2} + \frac{\varepsilon_1^2}{2} \\ &= \frac{1}{2a_1^2} z_1^2 \theta_1 \phi_1^T(Z_1) \phi_1(Z_1) + \frac{a_1^2}{2} + \frac{z_1^2}{2} + \frac{\varepsilon_1^2}{2}, \end{aligned} \quad (26)$$

Designing virtual control law α and adaptive law $\dot{\hat{\theta}}$ proceeds as follows:

$$\alpha = -\frac{1}{v_1} \left(k_{1,1} z_1^{2p-1} + \frac{1}{2a_1^2} z_1 \hat{\theta}_1 \phi_1^T(Z_1) \phi_1(Z_1) \right), \quad (27)$$

$$\dot{\hat{\theta}}_1 = \frac{\gamma_1}{2a_1^2} z_1^2 \phi_1^T(Z_1) \phi_1(Z_1) - 2\sigma_1 \hat{\theta}_1, \quad (28)$$

where $k_{1,1}$, a_1 , σ_1 are positive parameters that can be designed.

Substituting (26)-(28) into (24), we can obtain:

$$\dot{V}_1 \leq -k_{1,1} z_1^{2p} - k_{2,1} z_1^{2q} + v_1 N z_2 + \frac{2\sigma_1}{\gamma_1} \tilde{\theta}_1 \hat{\theta}_1 + \frac{a_1^2}{2} + \frac{\varepsilon_1^2}{2}. \quad (29)$$

Step2. Considering the second subsystem of (21) and (19), the derivative of tracking error z_2 is given:

$$\dot{z}_2 = \dot{x}_2 - \dot{\alpha} = f + gu - \dot{\alpha}, \quad (30)$$

Next, choose the following Lyapunov function:

$$V_2 = V_1 + \frac{1}{2} z_2^2 + \frac{1}{2\gamma_2} \tilde{\theta}_2^2, \quad (31)$$

where γ_2 is a positive parameter that can be designed,

$\theta_2 = \|W_2\|^2$, $\tilde{\theta}_2 = \theta_2 - \hat{\theta}_2$, $\hat{\theta}_2$ is the estimated value of θ_2 and $\tilde{\theta}_2$ is the estimated error.

Calculation of the time derivative of V_2 yields:

$$\begin{aligned} \dot{V}_2 &= \dot{V}_1 + z_2 \dot{z}_2 + \frac{1}{\gamma_2} \tilde{\theta}_2 \dot{\tilde{\theta}}_2 \\ &\leq -k_{1,1} z_1^{2p} - k_{2,1} z_1^{2q} + v_1 N z_2 + z_2 (f + gu - \dot{\alpha}) \\ &\quad - \frac{1}{\gamma_2} \tilde{\theta}_2 \dot{\tilde{\theta}}_2 + \frac{2\sigma_1}{\gamma_1} \tilde{\theta}_1 \hat{\theta}_1 + \frac{a_1^2}{2} + \frac{\varepsilon_1^2}{2} \\ &= -k_{1,1} z_1^{2p} - k_{2,1} z_1^{2q} + z_2 (-k_{2,2} z_2^{2q-1} + gu + h_2(Z_2)) \\ &\quad - \frac{1}{\gamma_2} \tilde{\theta}_2 \dot{\tilde{\theta}}_2 - \frac{1}{2} z_2^2 + \frac{2\sigma_1}{\gamma_1} \tilde{\theta}_1 \hat{\theta}_1 + \frac{a_1^2}{2} + \frac{\varepsilon_1^2}{2}. \end{aligned} \quad (32)$$

In (32), there is $h_2(Z_2) = f - \dot{\alpha} + \frac{1}{2} z_2$, and $k_{2,2}$ is a positive parameter that can be designed.

Fuzzy logic system $W_2^T \phi_2(Z_2)$ can approximate $h_2(Z_2)$ as follows:

$$h_2(Z_2) = W_2^T \phi_2(Z_2) + \delta(Z_2), \quad \|\delta(Z_2)\| \leq \varepsilon_2. \quad (33)$$

Then, by using Young's inequality [11, b2] for scaling, the following inequality can be obtained:

$$\begin{aligned} z_2 h_2(Z_2) &= z_2 (W_2^T \phi_2(Z_2) + \delta(Z_2)) \\ &\leq \frac{1}{2a_2^2} z_2^2 \|W_2\|^2 \phi_2^T(Z_2) \phi_2(Z_2) + \frac{a_2^2}{2} + \frac{z_2^2}{2} + \frac{\varepsilon_2^2}{2} \\ &= \frac{1}{2a_2^2} z_2^2 \theta_2 \phi_2^T(Z_2) \phi_2(Z_2) + \frac{a_2^2}{2} + \frac{z_2^2}{2} + \frac{\varepsilon_2^2}{2}, \end{aligned} \quad (34)$$

Substituting (34) into (32), and sorting out further rearrangements, it follows:

$$\begin{aligned} \dot{V}_2 &\leq -k_{1,1} z_1^{2p} - k_{2,1} z_1^{2q} \\ &\quad + z_2 \left(-k_{2,2} z_2^{2q-1} + gu + \frac{1}{2a_2^2} z_2 \hat{\theta}_2 \phi_2^T(Z_2) \phi_2(Z_2) \right) \\ &\quad + \frac{1}{\gamma_2} \tilde{\theta}_2 \left(\frac{1}{2a_2^2} z_2^2 \phi_2^T(Z_2) \phi_2(Z_2) - \dot{\tilde{\theta}}_2 \right) + \frac{2\sigma_1}{\gamma_1} \tilde{\theta}_1 \hat{\theta}_1 \\ &\quad + \frac{a_1^2}{2} + \frac{\varepsilon_1^2}{2} + \frac{a_2^2}{2} + \frac{\varepsilon_2^2}{2}. \end{aligned} \quad (35)$$

The AQM/TCP congestion control law u and the adaptive law $\dot{\hat{\theta}}_2$ are designed as follows:

$$u = -\frac{1}{g} \left(k_{1,2} z_2^{2p-1} + k_{2,2} z_2^{2q-1} + \frac{1}{2a_2^2} z_2 \hat{\theta}_2 \phi_2^T(Z_2) \phi_2(Z_2) \right). \quad (36)$$

$$\dot{\hat{\theta}}_2 = \frac{\gamma_2}{2a_2^2} z_2^2 \phi_2^T(Z_2) \phi_2(Z_2) - 2\sigma_2 \hat{\theta}_2. \quad (37)$$

where $k_{1,2}$, a_2 , σ_2 are positive parameters that can be designed.

Substituting (36)-(37) into (35) yields the result:

$$\begin{aligned} \dot{V}_2 &\leq -k_{1,1} z_1^{2p} - k_{2,1} z_1^{2q} - k_{1,2} z_2^{2p} - k_{2,2} z_2^{2q} + \frac{2\sigma_2}{\gamma_2} \tilde{\theta}_2 \hat{\theta}_2 \\ &\quad + \frac{2\sigma_1}{\gamma_1} \tilde{\theta}_1 \hat{\theta}_1 + \frac{a_1^2}{2} + \frac{\varepsilon_1^2}{2} + \frac{a_2^2}{2} + \frac{\varepsilon_2^2}{2}. \end{aligned} \quad (38)$$

As shown above, the so far derivation analysis has yielded the design sought for the AQM/TCP adaptive fuzzy congestion controller based on a fixed time. Thus the design derivation is completed.

3.2. Stability analysis

The adaptive fuzzy fixed time congestion controller for the AQM/TCP network system has been designed, however, the stability issue is yet to be explored and guaranteed. In what follows next, the stability analysis of the overall closed-loop system is described in Theorem 1 below. The proof is omitted and the interested reader is suggested to consult the source background article [9].

Theorem 1. For the considered TCP/AQM network system, if the initial condition of tracking error $e_1(t) = x_1(t) - q_{ref}$ satisfies $|e_1(0)| < \rho(0)$, then the designed actual controller (36), virtual controller (27) and adaptive laws (28) and (37) can achieve the control objectives (a) and (b) guaranteeing practical finite-time stability.

Proof. The complete proof is found presented in article [9]. In here we draw the attention focus that according to Young's inequality it may well be established:

$$\frac{2\sigma_1}{\gamma_1} \tilde{\theta}_1 \hat{\theta}_1 \leq -\frac{\sigma_1}{\gamma_1} \tilde{\theta}_1^2 + \frac{\sigma_1}{\gamma_1} \theta_1^2, \quad (39)$$

$$\frac{2\sigma_2}{\gamma_2} \tilde{\theta}_2 \hat{\theta}_2 \leq -\frac{\sigma_2}{\gamma_2} \tilde{\theta}_2^2 + \frac{\sigma_2}{\gamma_2} \theta_2^2. \quad (40)$$

Upon Substituting (39)-(40) into (38), and via using Lemma 2 and Lemma 3 by noting $V=V_2$, the negative definiteness of Laypunov function derivative can be proven. Then according to Lemma 1 it follows all signals z_i and $\tilde{\theta}_i$ ($i=1,2$) are practically fixed-time stable because they all become bounded. In equation (20), ζ is also actually fixed-time stable, and the boundedness of ζ also ensures the prescribed performance of tracking error $e_1(t)$ becomes constant too (these are confirmed simulation example). Because of $z_i = x_i - \alpha_i$, all state variables and control law $u(t)$ of the closed-loop system are also bounded. Thus, the tracking error $e_1(t) = x_1(t) - q_{ref}$ converges to the small region around of the state-space origin, which proves the practical stability of the overall network system.

Remark 2. Theoretically, inspired by article [35], in this manuscript we can increase the parameters $k_{1,1}$, $k_{1,2}$, $k_{2,1}$, $k_{2,2}$ and also decrease the parameter a_i to obtain even better overall system performance. For example, the tracking error can be reduced even further. However, the amplitude of control signal will increase with the increase of $k_{1,1}$, $k_{1,2}$, $k_{2,1}$, $k_{2,2}$. It is therefore that, in practice, it appears necessary to carry the benefit-cost trade-offs out based on considerations of the actual TCP/AQM network case investigated.

IV. APPLICATION EXAMPLE AQM/TCP NETWORK

In this section, the network topology structure of a bottleneck router shared by the number of N TCP flow sessions is shown in Figure 1. The active queue management algorithm, which is proposed in this paper, has been applied and simulation results obtained by using the *Matlab-Simulink* platform [39-42].

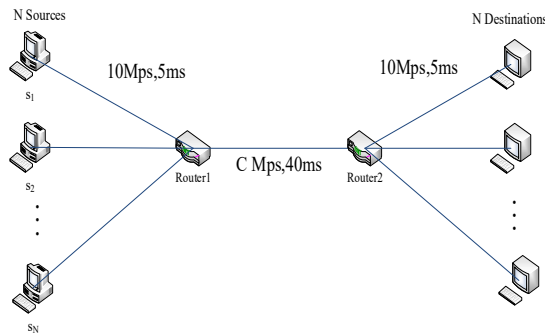


Fig. 1 Topology of the example in simulation study

The corresponding design parameters, system parameters of system (4), external interference and uncertainty are given below.

$$\begin{aligned} N &= 60, \quad C = 1000 \text{ packets/s}, \quad R = 0.1 \text{ s}, \\ q_{ref} &= 100, \quad \rho_0 = 0.5, \quad \rho_\infty = 0.005, \quad a = 4, \\ x(0) &= [100.4 \ 0 \ 0.3 \ 0.3]^T, \quad \omega = 0.2e^{-0.5t}, \\ \Delta &= 0.1 \sin(t)(\sqrt{x_1^2 + x_2^2} - 100). \end{aligned}$$

In order to show the effectiveness of the method in this paper, in the sequel we present and comment the simulation results in Figure 2 through Figure 7. The design parameters are shown in Table 1 further below.

Figure 2 presents the dynamic response change of the queue length. It can be seen that, despite the presence of external interference and uncertainty, the queue length is stable near the expected value. Figure 3 shows the dynamic response trajectory of the AQM/TCP network system tracking error. It can be seen from the Fig. 3 that the tracking error is always kept within the specified range. Figure 4 shows the change trajectory of the packet loss probability of the system which has stabilized at 0.01184 at the $t \approx 0.05$ s. The adaptive laws are shown in Figure 5 and Figure 6, respectively, and it can be seen that the weights of the fuzzy logic system are bounded. Therefore, the fuzzy logic system has a good compensation effect for external disturbance and un-modeled uncertainty.

Table 1. Design parameters

Parameter names	Values
Design parameters of virtual control law (27)	$\gamma_1=1, \sigma_1=0.5, a_1=0.02$
Design parameters of adaptive law (28)	$k_{1,1}=10, a_1=0.02$
Design parameters of actual control law (36)	$\gamma_2=0.05, \sigma_2=5, a_2=0.002$
Design parameters of adaptive law (39)	$k_{1,2}=k_{2,2}=20,$ $p=107/100, q=97/101$

Because the disturbing external interference and un-modeled uncertainties of the system are time-varying, and also there may be existing uncertainties in the system states, which shall have a direct and bad impact on the network system functioning. As it can be seen from the Figure 2 and Figure 4, the routing queue can still tracks well the expected queue, and still has a small packet loss probability.

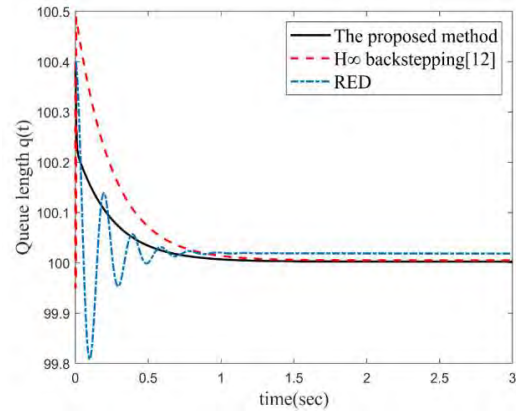


Fig. 2 Queue length dynamic response

In order to show the superiority of the proposed method, we compare it with results in [15] and the control design method based on RED. The comparison results of the response dynamics of both queue lengths are shown in Fig. 2. It can be seen that the here proposed fixed time congestion control method has better steady-state performance and also transient performance when there are external disturbances and uncertainties in the TCP/AQM network system. Because the convergence speed of the queue length in [15] is slower than that of our proposed method, and it does a large overshoot, the queue length response by RED based method has a longer-time fluctuation. Moreover, via the method proposed in this paper, the overall network system response is faster than those of other controllers as approaching closer to expected value. Compared with [15] and RED method, under the same external interference and uncertainty, the queue length in this paper is closer to the expected queue length. In turn, this fact demonstrates the here proposed method guarantees higher accuracy in practical network operation.

In order to show this method indeed has higher accuracy, the relative error ($\frac{x_1 - q_{ref}}{q_{ref}} \times 100\%$) has been calculated and arranged in the next Table 2.

Table 2. Comparison of relative errors at $t = 2.5s$

	The proposed method	H^∞ backstepping [12]	RED
relative error	0.0025%	0.0049%	0.186%

As it can be seen from Table 2, indeed the method proposed in this research study possess smaller relative error. Thus, apparently our proposed method guarantees higher accuracy.

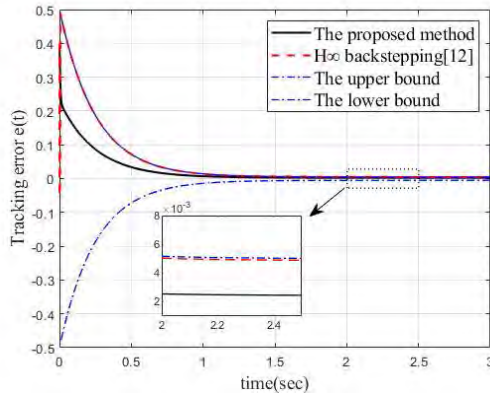


Fig. 3 Response dynamics of the tracking error

From Figure 3 it can be seen further that both the here proposed method and the method in work [15] have the characteristics of limiting the error within the presetting value range. However, the here presented method has considerably faster speed of convergence, and thus also higher convergence accuracy under the same external interference $\omega(t)$ and uncertainty $\Delta(t)$. This is clearly seen when observing any specific time interval such as $2s < t < 2.5s$.

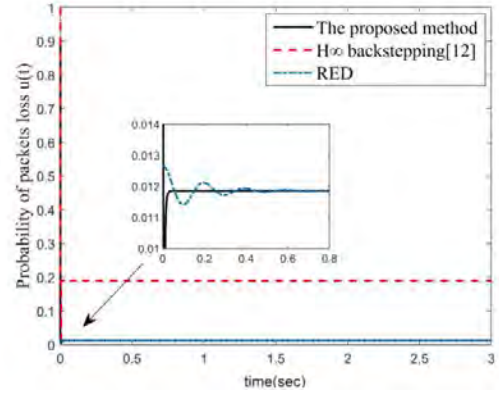


Fig. 4 Probability of packets loss

It can be seen from the simulation results in Figure 4 that, in comparison with those results achieved by means of the method in [15], the probability of packet loss with our method has smaller fluctuation and smaller packet loss rate between zero and one, $0 \sim 1$. Furthermore, the packet loss rate becomes stable when $t \approx 0.05s$, while the packet loss rate of the RED method [12] tends to be stable when $t \approx 0.6s$. Thus, indeed the control design in this paper possesses faster convergence speed in comparison with the RED-based control design.

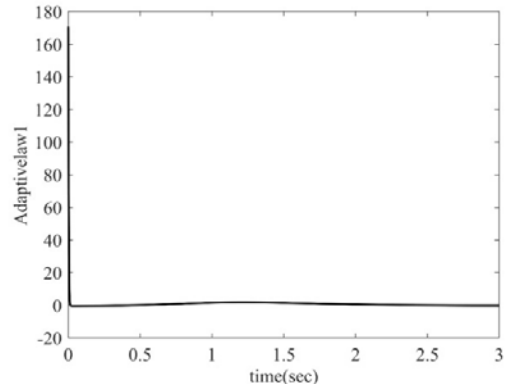


Fig. 5 Adaptive law $\dot{\theta}_1$

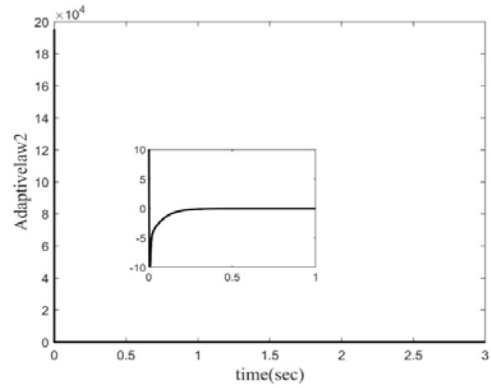


Fig. 6. Adaptive law $\dot{\theta}_2$

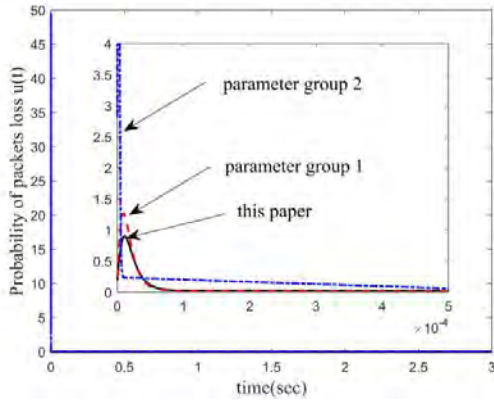


Fig.7. Comparison of the control performance under different controller parameters

In Figure 7, the influence of different controller parameters on the controller performance in terms of packet-loss probability is presented. Apparently, the group 1 of controller parameters $k_{i,j}$ ($i=1,2; j=1,2$) causes considerable higher magnitude than the designed controller of this article. In the case of controller with group 2 parameters $a_1=0.01$, $a_2=0.001$ the overall network systems tends to have enormous magnitude during the first half second. It can be seen from the Fig. 7 that, when compared with the controller of this paper, the increase of the parameters $k_{i,j}$ ($i=1,2; j=1,2$) and the decrease of the parameters a_i will also increase the amplitude of the packet-loss probability. Because the packet loss probability which is still is entailed by the controller in this paper, it is necessary to weigh the design of the controller parameters so as to achieve the ones which ensure avoiding high packet losses.

V. CONCLUDING REMARKS

This paper proposes a new design of adaptive fuzzy fixed-time AQM control algorithm for a class of AQM/TCP network systems under external disturbances and having un-modeled uncertainties. The controller does have certain favorable preset transient performance indicators. The universal approximation property of fuzzy-logic system models is used to deal with the external disturbance and the uncertainty appearing in the network system operation. Apparently it suppresses well the effects of system uncertainty and external disturbances. By making use of the prescribed performance method and technology, the transient and steady state performances of the tracking error have been enforced to meet the design requirements. By using the selected performance function and the error transformation function, the tracking error dynamics with inequality performance constraints is transformed into equivalent unconstrained error dynamics.

The simulation results have shown that the tracking control method, proposed in this paper, also has better adaptability to complex and sudden changes in the AQM/TCP network operation. Through a strict mathematical derivation, the given stability performance analysis of the network system in closed-loop has shown the feasibility of the proposed compound control synthesis design. The obtained simulation

results have further demonstrated the effectiveness and robustness of the designed controller.

Although the study in this paper uses the fixed-time control methodological approach to propose a new congestion control algorithm, still it solved precisely the single bottleneck link congestion control problem. In the real-world network environment complexity, there appear multiple bottleneck links. The next step of this research is solving the congestion control problem in networks with multiple bottleneck links.

ACKNOWLEDGEMENTS

This work has been supported by the National Natural Science Foundation of China [grant number 61773108].

REFERENCES

- [1] B. Braden, and D. Clark, et al. "Recommendations on queue management and congestion avoidance in the Internet." *RFC 2309*, April 1998.
- [2] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, no.4, pp. 397–413, August 1993.
- [3] J. Aweya, M. Ouellette, and D. Y. Montuno, "Enhancing TCP performance with a load-adaptive RED mechanism." *International Journal of Network Management*, vol. 11, no. 1, pp. 31–50, January 2001.
- [4] K. Zhou, K. L. Yeung, and V. O. Li, "Nonlinear RED: a simple yet efficient active queue management scheme," *Computer Networks*, vol. 50, no. 18, pp. 3784–3794, December 2006.
- [5] V. Misra, W. B. Gong, and D. Towsley, "Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED." *Proceedings of the ACM/SIGCOMM'00, Sweden*, pp. 151-160, August 2000.
- [6] C. V. Hollot, V. Misra, D. Towsley, and W. B. Gong, "Analysis and design of controllers for AQM routers supporting TCP flows," *IEEE Trans. on Automatic Control*, vol. 47, no. 6, pp. 945–959, June 2002.
- [7] L. J. Wang, L. Cai, X. Z. Liu, X. M. Shen, and J. S. Zhang, "Stability analysis of multiple-bottleneck networks." *Computer Networks*, vol. 53, no. 3, pp. 338–352, February 2009.
- [8] D. Melchor and S. I. Niculescu, "Computing non-fragile PI controllers for delay models of TCP/AQM networks." *International Journal of Control*, vol. 82, no. 12, pp. 2249–2259, October 2009.
- [9] J. Shen, Y. Jing, G. M. Dimirovski, "Fixed-time congestion tracking control for a class of uncertain TCP/AQM networks." *International Journal of Control, Automation and Systems*, May 2021, Accepted article DOI 10.1007/s12555 (Springer)
- [10] J. Kacprzyk, W. Pedrycz, *Springer Handbook of Computational Intelligence*. Springer Nature, Cham, 2015.
- [11] L. A. Zadeh, J. Kacprzyk, *Fuzzy Logic for the Management of Uncertainty*. John Wiley and Sons, New York, 1992.
- [12] S. K. Bisoy and P. K. Pattnaik, "Design of feedback controller for TCP/AQM networks." *Engineering Science and Technology International Journal*, vol. 20, no. 1, pp. 116–132, 2017.
- [13] W. M. Zheng, Y. X. Li, X. W. Jing, and S. K. Liu, "Adaptive Finite-Time Congestion Control for Uncertain TCP/AQM Network with Unknown Hysteresis." *Complexity*, July 2020.
- [14] Z. H. Li, Y. Liu, and Y. W. Jing, "Active queue management algorithm for TCP networks with integral backstepping and minimax." *International Journal of Control Automation and Systems*, vol. 17, no. 4, pp. 1059-1066, February 2019.
- [15] Y. Liu, X. Liu, and Y. Jing, "Adaptive backstepping H_∞ tracking control with prescribed performance for internet congestion." *ISA Transactions*, vol. 72, pp. 92-99, January 2018.
- [16] L. J. Ma, X. P. Liu, H. Q. Wang, and X. P. Deng, "Congestion tracking control for multi-router TCP/AQM network based on integral backstepping." *Computer Networks*, vol. 175, July 2020.
- [17] P. Wang, H. Chen, X. P. Yang, and Y. Ma, "Design and analysis of a model predictive controller for active queue management." *ISA Transactions*, vol. 51, no. 1, pp. 120-131, January 2012.
- [18] K. Wang, Y. W. Jing, Y. Liu, X. P. Liu, and G. M. Dimirovski, "Adaptive finite-time congestion controller design of TCP/AQM systems based on neural network and funnel control." *Neural Computing and Application*, vol. 32, no. 13, pp. 9471-9478, August 2020.

- [19] Y. Liu, X. P. Liu, Y. W. Jing, and Z. Y. Zhang, "Congestion tracking control for uncertain TCP/AQM network based on integral backstepping." *ISA Transactions*, vol. 89, pp. 131-138, June 2019.
- [20] L. X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least squares learning." *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 807-814, September 1992.
- [21] S. C. Tong and Y. M. Li "Adaptive fuzzy output feedback control for switched nonlinear systems with unmodeled dynamics." *IEEE Transactions on Cybernetics*, vol. 47, no. 2, pp. 295-305, February 2017.
- [22] X. P. Liu, H. Q. Wang, C. Gao, and M. Chen, "Adaptive fuzzy funnel control for a class of strict feedback nonlinear systems." *Neurocomputing*, vol. 241, pp. 71-80, June 2017.
- [23] K. Wang, Y. Liu, X. P. Liu, Y. W. Jing, and S. Y. Zhang, "Adaptive fuzzy funnel congestion control for TCP/AQM network." *ISA Transactions*, vol. 95, pp. 11-17, December 2019.
- [24] Y. X. Su and C. H. Zheng, "Robust finite-time output feedback control of perturbed double integrator." *Automatica*, vol. 60, pp. 86-91, October 2015.
- [25] S. P. Bhat and D. S. Bernstein, "Lyapunov analysis of finite-time differential equations." *Proceedings of 1995 American Control Conference*, vol. 3, pp. 1831-1832, June 1995.
- [26] Z. Y. Sun, L. R. Xue, and K. M. Zhang, "A new approach to finite-time adaptive stabilization of high-order uncertain nonlinear system," *Automatica*, vol. 58, pp. 60-66, August 2015.
- [27] Y. M. Sun, B. Chen, C. Lin, and H. Wang, "Finite-time adaptive control for a class of nonlinear systems with non-strict feedback structure." *IEEE Transactions on Cybernetics*, vol. 48, no. 10, pp. 2774-2782, October 2018.
- [28] Y. Liu, Y. W. Jing, and X. Y. Cheng, "Adaptive neural practically finite-time congestion control for TCP/AQM network. *Neurocomputing*, vol. 351, pp. 26-32, July 2019.
- [29] K. Wang, X. P. Liu, and Y. W. Jing, "Robust finite-time H_∞ congestion control for a class of AQM network systems," *Neural Computing and Applications*, vol. 32, no. 11, 7261-7268, July 2020.
- [30] A. Polyakov, "Nonlinear feedback design for fixed-time stabilization of linear control systems." *IEEE Transactions on Automatic Control*, vol. 57, no. 8, pp. 2106-2110, August 2012.
- [31] Z. Y. Zuo, "Nonsingular fixed-time consensus tracking for second-order multi-agent networks." *Automatica*, vol. 54, pp. 305-309, April 2015.
- [32] J. P. Li, Y. N. Yang, C. C. Hua, and X. P. Guan, "Fixed-time backstepping control design for high-order strict-feedback nonlinear systems via terminal sliding mode." *IET Control Theory Applications*, vol. 11, no. 8, pp. 1184-1193, May 2017.
- [33] J. K. Ni, C. K. Ahn, L. Liu, C. X. Liu, "Prescribed performance fixed-time recurrent neural network control for uncertain nonlinear systems." *Neurocomputing*, vol. 363, pp. 351-365, October 2019.
- [34] Q. Chen, S. Xie, M. Sun, and X. He, "Adaptive Nonsingular Fixed-Time Attitude Stabilization of Uncertain Spacecraft." *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 5, pp. 2937-2950, May 2018.
- [35] P. F. Guo, Z. Y. Liang, X. Wang, and M. W. Zheng "Adaptive trajectory tracking of wheeled mobile robot based on fixed-time convergence with un-calibrated camera parameters." *ISA Transactions*, vol. 99, pp. 1-8, April 2020.
- [36] J. Q. Zhang, S. H. Yu, and Y. Yan, "Fixed-time output feedback trajectory tracking control of marine surface vessels subject to unknown external disturbances and uncertainties." *ISA Transactions*, vol. 93, pp. 145-155, October 2019.
- [37] Q. J. Yao, "Fixed-time trajectory tracking control for unmanned surface vessels in the presence of model uncertainties and external disturbances," *International Journal of Control*, November 2020.
- [38] D. S. Ba, Y. X. Li, and S. C. Tong, "Fixed-time adaptive neural tracking control for a class of uncertain non-strict nonlinear systems." *Neurocomputing*, vol. 363, pp. 273-280, October 2019.
- [39] B. Y. Jiang, Q. L. Hu, and M. I. Friswell, "Fixed-time rendezvous control of spacecraft with a tumbling target under loss of actuator effectiveness." *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 1576-1586, April 2016.
- [40] Y. M. Li, S. C. Tong, L. Liu, and G. Feng, "Adaptive output-feedback control design with prescribed performance for switched nonlinear systems," *Automatica*, vol. 80, pp. 225- 231, June 2017.
- [41] Z. Zhu, Y. Xia, and M. Fu, "Attitude stabilization of rigid spacecraft with finite-time convergence," *International Journal of Robust and Nonlinear Control*, vol. 21, no. 6, pp. 686-702, February 2011.
- [42] MathWorks, *The Matlab, Control Toolbox*. Natick, MA: The MathWorks, Inc. 1992.
- [43] MathWorks, *The Matlab, Fuzzy Toolbox*. Natick, MA: The MathWorks, Inc. 1994.
- [44] MathWorks, *The Matlab, LMI Toolbox*. Natick, MA: The MathWorks, Inc. 1994.
- [45] MathWorks, *The Matlab - Simulink*. Natick, MA: The MathWorks, Inc. 1993.

Wireless Powered ALOHA Networks with Fixed User Rates and UAV-mounted Base Stations

Slavche Pejovski, Zoran Hadzi-Velkov

Faculty of Electrical Engineering and Information Technologies
Ss Cyril and Methodius University in Skopje
Skopje, Macedonia
{slavchep, zoranhv}@feit.ukim.edu.mk

Abstract— In this paper, we propose a resource allocation scheme for a wireless powered communication network (WPCN) whose base station (BS) is mounted on an unmanned aerial vehicle (UAV). The UAV flies along a circular trajectory and maintains a line of sight (LoS) link between the BS and each energy harvesting user (EHU). The EHUs utilize a common random access channel by employing the slotted ALOHA protocol. The BS broadcast RF energy to the EHUs. The EHUs transmit information codewords to the BS at fixed rates, and so the BS cannot successfully decode these codewords if the path loss of the BS-EHU channel is too high. For such WPCN, we determine the optimal fixed rates of the EHUs and the corresponding optimal channel access probabilities that guarantee proportionally fair resource allocation. Also, using numerical results, we determine the conditions where nonzero outage probability is optimal.

Keywords— *Wireless powered communication networks; slotted ALOHA; unmanned aerial vehicles; fixed rate transmission; outage probability*

I. INTRODUCTION

The RF power transfer technology have become very important due to its ability to offer almost infinite and independent energy supply to the Internet of Things (IoT) devices [1]. The wireless powered communication networks (WPCN) contain one or more base stations (BS) that transmit the energy, and multiple energy harvesting (EH) users (EHUs). Limited by their low cost and low complexity, the EHUs are usually incapable of performing channel estimation or support the signaling required for mutual coordination [2]. Therefore, the random access may be the preferred choice for medium access in WPCNs. Thus, the slotted ALOHA [3] have been relaunched as an inevitable technology for many existing and emerging wireless networks, such as RFID systems, LTE networks, and massive machine-type communications. Recently, the authors in [4] proposed the use of slotted ALOHA in WPCNs for *proportionally fair* (PF) [5] resource allocation.

One of the major limitations of WPCNs is the short distance between a fixed ground-based BS and the EHUs over which sufficient RF power can be transferred. This limitation is due to the *doubly near far effect* [1]. To alleviate it, as a cost-effective solution that does not require infrastructure

coverage, the mounting of a BS on an unmanned aerial vehicle (UAV) have emerged [6]. This way the BS can maintain line of sight (LoS) transmission of energy and information to/from the EHUs, allowing EHU deployment across larger service areas [7]-[9]. Recently, a combination of slotted ALOHA random access with the LoS transmission provided by the UAV mounted BS, was proposed in [10]. The authors in [10] use PF based resource allocation among the EHUs. Specifically, they assume circular UAV trajectory such that the BS periodically approaches each of the EHUs. The authors analyze two scenarios: a scenario where each EHU internally keeps track of the corresponding channel gains without the use of conventional channel estimation techniques, and a second scenario where the EHUs do not exploit the periodicity of the channel gains, and employ fixed-power fixed-rate transmission over the random access channel. In the second scenario, the authors find the optimal solution using an exhaustive search method over two variables: the radius of the circular trajectory and an auxiliary variable associated with the EHUs transmission rate. In this paper we analyze the second scenario and extend [10] by showing an analytical solution for the variables in the problem (that does not include exhaustive search) and thus find the optimal outage probability leading to the optimal transmission rate in such system. Additionally, we numerically evaluate the importance of the existence of nonzero outage probability in such system. To be able to get better insight about the importance of the nonzero outage probability in the system, differently from [10], in this paper we do not consider optimization over the radius of the UAV circular trajectory (thus, the proposed analytical solution does not apply to the radius of the circular trajectory).

The paper is organized as follows: in Section II we present the system model, in Section III we present the resource allocation problem and its solution. In Section IV we present the numerical results and we give the conclusion in Section V.

II. SYSTEM MODEL

We consider a scenario where the UAV transmits energy in downlink and receives information in the uplink from K EHUs. The energy and the information transmissions are carried out at different frequencies. Additionally, the information transmission is carried out using slotted ALOHA

random access. We assume that the UAV flies at constant speed over a circular trajectory with radius r_0 and height H . The EHU k is placed at distance r_{0k} from the center of the UAV trajectory projected on the ground. Each EHU and the UAV is equipped with a single antenna.

A. Channel model

We assume LoS channel between the UAV and all the EHUs at any time. Thus the channel gain between the UAV and the EHU at any time is given as:

$$x_k = B_0 d_k^{-2}, \quad (1)$$

where B_0 is the channel gain at distance of 1m, and:

$$d_k^2 = r_{0k}^2 + r_0^2 + H^2 - 2r_{0k}r_0 \cos \alpha_k, \quad (2)$$

where α_k is the angle between the line connecting the center of the UAV trajectory projected on the ground and the position of the EHU, and the line between the projections of the center of the UAV trajectory and the current UAV position. The communication system from a bird perspective is illustrated in Fig. 1.

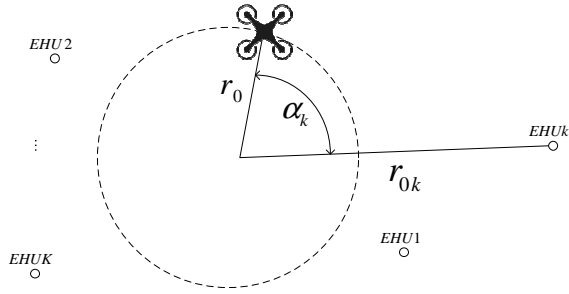


Fig. 1. The analyzed system from bird perspective with illustrated parameters for EHU k

B. EHU transmission rate

We assume that the EHUs don't know the current position of the UAV but know their own coordinates, the coordinates of the center of the UAV trajectory and H . This can be justified with the high speed of the UAV and the low processing capabilities of the EHUs. Thus, the EHU have only statistical knowledge of the channel. Additionally, we assume that the EHU has no knowledge of the position of the other EHUs, and knows only the number of EHUs K .

We assume that each EHU transmits with fixed power P_k

at a fixed rate $R_k = \log_2 \left(1 + \frac{P_k}{N_0} \frac{B_0}{d_{km}^2} \right)$ with access

probability q_k in all random access slots. The constant d_{km}

is the largest distance between the EHU k and the UAV at which the EHU k transmission is received correctly under the assumption that the EHU k is the only EHU that transmits in that slot. Based on the systems geometry we obtain

$d_{km} = \sqrt{r_{0k}^2 + r_0^2 + H^2 - 2r_{0k}r_0 \cos \alpha_{km}}$, where α_{km} is the angle for which we obtain d_{km} . Notice that for $d_k > d_{km}$

which happens for $\alpha_k > \alpha_{km}$ the transmission fails since the communication rate, at which the user transmits, is higher than the capacity of the channel, and thus, the communication system is in outage. For $\alpha_k < \alpha_{km}$ the transmission is successful. Since the UAV travels the circular trajectory at a constant speed, using the geometry of the system, the time duration when this transmission is in outage is equal to $\frac{2(1-\alpha_{km})}{2\pi}$. Thus, the probability of successful transmission

of the k -th EHU, when no other EHU transmits is given as

$$1 - P_{out_k} = \frac{\alpha_{km}}{\pi}. \quad (3)$$

C. EHU Average Harvested Power

Each EHU ($1 \leq k \leq K$) has a rechargeable battery with infinite storage capacity. It harvests RF energy from the BS and uses the harvested energy for information transmission. The average power collected by the EHU k from the UAV can be calculated as [10]:

$$P_{rk} = \frac{\eta_k B_0 P_0}{\sqrt{(r_{0k}^2 + H^2)^2 - 2(r_{0k}^2 - H^2)r_0^2 + r_0^4}}, \quad (4)$$

where η_k is the energy conversion efficiency of the k -th EHU and P_0 is the transmit power of the BS.

III. RESOURCE ALLOCATION

A. Problem formulation

We assume that there are many random access opportunities (slots) during one flight over the UAV circle trajectory and that the UAV will repeat its trajectory many time. In each access opportunity the EHU k will access the channel with access probability q_k . Thus the average rate per slot for user k is:

$$\bar{R}_k = R_k (1 - P_{out_k}) q_k \prod_{j \neq k} (1 - q_j) \quad (5)$$

Our goal is to maximize the proportional fairness of the average user's rates, and thus, we define the following optimization problem:

$$\begin{aligned}
& \max_{q_k, \alpha_{km}, P_k, R_k, P_{out_k}} \sum_{k=1}^K \log(R_k (1 - P_{out_k}) q_k \prod_{j \neq k} (1 - q_j)) \\
& s.t. \\
& c1: P_k q_k \leq P_{rk} \quad \forall k \\
& c2: R_k = \log \left(1 + \frac{B_0}{N_0} \frac{P_k}{d_{km}^2} \right) \quad \forall k \\
& c3: 1 - P_{out_k} = \frac{\alpha_{km}}{\pi} \quad \forall k \\
& c4: d_{km}^2 = r_{0k}^2 + r_0^2 + H^2 - 2r_{0k}r_0 \cos \alpha_{km} \quad \forall k
\end{aligned} \tag{6}$$

where c1 is due to the energy conservation law (on the left hand side of c1 is the average transmit power spend by the k-th EHU and the right hand side of c1 is the average harvested energy by the k-th EHU), c2 is associated with the EHU fixed transmission rate, c3 is due to (3) and c4 is an auxiliary constraint to associate α_{km} and d_{km} .

Observing the constraints c1 and c2, it is obvious that no matter the choice of q_k increased value of P_k leads to an increased value of R_k which increases the objective function. Thus the constraint c1 can be transformed to equality. By using all the equalities in the constraint to replace R_k , P_k and P_{out_k} in the objective, problem (6) transforms to:

$$\begin{aligned}
& \max_{q_k, \alpha_{km}} \sum_{k=1}^K \log \left(\log \left(1 + \frac{B_0 P_{rk} / (q_k N_0)}{r_{0k}^2 + r_0^2 + H^2 - 2r_{0k}r_0 \cos \alpha_{km}} \right) \right. \\
& \quad \left. \times \frac{\alpha_{km}}{\pi} q_k \prod_{j \neq k} (1 - q_j) \right)
\end{aligned} \tag{7}$$

Using the properties of the logarithm function we obtain:

$$\begin{aligned}
P_g : \max_{q_k, \alpha_{km}} \sum_{k=1}^K \log \log \left(1 + \frac{B_0 P_{rk} / (q_k N_0)}{r_{0k}^2 + r_0^2 + H^2 - 2r_{0k}r_0 \cos \alpha_{km}} \right) \\
+ \log \alpha_{km} + \log(q_k) + (K-1) \log(1 - q_k)
\end{aligned} \tag{8}$$

which can be solved independently for each user.

B. Solution of the optimization problem

Lets define $a_k = \frac{P_{rk} B_0}{N_0 ((r_{0k} + r_0)^2 + H^2)}$ and q_k^* as a solution to the following transcendent equation:

$$\frac{1 - K q_k^*}{1 - q_k^*} = \frac{1}{\log(1 + a_k / q_k^*)} \frac{1}{1 + a_k / q_k^*} \frac{a_k}{q_k^*} \tag{9}$$

Lets define $C_k = \frac{r_{0k}^2 + r_0^2 + H^2}{2r_{0k}r_0}$. Assuming that $C_k \leq \frac{\pi}{2}$

and setting $\alpha_{k \min}$ and $\alpha_{k \max}$ to be the solutions of the equation $\alpha_{km} \sin \alpha_{km} + \cos \alpha_{km} = C_k$ such that

$\alpha_{k \min} \in \left[0, \frac{\pi}{2} \right]$ and $\alpha_{k \max} \in \left[\frac{\pi}{2}, \pi \right]$, we define α_{km}^{**} as

the solution with lower value in the interval $[\alpha_{k \min}, \alpha_{k \max}]$ of the following equation (if solution exists):

$$\begin{aligned}
1 = \frac{1}{V(\alpha_{km}^{**})} \frac{1}{\log \left(1 + \frac{P_{rk} B_0 / N_0}{\frac{1 - V(\alpha_{km}^{**})}{K - V(\alpha_{km}^{**})} d_{km}^{2**}} \right)} \\
\frac{1}{1 + \frac{P_{rk} B_0 / N_0}{\frac{1 - V(\alpha_{km}^{**})}{K - V(\alpha_{km}^{**})} d_{km}^{2**}}} \frac{P_{rk} B_0 / N_0}{\frac{1 - V(\alpha_{km}^{**})}{K - V(\alpha_{km}^{**})} d_{km}^{2**}}
\end{aligned} \tag{10}$$

where, $V(\alpha_{km}^{**}) = \frac{d_{km}^{2**} / (2r_{0k}r_0)}{\alpha_{km}^{**} \sin \alpha_{km}^{**}}$ and d_{km}^{2**} is obtain from c4

in (6) for α_{km}^{**} . Then, we define $q_k^{**} = \frac{1 - V(\alpha_{km}^{**})}{K - V(\alpha_{km}^{**})}$.

Additionally, we define the difference in the objective function for the two possible solutions of q_k :

$$S_k = \log \frac{\log \left(1 + \frac{P_{rk} B_0 / N_0}{q_k^* d_{km}^{2**}} \right)}{\log \left(1 + \frac{a_k}{q_k^*} \right)} + \log \frac{q_k^{**}}{q_k^*} + (K-1) \log \frac{1 - q_k^{**}}{1 - q_k^*}$$

The optimal access probability of the k-th EHU is:

$$q_{ko} = \begin{cases} q_k^{**}, & \text{if } \alpha_{km}^{**} \text{ exist, and } S_k > 0 \\ q_k^*, & \text{otherwise} \end{cases} \tag{11}$$

the optimal value of α_{km} is:

$$\alpha_{kmo} = \begin{cases} \alpha_{km}^{**}, & \text{if } \alpha_{km}^{**} \text{ exist, and } S_k > 0 \\ \pi, & \text{otherwise} \end{cases} \tag{12}$$

and the optimal transmission rate of the k-th EHU is

$$R_{ko} = \log_2 \left(1 + \frac{\eta_k B_0^2 P_0 / (r_{0k}^2 + r_0^2 + H^2 - 2r_{0k}r_0 \cos \alpha_{kmo})}{q_{ko} N_0 \sqrt{(r_{0k}^2 + H^2)^2 - 2(r_{0k}^2 - H^2)r_0^2 + r_0^4}} \right) \tag{13}$$

Remark: The existence of a solution for (10) depends on the system parameters: C_k , K and $P_{rk} B_0 / N_0$, and is hard to be rigorously characterized. Nonetheless, the maximum of the

function $f_{kg}(\alpha_{km})$ (which is the right hand side of (10)) is found in the vicinity of $\alpha_{km} = \pi/2$. Thus a simple test to find if α_{km}^{**} exist, is to check if $f_{kg}(\pi/2) > 1 - \Delta$. We have found that a good value for Δ is $\Delta = 0.02$. Please note that $f_{kg}(\pi/2)$ depends only on C_k , K and $P_{rk} B_0 / N_0$, and thus, it is a system parameter that can be easily calculated.

IV. NUMERICAL RESULTS

In this section, we assume a WPCN with K EHUs and a BS mounted on the UAV. We assume that all EHUs are placed on a circle with radius r_{0k} which is centered at the same position as the trajectory of the UAV. We set $B_0 = 10^{-3}$, $N_0 = 10^{-12}W$ and $P_0 = 0.5W$ if not stated differently. As a benchmark we use the system that allows no outage i.e. it uses only the solution to (9) but is not aware of the solution q_k^{**} . This benchmark is termed “Solution with no outage”. On the other hand the solution represented by (11) and (12) (which is the optimal solution of problem P_g) is termed “Solution that includes nonzero outage”. As a performance metric we use the network sum throughput calculated as $\sum_{k=1}^K \bar{R}_k$, where \bar{R}_k is defined in (5). Also in order to obtain an insight in the existence of the solution to (10), when necessary, we show additional figures for the “Solution that includes nonzero outage” that show if the solution to (10) exist and which of the solutions q_{km}^* or q_{km}^{**} is the optimal one.

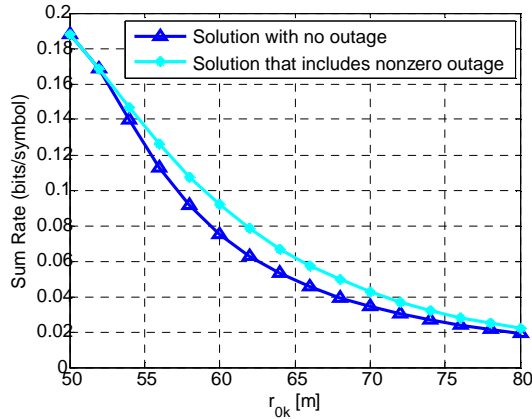


Fig. 1. Sum rate for different r_{0k} for $r_0 = 50m$, $K = 2$ and $H = 5m$

Fig. 1 (a) shows the sum rate as a function of r_{0k} . It can be observed that for values of r_{0k} sufficiently higher than r_0 , the existence of nonzero outage probability leads to increased sum rate. Nonetheless it should be observed that the values of

the sum rates, for which the nonzero outage probability brings performance improvement, are low. This is expected since the nonzero outage probability is advantageous for systems with low harvested energy and thus low SNR.

Fig. 2 shows the sum rate for different number of EHUs. As the number of EHUs increase the solution that uses zero outage probability becomes dominant. The numerical results show that q_{km}^{**} is always larger than q_{km}^* and with the increase of K the range of values for q_{km}^{**} become narrower, and, thus, its influence diminishes.

In Fig. 3a the sum rate for different values of r_0 is shown.

For values of r_0 closer to the center of the circular trajectory, the harvested energy is higher, and, thus, the sum rate is higher. It is obvious that in such scenarios the solutions without outage is more appropriate. For high values of r_0 , the harvested energy becomes low, and thus, the nonzero outage probability brings performance improvement. Additionally, Fig. 3b shows that in this region (with relatively high harvested energy) the solution of (10) does not exist. As r_0 increases, the conditions in the system are such that q_k^{**} exist but $S_k < 0$. At the end of the considered range for r_0 , q_k^{**} exist and $S_k > 0$, thus making the outage aware solution dominant.

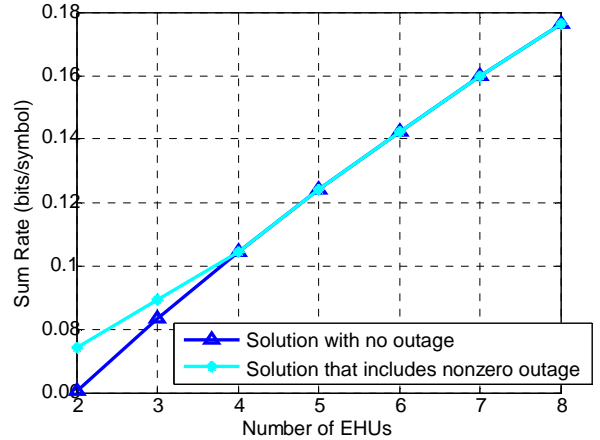
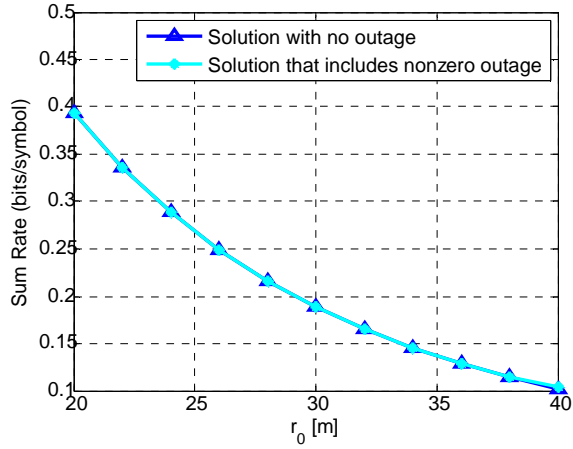
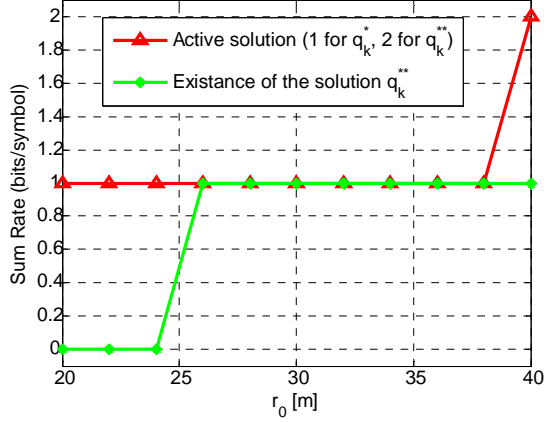


Fig. 2. Sum rate for different K for $r_0 = 50m$, $r_{0k} = 60$ and $H = 10m$

In Fig. 4a the sum rate for different values of P_0 is shown. For lower values of P_0 , the harvested energy is low and the “Solution that includes nonzero outage” shows better performance. On the contrary as P_0 increases, the harvested energy becomes higher which leads to higher sum rate and domination of the “Solution with no outage”. Additionally, Fig. 4b shows that in this region (with relatively high harvested energy) the solution of (10) does not exist.



(a) Sum rate

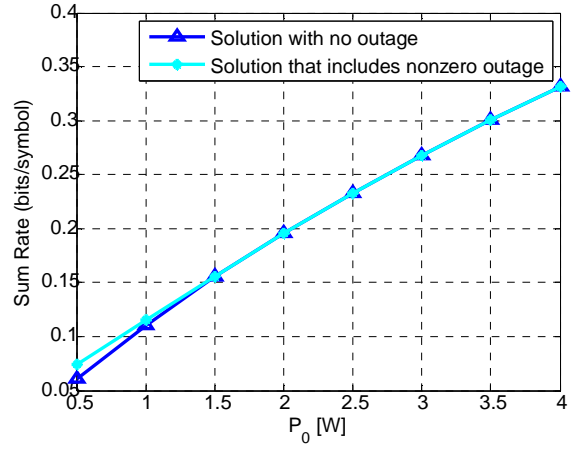


(b) Solutions distribution

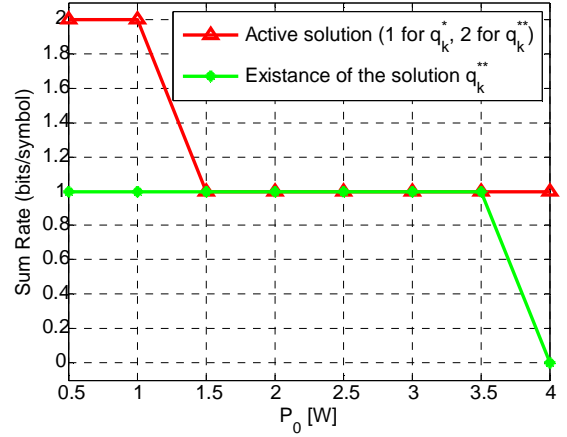
Fig. 3. Sum rate and solutions distribution for different values of r_0 for $r_{0k} = r_0 + 10m$, $K = 2$ and $H = 10m$

V. CONCLUSION

In this paper we presented a communication system where the users (EHUs) harvest energy from the base station which is mounted on a UAV. The UAV has circular trajectory and the EHUs randomly access the UAV for information transmission. The EHUs have only statistical channel state information, and thus, outage can occur. The outage probability is found to be nonzero in regions with low harvested energy and its impact on the performance in that region is significant. Thus, low values of the UAV transmit power, high UAV circular trajectory radius, high values for distance of the EHUs from the center of the UAV trajectory and low number of EHUs are factors that promote the nonzero outage probability. In regions where the harvested energy is high, the zero outage probability is optimal.



(a) Sum rate



(b) Solutions distribution

Fig. 4. Sum rate and solutions distribution for different values of P_0 for $r_0 = 50m$, $r_{0k} = 60m$, $K = 2$ and $H = 10m$

VI. APPENDIX

Here we give a sketch of the solution of problem P_g . Observing problem P_g for each user, two optimization variables are present: $q_k \in [0, 1]$ and $\alpha_{km} \in [0, \pi]$. The optimal values for the optimization variables can have values in the allowed intervals or at the boundaries of these intervals.

A. Solution at the boundaries

The solutions using $q_k = 0$, $q_k = 1$ and $\alpha_{km} = 0$ are impossible since they result in infinite negative value of the objective function. The boundary solution for $\alpha_{km} = \pi$ can be obtained by replacing $\alpha_{km} = \pi$ in P_g , finding the first

derivative of the objective function in P_g over q_k and setting it to zero. After several manipulations we arrive at equation (9), which has a single solution for $q_k \in [0, 1]$. Namely, since the right hand side of (9) is always positive, by observing the left hand side of (9) we can additionally restrict q_k to the interval $q_k \in [0, 1/K]$. By increasing q_k in its interval, the right hand side is monotonically increasing function from 0 to 1 for all $a_k > 0$, and the left hand side monotonically decreases from 1 to 0. Thus the two sides intercept at exactly one point.

B. Solutions in the interior of the allowed intervals

By taking the first derivative of the objective function in P_g with respect to q_k and α_{km} , and setting them to zero we obtain the following system of two equations with two unknowns:

$$\frac{1 - Kq_k}{1 - q_k} = \frac{1}{\log \left(1 + \frac{P_{rk} B_0 / N_0}{q_k d_{km}^2} \right)} \times \frac{1}{1 + \frac{P_{rk} B_0 / N_0}{q_k d_{km}^2}(\alpha_{km})} \quad (14)$$

$$V(\alpha_{km}) = \frac{1}{\log \left(1 + \frac{P_{rk} B_0 / N_0}{q_k d_{km}^2} \right)} \times \frac{1}{1 + \frac{P_{rk} B_0 / N_0}{q_k d_{km}^2}} \quad (15)$$

where d_{km}^2 is obtained using constraint c4 of (6) and $V(\alpha_{km}) = \frac{d_{km}^2 / (2r_{0k}r_0)}{\alpha_{km} \sin \alpha_{km}} = \frac{C_k - \cos \alpha_{km}}{\alpha_{km} \sin \alpha_{km}}$. Since the right hand side of (15) is positive number for the considered intervals of q_k and α_{km} we have $V(\alpha_{km}) \geq 0$. From the equality of the right hand sides of (14) and (15) we obtain:

$$\frac{1 - Kq_k}{1 - q_k} = V(\alpha_{km}) \quad (16)$$

Using that $V(\alpha_{km}) \geq 0$ from (16) we obtain that $q_k \in [0, 1/K]$. By rearranging (16) we obtain:

$$q_k = \frac{1 - V(\alpha_{km})}{K - V(\alpha_{km})} \quad (17)$$

By substituting (17) into (15) and using a simple transformation we arrive at equation (10).

The rest of the proof is aimed to show that the solution of (10) is unique. This part is omitted due to space restrictions. Nonetheless the proof first shows that in the $\alpha_{km} \in [\alpha_{\min}, \alpha_{\max}]$ there are at most 2 solutions for (10). Then, at the end of the proof there is an additional step to show that the first solution (solution with lower value) is local maximum and the second solution is local minimum.

REFERENCES

- [1] H. Ju and R. Zhang, "Throughput maximization in wireless powered communication networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 418-428, Jan. 2014
- [2] M. Poposka, Z. Hadzi-Velkov and S. Pejoski, "Fairness Optimization of Fixed-Rate Wireless Networks With RF Energy Harvesting Transmitters," in *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2859-2863, Dec. 2020
- [3] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. 25, pp. 117-128, Jan. 1977
- [4] Z. Hadzi-Velkov, S. Pejoski, N. Zlatanov, and R. Schober, "Proportional fairness in ALOHA networks with RF energy harvesting," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 277-280, Feb. 2019
- [5] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Info. Theory*, vol. 48, no. 6, Jun. 2002
- [6] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 3642, May 2016
- [7] J. Xu, Y. Zeng, and R. Zhang, "UAV-enabled wireless power transfer: trajectory design and energy optimization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5092-5106, Aug. 2018
- [8] H. Dai et. al., "How to deploy multiple UAVs for providing communication service in an unknown region?," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1276-1279, Aug. 2019
- [9] H. Dai, H. Zhang, B. Wang, and L. Yang, "The multi-objective deployment optimization of UAV-mounted cache-enabled base stations," *Physical Commun.*, vol. 34, pp. 114-120, Jun. 2019
- [10] Z. Hadzi-Velkov, S. Pejoski, R. Schober and N. Zlatanov, "Wireless Powered ALOHA Networks With UAV-Mounted-Base Stations," in *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 56-60, Jan. 2020

Performance Investigation of Bidirectional Optical IM/DD OFDM WDM-PON using RSOA as a Colorless Transmitter

Mahmoud Alhalabi

Institute of Natural and Applied Science,
Erciyes University,
Kayseri, Turkey
eng.halabi@hotmail.com

Necmi Taşpınar

Department of Electrical and Electronics
Engineering, Erciyes University,
Kayseri, Turkey
taspinar@erciyes.edu.tr

Fady El-Nahal

Electrical Engineering Department
Islamic University of Gaza
Gaza, Palestine
fnahal@iugaza.edu.ps

Abstract— In this paper, we have designed and simulated bidirectional hybrid Intensity Modulated and Direct Detected Optical Orthogonal Frequency Division Multiplexing Wavelength Division Multiplexing Passive Optical Network (IM/DD-OFDM-WDM-PON) with 40 Gbps 16-QAM downstream and 2.5 Gbps On-Off keying (OOK) upstream signal by using Optisystem software. The simulated system is considered as a simple, low cost and colorless network as a Reflective Semiconductor Optical Amplifier (RSOA) transmitter was used at the Optical Network Unit (ONU) and no Dispersion Compensating Fiber (DCF) is needed. Obtained results show that the simulated IM/DD-OFDM-WDM-PON can achieve good Bit Error Rate (BER) performance over propagation length of 20 km. For comparison and analysis, BER performance of the simulated network is analyzed and studied as well as the effect of the propagation length on the constellation diagram, and the relation of BER and bit energy and noise density ratio (E_b/N_0).

Keywords—component; Intensity Modulated and Direct Detected (IM/DD), Orthogonal Frequency Division Multiplexing (OFDM), Wavelength Division Multiplexing (WDM), Passive Optical Network (PON), Quadrature Amplitude Modulation (QAM), Bit Error Rate (BER), Reflective Semiconductor Optical Amplifier (RSOA).

I. INTRODUCTION

Wavelength Division Multiplexing Passive Optical Network (WDM-PON) is considered as a cost-effective solution in broadband optical communication since it satisfies several advantages such as higher capacity, better performance and long reach of optical fiber [1-2]. As well as, an Orthogonal Frequency Division Multiplexing (OFDM) technique adds more advantages to WDM-PONs as it provides more flexible bandwidth allocation and a higher data rate for Optical Network Units (ONUs) [3]. OFDM is a preferred solution for WDM-PON because it can provide high spectral efficiency, high tolerance to chromatic dispersion and extend the transmission distance. For reducing implementation cost of PONs, colorless optical sources at ONU are used to re-modulate downstream signal with lower upstream bit rate such as Reflective Semiconductor Optical Amplifier (RSOA) and injection-locked Fabry-Perot laser diode (FP-LD) [4]. RSOA is considered as reflector, modulator, amplifier and colorless transmitter simultaneously, and it is used to re-modulate

downstream signal for generating upstream signal without adding any additional system cost. Recent years, RSOA is preferred for using with PONs due to cost effective [5]. Intensity modulated direct detected OFDM system is considered as cost effective optical OFDM system compared to Coherent detection OFDM system because local laser is not be used at ONU in IM/DD system [6]. Our simulated system combines all mentioned techniques to achieve high data rate and cost-effective optical OFDM system.

M-Quadrature Amplitude Modulation (QAM) is used to increase both the capacity and efficiency of optical systems to support low Bit Error Rate (BER) and high data rate. IM/DD optical OFDM system is considered as a cost-effective optical communication system which can be found in a lot of optical applications [7-12]. According to [13], a bidirectional WDM-PON system based on modulated OFDM signal for 40 Gbps downstream signal and RSOA reused scheme with OFDM signal for 10 Gbps upstream signal was demonstrated. In [14], Bidirectional long reach WDM-PON system provided 20 Gbps downstream data and 10 Gbps upstream data and RSOA was presented and implemented over 45 km optical fiber. In [15], a novel cost-effective RSOA based bidirectional WDM-Ro-FSO-PON was established for next-generation free space optics (FSO) network. 10 Gbps downstream, 1.25 Gbps upstream signal and 1.49 Gbps video signal were sent over 500 m FSO channel. In [16], a 10 Gbps bidirectional RSOA based WDM-PON was analyzed by using Differential Phase Shift Keying (DPSK) downstream signal and OFDM modulated upstream signals over 25 km fiber transmission. In [17], an article wavelength reuse WDM-PON architecture based on incoherent unpolarized light was demonstrated. RSOA was used as a simple, low-cost and colorless reused optical source. In [18], a bidirectional RSOA based WDM-PON using 10 Gbps DPSK downstream signal and 5 Gbps OOK signal re-modulated for with high extinction-ratio in both directions was demonstrated over 20 km optical range. In [19], an EDFA-based 40 Gbps downstream and 10 Gbps upstream signals long-haul WDM-PON scheme was achieved by using QPSK downstream and FBG optical equalizer-based RSOA IM-DD for upstream signal over 40 km fiber transmission. In [20], a novel architecture of WDM OFDM-PON was demonstrated for sending 10 Gbps data both in downstream and upstream transmissions up to 50 km. Direct detection OFDM was used

for downstream signal and simple OOK data was used for upstream transmission. In [21], a long-haul OFDM WDM-PON that provided 100 Gbps downstream data and 2 Gbps upstream data on a single wavelength was simulated. CW laser was used for downstream signal at a central office and a RSOA was used for upstream signal at Gbps long reach IM/DD OFDM system with different M-QAM was analyzed and simulated to provide high data rate downstream signal by using DCF. 4-QAM OFDM system demonstrated the best BER performance compared to other simulated systems for long reach transmission distance. In [22], 100 Gbps IM/DD OFDM system with high order modulation techniques was designed and simulated to achieve good BER performance. As a result, 4-QAM OFDM system demonstrated the best BER performance for long haul transmission distance.

II. SYSTEM MODEL AND DESCRIPTION

The block schematic of the simulated bidirectional system is shown in Fig. 1. The system was designed and simulated using the OptiSystem software [23] as shown in Fig. 2. At the central office (CO), data rate of 40 Gbps has been generated. The number of OFDM subcarriers is 512 and number of Fast Fourier Transform (FFT) points is 1024 so the generated bit rate is 20 Gbps. To avoid inter symbol interference (ISI) between OFDM symbols, cyclic prefix of 100 is inserted for each OFDM symbol after IFFT. QAM sequence generator is used to convert generated bits into symbols according to the number of bits per symbol. M-ary sequence generator converts the generated symbols to multilevel pulses. In OFDM modulator, QAM symbols are modulated into multiple orthogonal sub-carriers. However, the numbers of subcarrier must be half of FFT points.

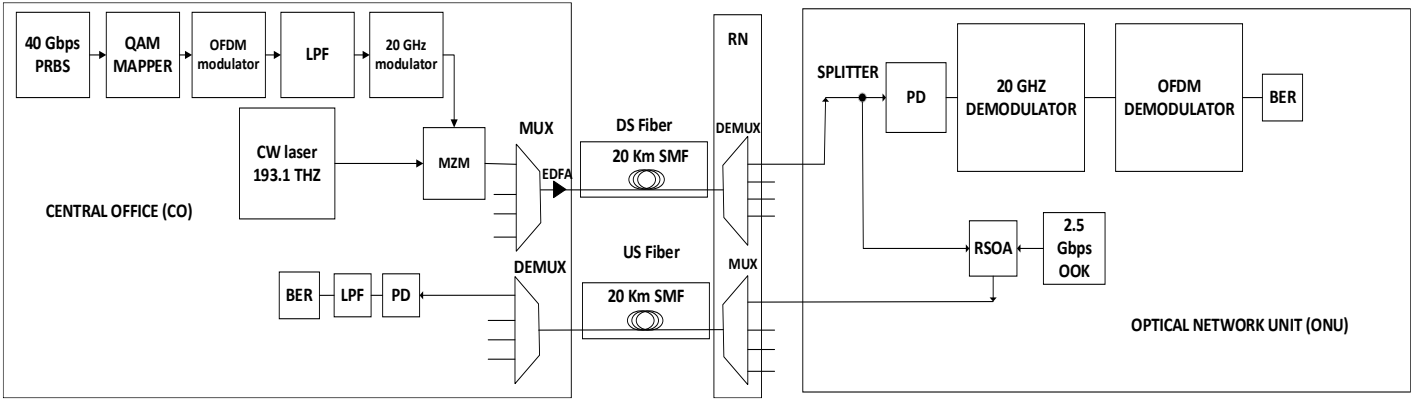


Fig. 1. Simulated Bidirectional IM/DD OFDM WDM-PON based RSOA block diagram

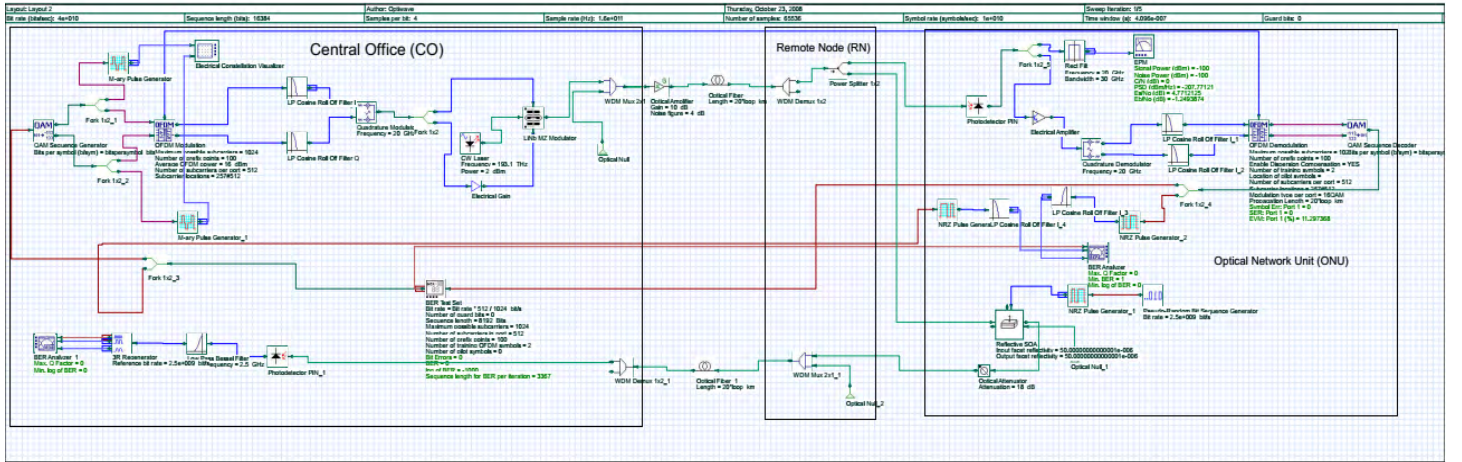


Fig. 2. Simulated Bidirectional IM/DD OFDM WDM-PON based RSOA block diagram on Optisystem software

In this paper, the system providing 40 Gbps 16-QAM OFDM downstream data rate and 2.5 Gbps OOK upstream data rate is designed and simulated. The paper is organized as follows: In Section II, the system is described. In Section III, the simulation results are given. Section IV includes conclusions.

Lower-rate subcarrier tones will take frequency ranges from 0 to 2.5 GHz. The position array in OFDM modulator is used to specify the locations of subcarriers being used and it must be half of the number of total subcarriers. Then, I-Q OFDM modulated signals will pass to low pass filter that has cutoff frequency of $(0.65 \times \text{symbol rate})$. Quadrature modulator is used to raise the frequency of OFDM signals up to 20 GHz. Mach-Zehnder Modulator (MZM) is used to modulate the RF electrical signal to the optical domain by using Continuous wave

laser at 193.1 THz, linewidth of 0.001 MHz, and power of 2 dBm. MZM is operating at quadrature point because of the applied voltage to its arms equal to half switching RF voltage.

After MZM, the optical signal is multiplexed by WDM multiplexer that multiplexes a downstream signal of different wavelength then is amplified by using optical amplifier that has 10 dB gain and 4 dB noise floor as main parameters. The resulting optical signal is then transmitted over $(20 \times \text{Number of loops})$ SMF with a dispersion of 16 ps/nm/km, attenuation of 0.2 dB/Km, a nonlinearity coefficient of 2.6×10^{-20} and a dispersion slope of 0.08 ps/nm²/km. At Remote Node (RN), transmitted optical signal is demultiplexed by using WDM demultiplexer that has a bandwidth of 60 GHz and insertion loss of 1.5 dB

At the ONU side, optical splitter is used to split optical downstream signal. Half of the optical signal is transmitted to PIN photodetector that converts the optical signal to an electrical signal. PIN photodetector has a dark current of 10 nA, a thermal power density (15×10^{-24} W/Hz) and a center frequency of 193.1 THz. Electrical amplifier is used to amplify the received electrical signal. Quadrature demodulator is used to down convert and recover the amplified signal to be in range 0 and 2.5 GHz. OFDM modulator and demodulator parameters are same in order to recover the transmitted QAM symbols correctly. QAM sequence detector decode the received symbols to bits according to number of bits per symbol.

The other half of optical signal is sent to RSOA for re-modulation with 2.5 Gbps OOK upstream signal. Table 1 lists the design parameters of the RSOA. The re-modulated signals of these RSOAs are combined again by WDM multiplexer at RN. These combined signals are sent back to upstream fiber (US fiber) that has the same parameters of Downstream fiber (DS fiber). At CO, upstream signal is then demultiplexed then it is detected by PIN photodetector (PD). Then, received upstream signal is sent through a low pass filter. BER for upstream signals is calculated by using BER analyzer.

Table 1: Main RSOA parameters

Parameters	Values
Input Facet Reflectivity	50×10^{-6}
Output Facet Reflectivity	50×10^{-6}
Active Length	0.0006 m
Taper Length	0.0001 m
Width	0.4e-006
Height	0.4e-006

III. SIMULATION RESULTS AND DISCUSSION

The system was simulated and analyzed for different propagation length. QAM coding is used to generate symbols and is considered the best coding compared to others. 16-QAM is used and applied to IM/DD-OFDM-WDM-PON to check the transmission performance and improve system budget. The laser linewidth is set to 0.001 MHz and its power to 2 dBm to minimize the effect of fiber nonlinearity. Propagation length is increased to check its effect on

transmission performance and BER results so our goal is to find the optimum propagation length to achieve the FEC limit which is a BER of 10^{-3} for downstream signal. Fig. 3 shows the transmitted and received electrical spectrum after passing through low pass filter.

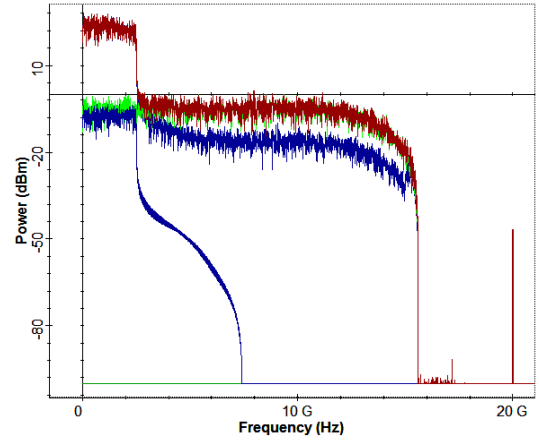


Fig. 3. Transmitted (red) and received (blue) signal after low pass filter

Fig. 4 shows the modulated and detected electrical signal at both CO and ONU, respectively. As shown in fig. 4, downstream signal is modulated at 20 GHz. Fig. 5 shows the optical spectrum of transmitted signal at both CO and ONU. Fig. 6 shows the constellation diagram of the transmitted and received 16-QAM signal after propagation length of 20 km at BER of 10^{-3} .

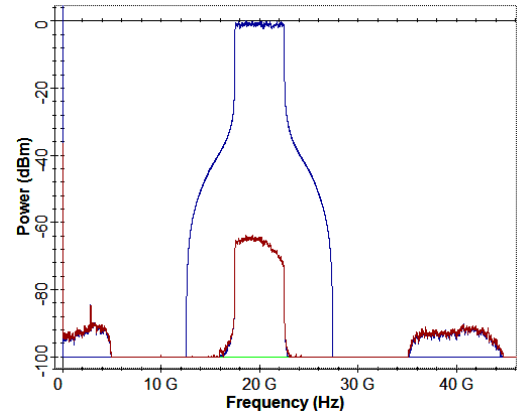


Fig. 4. Transmitted (blue) electrical signal after modulator and received (red) signal after PIN

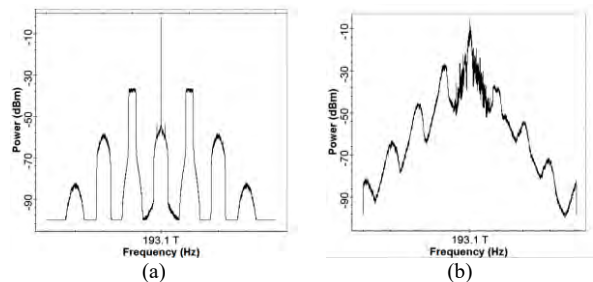


Fig. 5. Optical spectrum of (a) transmitted downstream and (b) upstream signal

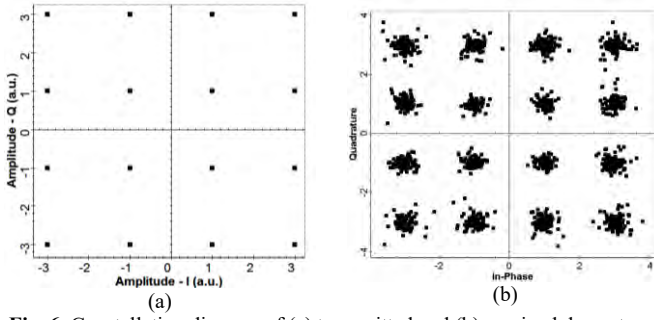


Fig. 6. Constellation diagram of (a) transmitted and (b) received downstream OFDM signal at BER of 10^{-3}

Eye diagram is used to calculate the combined effects of channel noise and ISI on the performance of a baseband pulse-transmission system. For minimum propagation length, Fig. 7 shows the eye diagram of received downstream and upstream signal, respectively. Eye opening can be represented by one-bit period and according to Fig. 7 the result of opening seems very clear.

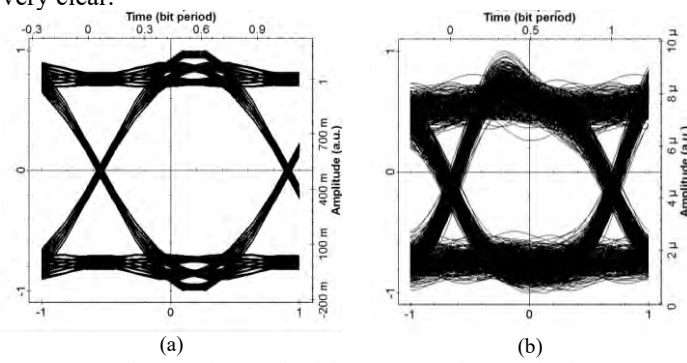


Fig. 7. Eye diagram of (a) received downstream and (b) received upstream

To check the BER performance of received downstream signal, Fig. 8 shows the variation of BER for received downstream signal with propagation length. Due to chromatic dispersion of fiber, BER value degrades as propagation length increases. 16-QAM OFDM WDM-PON achieved the best BER performance at 20 Km fiber length as shown in Fig. 8.

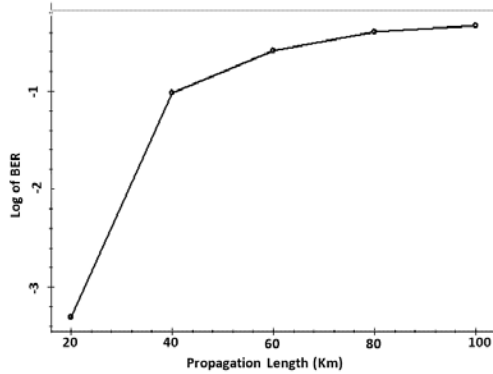


Fig. 8. BER results of received downstream versus propagation length

Fig. 9 shows the variation of BER for received upstream signal with propagation length. The best BER value is obtained at 20 km fiber length as shown in fig. 9.

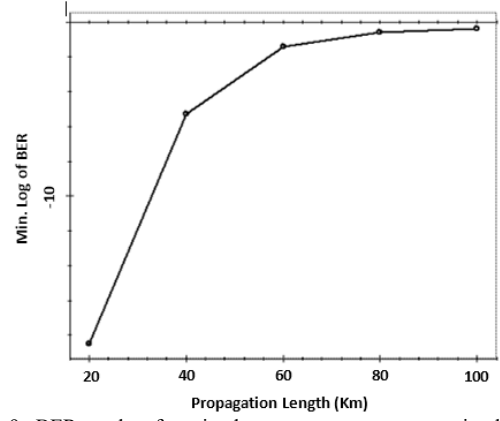


Fig. 9. BER results of received upstream versus propagation length
Fig. 10 shows the relationship between BER and E_b/N_0 for the simulated system. At BER of 10^{-3} and Q factor of 11.6, E_b/N_0 ratio is calculated as 29.5 dB at 20 km fiber length.

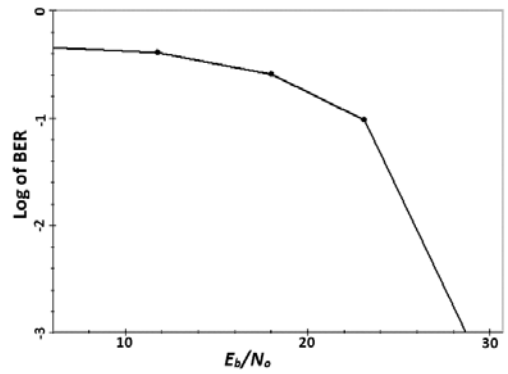


Fig. 10. BER as a function of E_b/N_0 for simulated system
It can be seen that 40 Gbps IM/DD OFDM WDM-PON system achieved the best results for 16-QAM at 20 km fiber length without using DCF.

IV. CONCLUSION

We have designed and simulated the bidirectional cost-effective IM/DD OFDM WDM-PON based on RSOA with 16-QAM for 40 Gbps downstream signal and 2.5 Gbps upstream signal. IM/DD OFDM WDM-PONs do not need another laser at the receiver like Coherent Detection OFDM systems. A 40 Gbit/s 16-QAM intensity modulated direct Detected OFDM WDM-PON has been transmitted over 20 km SMF and achieved BER of 10^{-3} . At BER of 10^{-3} , 16-QAM OFDM system achieved the best simulation results at propagation length of 20 km. Eye diagram, Q factor and E_b/N_0 results are explained against propagation length for the simulated system. This system is simulated and analyzed to achieve high BER performance and cost effective 40-Gbps optical IM/DD OFDM WDM-PON.

Acknowledgements

This work was supported by the Erciyes University Scientific Research Projects Coordination Unit (Project No: FDK-2019-8750).

REFERENCES

- [1] Kani, J.I., Bourgart, F., Cui, A., Rafel, A., Campbell, M., Davey and R., Rodrigues, "Next-generation PON Part I: technology roadmap and general requirements," *Journal of Electrical Engineering, IEEE Commun. Mag.* 47, 43–49, 2009.
- [2] Fady El-Nahal and Norbert Hanik, "Technologies for Future Wavelength Division Multiplexing Passive Optical Networks", *IET Optoelectronics*, Vol. 14, No. 2, pp: 53-57, 2020.
- [3] I. Djordjevic and B. Vasic, "Orthogonal frequency division multiplexing for high-speed optical transmission," *Opt. Exp.* Vol. 14, Issue 9, pp. 3767-3775, 2006.
- [4] C. Chow, C. Yeh, C. Wang, F. Shih and S. Chi, "Signal remodulation of OFD16-QAM for long reach carrier distributed passive optical networks," *IEEE Photon. Technol. Lett.* 21, 715–717, 2009.
- [5] A.S. Das, A.S. Patra, "Bidirectional transmission of 10 Gb/s using RSOA based WDM-PON and optical carrier suppression scheme", *J. Opt. Commun.* 35, 239–243, 2014.
- [6] A. El Kabil and M. Faqih, "Optical OFDM (O-OFDM) for Intensity Modulated/Direct Detection Optical Systems," 2018 IEEE International Conference on Communication, Networks and Satellite (Comnetsat)
- [7] J. Dang, L. Wu and Z. Zhang, "OFDM systems for optical communication with intensity modulation and direct detection", *Optical Fiber and Wireless Communications*, pp. 85-104, 2017.
- [8] T. Nguyen, S. Mhatli, E. Giacomidis, L. Compemolle, M. Wuilpart and P. Mégret, "Fiber nonlinearity equalizer based on support vector classification for coherent optical OFDM", *IEEE Photon. J.*, vol. 8, no. 2, 2016.
- [9] S. Selvendran, A. S. Raja, K. Esakki Muthu and A. Lakshmi, "Certain investigation on visible light communication with OFDM modulated white LED using optisystem simulation," *Wirel. Pers. Commun.*, vol. 109, pp.1377-1394, 2019.
- [10] H. Oubei, C. Shen, A. Kammoun and E. Zedini, "Light based underwater wireless communications", *Jpn J Appl Phys*, vol. 57, pp. 1-18, 2018.
- [11] J. Dang, L. Wu and Z. Zhang, "Optical Fiber and Wireless Communications," 2017.
- [12] Fady El-Nahal, "Bidirectional OFDM-WDM-PON system employing 16-QAM intensity modulated OFDM downstream and OOK modulated upstream", *Photonics Letters of Poland*, Vol. 8, No. 2, pp: 60-62, 2016.
- [13] T. Dong, Y. Bao, Y. Ji, A. Pak Tao Lau, Z. Li, and Chao Lu, "Bidirectional Hybrid OFDM-WDM-PON System for 40-Gb/s Downlink and 10-Gb/s Uplink Transmission Using RSOA Remodulation," *IEEE Photon. Technol. Lett.*, Vol. 24, No. 22, November 2012.
- [14] L. Tawade, S. Mhatli and R. Attia, "Bidirectional long reach WDM-PON delivering downstream data 20 Gbps and upstream data 10 Gbps using mode locked laser and RSOA," *Opt Quant Electron*, 47, 779-789, 2014.
- [15] G. Mandal, R. Mukherjee, B. Das, A. Sekhar Patra, "Next-generation bidirectional Triple-play services using RSOA based WDM Radio on Free-Space Optics PON," *Opt. Commun.*, 411 138–142, 2018.
- [16] P. Choudhury and T. Khan, "Symmetric 10 Gb/s wavelength reused bidirectional RSOA based WDM-PON with DPSK modulated downstream and OFDM modulated upstream signals", *Opt. Commun.*, Vol. 372, pp. 180–184, 2016
- [17] M. Sperm and D. Babic, "Wavelength reuse WDM-PON using RSOA and modulation averaging," *Opt. Commun.*, Vol. 451 pp. 1–5, 2019.
- [18] J. Zhang¹, X. Yuan¹, Y. Gu² and Y. Huang, "A Novel Bidirectional RSOA Based WDM-PON with Downstream DPSK and Upstream Remodulated OOK Data," *ICTON 2009*
- [19] Z. Zhang, X. Chen, L. Wang and M. Zhang, "40-Gb/s QPSK Downstream and 10-Gb/s RSOA based Upstream Transmission in Long-Reach WDM PON Employing Remotely Pumped EDFA and FBG Optical Equalizer," 2013 8th International Conference on Communications and Networking in China (CHINACOM).
- [20] G. Pandey and A. Goel, "Long reach colorless WDM OFDM-PON using direct detection OFDM transmission for downstream and OOK for upstream," *Opt Quant Electron*, 2014.
- [21] S. Mhatli, M. Ghanbarisabagh and L. Tawade, "Long-reach OFDM WDM-PON delivering 100 Gb/s of data downstream and 2 Gb/s of data upstream using a continuous-wave laser and a reflective semiconductor optical amplifier," *Opt. Lett.*, Vol. 39, No. 23, 2014.
- [22] N. Taspmar and M. Alhalabi, "Performance investigation of long-haul high data rate optical OFDM IM/DD system with different QAM modulations," *J. Electr. Eng.*, Vol. 72, No. 3, pp. 192-197, 2021
- [23] "OptiSystem Package from Optiwave"

Development and Deployment of a LoRaWAN Performance Test Setup for IoT Applications

Simeon Trendov

Master Student at the Double Degree
program Dedicated Embedded
Computer Systems and IoT
Faculty of Electrical Engineering and
IT, Univ. "Ss. Cyril and Methodius"/
Anhalt University of Applied Sciences
Skopje, R. N. Macedonia
trendov.s@yahoo.com

Prof. Dr. Eduard Siemens
Communication Systems
Department, Future Internet Lab
Anhalt (FILA)
Anhalt University of Applied
Köthen, Germany
eduard.siemens@hs-anhalt.de

Prof. Dr. Marija Kalendar
Computer Technologies and
Engineering Department
Faculty of Electrical Engineering and
IT, Univ. "Ss. Cyril and Methodius"
Skopje, R. N. Macedonia
marijaka@feit.ukim.edu.mk

Abstract—Emerging Internet of Things (IoT) trends including smart cities, smart factories, smart farming and many other applications comprising connected "things" are imposing more and more demand on the Radio Access Network (RAN) in terms of power consumption, coverage and scalability. Most of the technologies utilized for the Internet of Things are not able to manage all of the needed and upcoming requirements, hence they are continuously upgrading and developing. LoRaWAN is an emerging Low-Power Wide Area Network (LPWAN) technology. In order to enhance the opportunities for the new requirements, this paper covers the development and deployment of a LoRaWAN performance test setup including two end nodes and a gateway to test the round-trip time, range, packet loss and signal strength of the network.

Keywords— *LoRa, LoRaWAN, IoT, End Node, Gateway, Network Server, Application Server, Join Server*

I. INTRODUCTION

LoRaWAN is a Long Range Wide Area Network and it is one of the LPWAN technologies. For using the LoRaWAN there must be at least one end node and one gateway, able to communicate on the same frequency bands depending on the local frequency regulations [1].

LoRaWAN is a technology composed of End Nodes, Gateways, Network Servers, Application Server and Join Server [2]. A LoRaWAN-enabled end device is a sensor or an actuator which is wirelessly connected to a LoRaWAN network through radio gateways [3]. All of the data collected by the end devices are sent by low power LoRaWAN network to a gateway within a listening distance, the data is then forwarded to the network server. The network server manages the entire network, dynamically controls the network parameters to adapt the system to ever-changing conditions [4]. When a message is received, the network server will forward it to a specific application server. The Join server manages the connection process for the first time for an end device to be added to the network [5].

The main objective of this paper is to analyze and test the LoRaWAN technology performance, primarily in terms of packet transmission delay, data throughput and other QoS parameter. The paper aims to cover the development and

deployment of a LoRaWAN performance test. The LoRaWAN network will be tested under real life conditions. Another objective is to give special attention to the architecture of this technology, for the needs of the tests, which will be covered in part three. This also puts as objective to describe in detail all of the used equipment needed for accomplishing the tests. The final objective is the detailed analysis that will be made for the LoRaWAN technology and testing of its range, signal strength, packet loss and message transmission delay. In part four the results of all of the tested parameters are given together with the conclusions and recommendations for future work.

II. STATE OF THE ART

The story about LoRaWAN began in 2009, when Nicolas Sornin and Olivier Sella started working on their idea to develop a long range, low power modulation technology. Together with Francois Sforza, in 2010 they started the company Cycleo. The company Semtech acquired Cycleo. The LoRaWAN alliance was founded in 2015 [6].

There are many LoRa deployment experiments described in the literature, but some were more influential for this work. All these examples of experiments contributed for the experimental setup of the research for this paper. Some of them are described below.

The first example is the LW003-B is a LoRaWAN Bluetooth gateway integrating LoRa and Bluetooth wireless communication. It is used to realize environmental monitoring and indoor positioning [7]. For achieving low power, long range and high data transmission Kim [8] designed a module that combines Wi-Fi and LoRa. For regulating the power usage the system is integrated with a power and data scheduler that can choose between Wi-Fi and LoRa according to the priority of gathered data.

The authors in [9] tested the low-power wide-area communication protocol LoRaWAN with modelling in Matlab. For measuring urban greenhouse gas emissions in cities Ahlers in [10] used the LoRaWAN technology. It is a low cost automated system for greenhouse gas emissions

monitoring network around the city. Ruobing Liang [11] tested the performance of LoRa modules in a concrete office building. The tests were done on different floors with different distances, while using eight receiving LoRa modules.

Finally, Neumann in [12] wanted to test LoRaWAN performance and its limitations in indoor environments and to define its use in 5G networks. The biggest limitation were the ISM (Industrial, Scientific and Medical) bands regulations, that define the maximum amount of data sent per day on those frequency bands.

The development and the deployment of LoRa does not stop with these examples and its implementation continues to grow with the numerous opportunities for different applications. This research is also tending to contribute to its recognition and advantages.

Two end nodes and a gateway were used to test the round trip time, range, packet loss and signal strength of the network. The used equipment and the experimental set up is explained in detail. The summary includes the results of all of the tested parameters given together and discussed.

III. LORAWAN TEST PLATFORM ARCHITECTURE

In this paper two end devices and one gateway have been used to test the network parameters. The first end device will be used to test the round-trip time, packet loss, and the signal strength, the second device will be used to test the range. These two devices will have to establish a connection with the gateway and send messages to it. The gateway has to collect the data from the end devices, send back acknowledgments for every received message and forward the messages to the network server. The network server will send the messages to the application server through which it will be displayed on a computer monitor.

A. Equipment deployed in the test architecture

As end devices for the test setup, two Arduino UNO devices have been used. One of them is connected to a LoRa shield. The LoRa shield has an antenna connector and a RFM95W LoRa module [13]. This LoRa module is able to communicate on the 868 MHz frequency with the gateway. This setup together with the appropriate software defines one of the end nodes used for measuring the round-trip time, the packet loss and the signal strength [14].

The second end node is composed of the other Arduino UNO and a LoRa/GPS shield. The LoRa/GPS shield is composed of a LoRa part and a GPS part. This second setup, together with the appropriate software defines the second end node for testing the operation range of LoRaWAN [15].

A gateway named MultiTech Conduit AP was also used for the tests. Fig. 1 shows the used equipment for the needs of the tests with the connections and communication between them.

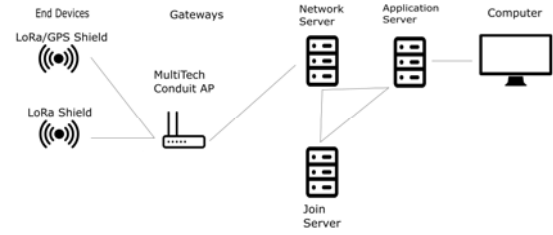


Fig. 1 LoRa Architecture of the performance test setup

B. End device with LoRa Shield

The Dragino LoRa Shield is a long range transceiver. It is capable of FSK (Frequency Shift Keying) and LoRa modulation. It is powered with 5 V through the Arduino board. This shield is capable to communicate on the 868 MHz frequency band [16]. The shield has an antenna connection through which it communicates with the gateway. On top of it there is the RFM95W module, soldered at the center.

C. End device with a LoRa/GPS Shield

The second end device used is composed of Arduino UNO and LoRa/GPS Shield. The LoRa/GPS Shield has a LoRa module and a GPS module. Two antennas can be connected to it, one to the LoRa Bee and one directly to the shield [14]. It needs power supply of 3.3 V and is capable for transmitting and receiving on the 868 MHz frequency band. This end device is powered by a 20 Ah AUKEY power bank in the field. The LoRaWAN network was tested with this end device from different distances and with different amounts of obstacles in the Fresnel zone [15].

IV. EXPERIMENTS, MEASUREMENTS AND RESULTS

For the purposes of the experiments the gateway was placed outdoor on the top of a four floors building with an approximate height of 15 meters to maximize the range. The received packets were available on a computer. The measurements with the LoRa Shield end device were done in a laboratory. The LoRa/GPS Shield end device was connected to a power bank and it was moved around the city of Kothen, Germany. All of the experiments are done in non-line of sight communication with a lot of objects in the Fresnel zone. The gateway was only communicating with the two end devices. It was set up to be private and not to accept any other messages from other end devices. Both of the devices have transmitted a new message every 20 seconds to the gateway. In such a setup, big amount of data was collected and the analysis becomes more accurate. The measurements of the round-trip time, packet loss, signal strength and the range are presented in the following sections. In all these measurements only one parameter was changed at a time, and the relationship between that parameter and the round-trip time or the range can be noticed in the results below.

A. Measurements of the Round-trip Time

For measuring the round-trip time, the packet loss and the signal strength, the gateway was placed outdoor on top of the same as above and the LoRa Shield end device was placed in a laboratory with an approximate distance between them of 120

meters and with several objects (especially thick solid walls) in the Fresnel zone. The behavior of the round-trip time was measured according to packet size, transmission power and the spreading factor.

For each case the round-trip time was measured ten times and the minimum, maximum and average round-trip times were calculated. For each case the average RSSI level is noted and given in the tables as registered by the gateway. For each message sent the end device is expecting an acknowledgement which is used to calculate the packet loss. The measurements for each case of the packet loss are also given in the tables below.

For the consistency of the test results, tests have been arranged in that way that a few assumptions can be made.

- When testing the round-trip time the distance between the end device and the gateway does not change.
- The objects in the Fresnel zone between the devices are the same and do not change.
- The processing delay is the same for all measurements.

1) Relationship Between the Round-trip Time and the Packet Size

The behavior of the round-trip time has been tested according to the amount of data sent. There are four groups of measurements done, for packet sizes of 13, 23, 39 and 50 Bytes. The messages have been sent with a transmission power of 14 dBm, a spreading factor of 7 and the bandwidth remains 125 KHz.

The difference in the round-trip time between the four groups is small but it can be easily spotted, and an increasing trend in the time can be noticed as the packet size increases. This increasing trend is not with a big friction.

Because of the small changes in the packet size the round-trip time is not affected a lot. The maximum packet size varies from 51 to 242 Bytes according to the spreading factor that is used. When the end node is sending data, its packet size can vary depending on the message that needs to be send to the gateway. The results show that the variation of the packet sizes will not have a great effect on the interference and collision of messages. All the results are presented on Fig. 2.

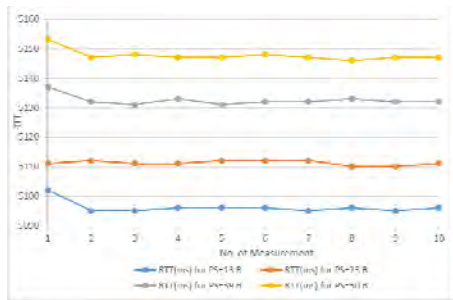


Fig. 2 Relationship between the Round-Trip Time and the Packet Size

For each of the four groups of measurements in Table I, the packet loss, average RSSI, minimum RTT, maximum RTT and average RTT are given. The information about the RSSI level is the average RSSI level registered by the gateway. The

differences in the round-trip time between the four groups can be easily spotted. The Time on Air (ToA) is calculated according to the formula (1) The Subtracted value from the round-trip time is the delay added by the gateway [10]. The processing time is not taken into consideration because of its variation and its small impact.

$$ToA = \frac{RTT - 530ms}{2} \quad (1)$$

TABLE I. BER, AVERAGE RSSI, MIN, MAX, AVERAGE RTT AND AVERAGE TIME ON AIR - MEASURING RTT WITH CHANGING THE PACKET SIZE

Packet Size (Bytes)	13	23	39	50
Packet Loss	10/10	10/10	9/10	10/10
Average RSSI (dBm)	-89.5	-89.6	-89.5	-88.8
Minimum RTT (ms)	5 095	5 110	5 131	5 146
Maximum RTT (ms)	5 102	5 112	5 137	5 153
Average RTT (ms)	5 096.2	5 111.2	5 132.5	5 147.7
Average ToA (ms)	2 283.1	2 290.6	2 301.25	2 308.85

As seen by the results from the packet loss, the data transmission is stable. Usually, all of the messages get to the gateway and an acknowledgement is received by the end device. The RSSI levels are stable, no difference can be seen when changing the packet size that is sent.

2) The Relationship Between the Round-trip Time and the Transmission Power

The behavior of the round-trip time is tested according to the transmission power. The round-trip time is tested when using a transmission power of 10, 14 and 20 dBm. The packet size is always 13 Bytes and a spreading factor of 7 is used for these three groups of measurements. A bandwidth of 125 kHz is used for all the measurements.

The results are presented on Fig. 3. There is no big difference between the groups of measurements when changing the transmission power. The power used for transmitting a message does not have a big effect on the round-trip time.

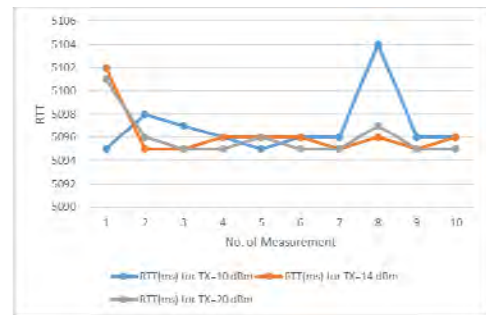


Fig. 3 The relationship between the Round-trip Time and the Transmission Power

Table II gives the values for the packet loss, average RSSI, minimum, maximum and the average RTT. It can be seen that the differences between the three groups of measurements are very small and the transmission power does not have big

effect on the round-trip time. The time on air is calculated as presented in (1).

TABLE I BER, AVERAGE RSSI, MIN, MAX, AVERAGE RTT AND AVERAGE TIME ON AIR - MEASURING RTT WITH CHANGING THE TRANSMISSION POWER

Transmission Power (dBm)	10	14	20
Packet Loss	9/10	10/10	10/10
Average RSSI (dBm)	-98.6	-89.5	-88.6
Minimum RTT (ms)	5095	5095	5095
Maximum RTT (ms)	5104	5102	5101
Average RTT (ms)	5097.5	5096.2	5096
Average ToA (ms)	2283.75	2283.1	2283

The data transmission is stable. Usually, all of the messages get to the gateway and an acknowledgement is received by the end device. A change can be noticed in the RSSI levels when using a transmission power of 10 dBm compared to the transmission power of 14 dBm. When sending a message with smaller transmission power, weaker signal arrives at the receiver. A small change can be also seen between the transmission powers of 14 dBm and 20 dBm.

3) The Relationship Between the Round-trip Time and the Spreading Factor

The behavior of the round-trip time is tested according to the spreading factor. In this cycle of measurements the behavior of the round-trip time is tested when using a spreading factor of 7, 8, 9, 10, 11 and 12 witch correspond to data rates of 5469 bps, 3125 bps, 1758 bps, 977 bps, 537 bps and 293 bps respectively. The packet size is always 13 Bytes and a transmission power of 14 dBm is used. A bandwidth of 125 kHz is used for all of the measurements. All of the results are presented on Fig. 4.

The differences between the six groups of measurements are easily noticeable. As the spreading factor is increasing, the message needs more time to arrive at the gateway and for an acknowledgement to be received by the end node. The measurements of the round-trip time for the first four groups for the spreading factors 7, 8, 9 and 10 are closer than the measurements for the spreading factors 11 and 12. The round-trip time significantly increases for the last two groups of measurements. This is because of the corresponding bits per second of each spreading factor.

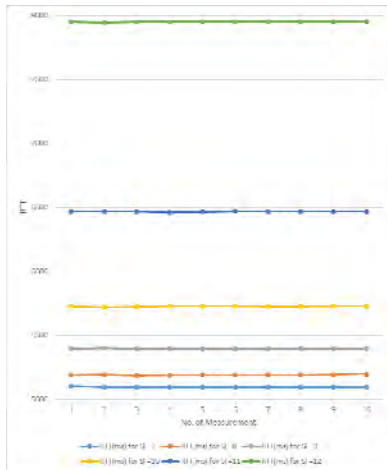


Fig. 4 The relationship between the Round-Trip Time and the Spreading Factor

Big changes in the round-trip time can be seen. When increasing the spreading factor, the message is transmitted slower and it is spending more time on air. This can cause more collisions and data loss.

Table III gives the packet loss, average RSSI, minimum, maximum and the average RTT. In the first four groups, from spreading factor 7 to 10 the time difference is noticeable but it is not that significant. A very big difference in the round-trip time can be noticed as the spreading factor receives a bigger value. The time on air is calculated as shown in (1). The packet loss is stable and usually all of the messages are sent to the gateway and an acknowledgement is received by the end device. The RSSI levels are stable, only small variations can be seen when changing the spreading factor.

TABLE II BER, AVERAGE RSSI, MIN, MAX, AVERAGE RTT AND AVERAGE TIME ON AIR - MEASURING RTT WITH CHANGING THE SPREADING FACTOR

Spreading Factor	7	8	9	10	11	12
Packet Loss	10/10	10/10	9/10	10/10	9/10	10/10
Average RSSI (dBm)	-89.5	-88.7	-90.3	-89.6	-89.4	-90.1
Minimum RTT (ms)	5095	5188	5391	5720	6460	7945
Maximum RTT (ms)	5102	5195	5399	5728	6469	7952
Average RTT (ms)	5096.2	5190.9	5394.4	5725.7	6465.9	7949.6
Average ToA (ms)	2283.1	2330.4	2432.2	2597.8	2967.9	3709.8

B. Measurements of the Range

For measuring the range of the LoRaWAN network when using the LoRa/GPS Shield end-device and the MultiTech gateway. The end device was connected to the power bank and it was carried around the city. The behavior of the range was measured when changing the transmission power and the spreading factor.

In order to ensure the accuracy of the test results, a few assumptions were made. Leaving these assumptions opens a space for future research in this direction.

- The weather conditions are the same.
- The humidity in the air is the same.
- The processing delay is the same for all measurements.

1) The relationship between the Range and the Transmission Power

This part covers the tests for the behavior of the range according to the transmission power. Three measurements were executed where the LoRa/GPS shield end device was using different transmission power. In all of the three cases the end device was sending 32 Bytes messages to the gateway containing the GPS coordinates of the end device. In all the cases the device was shifted (moved) in one direction, so the surrounding conditions don't change a lot compared with the other tests. For these tests only a spreading factor of 7 and a bandwidth 125 kHz was used. When using a transmission power of 10 dBm, 14 dBm and 20 dBm it was tested if for the

specific range there is communication in-between the end-node and the gateway.

The differences in the distances are clearly noticeable. As the transmission power was increased, longer distances are achieved, in which data transmission can be acquired between the end-device and the gateway. All the measurements are presented on Fig. 5. This is of great importance when the range needs to be controlled. In big smart factories where a lot of end nodes need to be connected, data can be lost due to interference and collision of packets. Because of this, they can be divided into smaller groups by their placement in the smart factory with a separate gateway in the middle of each group. With lowering the transmission power, the messages from one group of end-devices would not reach the other group. This way the number of collisions will drop to a minimum and important data will be saved. For this example, the spreading factor should also be kept to a minimum. The relationship between the spreading factor and the range is presented on Fig.6. This can function like a hexagonal honeycomb structure as in mobile communications.

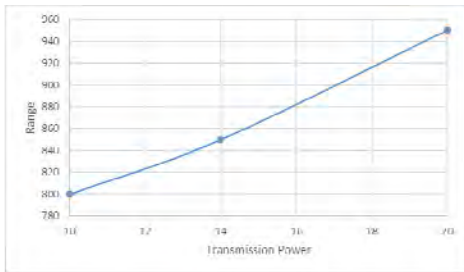


Fig. 5 The relationship between the Range and the Transmission Power

2) The relationship between the Range and the Spreading Factor

In this cycle of measurements the behavior of the range is tested when using a spreading factor of 7, 8, 9, 10, 11 and 12 which correspond to data rates of 5469 bps, 3125 bps, 1758 bps, 977 bps, 537 bps and 293 bps respectively. The LoRa/GPS Shield end-device was shifted (moved) in the same direction as for the tests done in part 6.2.1, so the surrounding conditions don't change considerably. In all of the cases the end device was sending 32 Bytes messages to the gateway, containing the GPS coordinates of the end device. For these tests only a transmission power of 14 dBm and a bandwidth 125 kHz was used. All the measurements are presented on Fig.6. The difference in the achieved distances in which data transmission can be acquired between the end device and the gateway is easily noticeable. As the spreading factor is increased, data can be transferred on longer distances.

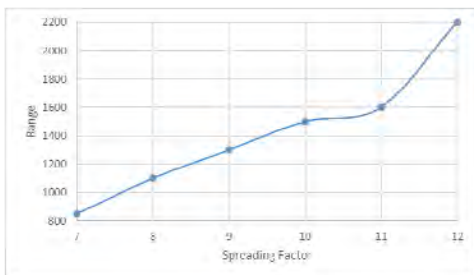


Fig. 6 The relationship between the Range and the Spreading Factor

The spreading factor is important to be controlled when data needs to be sent on big distances. As shown by the results when the spreading factor is increased, data transfer is possible on longer distances. For example, this can find a big use in smart farming, where there is a small number of end-devices but they are spread on a big field, so data needs to travel on long distances. Additionally the small amount of obstacles in the Fresnel zone will increase the maximum range in which data transfer is possible. The data can be gathered by one gateway without any collision problems because of the small number of end devices. The spreading factor is also important in urban environments. With increasing the spreading factor there is a bigger chance that a message will reach the receiver.

The biggest range in which data transfer was possible between the transmitter represented with the black arrow in the upper corner and the receiver represented with the black arrow in the down corner are shown on Fig.7. When transmitting the messages, the end device was using a spreading factor of 12 and a transmission power of 14 dBm.

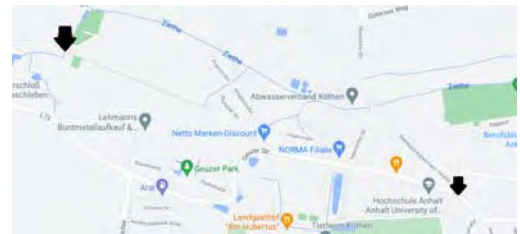


Fig. 7 Range of 2200m while using SF 12 and TX power 14 dBm

V. CONCLUSION AND FUTURE WORK

The research in the paper is a relatively small part of the possible research area including these new technologies, but it confirms some of the assumptions, sets the basics of an experimental set up, by opening new application horizons.

With the results obtained during the research in this paper we can conclude that:

- When talking about the round-trip time and the time on air, it is good to keep them to a minimum. When a device is spending more time on air it keeps the channel busy, which can lead to more collisions in data transfer and the battery life is shortened. The spreading factor should only be increased when there is the need to send the data on big distances or for bigger reassurance that the message will arrive at the gateway. When using spreading factor 7, a distance of 850m was reached. In this case the data reaches the gateway faster and the end device is spending an average of 2283.1ms time on air. In this case less battery is spend by the end device and the channels get released faster. If a bigger distance needs to be acquired a bigger spreading factor should be used. When using a spreading factor of 12 a distance of 2200m was achieved between the gateway and the end device. In this case the data is send slower and the device spends 3709.8ms time on air. When comparing this measurements for spreading factor 12 to the measurements for spreading factor 7 it can be concluded that the device is occupying the channels

longer. Because of the longer transmission it is spending more battery, but achieves bigger range in return.

- The transmission power used for sending a message only affects the range but not the time on air. When using a transmission power of 14dBm, an average of 2283.1ms was measured. There is a really small difference of 0.1ms when comparing this results to the results when using transmission power of 20 dBm. When sending a message with higher values of the transmission power it can achieve longer distances without spending a lot of time on the air. A key factor not to be forgotten here is the battery usage. When using higher transmission power more energy is needed and the end node will spend its battery faster. The LoRaWAN technology finds big use where data needs to be collected. Data transfer is possible on short ranges, when there is the need for it because of the presence of many end-devices. This will help the devices not to interfere with each other and the data collision will be dropped to a minimum. Data transfer is also possible on very long ranges. The numbers can be controlled depending on the use case.

Future applications of the LoRaWAN technology can be seen in combination with other technologies. If an end node needs to send small messages it can use the LoRa standard, but if greater amounts of data are collected it can use some other technology that provides transfer of greater amounts of data. With this combination power will be saved and the channels used for data transfer will not be overloaded.

Improvement of the LoRaWAN technology can be also done by creating better libraries for programing the end devices. A couple of problems were faced when one of the end devices was programed with the *lmic* library. One of the biggest issues was the short receiving window when increasing the spreading factor for testing the round-trip time. The library works well for the spreading factors 7 and 8, but for bigger values the receiving window is closed before an acknowledgement can be received. The library shows no problems for the values of the round-trip time around the average of 5190.9 ms for spreading factor 8, but problems start to appear for the spreading factor 9. When using a spreading factor 9, an average of 5394.4 ms was measured for the round-trip time. In this case some of the acknowledgements were received but most of them were lost. This problem was fixed by increasing the receiving window of the library, so the end devices can wait for the acknowledgement to arrive.

More research can be done with the integrations available for The Things Stack Enterprise which is a paid version of The Things Stack. Using these integrations, the collected data can be immediately analyzed and made useful to be presented to the user. This way the user can react immediately to the data gathered by the end-nodes. For future applications the behavior of the LoRaWAN network can be tested when combining it with other technologies. Tests can be done on the usage of the battery by the end device, and on the overloading of network. . Such research and the obtained results can leads to new ideas for future applications of the LoRaWAN technology.

This technology is relatively new, and has a lot of potential. With increasing the amount of public gateways it can become irreplaceable when talking about wireless communication between sensors and actuators. The LoRaWAN technology can be used for tracking goods, logistics, agriculture and farming, smart buildings, smart city and parking, even monitoring things like trash cans, pipes and a lot more. Using it in combination with other technologies it can present even more practical and powerful use-cases.

REFERENCES

- [1] K. E. John Koon, LoRaWAN Empowers Very Low-power, Wireless Applications: The Future of Low-power Wireless Network, John Koon, Kindle Edition.
- [2] E. M. O. Maximiliano Santos, Hands-On IoT Solutions with Blockchain.
- [3] P. Lea, Internet of Things for Architects, 2018.
- [4] B. Wiegmann, IoT Networks with LoRaWAN.
- [5] Semtech Corporation, "semtech," December 2019. [Online]. Available: <https://loro-developers.semtech.com/library/tech-papers-and-guides/loro-and-lorawan/>.
- [6] L. Slats, "semtech," semtech, 08 January 2020. [Online]. Available: <https://blog.semtech.com/a-brief-history-of-lora-three-inventors-share-their-personal-story-at-the-things-conference>.
- [7] mokosmart, "mokosmart," 01 2021. [Online]. Available: <https://www.mokosmart.com/lorawan-probe-lw003-b/>.
- [8] J. Y. L. a. J. D. K. D. H. Kim, "Low-Power, Long-Range, HighData Transmission Using Wi-Fi and LoRa," in *the 6th International Conference on IT Convergence and Security*, 2016.
- [9] L. B. M. J. B. a. J. A. M. Ivana Tomić, "The Limits of LoRaWAN in Event-Triggered Wireless Networked Control Systems," 2020.
- [10] A. Dirk, "A measurement-driven approach to understand urban greenhouse gas emissions in Nordic cities," NIK.
- [11] L. Z. a. P. W. Ruobing Liang, "Performance Evaluations of LoRa Wireless Communication in Building Environments," MDPI, 2020.
- [12] J. M. a. T. N. P. Neumann, "Indoor deployment of lowpower wide area networks (LPWAN): A LoRaWAN case study," in *IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications*, 2016.
- [13] P. P., "3glteinfo," 2020. [Online]. Available: <http://www.3glteinfo.com/loro/loro-architecture/>.
- [14] Dragino Technology Co., LTD., "dragino," 01 05 2018. [Online]. Available: https://wiki.dragino.com/index.php?title=Lora/GPS_Shield.
- [15] Dragino Technology Co., LTD., "dragino," 03 04 2018. [Online]. Available: https://www.dragino.com/downloads/downloads/UserManual/LG01_LoRa_Gateway_User_Manual.pdf.
- [16] Dragino Technology Co., LTD., "dragino," 18 10 2020. [Online]. Available: https://wiki.dragino.com/index.php?title=Lora_Shield.



ETAI 10: ARTIFICIAL INTELLIGENCE IN BIOMEDICINE

The Representation of Spoken Vowels in High Gamma Range of Cortical Activity

Daniela Janeva^{1,2}, Andrijana Kuhar², Lidija Olooska-Gagoska², Branislav Gerazov²

¹Macedonian Academy of Sciences and Arts, Skopje, Macedonia

²Faculty of Electrical Engineering and Information Technologies, UKIM, Skopje, Macedonia
danielajaneva@hotmail.com

Abstract - Spoken language is what distinguishes humans from other species. It is the easiest, yet the most complex action we perform. Precise and coordinated movement of multiple articulators is required in order to generate spoken sounds or phonemes. The organization of phonemes into complex semantic structures is currently a system of communication that people use to understand each other. For pronunciation of different phonemes, different articulators are involved in sound production. Speech motor control and its manifestation in the central nervous system is still under intense research. Although it is such an important topic, the spectro-temporal representation of speech in the cortex is still a mystery. Understanding how different phonemes are encoded is important for the further development of speech synthesis systems based on brain activity. The main focus of this paper is to create a feature space for the representation of vowels that will further contribute to the realization of a system mapping vowel articulation to characteristics of brain activity. This approach is a step towards a system for speech synthesis based on brain activity, which will potentially help people with speech impairments and contribute to research in the field of mental illnesses that impair speech ability. The created feature space of brain activity representing vowels is a set of extracted features from each electrode. For that purpose, the medium high gamma power of each spoken vowel from the ECoG electrodes is computed. A random forest classifier was trained on the extracted features and the model achieved 54% accuracy. Furthermore, feature importance has given us a perspective for the most important electrodes for encoding the vowels.

Keywords— speech; ECoG; cortex; consonant-vowel syllables; phonemes;

I. INTRODUCTION

The ability to articulate sounds in complex structures is what differentiates humankind from other species of the animal kingdom. Communication through spoken language is the ability to articulate sounds that are meaningful units that the listener can identify [1]. These speech units are called phonemes and according to their characteristics they are divided into consonants and vowels. The wide range of spoken sounds is a result of the flexible vocal tract configurations whose role is to filter sounds using the highly coordinated movement of the jaws, lips and tongue [2]. It is still a mystery how humans exert such exquisite control in terms of the highly variable movement possibilities. Each articulator has several extensive degrees of freedom, allowing for a large number of different realizations of speech sound sequences [3].

To maximize the clarity in vocal communication, a speaker presumably generates motor commands that differ greatly across distinct phonemes. In fluent speakers, the ventral (frontal) half of the sensory-motor cortex (vSMC) is thought to exert precise control of the vocal tract – control that has likely been optimized through evolution, learning, and extensive practice [4]. Understanding speech motor control requires analysis of the vSMC activity while speaking, which is difficult to achieve with noninvasive methods due to poor spatio-temporal resolution. On the other hand, direct cortical recordings through electrocorticography (ECoG) have a sufficient signal-to-noise ratio to resolve single-trial activity. Previous studies have shown that vSMC represents vocal tract articulators, i.e., the lips, tongue, jaw, and larynx, but the measuring the vocal tract's movements is hard. As an alternative, the vocal tract shape is directly reflected by the produced acoustics, especially vowel formants, which can be easily studied.

We use the ECoG signals to study the relationship between cortical activity and spoken vowels. Our goal is to determine the spoken vowels based on the cortical activity. Previous research implies that speech is not stored in discrete cortical areas, instead the production of phonemes and syllables is thought to arise from a coordinated motor pattern involving multiple articulator representations. We examined the degree to which cortical activity is differentiable of the production of the three different vowels. In order to understand the functional organization of the cortex in articulatory sensorimotor control we used the ECoG recordings from a dataset available online. The neural activity is recorded directly from the cortical surface in four human subjects. Each patient is given a task of uttering consonant-vowel (CV) syllables. Our goal is to differentiate the cortical activity of the three different vowels.

We propose a classification-based method based on high gamma power per spoken vowel. In order to be able to synthesize a speech segment we must first be able to understand how speech is encoded in the brain activity. Previous research shows that that speech is encoded in high gamma range of the cortical activity [5]. We explored the average power in the range in order to differentiate between the different spoken vowels. In this paper we introduce the dataset we are working with. In section II, we present the applied classification algorithm for differentiation between spoken consonants and vowels. In section III, we observe the results achieved with random forests model.

A. Data

In this work we used the Human ECoG speaking consonant vowel dataset [6]. The dataset consists of four native English-speaking participants that underwent an implantation of a high-density, subdural electrocorticographic (ECoG) array. The array contains 256 electrodes, ordered in a 16x16 matrix. The recordings of brain activity are from the language dominant hemisphere, i.e., the left hemisphere for right-handed subjects and the right for left-handed ones. Each participant was instructed to read out loud consonant-vowel syllables (CVs) composed of 18 consonants followed by one of 3 vowels. Each CV was produced between 15 and 100 times in one recording session. There are a different number of recording session per subject, making a total of 30 sessions and 10506 spoken syllables.

The timeseries and additional information are stored in a Neurodata Without Borders (NWB) file [7]. This is a data standard for neurophysiology, providing neuroscientists with a common standard to share, archive, use, and build analysis tools for neurophysiology data. Time series, annotations, channel position and additional information are provided in the dataset. We used the free software MNE for preprocessing the data and its visualization [8].

B. Brain anatomy

The speech part of the sensory motor cortex is the ventral portion of the sensory motor cortex. It is anatomically defined as the ventral portion of the precentral and postcentral gyri as well as the termination of the subcentral gyrus. The precentral gyrus is thought to be functionally divided a “premotor” and a “primary motor cortex. Cortical surface field potentials were recorded with ECoG array and a multichannel amplifier optically connected to a digital signal sensor. The sampling frequency of ECoG signals is 3052Hz. The total number of electrodes is 256. Three crucial electrodes in representing to different brain regions are observed

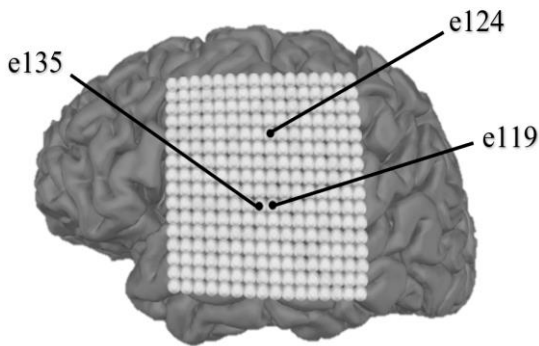


Fig 1. Cortical placement of electrodes observed in Fig 2. corresponding to language area of precentral a postcentral gyrus.

C. Methodology

Firstly, we removed the DC offset from signals of each electrode. For obtaining the high gamma power, we filtered the signals with a bandpass IIR filter in the range of 70-150Hz. The filter is obtained with the butterworth method. After filtering, we applied a Hilbert transformation and obtained the envelope representation of high gamma power as an absolute value of the real part of Hilbert’s transformation [9]. The representation of high gamma power in three different electrodes achieved through the mentioned procedure is shown in Fig. 2 in plots D, E and F. There are three electrode plots per electrode or total of nine plots. Red represents the signal in e124, blue represents e135 and green represents electrode e119 as shown on Fig 1.

The raw ECoG signals are filtered with butterworth IIR filter, and the spectrogram is computed with 50% overlapping Hamming windows. The spectrogram represents the time dependence of the frequency spectrum of high gamma range of the neural signal. The time duration of each window is 0.08 s or 256 samples. The spectrograms for syllables /ra/, /ri/, /ru/ in three different electrodes are shown in plots G, H, I in Fig 2.

The cortical recordings are aligned to acoustic onsets of the consonant to vowel transition in order to provide a common reference point, corresponding to the annotations of the time points provided in the dataset. We focus our analysis on the high gamma range, as previous studies have shown that speech production information is mostly encoded there, because of its correlation with multi-unit firing rates.

For speech production of different phonemes, different articulators are involved. This is expected to produce different cortical activity which would encode as a certain pattern among the ECoG electrodes. On Fig.2 the average high gamma power during 4 different phases is shown. The difference between the uttered consonant versus the vowel phase is mainly in the ventral (frontal) central gyrus as well as the postcentral gyrus which represent the sensory and motor cortex, accordingly.

The average gamma power in the pause before speaking, shown in the first image in Fig. 3, is mostly widespread in the frontal lobe, which is due to memory recall. Because the articulator movements are on the scale of tens of milliseconds, precious approaches have been unable to resolve temporal properties associated with individual articulatory representations. The hypothesis is that coordination of the multiple articulators required for speech production would manifest as spatial patterns of cortical activity. We explore the importance of electrodes for determining the class of a spoken vowel, based on the average gamma power computed of each vowel utterance. In section

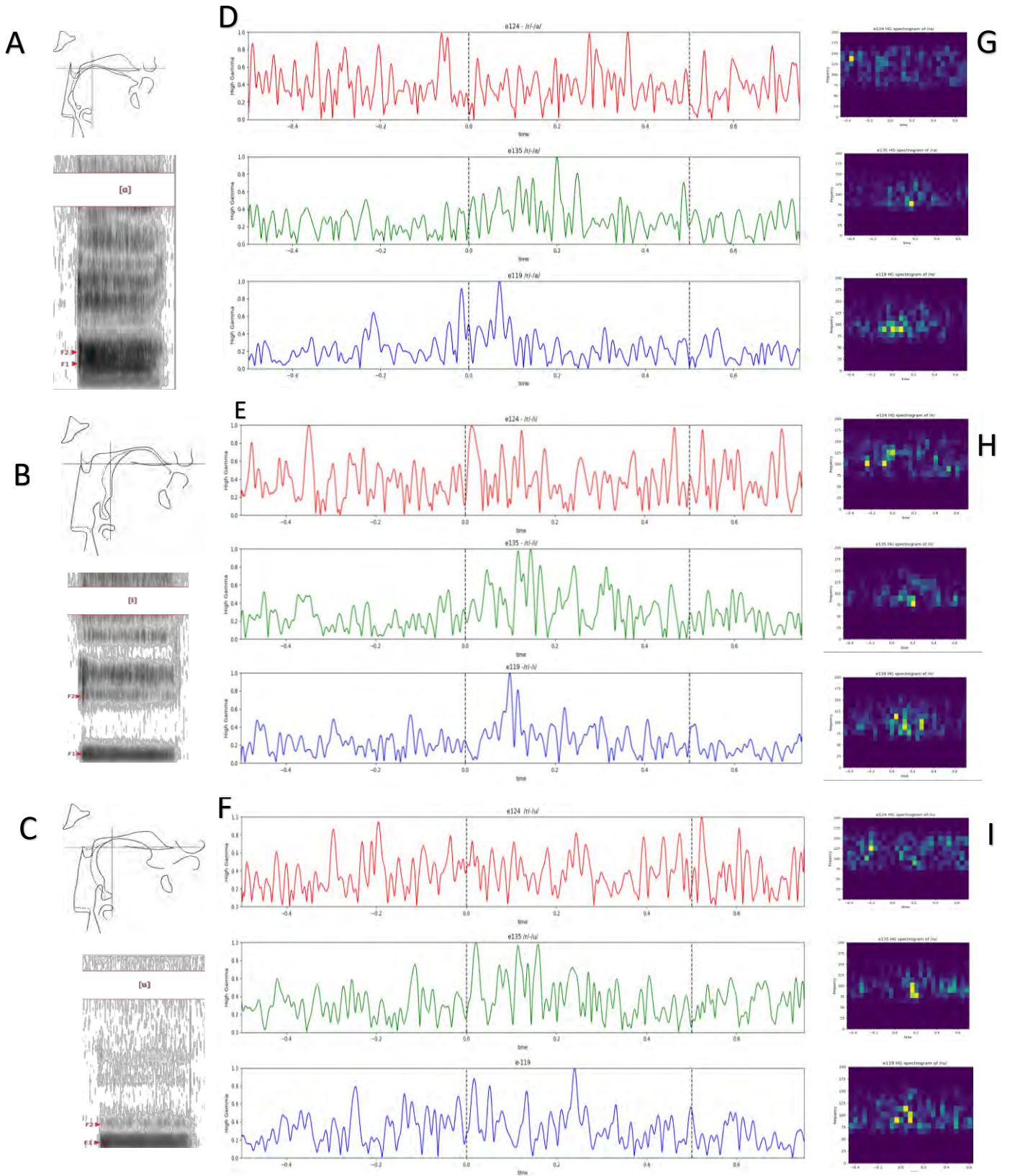


Fig 2. A, B and C is a representation of the articulatory position of the observed phonemes, and above the articulatory position each image shows the formant of the spoken vowels /a/, /i/, /u/ from top to bottom, respectively. D, E, F represents plots of normalized high gamma power in three electrodes for each spoken vowel. Time is measured from the acoustic consonant to vowel transition time. Red is e124, green e135 and blue is e119. G, H, I are spectrograms of high gamma range for each of the observed electrodes for each of the spoken vowels accordingly. Each spectrogram is aligned to the corresponding high gamma power.

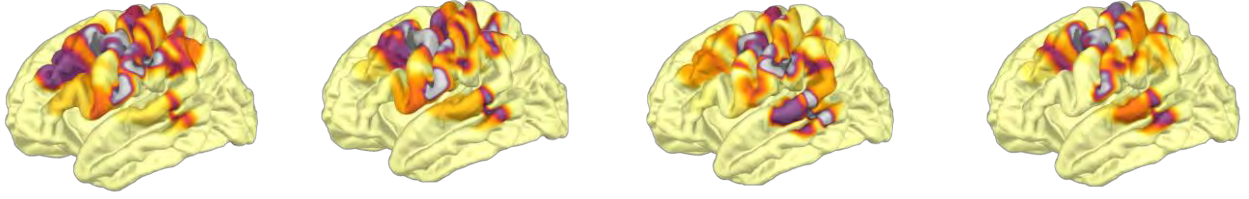


Fig 3. Approximative positioning estimation of lateral view of the left (language dominant hemisphere). Mean value of high gamma power for different phases. The first image represents the mean high gamma power of the pause before speaking, the second is the consonant phase, the third is the vowel phase and the last image is the mean gamma power of the pause after speaking.

II. CLASSIFICATION

After analysis of high gamma power over time, we computed its mean value of each vowel utterance per electrode. The duration of the signal segment is 1.75s. The brain signal onset is the acoustic consonant to vowel transition time. Due to the nature of the brain signal the observed signal segments are starting from 0.5s before the onset and 1.25s after the onset. For each segment the average high gamma power is computed.

The created feature space's dimension for three-class classification is specified by the number of different electrodes, which is 256. Furthermore, we computed 10032 values representing average high gamma power per vowel utterance for each of the electrodes. We used a random forests model for classification. The classification model was trained with 80% of the total dataset and the remaining 20 % were used for testing. For tuning the hyperparameters, we applied the grid search cross validation algorithm.

For gaining an insight into the differentiating regions of the brain during vowel utterance, we looked into the feature importance. This represents the significance of electrodes for classification of the average gamma power of /a/, /i/ and /u/.

III. RESULTS AND DISCUSSION

After applying the classification model to the testing data, we evaluated the results. For the three class classification based on the average gamma power of the brain signal per each vowel utterance we achieved an accuracy score of 54%. The results indicate that the computed values per electrode are somewhat significant for determining the vowel class.

Among the three spoken vowels /a/, /i/, /u/ represented as 0,1,2 correspondingly, on the confusion matrix in Fig 5, we can conclude that based on the average high gamma power, articulation of /a/ or 0 as shown on the image, can be differentiated with highest certainty because. This conclusion can be made due to the highest number of TP (true positive values).

After exploring the electrode importance we can conclude that for classifying the spoken vowels based on average high gamma power, most significant are electrodes numbered 103, 119, 124, 135, 153, and 222 as shown on Fig. 4. Electrode 103 has not been mentioned in previous research studies regarding this topic, which is a novelty worth exploring further. The electrode is placed on postcentral gyrus, which is the location of the primary sensory motor cortex.

We have seen different representations of the spoken vowels encoded in brain activity, and computed the average high gamma power. Differences between different spoken vowels are clearly evident.

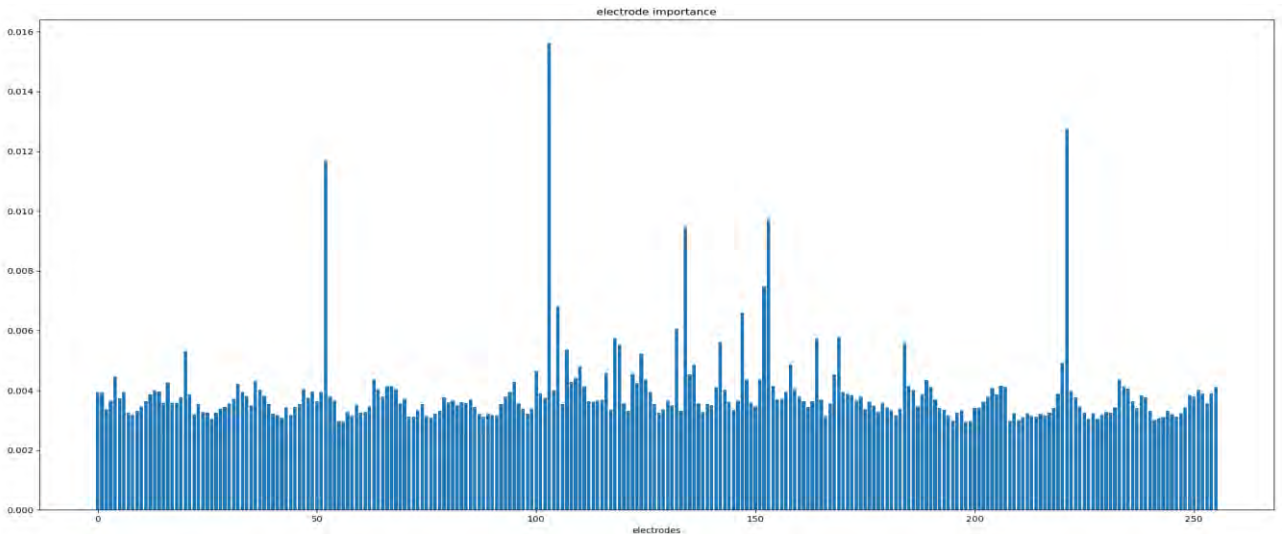


Fig 1: Feature importance per electrode for classification of the mean gamma power

IV. CONCLUSION

We can conclude that there is certain difference between differently spoken syllables in the cortical activity. However, the results observed on in Fig 5. indicate that high gamma power is not a reliable feature for differentiating between the syllable classes. Observing the electrode importance gives us an insight for further research about electrodes 103, 119, 124, 135, 153, and 222 specifically and determine their placement and cortical involvement in speech production.

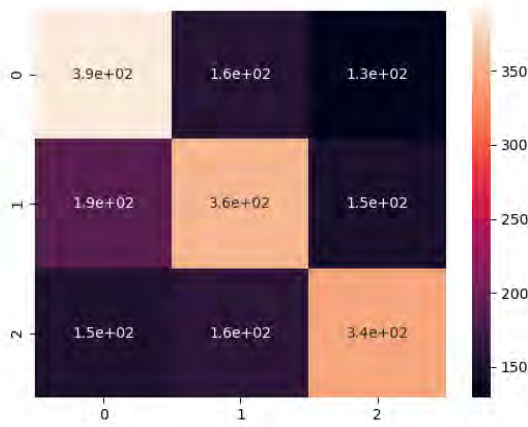


Fig 2 Confusion matrix /a/=0, /i/=1 /u/=2. Based on the average high gamma power, vowel /a/ is easiest to be determined.

V. REFERENCES

- [1] W. Levelt, *Speaking: From intention to Articulation*, MIT Press, 1993.
- [2] Gracco, V. Lofquist, A., "Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements," *J Neurosci.*, no. [PubMed: 7965062], p. 14:6585–6597., 1994;.
- [3] Bickerton, Derek., *Language and Species*, Chicago: Chicago Press, 1990.
- [4] A. Damasio, "Brain and Language," *Scientific American*, pp. 89-95, Sept. 1992.
- [5] Babajani-Feremi, Abbas, et al. "Variation in the topography of the speech production cortex verified by cortical stimulation and high gamma activity." *Neuroreport* 25.18 (2014): 1411.
- [6] Bouchard, Kristofer E.; Chang, Edward F (2019): Human ECoG speaking consonant-vowel syllables. figshare. Collection. <https://doi.org/10.6084/m9.figshare.c.4617263.v4>
- [7] Rübél, O., Tritt, A., Dichter, B., Braun, T., Cain, N., Clack, N., Davidson, T. J., Dougherty, M., Fillion-Robin, J.-C., Graddis, N., Grauer, M., Kiggins, J. T., Niu, L., Ozturk, D., Schroeder, W., Soltesz, I., Sommer, F. T., Svoboda, K., Ng, L., Frank, L. M., Bouchard, K. NWB:N 2.0: An Accessible Data Standard for Neurophysiology.
- [8] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*
- [9] Herff, Christian, et al. "Towards direct speech synthesis from ECoG: A pilot study." 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2016.
- [10] Bouchard, Kristofer E.; Chang, Edward F (2019): Human ECoG speaking consonant-vowel syllables. figshare. Collection. <https://doi.org/10.6084/m9.figshare.c.4617263.v4>

Scorpiano – A System for Automatic Music Transcription for Monophonic Piano Music

Bojan Sofronievski and Branislav Gerazov

*Faculty of Electrical Engineering and Information Technologies
Ss Cyril and Methodius University in Skopje, Macedonia
bojan.sof@hotmail.com, gerazov@feit.ukim.edu.mk*

Abstract—Music transcription is the process of transcribing music audio into music notation. It is a field in which the machines still cannot beat human performance. The main motivation for automatic music transcription is to make it possible for anyone playing a musical instrument, to be able to generate the music notes for a piece of music quickly and accurately. It does not matter if the person is a beginner and simply struggles to find the music score by searching, or an expert who heard a live jazz improvisation and would like to reproduce it without losing time doing manual transcription. We propose Scorpiano – a system that can automatically generate a music score for simple monophonic piano melody tracks using digital signal processing. The system integrates multiple digital audio processing methods: notes onset detection, tempo estimation, beat detection, pitch detection and finally generation of the music score. The system has proven to give good results for simple piano melodies, comparable to commercially available neural network based systems.

Index Terms—automatic music transcription, musical note recognition, pitch detection, onset detection, audio processing

I. INTRODUCTION

Automatic music transcription (AMT) is a process that automatically identifies the performed notes in a given melody track, with the goal of generating a music score. Besides automatic generation of music scores, AMT applications include interactive music systems and automated music tutors that teach playing an instrument [1], [2].

One of the main determinants for the difficulty of the task of AMT is whether the music to be transcribed is monophonic or polyphonic. Monophonic music simply means that the performer plays only one note at a time. There must not be sounds which overlap and this sounds are characterized by only one pitch. Contrary to monophonic music, polyphonic music consists of two or more simultaneous lines of independent melodies. This means that multiple notes are played at the same time and this sound is characterized with multiple pitches. AMT for polyphonic music is a non-trivial task and is a very active field of research [2]–[4].

There are few different approaches to AMT. One approach is based purely on digital signal processing and it is mainly used for monophonic music. These systems basically integrate methods for pitch and onset/beat detection. There are

mainly two approaches for pitch detection: *i)* the time-domain approach, which is primarily based on the autocorrelation function and its modifications [5], [6], *ii)* the frequency-domain approach, which is primarily based on the STFT (Short Time Fourier Transform) [7]. The onset detection algorithms are based on finding the peaks of a novelty function, i.e., a function whose peaks should coincide, within a tolerance margin, with note onset times [8].

The second approach is based on machine learning algorithms, mainly neural networks and these systems are the main choice for transcribing polyphonic music [9]. There are a few commercially available applications which follow this approach, such as AnthemScore¹ and Melody Scanner², and they advertise over 80% accuracy. However, neural networks in general require more processing power and a big set of data for training.

We propose the system for AMT of monophonic music, based on the digital signal processing approach, named Scorpiano. The piano is chosen to be the source instrument for transcription, because it produces sound by hitting the strings with hammers, giving the piano better frequency stability of the pitch, compared to other string instruments which produce sounds by plucking the strings. Our systems generates scores with high accuracy and fast transcription speeds, and is comparable to commercially available neural network based systems. The system also only has a small number of parameters that can be easily adjusted for different piano sources. Scorpiano is made available as free software.³

II. SYSTEM ARCHITECTURE

Fig. 1 shows Scorpiano's system architecture. The input to the system is an audio file of a monophonic piano recording. The system consists of five modules for: onset detection, tempo estimation, beat detection, pitch detection and music score generation. The system outputs an image file of the generated music score.

A. Onset Detection

Onset detection is the task of determining the starting times of musical notes in a music recording. Onsets correspond to a

¹<https://www.lunaverus.com>

²<https://melodyscanner.com>

³<https://gitlab.com/BojanSof/scorpiano>

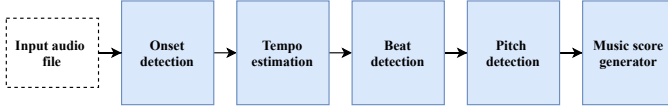


Fig. 1. System architecture.

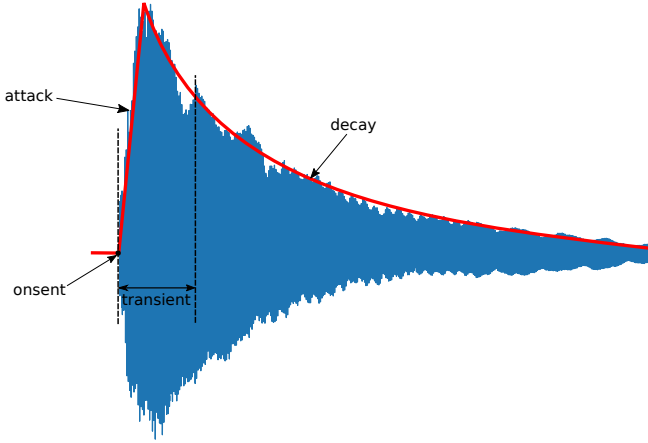


Fig. 2. The onset, attack, transient, and decay of an isolated note.

sudden increase of energy at the beginning of musical notes. There are a number of closely related concepts defined for each note realisation [10], as illustrated in Fig. 2:

- the *onset* of the note is the time moment when the transient starts,
- the *attack* of the note is the time interval during which the amplitude envelope increases, and
- the *transient*, in the case of the piano or other acoustic instruments, corresponds to the period during which the excitation is applied and then damped, leaving only the slow decay at the resonance frequency of the body of the instrument, and
- the *decay* is defined as the time interval in which the amplitude decreases gradually until the sound vanishes.

To detect a sudden increase of the energy, we compute the energy novelty function. The energy novelty function is a function that describes local changes in signal energy. To compute the novelty function, first we compute the local energy function of the signal x , using a bell-shaped window function w with length $2M + 1$, e.g. a Hann window, given with:

$$\begin{aligned}
 E[n] &= \sum_{m=-M}^{m=M} |x[n+m]w[m]|^2 \\
 &= \sum_{m \in \mathbb{Z}} |x[m]w[n-m]|^2 \\
 &= x^2[n] * w^2[n]
 \end{aligned}$$

To compute the changes of the energy, we take the first derivative of the local energy, which in the discrete case can be approximated by taking the difference between subsequent

energy values. Before taking the derivative, we apply logarithmic compression to the local energy function, taking into account the fact that human perception of sound intensity is logarithmic. Because we are interested only in energy increases, we apply a half-wave rectification of the derivative. The half-wave rectification function is given by:

$$r_{\text{half}}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

Thus, we obtain the local energy novelty function:

$$E_{\text{novelty}}[n] = r_{\text{half}}(E_{\gamma}[n+1] - E_{\gamma}[n])$$

where $E_{\gamma}[n]$ is the logarithmic compression of $E[n]$ for a positive constant γ :

$$E_{\gamma}[n] = \log(1 + \gamma \cdot E[n])$$

Fig. 3 shows the local energy, the energy novelty function and the marked maxima of the novelty function for a part of “Twinkle twinkle little star” melody, using a Hann window with a window length of 46 ms. Note that the energy novelty function is normalised by dividing it with its maximum value.

To detect onsets, we mark the local maxima of the energy novelty function. A simple method for finding local maxima is employed, keeping only maxima above a set amplitude threshold, and discarding maxima that are too close together. We set the amplitude threshold to be 0.1, and the time threshold to be 0.1 s.

B. Tempo Estimation

To find the correct timing of the musical notes and their beat duration, we must estimate the tempo of the music. For this purpose, we use the `beat.tempo` function from *librosa*⁴ library [11] to obtain the beats per minute value for the melody.

C. Beat Detection

After detecting the onsets of the musical notes and estimating the tempo, the beat duration of each note should be calculated. For every note, except the last, assuming there are no breaks, note duration can be computed by taking the difference between two subsequent onset moments. Then, the beat duration of the note is computed by multiplying the note duration with the tempo. To find the duration of the last note we can simply track backwards the normalised energy of the signal which represents the last note, and estimate the end of the note using a threshold.

D. Pitch Detection

According to the Fourier representation, each musical sound is a weighted sum of an infinite number of sinusoidal components. The frequency values of these components are integer multiples of the first one, called the fundamental frequency, denoted as F_0 , which is perceived as pitch.

We can apply pitch detection using the notes’ starting and ending times determined by their onset. We use the YIN

⁴<https://librosa.org>

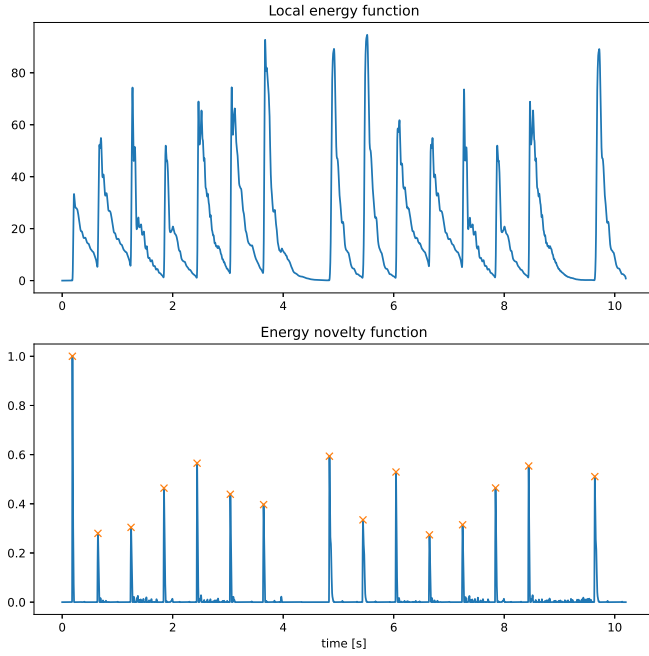


Fig. 3. Local energy function and energy novelty function.

algorithm for pitch detection [6]. It is an improved autocorrelation method for pitch detection, with a few modifications to minimize the error. The YIN algorithm can be sublimed in 6 steps:

- 1) calculate the autocorrelation function:

$$r_t[\tau] = \sum_{j=t+1}^{t+M-\tau} x[j]x[j+\tau]$$

- 2) calculate the difference function, over a window with size M samples (corresponding to 68 ms window length in our case):

$$d_t[\tau] = \sum_{j=1}^M (x[j] - x[j+\tau])^2$$

which can be expressed using the autocorrelation function as:

$$d_t[\tau] = r_t[0] + r_{t+\tau}[0] - 2r_t[\tau]$$

- 3) calculate the cumulative mean normalized difference function:

$$d'_t[\tau] = \begin{cases} 1, & \text{if } \tau = 0 \\ \frac{d_t[\tau]}{\frac{1}{\tau} \sum_{j=1}^{\tau} d_t[j]}, & \text{otherwise} \end{cases}$$

- 4) threshold – set the absolute threshold and choose the smallest value of τ that gives minimum of the cumulative mean normalized difference function, smaller than the threshold,
- 5) calculate parabolic interpolation – each local minimum of the cumulative mean normalized difference function

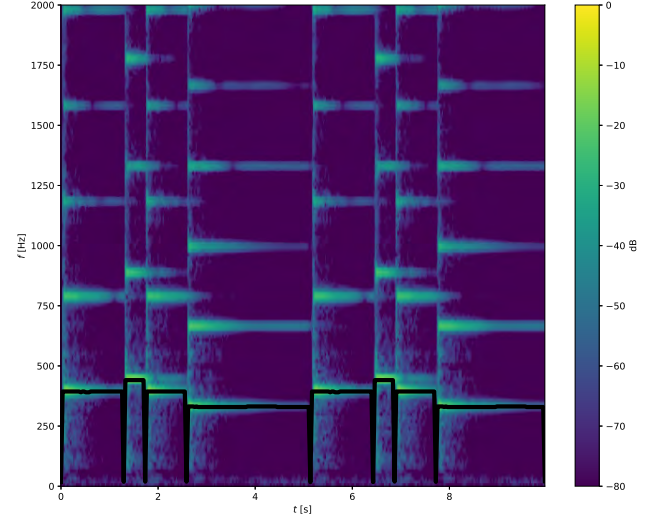


Fig. 4. Pitch contour obtained with the YIN algorithm for part of “Silent Night”

and its neighbors is fit by a parabola and the abscissa of the interpolated minimum is used as the period estimate.

- 6) calculate the best local estimate – repeat the period estimation in a shrinking time interval.

To estimate the fundamental frequency, the median of the estimated pitch periods for each window is computed, due to the fact that the median is more immune to single or few erroneous extreme values in the list of numbers, compared to the average of the numbers.

Figure 4 shows the pitch contour obtained with the YIN algorithm for part of the “Silent Night” melody. The peaks at beginning and ending of each note show why the fundamental frequency is computed using the median of the estimated pitch periods for each window.

E. Music Score Generator

Using the obtained beat durations and pitch of the musical notes, we can generate the music score. The library used for this step is *music21*⁵, with the *LilyPond* engine⁶. The time signature of the score must be given by the user. The output is an image file.

III. EXPERIMENTS

To evaluate the performance of the system, ten popular monophonic piano melodies are chosen.

In the first experiment, the system is tested with the melodies generated using the *MuseScore*⁷ application, which allows the generation of audio files from the music score for the melody.

In the second experiment, the system is tested with recordings of melodies played on a Miditech Midistart 3 MIDI keyboard, capable of modelling the dynamics of a piano.

⁵<https://web.mit.edu/music21>

⁶<http://lilypond.org>

⁷<https://musescore.org>

For simulating the piano sound, the *Addictive Keys*⁸ software was used and the audio output of the program was recorded using *Audacity*⁹. We will refer to this configuration as “digital piano”.

In the third experiment, the effect of the change of tempo on the system accuracy is evaluated with two melodies played on the digital piano, with three different tempos: slow, normal and fast, i.e. 80 bpm, 100 bpm and 120 bpm.

The performance of our system is compared with the commercial *AnthemScore* program. AnthemScore is an AMT software based on neural networks, advertised as being trained on millions of data samples.

To evaluate the performance of our system we define the following three errors:

- **note error rate** – the relative error for the number of original, n_{original} , and detected notes, n_{detected} :

$$\varepsilon_{\text{note}} = \frac{|n_{\text{detected}} - n_{\text{original}}|}{n_{\text{original}}} \times 100\%$$

- **pitch error rate** – the relative error for the number of incorrectly detected pitches, $p_{\text{incorrect}}$, excluding the extra detected notes:

$$\varepsilon_{\text{pitch}} = \frac{p_{\text{incorrect}}}{n_{\text{original}}} \times 100\%$$

- **beat error rate** – the relative error for the number of incorrectly detected beats, $b_{\text{incorrect}}$, excluding the extra detected notes:

$$\varepsilon_{\text{beat}} = \frac{b_{\text{incorrect}}}{n_{\text{original}}} \times 100\%$$

IV. RESULTS

A. MuseScore generated melodies

The number of notes of the original and automatically generated scores, and the number of incorrect pitches and beats excluding the extra detected notes, for each melody, are given in Table I.

The chosen parameter values give nearly perfect results for the chosen melodies. The incorrectly detected beats in the melodies mainly appear for the end note, as shown in Fig. 5 for “London Bridge”. The errors in the generated score are marked with red.

From a musical standpoint, the last note marked with red in the generated score, which is tied with the previous note and has the same pitch, means that the duration of the previous note is increased, so the total beat duration of the tie is $2\frac{3}{4}$. In fact, the system detects the last note as one note with duration $2\frac{3}{4}$, but the score generator renders it as a tie.

B. Digital piano recordings

Table II shows the error rates for generated scores using Scorpiano and AnthemScore for the selected melodies. The worst results using Scorpiano are obtained for the “Silent Night” melody. The main reason for this is because the tempo

TABLE I
COMPARISONS OF THE ORIGINAL AND GENERATED MUSIC SCORES FOR MELODIES GENERATED USING *MuseScore*.

Melody name	Number of notes (original)	Number of notes (detected)	Incorrect pitches	Incorrect beats
The Alphabet song	43	43	0	0
Auld Lang Syne	58	58	0	1
Cannon in D	46	46	0	0
Happy Birthday	25	25	0	1
Jingle Bells	49	49	0	1
London Bridge	24	25	0	0
Mary had a little lamb	25	25	0	1
Ode to Joy	62	62	0	0
Silent Night	47	47	0	0
Twinkle Twinkle Little Star	42	42	0	0



Fig. 5. “London Bridge” original score (top) and generated score (bottom).

of the melody is estimated to be 143 bpm, which is twice the actual of 70 bpm. Because of this, every note in the generated score has nearly twice the actual beat duration.

The errors in the scores generated with AnthemScore mostly comprise extra notes that are detected as played simultaneously with the correct note, forming a chord. The worst results are for “Canon in D” melody for which AnthemScore estimated the tempo incorrectly to be 113 bpm, versus the actual of 76 bpm. Despite of this, AnthemScore gives really good results for the digital piano recordings, showing the advantage of using neural networks for AMT. Fig. 6 shows a comparison of the errors in transcription for “Ode to Joy” generated with Scorpiano and AnthemScore. The notes colored red in the generated score with Scorpiano have incorrectly detected pitch. Their fundamental frequency was estimated to be two times smaller than the actual one. This type of error

⁸https://www.xlnaudio.com/products/addictive_keys

⁹<https://www.audacityteam.org>

TABLE II
ERROR RATES FOR THE AUTOMATICALLY GENERATED SCORES USING *Scorpiano* AND *AnthemScore* FOR MELODIES PLAYED WITH THE DIGITAL PIANO.

Melody name	Algorithm	Note error rate	Pitch error rate	Beat error rate
Auld Lang Syne	Scorpiano	3.45	1.72	0.00
	AnthemScore	1.72	0.00	0.00
Canon in D	Scorpiano	4.35	0.00	2.17
	AnthemScore	8.70	0.00	56.52
London Bridge	Scorpiano	0.00	0.00	16.67
	AnthemScore	0.00	0.00	16.67
Ode to Joy	Scorpiano	0.00	3.23	0.00
	AnthemScore	17.74	0.00	1.61
Silent Night	Scorpiano	70.21	0.00	100
	AnthemScore	8.51	0.00	0.00

TABLE III
ERROR RATES FOR GENERATED SCORES USING *Scorpiano* FOR MELODIES PLAYED ON A DIGITAL PIANO WITH THREE DIFFERENT TEMPOS.

Tempo	Note error rate	Pitch error rate	Beat error rate
Slow	0.00	0.00	1.02
Normal	0.00	0.00	1.02
Fast	2.38	0.00	0.00

in pitch detection, when the estimated fundamental frequency and the actual one form a ratio equal to a power of 2 is known as *octave error* [12].

C. Tempo effect

Table III shows the average error rates for generated scores with *Scorpiano*, for the melodies “Jingle Bells” and “Twinkle Twinkle Little Star” played on a digital piano with three different tempos: slow, normal and fast. From the results, it can be concluded that overall, the tempo has a small effect on the system performance, although there are cases, like with the “Silent Night” melody in the digital piano experiment, where the incorrect estimation of the tempo makes beat duration of the notes incorrect. This problem also happened with *AnthemScore* in the digital piano experiment for “Canon in D” melody.

D. Summary

Table IV summarises the average error rates for *Scorpiano* and *AnthemScore* for each experiment. Both systems score equally good for the MuseScore and Tempo experiment. *AnthemScore* shows slightly better results for the Digital piano experiment.

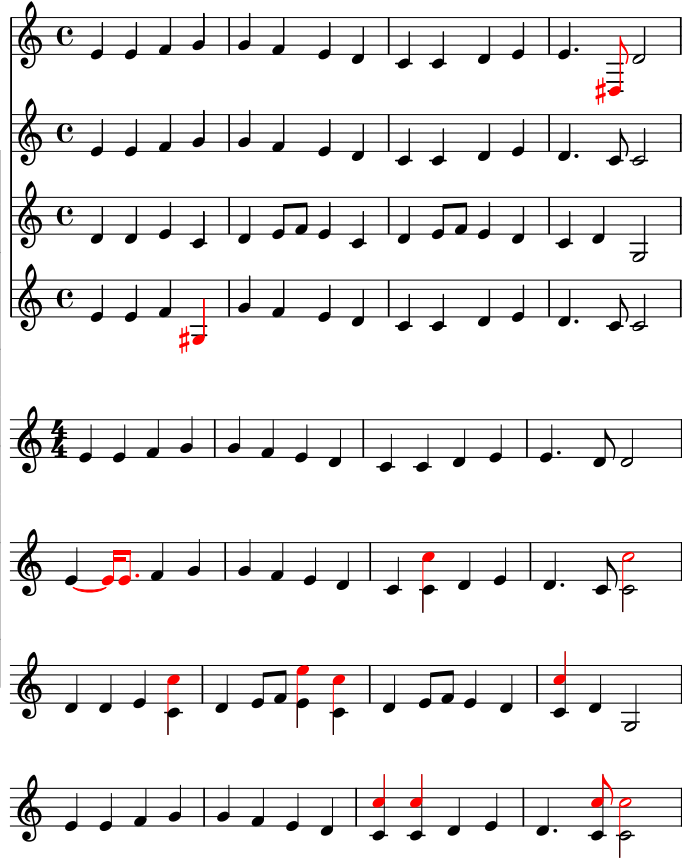


Fig. 6. *Scorpiano* (top) and *AnthemScore* (bottom) automatically generated scores for “Ode to Joy” played with a digital piano.

V. CONCLUSION

The problem of AMT for monophonic music can be effectively addressed using digital processing techniques. We propose a AMT system for monophonic piano music, that uses pure digital signal processing and has the advantage of being computationally inexpensive, fast, and does not need big training sets, whilst obtaining good results with low error rates, comparable to commercial neural network based systems.

In its current form, the system lacks detection of breaks and

TABLE IV
SUMMARY RESULTS.

Experiment	Algorithm	Note error rate	Pitch error rate	Beat error rate
MuseScore	Scorpiano	0.42	0.00	1.18
	AnthemScore	0.99	0.00	1.49
Digital piano	Scorpiano	8.20	0.89	12.89
	AnthemScore	5.94	0.00	8.35
Tempo	Scorpiano	0.79	0.00	1.02
	AnthemScore	1.70	0.00	1.19

time signatures. Sometimes mistakes can make the generated score look wrong, although most of the time the errors were obvious and could easily be corrected by human intervention.

The described system can be extended in the future. Post-processing the outputs of the modules can improve the performance of the system. For example, the estimated fundamental frequency for each note can be compared with the neighbouring notes to avoid octave errors. The system can also be modified to work with different musical instruments.

REFERENCES

- [1] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: Challenges and future directions," *Journal of Intelligent Information Systems*, vol. 41, 12 2013.
- [2] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic music transcription: An overview," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20–30, 2019.
- [3] G. Costantini, M. Todisco, and G. Saggio, "Automatic music transcription based on non-negative matrix factorization," 2010.
- [4] J. Sleep, "Automatic music transcription with convolutional neural networks using intuitive filter shapes," 10 2017.
- [5] P. S. Rao, S. Khoushikh, S. Ravishankar, R. A. Ananthkrishnan, and K. Balachandra, "A comparative study of various pitch detection algorithms," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, 2020, pp. 1–6.
- [6] A. de Cheveigné and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *Acoustical Society of America*, vol. 13, Apr. 2002.
- [7] B. Gerazov and Z. Ivanovski, "Building a basis for automatic melody extraction from macedonian rural folk music," 06 2010.
- [8] C. M. T. Rosão, "Onset detection in music signals," Ph.D. dissertation, University of Lisbon, 2012. [Online]. Available: <http://hdl.handle.net/10071/5991>
- [9] F. Saputra, U. G. Namyu, Vincent, D. Suhartono, and A. P. Gema, "Automatic piano sheet music transcription with machine learning," *Journal of Computer Science*, vol. 17, no. 3, pp. 178–187, Mar. 2021. [Online]. Available: <https://thescipub.com/abstract/jcssp.2021.178.187>
- [10] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, vol. 13, pp. 1035–1047, Sep. 2005.
- [11] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, vol. 8, 2015.
- [12] B. Kumaraswamy and P. G. Poonacha, "Octave error reduction in pitch detection algorithms using fourier series approximation method," *IETE Technical Review*, vol. 36, no. 3, pp. 293–302, 2019. [Online]. Available: <https://doi.org/10.1080/02564602.2018.1465859>
- [13] M. Miller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*, 1st ed. Springer Publishing Company, Incorporated, 2015.
- [14] Q. Gao, "Pitch detection based monophonic piano transcription," *Yankee*, vol. 7, no. 60, p. C4, 2015.

Facial Emotion Recognition Using Deep Learning

Gjorgji Smilevski, Tomislav Kartalov

Faculty of Electrical Engineering and Information Technologies

University of Ss. Cyril and Methodius

Skopje, Macedonia

smilevskigjorgji@gmail.com, kartalov@feit.ukim.edu.mk

Abstract—Emotions play a significant role in everyday communication and interaction between people. There are different modalities for predicting emotions, the most famous one being the face. Most of the work is done on predicting emotions in posed scenes, because the spontaneous facial emotions are much more unpredictable. In this paper, we propose an approach for prediction of the spontaneous facial emotions in pictures, using deep learning, more specifically, a VGG network. The FER+ dataset is used for the experiments, which consists of spontaneous images carrying multiple labels for each face image. The pictures were used as originals, without doing any additional modifications. The results show that using the proposed approach, the overall prediction accuracy can be achieved, which is close to the accuracy of the algorithms that use posed expressions. This brings a real life value and usability of this method.

Keywords — *Emotion Recognition, Facial Expression, Deep Learning, Convolutional Neural Network*

I. INTRODUCTION

Emotions are an inevitable portion of any inter-personal communication. They can be expressed in many different forms which may or may not be observed with the naked eye. Therefore, with the right tools, any indications preceding or following them can be subject to detection and recognition. There has been an increase in the need to detect a person's emotions in the past few years. The interest in human emotion recognition in various fields includes, but is not limited to: human-computer interface [1], animation [2], medicine [3], [4], security [5], [6], et cetera.

In psychology, emotion is often defined as a complex state of feeling that results in physical and psychological changes that influence thoughts and behavior. Emotionality is associated with a range of psychological phenomena, including temperament, personality, mood, and motivation. The major theories of motivation can be grouped into three main categories: physiological, neurological, and cognitive: [7]

- Physiological theories suggest that responses within the body are responsible for emotions.
- Neurological theories propose that activity within the brain leads to emotional responses.

- Cognitive theories argue that thoughts and other mental activity play an essential role in forming emotions.

Emotion recognition can be performed using different modalities, such as face, speech, EEG, and even handwriting. Among these features, facial expressions are one of the most popular, if not the most popular for emotion recognition, due to a number of reasons; they are visible, they contain many useful features, and it is easier to collect a large dataset of faces (than other means for human recognition) [8]. One of the pioneer works by Paul Ekman [9] identified 6 emotions that are universal across different cultures: anger, disgust, happiness, fear, sadness, and surprise, to which he later added contempt. Furthermore, Ekman developed the Facial Action Coding System (FACS) [10], which became the standard scheme for facial expression research. Facial expression analysis can thus be conducted by analyzing facial action units for each of the facial parts (eyes, nose, mouth corners, etc.), and map them into FACS codes [11]. Unfortunately, FACS coding requires professionally trained coders to annotate, and there are very few existing data sets that are available for learning FACS based facial expressions, in particular for unconstrained real-world images. With the latest advances in machine learning, it is more and more popular to recognize facial expressions directly from input images [12].

In this paper, a deep learning algorithm is proposed, for spontaneous facial expression prediction, based on VGG network and using the FER+ image dataset, manually annotated for experiments, and allowing multiple emotion flags to be set on each image. The rest of the paper is organized as follows: first, the Section II is discussing the elicitation methods, and the factual differences between posed, induced and spontaneous emotions, the last being the main focus of the work reported in this paper. Then, the Section III continues the review of the related work from the introduction section, this time in more narrow and specific fashion, targeting this particular field of research interests, and the Section IV gives brief description of the dataset used in the experiments. Section V provides details about the proposed approach, and at the end, Sections VI and VII present the achieved results and conclusions.

II. ELICITATION METHODS

An important choice to make in gathering data for emotion recognition databases is how to bring out different emotions in the participants. This is the reason why facial emotion databases are divided into three main categories: [13]

A. Posed emotions

Emotions acted out based on conjecture or with the guidance from actors or professionals are called posed expressions [14]. Most facial emotion databases, especially the early ones i.e. Banse-Scherer [15], Cohn-Kanade [16], and Chen [17], consist purely of posed facial expressions, as it is the easiest to gather. However, they also are the least representative of real world authentic emotions as forced emotions are often over-exaggerated or missing subtle details. Due to this, human expression analysis models created through the use of posed databases often have very poor results with real world data [18], [19].

B. Induced emotions

This method of elicitation displays more genuine emotions as the participants usually interact with other individuals or are subject to audiovisual media in order to invoke real emotions. Induced emotion databases have become more common in recent years due to the limitations of posed expressions. The performance of the models in real life is greatly improved, since they are not hindered by overemphasized and fake expressions, making them more natural. There are several databases that deal with audiovisual emotion elicitation like the SD [20], UT DALLAS [21] and SMIC [22], and some that deal with human to human interaction like the ISL meeting corpus [23], AAI [24] and CSC corpus [25]. Databases produced by observing two-way human-computer interaction on the other hand are a lot less common. The best representatives are the AIBO database [26], where children are trying to give commands to a Sony AIBO robot, and SAL [27], in which adults interact with an artificial chat-bot. Even though induced databases are much better than the posed ones, they still have some problems with truthfulness. Since the emotions are often invoked in a lab setting with the supervision of authoritative figures, the subjects might subconsciously keep their expressions in check [14, 19].

C. Spontaneous emotions

These datasets are considered to be the closest to actual real-life scenarios. However, since true emotion can only be observed when the person is not aware of being recorded [19], they are difficult to collect and label. The acquisition of data is usually in conflict with privacy or ethics, whereas the labeling has to be done manually and the true emotion has to be guessed by an annotator [14]. This arduous task is both time-consuming and erroneous [18], [28], having a sharp contrast with posed and induced datasets, where labels are either predefined or can be derived from the elicitation content. With that being said, there still exist a few databases out there that consist of data extracted from movies [29], [30], YouTube videos [31], or television series

[32]. However, these databases have inherently fewer samples in them than their posed and induced counterparts.

In this work, we will be using the FER+ dataset [33], which is part of the spontaneous emotion datasets.

III. RELATED WORK

With the great success of deep learning, and more specifically convolutional neural networks (CNNs) for image classification and other vision problems, several groups developed deep learning-based models for facial expression recognition [12]. To name some of the promising works, Khorrani in [34] showed that CNNs can achieve a high accuracy in emotion recognition and used a zero-bias CNN on the extended Cohn-Kanade dataset (CK+) and the Toronto Face Dataset (TFD) to achieve state-of-the-art results. Tang in [35] used the primal objective of an SVM as the loss function for training a CNN. He achieved the best accuracy of 71.162% in the competition that released the original FER2013 dataset. Emad et al in [12] gave four different schemes working on the FER+ dataset: majority voting, multi-label learning, probabilistic label drawing and cross entropy loss. They used a custom VGG13 network, while also applying affine transforms to improve the robustness of the model against translation, rotation and scaling, achieving accuracy of 83.852% for majority voting. Shervin in [8], proposed an end-to-end deep learning framework, based on attentional convolutional network. They applied their method on four different datasets: FER2013 with 70.02% accuracy, JAFFE with 92.8% accuracy, CK+ with 98.0% accuracy and FERG with 99.3% accuracy. Henrique et al [36] used wide ensemble-based convolutional neural networks to obtain an accuracy of 87.153% on FER+, by fine-tuning a network trained on the AffectNet dataset, where it achieved accuracy of 59.3%. Debin et al in [37] developed frame attention networks for facial expression recognition in videos and achieved a state-of-the-art accuracy of 99.69% on the CK+ dataset, as well as 51.18% accuracy on the AFEW 8.0 dataset.

IV. DATA

In this work, we used the FER+ dataset. It contains 35710 grayscale facial images, and it is separated in 3 subsets: training data – 28558 images, validation data – 3579 images and test data – 3573 images, which is a ratio of 80:10:10. It contains the 7 basic emotions by Ekman, joined by neutral, meaning there are 8 different classes: neutral, happiness, surprise, sadness, anger, disgust, fear and contempt. This dataset is based on the FER2013 dataset, which was created by web crawling face images with emotion related keywords, followed by cropping the regions that contain only faces, with a size of 48x48, and converting them to grayscale. The images were labeled by human annotators, but the label accuracy is not very high [35]. That is why [12] took the same images, and decided to re-label them with crowd sourcing. Instead of every image having only one class in FER2013, in FER+ every image was labelled by 10 different people, and there is a distribution of emotions. The difference between the image labels can be seen in Fig. 1.



Fig. 1. Difference in labels between FER2013 (top label) and FER+ (bottom label).

As discussed earlier, when labeling spontaneous datasets, the emotion has to be guessed by the annotator. There is a factor of subjectivity involved, which should be reduced by having 10 people label each image, as can be seen in Fig. 2 [12]:

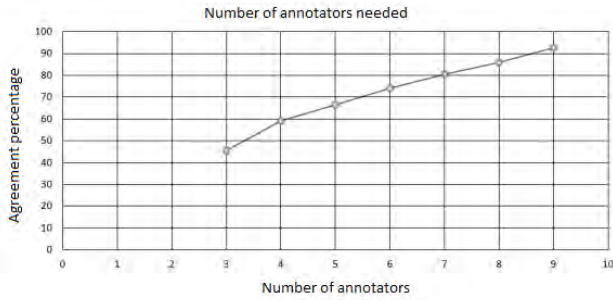


Fig. 2. Annotator count versus quality.

V. THE PROPOSED APPROACH

Before we begun training the model, we analyzed the images. The emotion shown in every image was classified by 10 people. However, for some of the images, the votes were diverse. For example, for the face shown in Fig. 3, four annotators voted for neutral, one for sadness, three for anger and two annotators voted for disgust. We realized that these images present an issue. If people cannot decide what emotion is shown on the image, it cannot be expected from the model to learn correctly. This is why we came to a decision to remove every image that does not have at least 6 votes for a single emotion.



Fig. 3. Example of an image that contains a face with inconclusive emotion.

TABLE I. THE USED DATASET

Emotion	Datasets		
	Train	Validation	Test
Neutral	7485	1062	931
Happiness	7080	845	874
Surprise	2748	381	344
Sadness	2388	277	297
Anger	1648	235	211
Disgust	52	16	9
Fear	346	41	50
Contempt	76	9	8
Total	21823	2866	2724

In the images that remained, the emotion with the most votes was made the only true class for the image. This reduced the dataset by 23.24%, leaving a total of 27413 images. However, we can expect that this selection is going to give better results and help create a more robust model. A more detailed representation of the dataset and division of images in classes can be seen in Table I.

As can be seen from the numbers, there is a similar class distribution for the three different subsets, which is important for correctly analyzing the results of the training process. Images that resemble neutral and happiness are most common, and it is to be expected that the model will learn those best, while it will make more mistakes when trying to predict disgust, fear and contempt, as they are represented in very few images.

In order to solve the problem, we tried multiple CNN models. The best result was achieved using a custom version of the VGG network [38]. The architecture of the network can be seen in Fig. 4. Before the training process is started, every image undergoes normalization for pixel values between 0 and 1, followed by a random shuffle of the images.

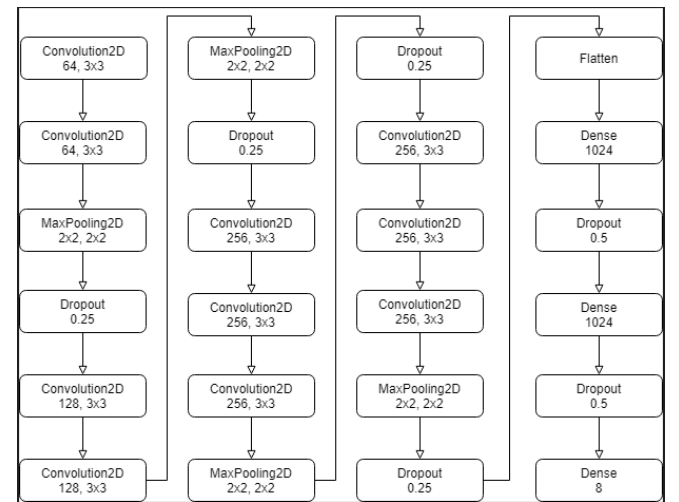


Fig. 4. Network Architecture

VI. RESULTS

The training process consisted of 100 epochs. The optimizer used is Adam, with a learning rate of 0.0001. The accuracy and the loss of the model can be seen in Fig. 5 and Fig. 6, respectively.

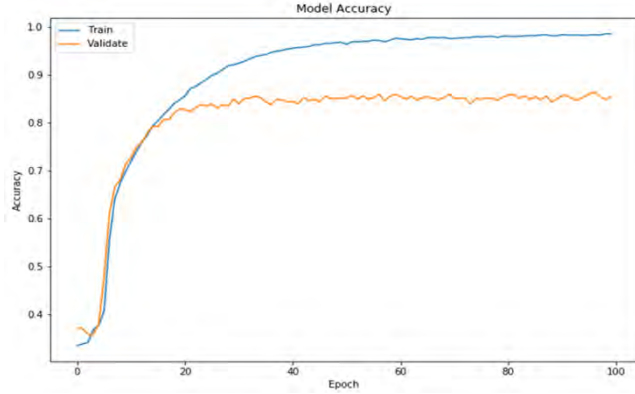


Fig. 5. Model Accuracy

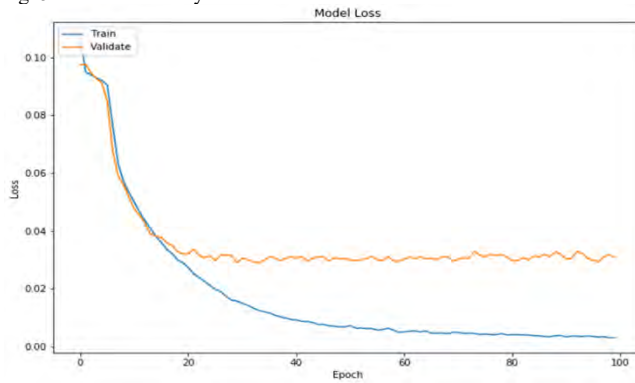


Fig. 6. Model Loss

The accuracy of the validation set reaches 86.390%, closely followed by 84.948% accuracy for the test set. This accuracy comes close to the 90% mark, which is the accuracy that models trained on posed pictures almost always reach [8].

In order to better analyze the results, we will use a confusion matrix without normalization, Fig. 7, and with normalization, Fig. 8.

As expected, the emotions that were most common in the dataset, neutral and happiness, joined by surprise, are also the ones that have the best prediction accuracy. Anger got correctly predicted in 78% of the cases, even though the number of images representing it was not as high. Furthermore, the emotions that were least represented, disgust, fear and contempt, have the worst prediction accuracy, however a difference in accuracy between fear and the other two can be observed, as fear had a tiny bit more images representing it. An interesting occurrence is that 34% of images that represented sadness were predicted as neutral. We put that down to similarities in the expressions shown when a face is neutral, and when it is sad.

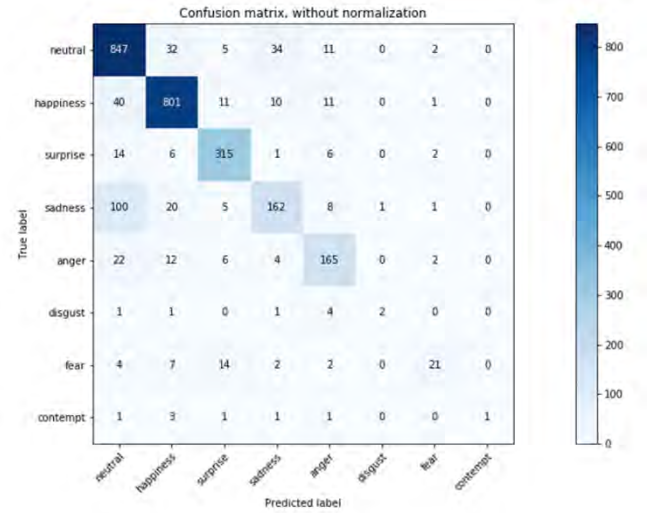


Fig. 7. Confusion matrix without normalization

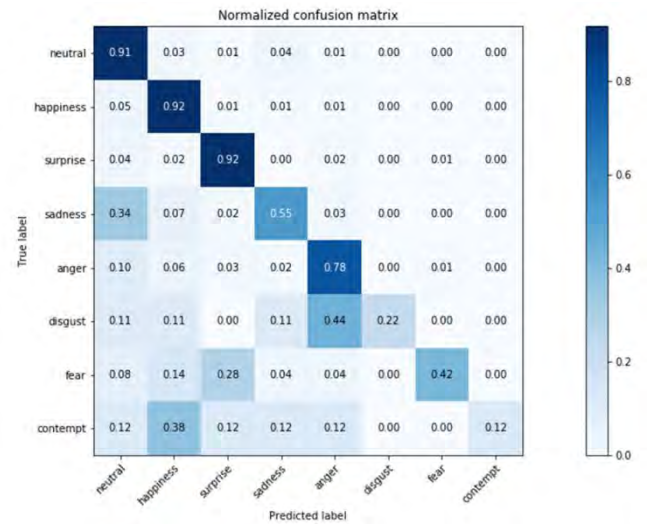


Fig. 8. Normalized confusion matrix

VII. CONCLUSION

In this work, we tried to analyze and predict facial emotion expressions by using a deep convolutional neural network. We see that working with datasets that contain spontaneous emotions comes with many difficulties, biggest one being the quality of the labels. However, using the FER+ dataset, we achieve results close to what models trained on datasets containing posed emotions achieve. We believe that with a better distribution of the emotion classes, a model that can work in real life situations can be created.

REFERENCES

- [1] Cowie, Roddy, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G. Taylor. "Emotion recognition in human-computer interaction." *IEEE Signal processing magazine* 18, no. 1: 32-80, 2001.
- [2] Aneja, Deepali, Alex Colburn, Gary Faigin, Linda Shapiro, and Barbara Mones. "Modeling stylized character expressions via deep

- learning." In Asian Conference on Computer Vision, pp. 136-153. Springer, Cham, 2016.
- [3] Edwards, Jane, Henry J. Jackson, and Philippa E. Pattison. "Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review." *Clinical psychology review* 22.6: 789-832, 2002.
 - [4] Chu, Hui-Chuan, William Wei-Jen Tsai, Min-Ju Liao, and YuhMin Chen. "Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning." *Soft Computing*: 1-27, 2017.
 - [5] Clavel, Chlo, Ioana Vasilescu, Laurence Devillers, Gal Richard, and Thibaut Ehrette. "Fear-type emotion recognition for future audio-based surveillance systems." *Speech Communication* 50, no. 6: 487-503, 2008.
 - [6] Saste, Sonali T., and S. M. Jagdale. "Emotion recognition from speech using MFCC and DWT for security system." In *Electronics, Communication and Aerospace Technology (ICECA)*, 2017 International conference of, vol. 1, pp. 701-704. IEEE, 2017.
 - [7] <https://www.verywellmind.com/theories-of-emotion-2795717>
 - [8] Shervin Minaee, Amirali Abdolrashidi. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Expedia Group. University of California, Riverside
 - [9] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.
 - [10] P. Ekman and W. V. Friesen. Facial action coding system. 1977.
 - [11] Y.-I. Tian, T. Kanade, and J. F. Cohn. Recognizing action units for facial expression analysis. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 23(2):97-115, 2001.
 - [12] Emad Barsoum, Cha Zhang, Cristian Canton Ferrer and Zhengyou Zhang. Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution. Microsoft Research. One Microsoft Way, Redmond, WA 98052
 - [13] Wu C-H, Lin J-C, Wei W-L. Survey on audiovisual emotion recognition: Databases, features, and data fusion strategies. *APSIPA Transactions on Signal and Information Processing*. 2014;3:e12.
 - [14] Sebe N, Cohen I, Gevers T, Huang TS. Multimodal approaches for emotion recognition: A survey. In: *Electronic Imaging 2005*; International Society for Optics and Photonics; 2005. pp. 56-67.
 - [15] Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*. 1996;70(3):614.
 - [16] Kanade T, Cohn JF, Tian Y. Comprehensive database for facial expression analysis. In: *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000; IEEE; 2000. pp. 46-53.
 - [17] Lawrence Shao-Hsien Chen. Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction [PhD thesis]. Citeseer; 2000.
 - [18] Jaimes A, Sebe N. Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*. 2007;108(1):116-134.
 - [19] Zeng Z, Pantic M, Roisman GI, Huang TS. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2009;31(1):39-58.
 - [20] Sebe N, Lew MS, Sun Y, Cohen I, Gevers T, Huang TS. Authentic facial expression analysis. *Image and Vision Computing*. 2007;25(12):1856-1863.
 - [21] O'Toole AJ, Harms J, Snow SL, Hurst DR, Pappas MR, Ayyad JH, Abdi H. A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005;27(5):812-816.
 - [22] Pfister T, Li X, Zhao G, Pietikäinen M. Recognising spontaneous facial micro-expressions. In: *IEEE International Conference on Computer Vision (ICCV)*, 2011; IEEE; 2011. pp. 1449-1456.
 - [23] Burger S, MacLaren V, Yu H. The ISL meeting corpus: The impact of meeting type on speech style. In: *INTERSPEECH*, Denver, Colorado, USA; 2002.
 - [24] Roisman GI, Tsai JL, Chiang K-HS. The emotional integration of childhood experience: Physiological, facial expressive, and self-reported emotional response during the adult attachment interview. *Developmental Psychology*. 2004;40(5):776.
 - [25] Hirschberg J, Benus S, Brenier JM, Enos F, Friedman S, Gilman S, Girand C, Graciarena M, Kathol A, Michaelis L, et al. Distinguishing deceptive from non-deceptive speech. In: *Interspeech*; 2005. pp. 1833-1836.
 - [26] Batliner A, Hacker C, Steidl S, Nöth E, D'Arcy S, Russell MJ, Wong M. "You stupid tin box"-children interacting with the AIBO robot: A cross-linguistic emotional speech corpus. In: *LREC*, Lisbon, Portugal; 2004.
 - [27] Athanaselis T, Bakamidis S, Dologlou I, Cowie R, Douglas-Cowie E, Cox C. ASR for emotional speech: Clarifying the issues and enhancing performance. *Neural Networks*. 2005;18(4):437-444.
 - [28] Kirouac G, Dore FY. Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. *Journal of Nonverbal Behavior*. 1985;9(1):3-7.
 - [29] Dhall A, Goecke R, Lucey S, Gedeon T. Acted facial expressions in the wild database. Australian National University, Canberra. Technical Report TR-CS-11, 2; 2011.
 - [30] Dhall A, Lucey S, Joshi J, Gedeon T. Collecting Large, Richly Annotated Facial-Expression Databases from Movies, IEEE MultiMedia, 2012;19(3):34-41.
 - [31] Rosas VP, Mihalcea R, Morency L-P. Multimodal sentiment analysis of Spanish online videos. *IEEE Intelligent Systems*. 2013;28(3):38-45.
 - [32] Douglas-Cowie E, Campbell N, Cowie R, Roach P. Emotional speech: Towards a new generation of databases. *Speech Communication*. 2003;40(1):33-60.
 - [33] Fer+ emotion label. <https://github.com/Microsoft/FERPlus>, 2016.
 - [34] Khorrami, Pooya, Thomas Paine, and Thomas Huang. "Do deep neural networks learn facial action units when doing expression recognition?." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2015.
 - [35] Challenges in Representation Learning: A report on three machine learning contests. URL [http:// https://arxiv.org/pdf/1307.0414v1.pdf](http://https://arxiv.org/pdf/1307.0414v1.pdf)
 - [36] Henrique Siqueira, Sven Magg and Stefan Wermter. Efficient Facial Feature Learning with Wide Ensemble-based Convolutional Neural Networks. Department of Informatics, University of Hamburg. Vogt-Koelln-Str. 30, 22527 Hamburg, Germany, 2020.
 - [37] Debin Meng, Xiaojiang Peng , Kai Wang, Yu Qiao. Frame Attention Networks for Facial Expression Recognition in Videos. University of Chinese Academy of Sciences, Beijing, China, 2019.
 - [38] [38] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

Automatic Composition of Lyrics and Music for a Song in Macedonian using Deep Learning

Angela Najdoska, Emilija Kotevska, Tamara Markachevikj, Hristijan Gjoreski
Faculty of Electrical Engineering and Information Technologies
Ss. Cyril and Methodius University in Skopje, Macedonia
angela.najdoska17@gmail.com, emilijakotevska@gmail.com, markacevic@gmail.com

Abstract - This paper presents an approach to develop deep learning models to automatically generate music and lyrics for a song in Macedonian language. The approach consists of two steps: (i) a deep learning model to generate text (lyrics) for songs using Gated Recurrent Unit approach, and (ii) a deep learning model for music generation for the lyrics of the song generated in the previous step. The model for lyrics generation is based on Gated Recurrent Units algorithm, and the model for music generation is based on a pre-trained Recurrent Neural Network and Variational Autoencoder model, which generates sequences of music that are later combined to produce the final music for the song. The music generation model interpolates between two-note sequences and thus completes the process of creating the music. The results show that our model generates understandable lyrics for songs and decent quality music for the same. The result songs are unique, new, and after some melodical post-processing, can be potentially used to help the process of song creation.

Keywords— Text generation; Lyrics generation; Macedonian song; GRU; Deep learning; RNN; Machine learning.

I. INTRODUCTION

Music is the art of arranging sounds in order to produce a composition throughout the elements of melody, harmony and rhythm. General definitions of music include common elements such as pitch, rhythm, dynamics, timbre and texture. In many cultures, music is an important part of people's way of life, as it plays a key role in social and cultural activities. The music industry includes the individuals who create new songs and musical pieces, individuals who perform and record music, who organize concert tours and who sell recordings, sheet music and scores to customers.

Machine Learning, as an important segment of Artificial Intelligence, is increasingly present in our daily lives. Text generation with the help of Machine Learning and Deep Learning is popular in every industry, especially in mobile phones, applications, data-science and even journalism [1]. Every day, we come across text generation: iMessage autocomplete, Google search, Gmail smart typing are just a few examples. Using automatic text generation, we can make many processes easier, for example: composing a text (lyrics) for a song and producing a music

for the same. Writing a song and a melody for the same is a challenging process that requires inspiration and artistic way of thinking. Having a system that can automatically produce several dozens of lyrics and music for the same, can be of great importance and help. Even if the composer and the music producer have to manually go through them, and analyze and; these artificially generated songs can still be used as a basis and inspiration and this way making the whole process easier.

In this paper we propose a Deep learning approach to generating lyrics for songs in Macedonian language. Additionally, we propose a Deep Learning approach for composing music for the same songs. Even though, there have been approaches for generating poems, news articles in Macedonian language [2] to the best of our knowledge, this is a first study that tackles the problem of generating songs and music in Macedonian language using Deep Learning.

II. DATASET AND PREPROCESSING

For the purposes of this study, we created our own dataset - because there is no publicly available repository and database for the same. We used the lyrics of almost all the songs performed by one of the most famous Macedonian singers - Toshe Proeski. We used 64 songs, each with a different number of verses, and a different metric. The total number of words in the database is 10731. Next, we performed data preprocessing and preparation for the model. This included, parsing, tokenization and equalizing the number of tokens in each verse. Then, we removed the punctuation marks and converted all letters to lowercase. Then, we transformed the corpus so that each verse of the database represents one element of the corpus. Finally, we performed padding, so that all the sequences have the same length. The data and the code are available online [3].

III. LYRICS GENERATION MODEL

To generate the lyrics of a song, we used a sequential type of Deep Learning approach, in particular Recurrent Neural Network (RNN). The model is shown in Figure 1, and consists of:

- Embedding Layer with number of inputs equal to the number of different words in the corpus (the number of tokens in sequence decreased by one) and 5 outputs.

- Two hidden layers of Gated Recurrent Unit (GRU) [4] which is an RNN variant that uses multiplicative connections which allow the current input character to determine the transition matrix from one hidden state vector to the next.
- Dropout layer is used to reduce overfitting [5]. This technique works by randomly dropping units (along with their connections) from the neural network during training. This prevents units from co-adapting too much. The number of neurons in both layers is 80 and 120 respectively.
- The last layer is a Dense layer which has outputs equal to the number of different words in the corpus and “softmax” activation function.

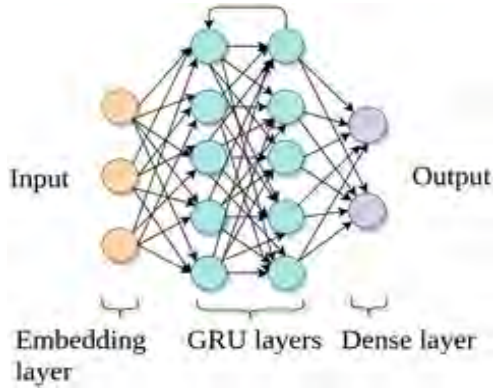


Figure 1: Model architecture: input, embedding layer, GRU layers, Dense layer and output.

We used Categorical cross-entropy as a cost function and optimize it with Adam optimizer (Adaptive Moment Estimation), which is method that computes adaptive learning rates for each parameter [6]. It is an algorithm for first-order gradient-based optimization of stochastic objective functions. For monitoring the accuracy of the model, in the training process, we use accuracy metrics.

We trained the model for 400 epochs and used EarlyStop callback in order to prevent overfitting. From the existing corpus, we make a new corpus whose elements are words, in order to choose a random word that we will use as a seed on which the whole song will be generated.

Finally, the text for the song was generated verse by verse. The first generated word use the one-to-one principal, while all subsequent words in the verse are generated by many-to-one principal. Moreover, the last word from the previous verse is the seed for the next verse. Arbitrarily, we chose the song to consist of 4 stanzas with 4 verses each, and each verse to contain 6 words.

IV. MUSIC GENERATION MODEL

After the lyrics was generated, the next step was to generate a music for the same. For music generation, we used the Magenta open-source Python library [7]. Despite Magenta, we also used the `node_seg` library which allows abstract representation of a series of notes, each with different pitches, instruments and strike velocities, much

like the MIDI standard (Musical Instruments Digital Interface).

In the Magenta Library we used a pre-trained Melody RNN model - a model of the Recurrent Neural Network for generating a sequence of notes based on an initially given sequence. We used this model to generate two different music sequences. First, we created the initial sequences that are actually part of two songs by TosheProeski: "Ledena" and "Po tebe". This is shown in Figure 2. Then, in the Melody RNN model we set certain parameters such as temperature (degree of randomness of note generation), number of steps (the length of the generated sequence depends on it), and tempo. This is shown in Figure 3.

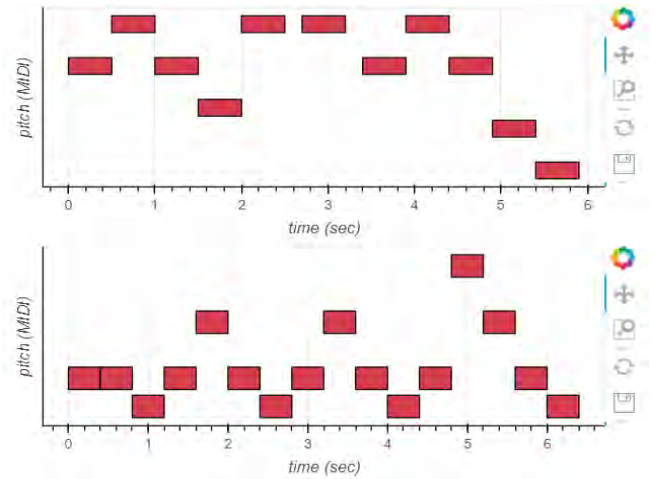


Figure 2: Initial sequences-parts of Toshe Proeski's songs.

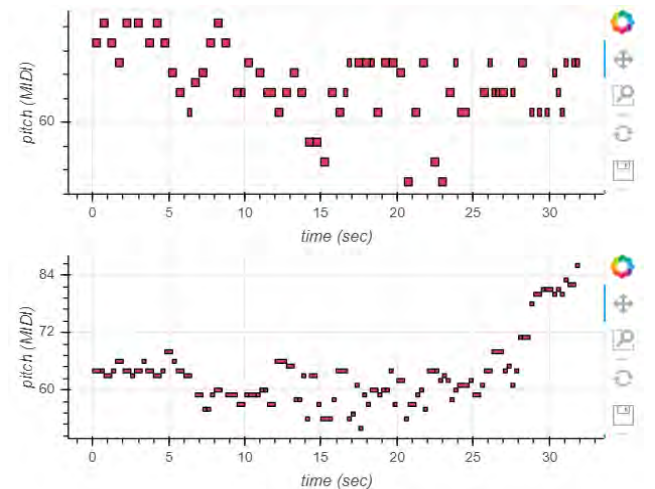


Figure 3: Sequences generated by Melody RNN.

In the next step we used the Magenta Studio, which is a collection of applications built on Magenta tools and models. This collection contains the applications: Continue, Groove, Generate, Drumify and Interpolate, which allow implementation of the models from the Magenta library on midi files. Interpolate is an application that takes two drumbeats or two tunes as inputs. It can create up to 16 files that combine the qualities of the two input files. It is useful

for merging musical ideas or for creating a smooth match between them.

We used Magenta Studio to merge the two previously generated musical sequences. This application is based on the Music VAE (Variational Autoencoder) model which has a hierarchical recurrent variational autoencoder for learning latent spaces for musical scores. Latent spaces are capable of learning the fundamental characteristics of a training dataset and they are able to represent the variation of real data in a lower-dimensional space [8]. Autoencoders are different from Recurrent Neural Networks since they use an encoder and a decoder to learn from the data. In the case of music, a sequence of music notes is encoded into a latent factor and then decoded into a sequence again [9]. One way of thinking about VAE is like mapping from MIDI to compressed space in which similar music patterns come together. Each of your input forms is represented by a position on this map. The interpolation draws a line between these two positions and returns clips along this line. A visual representation of the Music VAE model is shown in Figure 4.

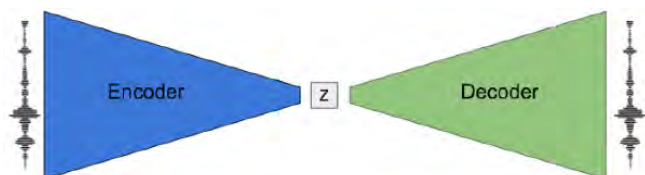


Figure 4: Structure-part of the Music VAE model

The number of recovered files is set with the steps slider. Each of the applications has a temperature slider. Temperature is a parameter used for taking samples in the last layer of the neural network. This parameter is used to generate randomness. Higher values create more variation, and sometimes even chaos, while lower values are more conservative in their predictions.

V. EXPERIMENTAL RESULTS

In our experiments, first we generated several lyrics for a particular song using the lyrics generation model. After that, we qualitatively inspected them and chose the most interesting one. In the next step, after applying interpolation on the two previously generated sequences, we got the music for the song. We attached the two melodic sequences, selected a temperature equal to 1.5 and generated 5 files, by setting the steps parameter. Then, we merged these files into one, with the help of an online converter for combining MIDI files into one midi file, in order to get a longer music file [10]. This way, we got music with a duration of 32 seconds. The result song is shown in Figure 5. The music for the same can be accessed online [3], where we also included most of the other generated songs.

The lyrics of the song (see Figure 5) is understandable, and the ones that know Toshe's songs can recognize some of the style and the phrases which are also used in his previous songs. However, the whole text is new, has new phrases and

combination of words, which makes the song unique and after some post-processing, can be further used.

во окото сјајот не поминува не
не е месечината во ноќва што
ме привлекува што во глава ти
фали си слатка но и студена

ми илјада и двеста преградки насмевка
и ледена и медена и да
ми илјада и двеста преградки насмевка
и ледена и медена и да

усни на усни да те води
познат и двеста дена суви со
чуда по тебе изгорев јас и
каде ли траг на вратот криеш

дали е доцна нешто да сменам
да те најдам ради нас патуваш
пак не е грубо ако е
од тебе во него на ум

Figure 5: The lyrics for the generated song.

In our experiments, the part with the automatic text (lyrics) generation was the most challenging. The music generation process was more of a usage of already existing tools and models. For the lyrics generation we tried different types of Deep Learning approaches and neural networks architectures, various number of hidden layers, various number of neurons, and also different activation functions. Here we explain the process, and the results achieved by the different combinations.

Firstly, we tried the Long Short Term Memory Neural Network (LSTM) approach [11], which consisted of an Embedding Layer with a number of inputs equal to the number of different words in the body, 7 outputs and input length equal to the number of tokens in sequence decreased by one. However, the results that we got with the GRU approach were better for several percentage points compared to the LSTM, therefore in the next optimization steps we used GRU.

Next, we tried several combinations of adding hidden layers. The experiments showed us that increasing the number of hidden layers, leads to a decrease of accuracy. We speculate that the reason for this is overfitting, i.e., additional layers contribute additional parameters and more complex model, which for a small dataset overfits more easily. Therefore, we fixed the number of hidden layers at two, which provided the best accuracy in the experiments.

Next, we tried two activation functions: RELU and softmax. The results showed us that the model with the RELU function achieves significantly lower accuracy compared to the softmax [12].

Finally, we tried reducing the number of outputs from the Embedding layer, and the highest accuracy was obtained with 5 outputs - this was used in the final model. The final accuracy of the model was 72% - which is in line with

similar tasks in the literature [2][11]. Please note, that the accuracy in this task is calculated as a ratio of the correctly predicted words from the total number of words.

VI. CONCLUSION

The paper described our research in the area of automatic song lyrics generation in Macedonian language and producing a music for the generated songs. In the process, we were able to develop two Deep Learning models for the two tasks: lyrics generation and music generation. The models were developed and tested using all the songs sang by one of the most famous Macedonian singers - Toshe Proeski. The results showed that the generated songs are unique, new, and after some post-processing can be potentially used to help the process of song creation.

The dataset, the models and the code are available for usage [3]. In this format, the models can be used by people that have basic understanding of Python and Machine Learning. However, in the future we plan to release a version with Graphical User Interface, so that music practitioners can freely use the software without prior knowledge of coding.

We envision that our models will be used as a tool that will help songwriters and producers in the process. In the future, the model could be further enhanced by adding more songs from other Macedonian performers to the database and by generating sounds from another instrument, such as drums.

To conclude, AI is very useful for song generation, but it cannot completely replace human resources. However, it is good enough to help in the process, and it has a predisposition to improve over time.

REFERENCES

- [1] Iqbal, Touseef, and Shaima Qureshi. "The survey: Text generation models in deep learning." *Journal of King Saud University-Computer and Information Sciences* (2020).
- [2] Ivona Milanova, Ksenija Sarvanoska, Viktor Srbinoski, Hristijan Gjoreski. Automatic Text Generation in Macedonian Using Recurrent Neural Networks. *ICT Innovations* 2019
- [3] The code and the dataset for this research. Online, accessed July 2021: https://github.com/ekotevska/Macedonian_song
- [4] Gao, Yuan & Glowacka, Dorota. Deep Gate Recurrent Neural Network. *Proceedings of The 8th Asian Conference on Machine Learning*, PMLR 63:350-365 (2016).
- [5] Nitish Srivastava and Geoffrey Hinton and Alex Krizhevsky and Ilya Sutskever and Ruslan Salakhutdinov.. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 1929-1958. (2014)
- [6] Kingma, Diederik & Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*.
- [7] Magenta library. Online, accessed July 2021: <https://magenta.tensorflow.org/>
- [8] MusicVAE: Creating a palette for musical scores with machine learning. Online, accessed July 2021: <https://magenta.tensorflow.org/music-vae>
- [9] Tikhonov, Alexey, and Ivan P. Yamshchikov. "Music generation with variational recurrent autoencoder supported by history." *arXiv preprint arXiv:1705.05458* (2017).
- [10] MIDI file processing. Online, accessed July 2021: <https://www.ofoct.com/merge-midi-files>
- [11] Pawade, D., et al. "Story scrambler-automatic text generation using word level RNN-LSTM." *International Journal of Information Technology and Computer Science (IJITCS)* 10.6 (2018): 44-53.
- [12] Steffen Eger, Paul Youssef, Iryna Gurevych. Comparing Deep Learning Activation Functions across NLP tasks

Machine Learning and Data Science Awareness and Experience in Vocational Education and Training High School Students

Stefan Zlatinov¹, Branislav Gerazov¹, Gorjan Nadzinski¹, Tomislav Kartalov¹,
Igor Atanasov², Jelena Horstmann³, Uroš Sterle⁴, and Matjaž Gams⁵

¹*Faculty of Electrical Engineering and Information Technologies*

Ss Cyril and Methodius University in Skopje, Macedonia

²*Secondary municipal vocational school "Ilinden"*

³*Electrical Engineering School "Mihajlo Pupin" in Novi Sad, Serbia*

⁴*School centre Kranj, Kranj, Slovenia*

⁵*Jožef Stefan Institute, Ljubljana, Slovenia*

zlatinov@feit.ukim.edu.mk, gerazov@feit.ukim.edu.mk, gorjan@feit.ukim.edu.mk, kartalov@feit.ukim.edu.mk

Abstract—Data Science Machine learning and are increasingly important in the world's economy and there is an increasing gap in the job market of skilled workers. To address this the Valence project proposes the design and implementation of a Data Science and Machine Learning focused curriculum in VET high schools. As a precursor to this, we designed a survey to assess the awareness/experience in these areas in students in Vocational Education and Training high schools. The survey was distributed across the three partner VET institutions. The analysis shows that While most students are aware of these two areas, only a small proportion of them have any practical experience in them or have followed an online tutorial. This reaffirms the need for the design and deployment of an accessible Data Science and Machine Learning curriculum.

Index Terms—machine learning; data science; high-school education; vocational education; survey

I. INTRODUCTION

The total amount of digital data that we create/generate each day is growing exponentially. According to estimates, to the end of 2021, there will be 74 zettabytes of generated data in the world [1]. That is expected to double by the end of 2024. This huge amount of data can be exploited by many industries and open new business opportunities. Indeed, according to the final study of the European Data Market tool, the value of the Data Economy exceeded the threshold of 400 Billion Euro in 2019 for the EU27 plus the UK, with a growth of 7.6% over the previous year [2]. This growth is complemented by an increase in the number of data professionals reaching 76 million in 2019, which is 3.6% of the total workforce - an increase of 5.5% over the previous year. The increased need of data professionals has led to an imbalance between the demand and the supply of data skills in Europe with a gap of approximately 459,000 unfilled positions corresponding to 5.7% of total demand. The data skills gap is forecast to

continue as demand will continue to outpace supply [2]. This is aggravated by the reported lack of analytical skills as a key challenge by 43% of employers [3].

To answer this imbalance, there is an urgent need for more widespread education of new generations of students in the disciplines of Data Science (DS) and Machine Learning (ML). At the University level, nearly every technical university has a series of courses dealing with data science and machine learning. However, there is also an increasing trend of having introductory courses in data science and machine learning at non-technical universities. One example is the introductory course for Data Science proposed by [4], implemented at Harvard College and in the School of Public Health, Boston, MA on a diverse group of students in terms of their knowledge of programming and statistics. Moreover, Data Science and Machine Learning are increasingly becoming an integral part of education at the elementary and high school levels. A comprehensive analysis of some 30 instructional units across different schools and curricula, shows a rising trend of courses covering machine learning per year, that is essentially skyrocketing since 2018 [5]. Some of these have explicitly addressed the need for introducing data science in Vocational Education and Training (VET) [6].

The Erasmus+ KA202 project VALENCE - Advancing machine learning in vocational education is focused on developing a curriculum and an integrated free and open-source software platform for teaching Machine Learning and Data Science.¹ The primary target audience are students attending VET high-schools, but the modularity of the platform will allow its wider usage. The project will deploy and test the curriculum in the three partner institutions: the Kranj School Centre in Kranj, Slovenia, the Electrical Engineering School "Mihajlo Pupin" in Novi Sad, Serbia, and the Vocational High

¹<https://valence.feit.ukim.edu.mk/>

School “Ilinden” in Skopje, Macedonia.

To assess the awareness of and experience with Data Science and Machine Learning, we designed an online survey that was deployed to the students attending these three VET high schools. The analysis of the results shows that although these topics are increasingly present in our daily lives, high school students have only partial awareness of them, and very few have any direct hands-on experience. These survey results will be of great importance in the development of the curriculum and the platform for teaching Machine Learning and Data Science to VET high-school students.

In Sections 2 and 3 of this paper we will present the design and the deployment of the survey, respectively. In Section 4 we will give an in-depth look into the survey results and conduct a thorough analysis of the answers, before using Section 5 to give a conclusion and outline the most important conclusions going forward.

II. SURVEY DESIGN

The survey was made comprehensive and thorough. It comprises 8 sections:

- 1) General questions - serve to obtain personal information about the student without jeopardizing their anonymity, and include: age, sex, school, specialization and grade average,
- 2) Awareness of DS and ML - a series of 6 questions to assess if the student has heard of or used DS or ML, do they grasp the usage potential of DS and ML, and do they know how to define them in their own words,
- 3) Contact with, and exposure to DS and ML - questions that assess the frequency and continuity of the student's use of modern IT platforms, including social media, video streaming services, video games and communication apps,
- 4) Interest in DS and ML - questions to assess the interest of the student in learning DS and ML, as well as their particular applications,
- 5) Experience with DS and ML - 2 short questions allowing students to express any practical experience they have with DS and ML,
- 6) General IT and language skills - questions to assess the programming experience of students as well as their proficiency in English,
- 7) Learning preferences - evaluation of the perceived benefit and personal preference of the use of various teaching methods including: traditional lectures, course books, homework projects, video lectures, online courses, interactive demos, and work on practical projects,
- 8) Extras - a group of miscellaneous questions on topics that include: cyberbullying, sports, music, movies, art, and languages.

In total the Survey contains 72 questions, 45 of which are questions in Sections 1 - 7 and directly relate to DS and ML and the development of the VALENCE curriculum, and 27 questions are in the Extras section. The latter, although unrelated to DS and ML, will serve as a rich source of practical

examples in the process of the design and development of the VALENCE curriculum.²

III. SURVEY DEPLOYMENT

The initial version of the Survey was designed in English, and after its finalization, it was translated into the languages of the partner VET High Schools: Slovenian, Serbian and Macedonian. Each translation was performed by a native speaker.

The Survey was deployed in mid-May 2021, in the three partner VET high schools to over 1,000 students of all levels and study profiles. There was a slight preference to distribute the Survey to Computer Science students, as they were identified as high performers. The schools were the Kranj School Centre in Kranj, Slovenia, the Electrical Engineering School “Mihajlo Pupin” in Novi Sad, Serbia, and the Vocational High School “Ilinden” in Skopje, Macedonia. The Survey was left open for nearly one month, in order to give more students a chance to participate. Moreover, since this was near the end of the school year, when students are busy with final projects and state exams, as well as preparations for the prom, they were reminded about filling in the Survey several times. At the end a total of 857 students responded to the Survey.

IV. RESULTS

There were a total of 857 participants in the Survey with the majority, 550 participants, coming from Serbia, then 170 from Macedonia, and 137 from Slovenia.

A. Preprocessing

We first preprocessed the data and identified participants that gave one or more inadequate answers to the questions with textual response. For example, many wrote jokes about their peers, few of them spammed the survey with arabic or chinese letters, while one pupil was determined enough to fill many fields with more than 5000 characters. A lot of them exploited the other option of the sex question in order to express their creativity. The participants identified using this criterion were eliminated from further analysis. Although seemingly strict, i.e. having one senseless answer discard all the other 72, we decided that this is necessary to maintain stronger validity of the overall analysis of the results. Nearly 25% of the participants were discarded in this way.

B. Demographic distribution

Figs. 1 – 3 show the demographic distribution of the students, i.e. age, sex and study profile, separated by high school (country). We can see that the students are aged 14 – 20, but most are 15 – 17 years old. They are predominantly male, which is not surprising for VET schools. However, we can see that there is a clear difference in the number of women across the three countries – female students make 23.5% of the students in Macedonia, 8.76% in Serbia, and only 2.26% in Slovenia. Regarding their study profiles, almost

²The Survey can be found on the following link <https://valence.feit.ukim.edu.mk>

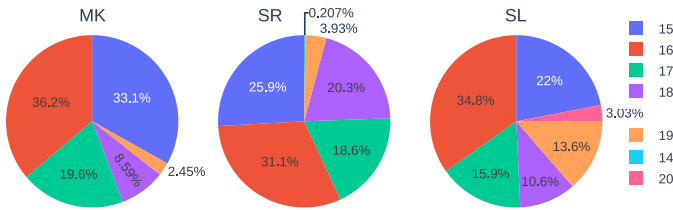


Fig. 1. Age distribution of the participants in the Survey for each country.

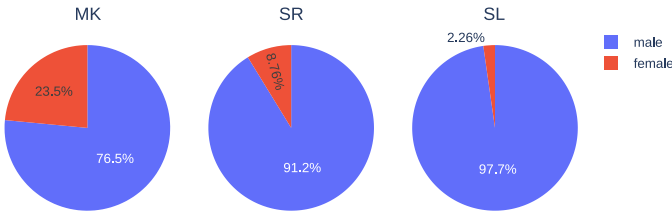


Fig. 2. Sex distribution of the participants in the Survey for each country.

50% of students are Computer Science students, followed by Electronics and Automatics, except in Macedonia where we have a more balanced distribution. This is in line with our preferences in distributing the Survey, but also reflects the number of students in the different study profiles.

C. Awareness of DS and ML

We gauge the level of awareness of DS and ML through asking “Where did you hear about ... ?” for DS and ML separately. Fig. 4 shows the analysis of the results. We can see that, even though these subjects are not part of the curriculum, most students have already heard about their existence, with ML being the more familiar term. This reflects their omnipresence in the high tech world we live in today. Despite these encouraging results, some 40% of the students have not yet heard about ML or DS, validating the need to have them included in the curriculum. As expected, the internet is the dominant source of information, and precedes the other sources of information in the chart.

When asked about the definition of the terms DS and ML, two patterns emerge in the student’s definitions. The first one is the classic “Data science is the science about data”. The second pattern, broadly exploited by the Macedonian pupils, is the first paragraph of the respected Wikipedia page: “Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains.”³

D. Experience with DS and ML

Delving deeper, we asked the students whether they have used DS or ML in a personal project. Only a handful of Serbian students responded affirmatively with the top three listed answers being neural networks, chatbots and image processing. Fig. 5 shows how many students followed an online

³https://en.wikipedia.org/wiki/Data_science

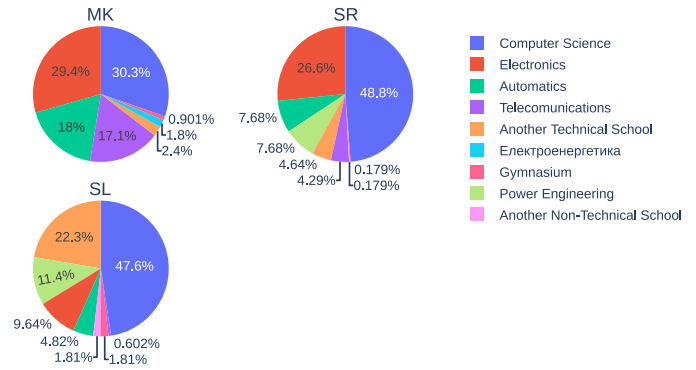


Fig. 3. Study profile distribution of the participants in the Survey for each country.

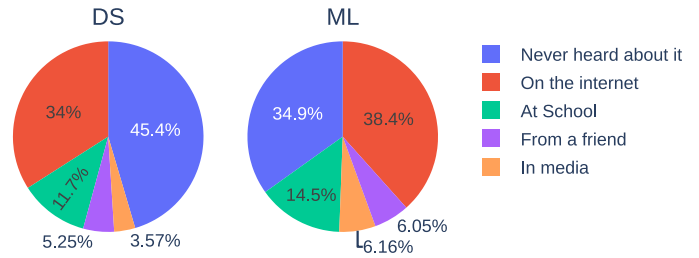


Fig. 4. Awareness of DS and ML of the participants in the Survey.

tutorial about DS or ML. We can see that Slovenian students are in the lead with Serbian students in second place. The most interesting topics of the tutorials were general programming and ML interest for the Serbian and Slovenian students, while robotics and ML application were the most interesting for the Macedonian students. However, upon closer analysis, almost a half of the provided responses clearly show that the students are still largely unfamiliar with what constitutes DS and ML, and have a hard time drawing the line between them and classical engineering.

E. Readiness for DS and ML

To indirectly evaluate the readiness for practical work in DS and ML we asked students whether they have any experience writing code in Python. Fig. 6 shows that the Slovenian students are ahead in Python programming experience, followed by the students in Serbia and Macedonia. In contrast, the use of Jupyter Notebooks is level at 5% in all of the three high schools. Speaking of general programming experience, Slovenian and Serbian students are much more versatile with many of the students ticking two or more programming language checkboxes. On the other hand, Macedonian students have exclusively and in large numbers marked themselves as knowledgeable in C++, reflecting the focus on this programming language in the Macedonian curriculum.

F. Interest in DS and ML

Next, the students were asked to express their curiosity about learning DS, ML, and statistics. Fig. 7 shows that interest in these areas is varied. The mean expressed interest,

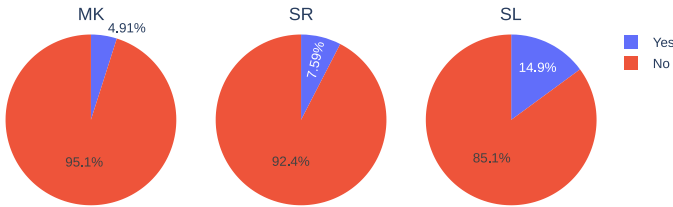


Fig. 5. Ratio of students that followed a DS or ML related online tutorial for each country.

Do you code in python?

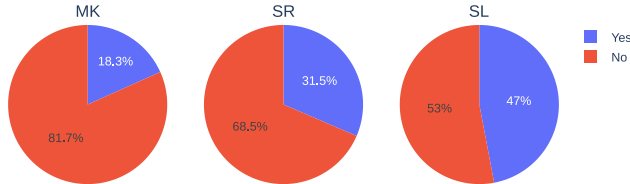


Fig. 6. Python experience of the participants in the Survey for each country.

as can be seen in Table I, is 4.4/10, and differs between the three areas. ML is the most interesting with more than 50% of students expressing interest at 5 and above, while DS and Statistics are slightly less interesting. In Fig. 7 we can see that the largest group of students, almost 20%, expressed themselves as neither being interested nor disinterested about learning DS and ML choosing 5. There are also pronounced groups giving positive scores of 7 and 10. It should be noted however, that even though there are a large number of positive responses, there is a significant portion of students (around 17%) that would not want DS, ML or Statistics included in their curricula at all. This is rather discouraging, but might correlate with the students that feel negative about the education process in general.

We also queried their particular interest in the most popular applications of DS and ML. The results show that the students exhibit a huge interest in ML applied to robotics, followed by image processing and speech recognition. On the other

V. CONCLUSION

The analysis shows encouraging results about the awareness of the fields of Data Science and Machine Learning of students from three representative VET high schools from Slovenia, Serbia and Macedonia. The number of students participating in the designed Survey, as well as the established quality of their answers, gives a high level of validity of the results. The analysis shows that students don't have ample awareness and experience in the fields of DS and ML even in VET high schools. Even if most students are aware of DS and ML, only a small proportion of them have any practical experience in them or have followed an online tutorial. This reaffirms the need for the design and deployment of an accessible DS and

How much would you like to learn about...

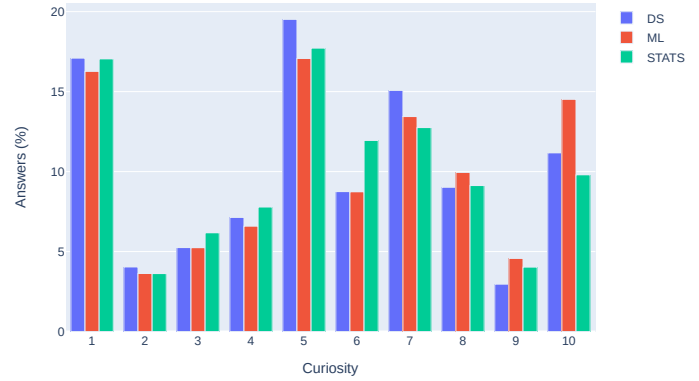


Fig. 7. Interest in learning Data Science, Machine Learning and Statistics.

TABLE I
STATISTICS OF THE INTEREST IN LEARNING DATA SCIENCE, MACHINE LEARNING AND STATISTICS.

	Data Science	Machine Learning	Statistics	Average
mean	4.35	4.63	4.3	4.43
median	4	5	4	4.33

end, w.r.t. DS applications, the majority of the answers are uniformly distributed between business intelligence, digital advertising, and internet search.

ML curriculum. In fact, their experience with Python, shows that a large number of students already have good prerequisites for following such a course.

VI. ACKNOWLEDGEMENT

The work has been financed by Erasmus+ KA202 project VALENCE: Advancing Machine Learning in Vocational Education.

REFERENCES

- [1] A. Holst. (2021) Amount of data created, consumed, and stored 2010 - 2025. [Online]. Available: <https://www.statista.com/statistics/871513/worldwide-data-created/>
- [2] European Centre for the Development of Vocational Training, "Learning outcomes approaches in VET curricula: A comparative analysis of nine European countries," https://www.cedefop.europa.eu/files/5506_en.pdf, 2010.
- [3] G. Press. (2010) The supply and demand of data scientists: What the surveys say. [Online]. Available: <https://www.forbes.com/sites/gilpress/2015/04/30/the-supply-and-demand-of-data-scientists-what-the-surveys-say>
- [4] S. C. Hicks and R. A. Irizarry, "A guide to teaching data science," *The American Statistician*, vol. 72, no. 4, pp. 382–391, 2018.
- [5] L. S. Marques, C. Gresse von Wangenheim, and J. C. Hauck, "Teaching machine learning in school: A systematic mapping of the state of the art," *Informatics in Education*, vol. 19, no. 2, pp. 283–321, 2020.
- [6] SEnDing Erasmus+ project. (2017) D2.3: Vocational curricula/educational modules for Data Science and Internet of Things VET program. [Online]. Available: http://sending-project.eu/attachments/article/71/SEnDing_DL2.3-1st_version.pdf

Towards a System for Converting Text to Sign Language in Macedonian

Stefan Spasovski¹, Branislav Gerazov¹, Risto Chavdarov¹, Viktorija Smilevska², Aneta Crvenkovska, Tomislav Kartalov¹, Zoran Ivanovski¹, and Toni Bachvarovski³

¹*Faculty of Electrical Engineering and Information Technologies
Ss Cyril and Methodius University in Skopje, Macedonia*

²*Elementary school "Kuzman Josifovski - Pitu", Skopje, Macedonia*

³*Association for Assistive Technologies "Open the Windows", Skopje, Macedonia
stefanspasovski11@gmail.com, gerazov@feit.ukim.edu.mk*

Abstract—The paper presents initial results in the design and development of a system for automatic conversion of text to sign language in Macedonian. The system will be an essential part of a larger system for the automatic generation of Macedonian sign language based on text. This system will facilitate the digital inclusion and will ease communication with the Macedonian deaf and hearing impaired community. The system is implemented as a web application which allows input text to be encoded in the equivalent sequence of sign language signs. The initial results show an average sign error rate of 4.49%. Online testing was also organized that confirmed these promising results.

Index Terms—natural language processing, assistive technology, text-to-sign language, sign language, deafness

I. INTRODUCTION

Deafness is defined as a condition of extreme hearing loss, i.e. having very little or no hearing at all. The American Speech-Language-Hearing Association (ASHA) defines profound hearing loss as only being able to hear sounds above 90 dB, with severe hearing loss ranging between 71 – 90 dB [1]. Macedonian Law, places the threshold at 80 dB. The estimated number of deaf and hearing impaired people in Macedonia is around 6000 according to information from 2006 [2], which is 0.3% of Macedonia's population [3]. This is comparable to the percentage of deaf people in places like the United States (0.38%) [4] and in Germany (0.28%) [5].

In today's high-tech information-rich world, the digital inclusion of people with disabilities is becoming increasingly important. The deaf and hearing impaired are generally able to directly use and interact with computers and smart devices, and can follow traditional visual media such as TV and newspapers. However, they find it hard to read text at speed. This is especially true for those born deaf or hearing impaired, as they have never heard the sounds of phonemes that phonetic orthography transcribes with graphemes in written text. As a consequence, for them phonetic transcription is not much different from logograms, such as the Chinese characters used to write Mandarin and other Asian languages. This problem can be mitigated through offering live sign language (SL) translation, but most TV broadcasters do not offer this service.

The problem is even more pronounced in Macedonia, where there are only around 30 certified SL translators.¹

One way to ease the digital inclusion of the deaf and hearing impaired is through assistive systems able to automatically convert text-to-sign language. One example of such a system is the HandTalk App in which a virtual avatar named Hugo converts text to sign language on the users smart device.² These systems are made up of two essential parts: *i*) a text-to-SL converter that transforms textual input to a sequence of signs or gestures, and *ii*) a SL generator that uses the sequence of signs to generate sign language, usually via 3D rendering of an animated character, i.e. avatar. Sign language differs from spoken language, in that it does not support inflection. For example, to create the future tense in Macedonian we add the future particle before the verb. Tense is not formed in that way in sign language. Instead the speaker uses the infinitive form of the verb together with signs like "later" to signify that they are speaking about a future event. The same holds true for verb conjugation, singular and plural, case and articles.

Although sign languages across the world do share signs, there are still different standardized sign languages, such as: American Sign Language (ASL), Italian Sign Language (LIS), Indian Sign Language (ISL), Vietnamese Sign Language (VSL), and Macedonian Sign Language (MSL). The text-to-SL converter can encode the signs in various formats. One famous format that dates back to 1984 is the Hamburg Notation System for Sign Languages (HamNoSys), which encodes signs through a set of pictograms or symbols [6]. As an extension to HamNoSys, the Signing Gesture Mark-up Language (SiGML) describes the symbols using XML tags [7]. This extension allows the storage and use of the transcriptions in computer based systems, such as 3D rendering software, that can be used to generate sign language via an animated avatar.

Text-to-SL systems have been developed for many of the world languages, such as English [8], German [9], Vietnamese

¹<http://www.deafmkd.org.mk/lista-na-tolkuvaci/>

²<https://handtalk.me/en/app/>

[10], Kurdish [11], Arabic [12], Brazilian Portuguese [13], Punjabi [14], Korean [15] etc. Most systems rely on a simple word-to-sign mapping, i.e. each word token from the input text is looked up in a lexicon of signs, and if no match is found it is spelled out using a sequence of alphabet letter signs [11]. More advanced rule-based-systems map the input text grammar to natural sign language grammar [15]. Recently, the application of machine learning has allowed improved performance in these systems, directly applying methods used in the area of Machine Translation [10].

In Macedonia, work has mostly been done on the generation of Macedonian Sign Language via virtual avatars. Koceski and Koceska [16] developed and evaluated a 3D virtual tutor for MSL. Joksimoski et al. [17] presented a 3D visualization system that extensively uses animation and game concepts for accurately generating sign languages using 3D avatars.

Here, we present a text-to-SL system that translates an input Macedonian text into an output sequence of Macedonian Sign Language signs. The system is built on a rule based algorithm, which analyses input text, comparing the input word tokens to a lexicon of some 200 signs. The performance of the system is evaluated with a set of test sentences and the results show a Sign Error Rate (SER) of 4.49%. We augment this analysis with online testing of the system, which confirms the validity of the initial results. The system can be used as the basis for building a complete system for text based sign language generation in Macedonian.

II. ALGORITHM

A. Sign mappings organization

We organize the data by placing the words and signs in five different files:

- list of signs,
- list of names,
- skip list,
- dictionary of word to sign mappings, and
- dictionary of phrases that map directly to sequences of signs.

In the presented system we have 221 signs.

B. Text preprocessing

Text input is first normalised by converting all upper case characters to lower case and removing all punctuation. We then divide the text string into into a list of word tokens. We initialise two empty lists for storing the sign sequence output and unrecognized word tokens, and define a flag to be used when a phrase has been recognized. With that we are ready to begin the translation process.

C. Main loop

We go through the word tokens in the input list one by one and run them through multiple checks. Firstly, we check if the word token is part of a phrase by concatenating the succeeding token from the list. If it is, then we append the sequence of signs corresponding to that phrase, skip the second word from the next iteration and move on from there. Next, if the token

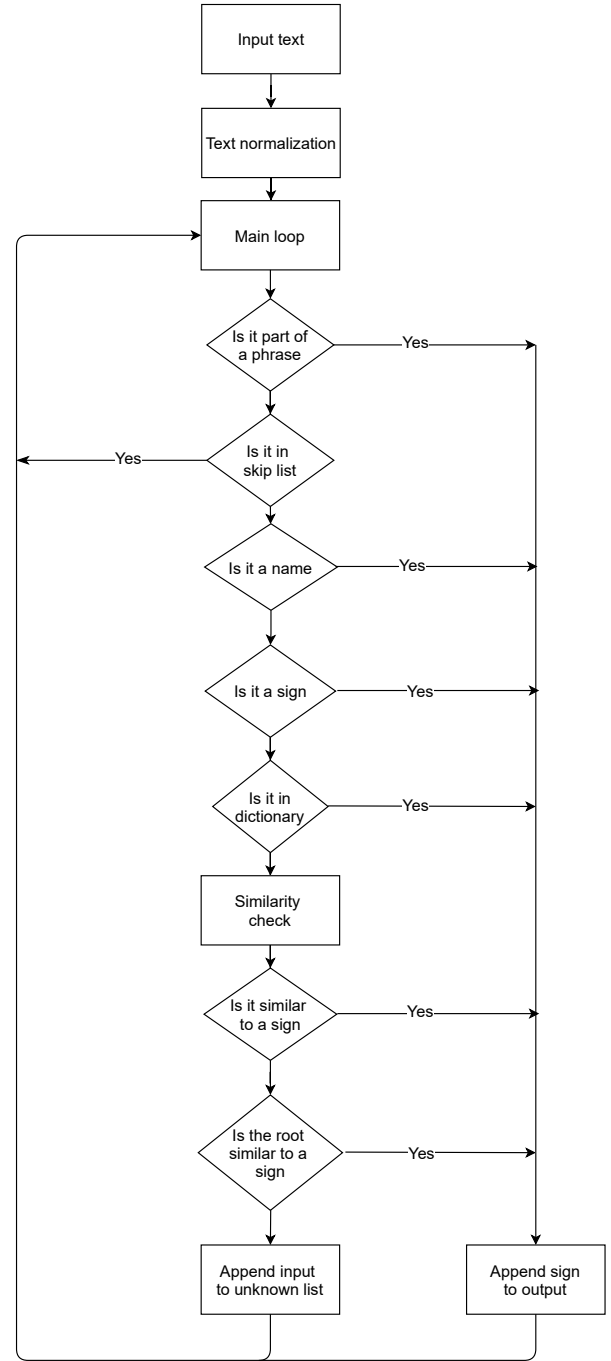


Fig. 1. Block diagram of the algorithm for converting text to sign language signs.

is in the skip list it is ignored and the next token is processed. If not in the skip list, we check if the input word token is part of the names or signs lists. If found the word token is appended as is to the output sign sequence. If not, the token is looked up in the dictionary of word to sign mappings, and if found it's sign mapping is appended to the output. If the word still has not been found in any of the checks then we continue with similarity checks.

D. Similarity checks

The final part of the main loop consists of similarity checks. The system comprises two different similarity checks. They are based on the “gestalt pattern matching” algorithm suggested by Ratcliff and Obershelp in the 1980s [18]. The idea being to find the longest contiguous matching subsequence that contains no “junk” elements. This is applied recursively to the pieces of the sequences to the left and to the right of the matching subsequence. We use the implementation of the algorithm in the `diffib`³ Python library.

The similarity check outputs a list of 3 strings sorted from the most likely to the list likely match. Based on our experiments we get the best results when the cutoff is equal to 0.7. One other problem that we encounter is the fact that sometimes the most likely match (the first element of the string) is not at all the most likely, and that the second or third string in the list is the correct answer. We augment the similarity search algorithm with a Character Error Rate (CER) to select one of the three offered outputs:

$$CER = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C} \quad (1)$$

where S is the number of character substitutions, D is the number of character deletions, I is the number of character insertions, C is the number of correct characters, and N is the total number of characters in the reference, i.e. $N = S + D + C$.

III. EXPERIMENTS

We used two experiments to evaluate the performance of the proposed system. The first was based on an internal test set, and the second was based on online testing.

A. Test set evaluation

We developed an internal test set comprising 124 sentences for which we provided reference translations to sign language sequences of signs. We took special care to have a varied test set both in terms of sentence length as well as ample coverage of the set of signs known to the system.

B. Online evaluation

To provide a platform for testing the proposed system, we developed a web application that provides a user interface for online testing. The web app was developed based on Flask⁴. HTML was used to build the site layout, while Flask was used to render the website, receive user input and return the results of the translation process.

The website lets the user input words for translation. After submitting the input the translation process starts and the output from the proposed system including the output sign sequence and the list of unrecognized word tokens. There is also a text form that the user can use to give feedback. For our online tests this was the correct sign sequence in case the system returned an erroneous one. To improve coverage of the known signs in the online tests, the signs known by the system are listed at the end of the web page.

³<https://docs.python.org/3/library/difflib.html>

⁴<https://flask.palletsprojects.com/en/2.0.x/>

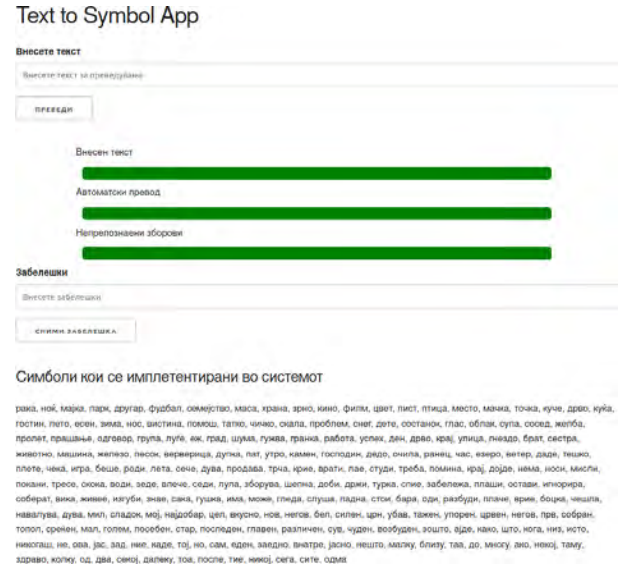


Fig. 2. The web app developed for the online testing.

C. Metrics

To assess the level of performance of the proposed system we use the Word Error Rate (WER) to compare the output sign sequence with the reference sign sequence translation, either defined in the test set, or input by the online participants. The WER is similar to the CER used in the similarity checks and is defined as:

$$WER = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C} \quad (2)$$

where S is the number of word substitutions, D is the number of word deletions, I is the number of word insertions, C is the number of correct words, and N is the number of words in the reference, i.e. $N = S + D + C$.

IV. RESULTS

The evaluation of the system's performance using the test set was an average of 4.49% WER, i.e. about 1 erroneous sign in 20 signs output. A small part of these errors are due to the difference between the output signs, as they are given in the system's dictionary and as they are provided in the reference translations. These are minor differences comprising one or two characters added in some of the signs.

Errors occurred most notably when input words were expanded compared to the way they are found in the system's database. For example, when a word is written in the diminutive plural form. In those cases either the algorithm finds a similar word which is not the correct sign corresponding to the word token, or appends the token to the unrecognized word token list. Such errors do not occur when the input word has a few changes, or the word is long enough that the relative number of changes is minimal.

The online testing results were also promising, with minimal errors in the output sequences for the input provided by the user. Most errors could be attributed to: *i*) the word translations

not being part of the signs listed in the web app or *ii*) being synonyms to words that are provided in the system's dictionary and which have not been added to the system. In some of these cases the words are appended to the unrecognized word token list, but a frequently enough a similar, but erroneous, sign is found.

V. CONCLUSION

The proposed system is able to translate an input text sequence in to an output sign language sign sequence. The system is based on a rule-based algorithm that converts the input word tokens sequentially into their sign language equivalents. Even though supporting a limited sign vocabulary, the results from the internal and online testing are promising. With future improvements the system can be used in combination with a sign language generator to create a complete text-to-sign language solution. This would be of great help for the digital inclusion and communication with the deaf and hearing impaired community.

REFERENCES

- [1] A. S.-L.-H. Association. (2021) Hearing loss (ages 5+). [Online]. Available: <https://www.asha.org/practice-portal/clinical-topics/hearing-loss/>
- [2] Dnevnik. (2006) Around 6,000 deaf people are asking for sign language news on mtv. [Online]. Available: <https://web.archive.org/web/20110727190206/http://star.dnevnik.com.mk/default.aspx?pbroj=1782&stID=10852>
- [3] Republic of North Macedonia State Statistical Office. (2002) Census of Population, Households and Dwellings in the Republic of North Macedonia. [Online]. Available: <https://www.stat.gov.mk/Publikacii/knigaXIII.pdf>
- [4] G. R. I. Charles Reilly. (2011) Snapshot of deaf and hard of hearing people, postsecondary attendance and unemployment. [Online]. Available: <https://www.gallaudet.edu/office-of-international-affairs/demographics/deaf-employment-reports/>
- [13] D. Stiehl, L. Addams, L. S. Oliveira, C. Guimarães, and A. Britto, "Towards a signwriting recognition system," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 26–30.
- [5] T. G. N. T. Board. (2021) Information for deaf visitors to germany. [Online]. Available: <https://www.germany.travel/en/accessible-germany/disability-friendly-travel-for/deafness.html>
- [6] T. Hanke, "HamNoSys-representing sign language data in language resources and language processing contexts," in *LREC*, vol. 4, 2004, pp. 1–6.
- [7] K. Kaur and P. Kumar, "Hamnosys to sigml conversion system for sign language automation," *Procedia Computer Science*, vol. 89, pp. 794–803, 2016.
- [8] M. Varghese and S. K. Nambiar, "English to sigml conversion for sign language generation," in *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*. IEEE, 2018, pp. 1–6.
- [9] S. Ebling and J. Glauert, "Building a swiss german sign language avatar with jasingning and evaluating it among the deaf community," *Universal Access in the Information Society*, vol. 15, no. 4, pp. 577–587, 2016.
- [10] N. C. Ngon and Q. L. Da, "Application of hamnosys and avatar 3d jasingning to construction of vietnamese sign language animations," *The University of Danang-Journal of Science and Technology*, pp. 61–65, 2017.
- [11] Z. Kamal and H. Hassani, "Towards kurdish text to sign translation," in *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, 2020, pp. 117–122.
- [12] N. Aouiti, "Towards an automatic translation from arabic text to sign language," in *Fourth International Conference on Information and Communication Technology and Accessibility (ICTA)*. IEEE, 2013, pp. 1–4.
- [14] A. S. Dhanjal and W. Singh, "An automatic conversion of punjabi text to indian sign language," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 7, no. 28, p. e9, 2020.
- [15] S. H. Kang and S. H. Park, "Toward korean text-to-sign language translation system (test)."
- [16] S. Koceski and N. Koceska, "Development and evaluation of a 3d virtual tutor for macedonian sign language," in *International Conference on Information Technology and Development of Education – ITRO 2015*, Zrenjanin, Serbia, 2015.
- [17] B. Joksimoski, I. Chorbev, K. Zdravkova, and D. Mihajlov, "Toward 3d avatar visualization of macedonian sign language," in *International Conference on ICT Innovations*. Springer, 2015, pp. 195–203.
- [18] J. W. Ratcliff and D. E. Metzner, "Pattern matching: the Gestalt approach," *Dr Dobbs Journal*, vol. 13, no. 7, p. 46, 1988.

ETAI 2021 TECHNICAL PROGRAMME AGENDA

THURSDAY, 23. IX. 2021

SESSION - 1: CIRCUITS AND SYSTEMS

09:30 – 11:00

Co-Chairs: Katerina Raleva and Tomislav Kartalov

ETAI 1-1	A SONAR-BASED OBSTACLE DETECTION SYSTEM FOR THE BLIND AND VISUALLY DISABLED
	Stefana Hristovska, Kristijan Lazarev and Branislav Gerazov
ETAI 1-2	LIGHTING DESIGN, AUTOMATION, EFFICIENCY AND ADVANTAGES MADE WITH ILLUMINATION LEVEL CONTROL IN INDUSTRIAL FACILITIES
	Mehmet Gürçan Gür and Yilmaz Uyaroğlu
ETAI 1-3	DETECTION OF INDIVIDUAL FINGER FLEXIONS USING TWO-CHANNEL ELECTROMYOGRAPHY
	Blagoj Hristov and Gorjan Nadzinski
ETAI 1-4	THE SELECTION OF BI-FRACTIONAL ORDER REFERENCE MODEL PARAMETERS FOR MINIMUM SETTLING TIME
	Ertuğrul Keçeci, Erhan Yumuk, Müjde Güzelkaya and İbrahim Eksin
ETAI 1-5	INTRA-NODAL CACHING ASSISTED UAV BASED DATA ACQUISITION FROM WIRELESS MOBILE AD-HOC SENSOR NETWORKS
	Umair Chaudhry and Chris Phillips

SESSION - 2: CYBER SECURITY AND MATHEMATICS

09:30 – 11:00

Co-Chairs: Danijela Efnusheva and Sanja Atanasova

ETAI 2-1	ANALYSIS OF SMART HOME SECURITY BY APPLYING MACHINE LEARNING ALGORITHMS
	Irina Senchuk, Ana Cholakoska and Danijela Efnusheva
ETAI 2-2	NETWORK SECURITY ANALYSIS BY APPLYING MACHINE LEARNING ALGORITHMS
	Martina Shushlevska, Ana Cholakoska and Danijela Efnusheva
ETAI 2-3	NUMERICAL SOLUTION OF LAPLACE DIFFERENTIAL EQUATION USING THE FINITE DIFFERENCE METHOD
	Bojana Petrovska, Daniela Janeva, Emilija Tasheva and Andrijana Kuhar
ETAI 2-4	MODELING POPULATION DYNAMICS AND ECONOMIC GROWTH AS COMPETING SPECIES FOR NORTH MACEDONIA
	Stefan Boshkovski and Sanja Atanasova
ETAI 2-5	PERFORMANCE OF GRADIENT ALGORITHMS FOR SOLVING LEAST SQUARES PROBLEM
	Naum Dimitrieski, Katerina Hadzi-Velkova Saneva and Zoran Hadzi-Velkov

SESSION - 3: CONTROL SYSTEMS AND AUTOMATION

13:30 – 15:00

Co-Chairs: Mile Stankovski, Georgi Dimirovski

ETAI 3-1	MULTI-OBJECTIVE OPTIMIZATION BASED FRACTIONAL ORDER PID CONTROLLER DESIGN
	Erhan Yumuk, Eda Budak, Mjude Güzelkaya and İbrahim Eksin
ETAI 3-2	FRACTIONAL INTEGRATING INTEGER ORDER PI CONTROLLER DESIGN FOR THE FIRST INTEGER ORDER PLUS TIME DELAY SYSTEM
	Erhan Yumuk, Mjude Güzelkaya and İbrahim Eksin
ETAI 3-3	FUZZY LOGIC BASED MAXIMUM POWER POINT TRACKING FOR PHOTOVOLTAIC SYSTEMS
	Zeynep Bala Duranay and Hanifi Guldemir
ETAI 3-4	FUZZY-LOGIC OUTPUT-TRACKING CONTROL FOR UNCERTAIN TIME-DELAY DYNAMICAL PROCESSES: EXPLORING TAKAGI-SUGENO FUZZY MODELS
	Yuanwei Jing, Xin-Jiang Wei, Janusz Kacprzyk, Imre J Rudas and Georgi Dimirovski
ETAI 3-5	DISCRETE-TIME UNSCENTED KALMAN FILTERS WITH OPERATING OF UNCERTAINTIES: STOCHASTIC STABILITY ANALYSIS
	Yuanwei Jing, Jiahe Xu, Peng Shi and Georgi Dimirovski
ETAI 3-6	COMPLEX MULTI-NETWORKS WITH FAULTY INTER-NETWORK CONNECTIONS: SYNCHRONIZATION VIA NOVEL PINNING-NODE CONTROL
	Yuanwei Jing, Guanrong Chen, Peng Shi and Georgi Dimirovski

SESSION - 4: E-HEALTH

13:30 – 15:00

Co-Chairs: Hristijan Gjoreski, Daniel Denkovski

ETAI 4-1	INSIEME: A UNIFYING ELECTRONIC AND MOBILE HEALTH PLATFORM
	Primoz Kocuvan, Erik Dovgan, Tine Kolenik and Matjaž Gams
ETAI 4-2	A SYSTEM FOR AUTOMATIC DETECTION OF MAJOR DEPRESSIVE DISORDER BASED ON BRAIN ACTIVITY
	Daniela Janeva, Silvana Markovska-Simoska and Branislav Gerazov
ETAI 4-3	PREDICTING TRENDS AND ANOMALIES IN DAILY ACTIVITIES
	Vito Janko and Mitja Luštrek
ETAI 4-4	FINDING EFFICIENT INTERVENTION PLANS AGAINST COVID-19
	Nina Reščič, Vito Janko, David Susič, Carlo De Masi, Aljoša Vodopija, Matej Marinko, Tea Tušar, Erik Dovgan, Anton Gradišek, Matej Cigale, Matjaž Gams and Mitja Luštrek
ETAI 4-5	MACHINE LEARNING BASED ANOMALY DETECTION IN AMBIENT ASSISTED LIVING ENVIRONMENTS
	Ana Cholakoska, Valentin Rakovic, Hristijan Gjoreski, Bjarne Pfitzner, Bert Arnrich and Marija Kalendar
ETAI 4-6	INVESTIGATING PRESENCE OF ETHNORACIAL BIAS IN CLINICAL DATA USING MACHINE LEARNING
	Bojana Velichkovska, Hristijan Gjoreski, Daniel Denkovski, Marija Kalendar, Leo Anthnoy Celi and Venet Osmani

SESSION - 5: COMMUNICATION NETWORKS - 5G

(SUPPORTED BY THE COMMUNICATIONS AND INFORMATION THEORY CHAPTERS OF IEEE
N. MACEDONIA SECTION)

13:30 – 15:00

Co-Chairs: Tomislav Shuminoski, Pero Latkoski

ETAI 5-1	EVALUATION OF DISTRIBUTED NFV INFRASTRUCTURES FOR EFFICIENT EDGE COMPUTING IN 5G
	Gjorgji Ilievski and Pero Latkoski
ETAI 5-2	PARTICLE SWARM OPTIMIZATION (PSO) BASED RESOURCE ALLOCATION FOR DEVICE TO DEVICE COMMUNICATION FOR 5G NETWORK
	Wisam Hayder Mahdi and Necmi Taspınar
ETAI 5-3	INVESTIGATION OF EFFECT OF THE PILOT REUSE FACTOR VIA INTELLIGENT OPTIMIZATIONS ON ENERGY AND SPECTRAL EFFICIENCIES TRADE-OFF IN MASSIVE MIMO SYSTEMS
	Burak Kürşat Gül and Necmi Taşpınar
ETAI 5-4	COMPUTING ON THE EDGE: A SYSTEM AND TECHNOLOGY OVERVIEW
	Marija Poposka and Zoran Hadzi-Velkov
ETAI 5-5	MOBILE EDGE COMPUTING SERVICES WITH QOS SUPPORT FOR BEYOND 5G NETWORKS – USE CASES
	David Nunev, Tomislav Shuminoski, Bojana Velichkovska and Toni Janevski

FRIDAY, 24. IX. 2021

SESSION - 6: INSTRUMENTATION AND MEASUREMENTS

12:00 – 13:30

Co-Chairs: Mare Srbinovska, Kiril Demerdziev

ETAI 6-1	POSITIONAL VALUE MEASUREMENT FOR A ROOK AND KING VS ROOK CHESS ENDGAME ALGORITHM
	Adrijan Bozinovski and Filemon Jankuloski
ETAI 6-2	OVERVIEW OF SECURITY AND SAFETY SYSTEMS IN THE AUTOMOTIVE INDUSTRY
	Aleksandra Gjorgjievska, Mare Srbinovska and Martin Gjorgjievski
ETAI 6-3	DESIGN AND EVALUATION OF COLLABORATIVE LEARNING PLATFORM WITH INTEGRATED REMOTE LABORATORY ENVIRONMENT
	Zivko Kokolanski, Bodan Velkovski, Tomislav Shuminoski, Dušan Gleich, Andrej Sarjaš, Ana B. Kokolanska, Anita K. Mijovska, Matjaž Šegula, Matic Podobnik, Zlatko Rušić and Tibor Kratofil
ETAI 6-4	ERROR EVALUATION IN REACTIVE POWER AND ENERGY MEASUREMENTS ADOPTING DIFFERENT POWER THEORIES
	Kiril Demerdziev and Vladimir Dimchev
ETAI 6-5	VIRTUAL REAL TIME POWER QUALITY DISTURBANCE CLASSIFIER BASED ON DISCRETE WAVELET TRANSFORM AND MACHINE LEARNING
	Petar Vidoevski, Dimitar Taskovski and Zivko Kokolanski
ETAI 6-6	IMPROVING THE EFFICIENCY OF GROUNDING SYSTEM ANALYSIS USING GPU PARALLELIZATION
	Bodan Velkovski, Blagoja Markovski, Vladimir Gjorgievski, Marija Markovska, Leonid Grcev, Stefan Kalabakov and Elena Merdjanovska

SESSION - 7: CLOUD AND IOT TECHNOLOGIES

**(SUPPORTED BY THE COMMUNICATIONS AND INFORMATION THEORY CHAPTERS OF IEEE
N. MACEDONIA SECTION)**

12:00 – 13:30

Co-Chairs: Toni Janevski, Simon Bojadzievski

ETAI 7-1	TECHNOLOGICAL, REGULATORY AND BUSINESS ASPECTS OF LPWAN IMPLEMENTATION IN IOT
	Atanas Godzoski, Toni Janevski and Aleksandar Risteski
ETAI 7-2	EXTENDED PERFORMANCE EVALUATION OF THE TENDERMINT PROTOCOL
	Jovan Karamachoski and Liljana Gavrilovska
ETAI 7-3	ANALYSIS OF SECURITY MECHANISMS OF CONTAINERS IN CLOUD
	Martina Janakieska and Aleksandar Risteski
ETAI 7-4	THE APPLICATION OF THE INTERNET OF THINGS IN EVERYDAY EQUIPMENT AFFECTS TO HAVE A MORE EFFICIENT AND QUALITY LIFE
	Avni Rustemi
ETAI 7-5	USER-TO-CLOUD LATENCY PERFORMANCE CHARACTERISTICS IN AN EUROPEAN CLOUD INFRASTRUCTURE
	Teodora Kochovska, Marija Kalendar and Simon Bojadzievski

SESSION - 8: ARTIFICIAL INTELLIGENCE IN AUTOMATION

12:00 – 13:30

Co-Chairs: Gorjan Nadzinski, Vesna Latkoska-Ojleska

ETAI 8-1	FORECASTING DYNAMIC TOURISM DEMAND BY ARTIFICIAL NEURAL NETWORKS
	Cvetko Andreeski and Biljana Petrevska
ETAI 8-2	FORECASTING POWER CONSUMPTION FOR RESIDENTIAL SECTOR
	Aleksandra Zlatkova, Aneta Buckovska and Dimitar Taskovski
ETAI 8-3	MODULATION CLASSIFICATION WITH DEEP LEARNING: COMPARISON OF DEEP LEARNING MODELS
	Selçuk Balsüzen and Mesut Kartal
ETAI 8-4	MACHINE LEARNING APPROACH FOR AUTONOMOUS CONTROL OF VERTICAL CEMENT ROLLER MILLS
	Othon Manis, Gorjan Nadzinski and Mile Stankovski
ETAI 8-5	SELECTING AN OPTIMISATION ALGORITHM FOR OPTIMAL ENERGY MANAGEMENT IN GRID-CONNECTED HYBRID MICROGRID WITH STOCHASTIC LOAD
	Natasha Dimishkovska, Atanas Iliev and Borce Postolov
ETAI 8-6	COMPARATIVE ANALYSIS OF DIFFERENT HELIOSTAT FIELD CONTROL ALGORITHMS
	Ivan Andonov, Vesna Ojleska Latkoska and Mile Stankovski

SESSION - 9: COMMUNICATION TECHNOLOGIES

(SUPPORTED BY THE COMMUNICATIONS AND INFORMATION THEORY CHAPTERS OF IEEE N. MACEDONIA SECTION)

14:00 – 15:30

Co-Chairs: Zoran Hadzi-Velkov, Slavce Pejovski

ETAI 9-1	UNCERTAIN AQM/TCP COMPUTER AND COMMUNICATION NETWORKS: FIXED-TIME CONGESTION TRACKING CONTROL USING GAUSSIAN FUZZY-LOGIC EMULATOR
	Jindong Shen, Yuanwei Jing, Janusz Kacprzyk, Georgi M. Dimirovski
ETAI 9-2	WIRELESS POWERED ALOHA NETWORKS WITH FIXED USER RATES AND UAV-MOUNTED BASE STATIONS
	Slavche Pejovski and Zoran Hadzi-Velkov
ETAI 9-3	PERFORMANCE INVESTIGATION OF BIDIRECTIONAL OPTICAL IM/DD OFDM WDM-PON USING RSOA AS A COLORLESS TRANSMITTER
	Mahmoud Alhalabi, Necmi Taşpınar and Fady El-Nahal
ETAI 9-4	DEVELOPMENT AND DEPLOYMENT OF A LORAWAN PERFORMANCE TEST SETUP FOR IOT APPLICATIONS
	Simeon Trendov, Marija Kalendar and Eduard Siemens

SESSION ETAI – 10: ARTIFICIAL INTELLIGENCE IN BIOMEDICINE

(SUPPORTED BY THE COMPUTATIONAL INTELLIGENCE CHAPTER OF IEEE N. MACEDONIA SECTION)

14:00 – 15:30

Co-Chairs: Branislav Gerazov, Andrijana Kuhar

ETAI 10-1	THE REPRESENTATION OF SPOKEN VOWELS IN HIGH GAMMA RANGE OF CORTICAL ACTIVITY
	Daniela Janeva, Andrijana Kuhar, Lidija Olooska-Gagoska and Branislav Gerazov
ETAI 10-2	SCORPIANO - A SYSTEM FOR AUTOMATIC MUSIC TRANSCRIPTION FOR MONOPHONIC PIANO MUSIC
	Bojan Sofronievski and Branislav Gerazov
ETAI 10-3	FACIAL EMOTION RECOGNITION USING DEEP LEARNING
	Gjorgji Smilevski and Tomislav Kartalov
ETAI 10-4	AUTOMATIC COMPOSITION OF TEXT AND MUSIC FOR A SONG IN MACEDONIAN USING DEEP LEARNING
	Angela Najdoska, Emilija Kotevska, Tamara Markachevikj and Hristijan Gjoreski
ETAI 10-5	MACHINE LEARNING AND DATA SCIENCE AWARENESS AND EXPERIENCE IN VOCATIONAL EDUCATION AND TRAINING FOR HIGH-SCHOOL STUDENTS
	Stefan Zlatinov, Branislav Gerazov, Gorjan Nadzinski and Tomislav Kartalov
ETAI 10-6	TOWARDS A SYSTEM FOR CONVERTING TEXT TO SIGN LANGUAGE IN MACEDONIAN
	Stefan Spasovski, Branislav Gerazov, Risto Chavdarov, Viktorija Smilevska, Aneta Crvenkovska, Tomislav Kartalov, Zoran Ivanovski and Toni Bachvarovski

PROGRAM

Day	Time	Virtual room 1	Virtual room 2	Virtual room 3
THURSDAY, 23. IX 2021	09:30-11:00	Session 1 (Circuits and systems) Zoom Link ID: 943 9993 9052 Pass:ETAI_1	Session 2 (Cyber security and Mathematics) Zoom Link ID: 962 3612 0428 Pass:ETAI_2	
	11:00-11:30	Conference Opening Zoom Link ID: 935 5274 2669 Pass:ETAI_OPEN		
	11:30-12:15	Plenary lecture 1 Prof. Peng Shi (Formation Control Design for Multi-agent Systems) Zoom Link ID: 935 5274 2669 Pass:ETAI_OPEN		
	12:15-13:00	Plenary lecture 2 Prof. Matjaz Gams (Is Web Transforming Our Minds and Where is Our Civilisation Going to?) Zoom Link ID: 935 5274 2669 Pass:ETAI_OPEN		
	13:00-13:30	Lunch Break		
	13:30-15:00	Session 3 (Control systems and automation) Zoom Link ID: 990 6823 4264 Pass:ETAI_3	Session 4 (eHealth) Zoom Link ID: 955 0609 0111 Pass:ETAI_4	Session 5 (Communication networks - 5G) Zoom Link ID: 921 5973 7239 Pass:ETAI_5
	15:00-16:30	Round Table 1: e-Health and Pervasive Technologies: Challenges and Opportunities (moderator Dr. Hristijan Gjoreski) Zoom Link		
FRIDAY, 24. IX 2021	10:00-10:45	Plenary lecture 3 Prof. Janusz Kacprzyk (Human-in-the-loop AI in decision and control systems: the role of linguistic data summaries) Zoom Link ID: 982 1242 8400 Pass:ETAI_PLEN		
	10:45-11:30	Plenary lecture 4 Dr. Tome Eftimov (Modern approaches for Benchmarking in Evolutionary Computation) Zoom Link ID: 982 1242 8400 Pass:ETAI_PLEN		
	11:30-12:00	Coffee Break		
	12:00-13:30	Session 6 (Instrumentation & Measurements) Zoom Link ID: 959 4581 0574 Pass:ETAI_6	Session 7 (Cloud and IoT technologies) Zoom Link ID: 985 2669 6825 Pass:ETAI_7	Session 8 (Artificial intelligence in Automation) Zoom Link ID: 941 5568 8584 Pass:ETAI_8
	13:30-14:00	Lunch Break		
	14:00-15:30	Session 9 Zoom Link (Communication technologies) ID: 979 4445 2721 Pass:ETAI_9	Session 10 (Artificial intelligence in Biomedicine) Zoom Link ID: 963 3489 3306 Pass:ETAI_10	
	15:30-17:00	Round Table 2: Next generation wireless communications - Opportunities and challenges for IoT (moderator Dr. Ljupco Jorguseski) Zoom Link ID: 910 8233 3630 Pass:ETAI_ROUND		
	17:00-17:30	Closing ceremony ETAI 40th anniversary celebration Best paper awards Zoom Link ID: 934 1811 1014 Pass:ETAI_CLOSE		